RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 24
(2000)

David B. Pisoni, Ph.D.
Principal Investigator

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405-1301

Research Supported by:

Department of Health and Human Services
U.S. Public Health Service

National Institutes of Health
Research Grant No. DC-00111

and

National Institutes of Health
Training Grant No. DC-00012

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)

Table of Contents

Introduction vii

Speech Research Laboratory Faculty, Staff, and Technical Personnel viii

I. Extended Manuscripts 1

• Working Memory Spans as Predictors of Word Recognition and Receptive Vocabulary in Children with Cochlear Implants
Miranda Cleary, David B. Pisoni and Karen I. Kirk 3

• The Influence of Short-term and Long-term Memory on the Identification and Discrimination of Non-native Speech Sounds
James D. Harnsberger 21

• Some Acoustic Cues for Categorizing American English Regional Dialects: An Initial Report on Dialect Variation in Production and Perception
Cynthia G. Clopper 43

• Prosodic and Morphological Effects on Word Reduction in Adults: A First Report
Allyson K. Carter and Cynthia G. Clopper 67

• Perception of "Elliptical Speech" by an Adult Hearing-Impaired Listener with a Cochlear Implant: Some Preliminary Findings on Coarse-Coding in Speech Perception
Rebecca Herman and David B. Pisoni 87

• Using Nonword Repetition to Study Speech Production Skills in Hearing-Impaired Children with Cochlear Implants
Caitlin M. Dillon and Miranda Cleary 113

• Speech Perception and Implicit Memory: Evidence for Detailed Episodic Encoding of Phonetic Events
Lorin Lachs, Kipp McMichael and David B. Pisoni 149

• Effects of Speaking Style on the Perceptual Learning of Novel Voices: A First Report
James D. Harnsberger, Richard Wright and David B. Pisoni 169

• Some Effects of Phonotactic Probabilities on the Processing of Spoken Words and Nonwords by Post-Lingually Deafened Adults with Cochlear Implants
Michael S. Vitevitch, David B. Pisoni, Karen I. Kirk, Marcia Hay-McCutcheon and Stacy Yount 189

• Effects of Talker Variability and Lexical Competition on Audiovisual Word Recognition

by Adult Users of Cochlear Implants
Adam R. Kaiser, Karen I. Kirk, Lorin Lachs and David B. Pisoni 219

• PET Imaging of Differential Cortical Activation to Monaural Speech and Nonspeech Stimuli
Donald Wong, David B. Pisoni, Jennifer Learn, Jack T. Gandour, Richard T. Miyamoto and Gary D. Hutchins 247

II. Short Reports and Work-in Progress 271

• Use of Partial Stimulus Information by Cochlear Implant Patients and Normal-Hearing Listeners in Identifying Spoken Words: Some Preliminary Analyses
Lorin Lachs, Jonathan W. Weiss and David B. Pisoni 273

• Talker Discrimination by Prelingually Deaf Children with Cochlear Implants: Some Preliminary Results
Miranda Cleary and David B. Pisoni 289

• Lexical Neighborhood Properties of the Original and Revised Speech Perception In Noise(SPIN) Tests
Constance M. Clarke 305

• Perceptual Adjustments to Foreign Accented English
Constance M. Clarke 321

• Auditory Learning and Adaptation after Cochlear Implantation: A Preliminary Study of Discrimination and Labeling of Vowel Sounds by Cochlear Implant Users
Mario A. Svirsky, Alicia Silveira, Hamlet Suarez, Heidi Neuburger, Ted T. Lai and Peter M. Simmons 337

• Early Word Learning Skills of Hearing-Impaired Children Who Use Cochlear Implants: Development of Procedures and Some Preliminary Findings
Derek M. Houston, Allyson K. Carter, Elizabeth A. Ying, Karen I. Kirk and David B. Pisoni 345

• Reduced, Citation, and Hyperarticulated Speech in the Laboratory: Some Acoustic Analyses
James D. Harnsberger and Lori A. Goshert 357

• Change Deafness: The Inability to Detect Changes in a Talker's Voice
Michael S. Vitevitch 369

• Speech Perception and Language Skills of Deaf Infants After Cochlear Implantation: A Review of Assessment Procedures and a Research Plan
Derek M. Houston 377

• Memory Span and Sequence Learning Using Multimodal Stimulus Patterns: Preliminary Findings in Normal-Hearing Adults
Jeff Karpicke and David B. Pisoni 393

III. Instrumentation and Software 407

• A Multi-Talker Dialect Corpus of Spoken American English: An Initial Report on Development
Cynthia G. Clopper, Allyson K. Carter, Caitlin M. Dillon, James D. Harnsberger, Rebecca Herman, Connie M. Clarke, David B. Pisoni and Luis R. Hernandez 409

IV. Publications: 2000 415

INTRODUCTION

This is the twenty-fourth annual progress report summarizing research activities on speech perception and spoken language processing carried out in the Speech Research Laboratory, Department of Psychology, Indiana University in Bloomington. As with previous reports, our main goal has been to summarize our accomplishments over the past year and make them readily available to granting agencies, sponsors and interested colleagues in the field. Some of the papers contained in this report are extended manuscripts that have been prepared for formal publication as journal articles or book chapters. Other papers are simply short reports of research presented at professional meetings during the past year or brief summaries of "on-going" research projects in the laboratory. From time to time, we also have included new information on instrumentation and software developments when we think this information would be of interest or help to others. We have found the sharing of this information to be very useful in facilitating research.

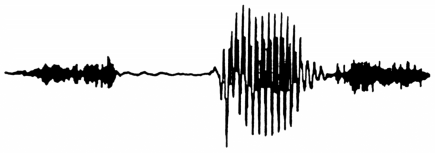
We are distributing progress reports of our research activities because of the ever increasing lag in journal publications and the resulting delay in the dissemination of new information and research findings in the field of spoken language processing. We are, of course, very interested in following the work of other colleagues who are carrying out research on speech perception and spoken language processing and we would be grateful if you and your colleagues would send us copies of any recent reprints, preprints and progress reports as they become available so that we can keep up with your latest findings. Please address all correspondence to:

Professor David B. Pisoni
Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405-1301
United States of America

Telephone: (812) 855-1155, 855-1768
Facsimile: (812) 855-1300
E-mail: pisoni@indiana.edu
Web: http://www.indiana.edu/~srlweb

Copies of this progress report are being sent primarily to libraries and specific research institutions rather than individual scientists. Because of the rising costs of publication and printing, it is not possible to provide multiple copies of this report to people at the same institution or issue copies to individuals. We are eager to enter into exchange agreements with other institutions for their reports and publications. Please write to the above address for further information.

The information contained in this progress report is freely available to the public and is not restricted in any way. The views expressed in these research reports are those of the individual authors and do not reflect the opinions of the granting agencies or sponsors of the specific research.



SPEECH – THE FINAL FRONTIER

SPEECH RESEARCH LABORATORY
FACULTY, STAFF, AND TECHNICAL PERSONNEL

(January 1, 2000–December 31, 2000)

RESEARCH PERSONNEL

David B. Pisoni, Ph.D. ..... Chancellors' Professor of Psychology and Cognitive Science[1,2]

Karen I. Kirk, Ph.D. ..... Associate Professor of Otolaryngology–Head and Neck Surgery[3,4]
Mario A. Svirsky, Ph.D. ..... Associate Professor of Otolaryngology–Head and Neck Surgery[3,5]
Steven B. Chin, Ph.D. ..... Assistant Scientist in Otolaryngology–Head and Neck Surgery[3]

Adam R. Kaiser, M.D. ..... NIH Postdoctoral Trainee[3]
Allyson K. Carter, Ph.D. ..... NIH Postdoctoral Trainee
James D. Harnsberger, Ph.D. ..... NIH Postdoctoral Trainee
Rebecca Herman, Ph.D. ..... NIH Postdoctoral Trainee
Derek Houston, Ph.D. ..... NIH Postdoctoral Trainee
Holly L. Storkel, Ph.D. ..... NIH Postdoctoral Trainee
Michael S. Vitevitch, Ph.D. ..... NIH Postdoctoral Trainee

Constance M. Clarke, M.A. ..... Visiting Research Associate
Sarah H. Ferguson, M.A. ..... NIH Predoctoral Trainee
Laura W. McGarrity, M.A. ..... NIH Predoctoral Trainee
Miranda Cleary, B.A. ..... NIH Predoctoral Trainee
Cynthia G. Clopper, B.A. ..... NIH Predoctoral Trainee
Caitlin M. Dillon, B.A. ..... NIH Predoctoral Trainee
Lorin Lachs, B.A. ..... NIH Predoctoral Trainee
Winston D. Goh, M.Soc.Sci. ..... Predoctoral Trainee[6]

John Weiss, B.S. ..... NIH Medical Student Trainee
William Walsh, B.A. ..... NIH Medical Student Trainee
Jon Hoversland, B.S. ..... NIH Medical Student Trainee

---

[1] Also Adjunct Professor of Linguistics, Indiana University, Bloomington, IN.
[2] Also Adjunct Professor of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.
[3] Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.
[4] Also Adjunct Associate Professor of Speech and Hearing Sciences, Indiana University, Bloomington, IN.
[5] Also Adjunct Associate Professor of Electrical Engineering, Purdue School of Engineering and Technology, Indianapolis, IN.
[6] Also Senior Tutor, Department of Social Work and Psychology, National University of Singapore.

TECHNICAL PERSONNEL

Luis R. Hernández, B.A. ..... Research Associate in Psychology/Systems Administrator
Darla J. Sallee ..... Administrative Assistant

Damon Stewart ..... Graduate Research Assistant
Mark Van Dam ..... Graduate Research Assistant

Jeff Karpicke ..... Undergraduate Research Assistant
Corey Yoquelet ..... Undergraduate Research Assistant

E-MAIL ADDRESSES

Allyson K. Carter ..... allcarte@indiana.edu
Steven B. Chin ..... schin@iupui.edu
Constance M. Clarke ..... clarke@u.arizona.edu
Miranda Cleary ..... micleary@indiana.edu
Cynthia G. Clopper ..... cclopper@indiana.edu
Caitlin M. Dillon ..... cmdillon@indiana.edu
Sarah H. Ferguson ..... safergus@indiana.edu
Winston D. Goh ..... wigoh@indiana.edu
James D. Harnsberger ..... jharnsbe@indiana.edu
Rebecca Herman ..... rebecca_h76@yahoo.com
Luis R. Hernández ..... hernande@indiana.edu
Derek Houston ..... dmhousto@indiana.edu
Adam R. Kaiser ..... arkaiser@iupui.edu
Jeff Karpicke ..... karpicke@indiana.edu
Karen I. Kirk ..... kkirk@iupui.edu
Lorin Lachs ..... llachs@indiana.edu
Laura W. McGarrity ..... lmcgarr@indiana.edu
David B. Pisoni ..... pisoni@indiana.edu
Darla J. Sallee ..... dsallee@indiana.edu
Holly L. Storkel ..... hstorkel@indiana.edu
Mario A. Svirsky ..... msvirsky@iupui.edu
Michael S. Vitevitch ..... mvitevit@indiana.edu

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Working Memory Spans as Predictors of Spoken Word Recognition and Receptive Vocabulary in Children with Cochlear Implants [1]

Miranda Cleary, David B. Pisoni,[2] and Karen Iler Kirk[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Research Grant DC00111 and NIH-NIDCD Training Grant DC00012 to Indiana University and by NIH Research Grant DC00064 to the Indiana University School of Medicine.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Working Memory Spans as Predictors of Spoken Word Recognition and Receptive Vocabulary in Children with Cochlear Implants

Abstract. The present study investigated whether individual differences in working memory could account for a significant proportion of the variance in prelingually-deafened pediatric cochlear implant users' open-set word recognition and receptive vocabulary skills, after the contribution of known predictors was taken into account. The contributions of four different measures of working memory were examined separately for Oral (N=32) and TC (N=29) children. WISC digit-spans, requiring immediate recall of auditory-only lists in both forwards and backwards (reversed order) directions were collected. Two versions of a novel "memory span game" were also administered: one required memory for sequences of colored lights, the other assessed memory for colored lights presented in conjunction with auditory color-names. The contribution of working memory differed depending on the particular memory span task and dependent measure being considered. Forward digit-span accounted for a significant amount of additional variance in open set word recognition scores for both the Oral and TC groups. Backwards digit-span accounted for additional variance in word recognition scores only for the TC group. In the case of receptive vocabulary as measured using the PPVT administered in the child's preferred communication mode, forward digit-span accounted for 27% in additional variance in the vocabulary scores of the Oral group, but very little additional variance in the vocabulary scores of the TC group. Backwards digit-span showed a small contribution to the receptive vocabulary scores of both groups. Scores in the "lights-plus-sound" version of the memory game accounted for 5% in additional variance in PBK word recognition and 5% in receptive vocabulary, but only for the Oral group. The "lights-only" condition accounted for no additional variance in either word recognition or vocabulary scores. The pattern of results suggests that the observed relationships between working memory and outcome measures are specific to the auditory modality, partially linked to communication mode, and not related to individual differences in a general-purpose component of working memory.

Introduction

Children who have lost their hearing at an early age and who have learned spoken language while using a cochlear implant show a great deal of variability in the speech and language skills they eventually attain. Some children benefit a great deal from use of their implant, developing excellent auditory-only listening skills and highly intelligible speech. Other children with similar medical histories do not show such progress. The factors that contribute to this wide variability in outcome are not yet fully understood. Although approximately 37-64% of the existing variance in outcome measures can be accounted for in terms of known demographic variables such as duration of deafness, length of device use, and age at implantation, there is still a large proportion of unexplained variability that remains (Blamey et al., 2001; Dowell, Blamey, & Clark, 1995; Miyamoto et al., 1994; Sarant, Blamey, Dowell, Clark, & Gibson, 2001; Snik, Vermeulen, Geelen, Brokx, & van den Broek, 1997).

Recent studies have begun to investigate the hypothesis that individual differences in working memory may contribute to the wide variability in outcome (e.g., Pisoni, 2000; Pisoni, Cleary, Geers, & Tobey, 2000; Pisoni & Geers, 2000). In particular, this line of research has targeted a potential role of auditory working memory—a short-term storage/maintenance mechanism for auditory verbal information. Pisoni and colleagues have argued that children with cochlear implants are an important

population to study in order to learn more about the role of early sensory experience on the development of working memory and its theorized component subsystems used to process input from different modalities.

The present study was motivated by the hypothesis that the role played by working memory may differ in fundamental ways depending on whether the implanted child develops language in an Oral versus a Total Communication environment. Children designated as "oral" are typically immersed in environments in which spoken language is used without the addition of manual sign language (Archibold, 2000). In contrast, children who use "total communication" utilize a combination of spoken language and manual sign language in their everyday communication (Archibold, 2000). The manual signs used in total communication are primarily supplements to spoken language and thus are not American Sign Language proper, but rather a form of Signed English which uses supplemental signs to facilitate communication of certain aspects of spoken English that are especially difficult for hearing-impaired persons to perceive. Given this background, we hypothesized that children who primarily use oral communication may rely more heavily on auditory memory processes than children who use total communication. In the analyses reported below, we examined whether the method of communication used by the child affects the degree to which the working memory measures were able to account for the remaining unexplained variance in language outcome measures.

Several different working memory measures were obtained in this study in order to determine if the proposed contribution of working memory is specific to auditory working memory, and not, for example, related simply to some "general purpose" modality-independent component of the working memory system (Baddeley, 1992; 1998). The four different memory span tasks used in this study differed in the degree to which they incorporated an auditory processing component. The four tasks included auditory-only digit-span requiring recall of items in the forward direction, auditory-only digit-span requiring recall of items in reversed order, memory span for visual sequences of colored lights, and memory span for sequences of colored lights presented in conjunction with auditory color-names matched to the light sequence. If the contribution of working memory were only observed for the memory tasks that require auditory processing, and not for the "lights-only" visual sequences, this finding would suggest that the effects are due to phonological processing components of working memory and not some general purpose, modality-independent attentional or control component.

The inclusion of working memory measures for both visual sequences of colored lights, and sequences of colored lights presented in conjunction with auditory color-names matched to the light sequence was motivated by previous research from our lab showing that children with cochlear implants benefit less in a "memory game task" from the presentation of the informationally redundant auditory color-names than do normal-hearing children of the same age (see Cleary, Pisoni, & Geers, in press). Even pediatric CI users who were able to accurately identify the recorded color-name stimulus tokens used in this task when these tokens were presented in isolation, demonstrated a smaller "redundancy gain" between the uni-modal "lights-only" condition and the multi-modal "color-names plus lights" condition than did an age- and gender-matched group of normal-hearing eight- and nine-year old children. In the present study, we sought to replicate this result in a new sample of children from a different testing center and to examine whether the child's communication mode affects the size of the observed redundancy gain. This aspect of the study, however, was secondary to the primary goal of using linear regression models to determine the degree to which measures of working memory accounted for variance in language outcome measures not already explained by "traditional" demographic predictors such as age at implantation, duration of use of the implant, and unaided PTA threshold prior to use.

Although there are numerous different outcome measures we might have selected, we chose auditory-only word recognition and receptive vocabulary administered via the child's preferred

communication mode as our two principal measures. The spoken word recognition and receptive vocabulary measures examined in this report may appear similar in that both are based on the processing of single lexical items, however, the two tasks differed in several potentially important ways. The word recognition tasks used in this study required the children to repeat isolated words presented to them using an auditory-only presentation format. To complete the word recognition tasks, it was necessary that the child both perceives the item and reproduces it accurately enough for the examiner to recognize it as an attempted version of the target. The ability of the child to produce an accurate imitation did not necessarily indicate, however, that the child knew the meaning of the test item since both normal-hearing and some hearing-impaired children have been shown to be able to do this same repetition task with unfamiliar nonsense words that obey the phonotactic characteristics of the child's language (Dillon & Cleary, this volume). By comparison, the receptive vocabulary task used in this study minimized the role of articulatory production by requiring the child to demonstrate knowledge of word meaning by pointing to an appropriately matched picture from a closed set of response alternatives. Since all vocabulary target items were presented via each child's preferred communication mode (oral language or total communication), this measure essentially reflects linguistic knowledge irrespective of the sensory modality by which it was acquired.

Two different word recognition tests were included in this study in order to investigate whether key differences in test administration might impact on the results of the regression analysis. The PBK word identification task (Haskins, 1949) was administered in this study using a clinician's live-voice presentation of each word. The version of the Lexical Neighborhood Test (Kirk, Pisoni, & Osberger, 1995) used in this study employed recorded tokens of each target word. These tokens were spoken by several different talkers, such that the child heard the voices of five different talkers over the course of each set of words. In order to do well on this test, a child must be able to cope with the cross-talker variability present in the speech signal. We were interested in what impact, if any, the use of the live-voice PBK versus the recorded multi-talker LNT would have on our assessment of the role of working memory in the two groups of children.

Method

Data Selection

All data included in this report were collected as part of a larger longitudinal study currently being conducted at the Indiana University School of Medicine in Indianapolis. Data using the new memory game span task were collected in total of 85 test sessions. Five visits involving hearing aid users, thirteen visits involving postlingually deafened children (onset of hearing loss after age three years), and two visits involving data collected before the child had begun to use his/her implant were eliminated from the data set. Two more visits were eliminated due to lack of memory scores in at least one condition along with missing recorded LNT scores. A further two visits were eliminated for missing both PBK and recorded LNT scores. Data from three children (one Oral, and two TC) who only lacked recorded LNT scores were retained in the analysis. Thus, data from 61 visits remained for our final analysis.

Forty-five of the sixty-one visits involved data gathered from a child completing the memory game task for the first time. Twelve of the remaining sixteen visits yielded data collected from children who had been tested once before on the memory game task and whose first set of data from this task was already included in the data set. The remaining four data points came from children who were being tested for the second time on the memory game task, but whose first data set was eliminated from consideration for one of the reasons already listed above. There were therefore 49 different children examined in this sample.

In our initial set of statistical analyses we decided to treat all sixty-one visits as independent cases. In actuality, since 12 of the visits were repeated visits from children already represented in the sample, this inclusion violates an assumption that the cases are independently sampled. However, discarding this data was not an attractive option either since it would reduce the power of the study. An argument for retaining the twelve repeated cases is that an interval of one year separated the two visits and because of this, the associated values of "device use" and "age of child" are necessarily different between the two visits, as are also, in most cases, the level of performance on the three dependent measures. Thus, including a particular child "twice" is not simply counting the "same" case twice in the analysis. We have conducted the multiple linear regression analyses both with and without discarding the 12 visits that were repeated measures from children already represented in the sample and obtained almost identical results, thus assuring us that the violation of independence assumption had little if any impact on the final results. Thus, in the rest of this report, the values for "N" refer to the "visit" sample size, unless otherwise indicated.

Participant and Device Characteristics

A summary of participant characteristics is shown in Table 1. Of the 49 children who were included in the study, 27 primarily used oral communication. These children contributed a total of 32 visits to the analysis. Twenty-two of the children used total communication and a total of 29 visits resulted from their participation. The children ranged in age from 5.2 years to 16.5 years at time of testing. The mean age at testing over the 61 visits was 9.16 years. Mean age at time of testing did not differ significantly between the Oral and TC groups.

PARTICIPANT CHARACTERISTICS	Group Mean (SD)		Min		Max		Group Means Significantly different by independent samples t-test?
	Oral (N=32)	TC (N=29)	Oral	TC	Oral	TC	
Age at Testing in Years	9.16 (2.48)	9.16 (2.74)	5.2	5.9	15.6	16.5	ns
Age at Onset of Deafness in Years	.33 (.74)	.44 (.86)	0	0	2.4	3.0	ns
Duration of Deafness in Years	4.35 (2.27)	3.61 (1.72)	0.5	0.8	9.6	8.9	ns
Age at Implantation in Years	4.69 (1.95)	4.06 (1.59)	1.4	2.2	10	8.9	ns
Duration of Implant Use in Years	4.47 (1.70)	5.10 (2.39)	0.1	1.6	9.0	11.6	ns
PTA Threshold Pre-Implantation	109.12 (6.03)	113.21 (5.23)	96.7	103.3	118.4	120.1	$t(59) = 2.81$, $p = .007$

Table 1. Descriptive statistics for demographic variables in the Oral and TC groups. The participant characteristics in boldface were included as predictor variables in the regression analyses reported below.

In 25 of the 32 cases in the Oral group, and 21 of 29 TC cases, the child was reported as congenitally deaf. Three cases in each of the Oral and TC groups were from children with an onset of deafness in the first year of life. Children deafened after age 1 year, but prior to age 3 years comprised four cases in the TC group and five cases in the Oral group. The two groups are clearly also comparable with respect to this variable. Although we initially considered using age at onset of severe to profound

deafness as a covariate in our analyses, because these data were severely skewed and limited in range, we decided not to do so.

The children in both groups had been deaf for an average of about four years prior to implantation. Age at implantation was approximately four years of age, on average, in both groups. The Oral group had used an implant for a mean duration of about four and a half years at time of testing, while the TC group averaged a little over five years of implant use. Statistical tests indicated that the groups did not differ significantly in terms of duration of deafness, age at implantation, or duration of implant use. The two groups did, however, differ significantly on their unaided pure tone average (PTA) threshold averaged over responses at 500, 1000, and 2000 Hz, measured prior to implantation. Although all children had a profound hearing loss, the TC group in the present study had somewhat poorer thresholds than the Oral group. This difference supported inclusion of PTA thresholds as one of the predictor variables in our analyses.

The etiology of hearing loss in 35 of the 49 individual children was unknown. Known etiologies consisted of 7 meningitis cases, 4 genetically related cases, 2 cytomegalovirus cases, and 1 case of Mondini deformation. Table 2 provides information regarding the children's devices, speech processors, and coding strategies for both the Oral and TC groups. Examination of Table 2 reveals that the two groups were quite comparable on characteristics related to the implant itself. Establishing the comparable nature of the Oral and TC groups was important to our selection of analyses and their interpretation.

DEVICE CHARACTERISTICS	Oral (N=32)	TC (N=29)
Device Type	# of cases	# of cases
Nucleus 22	26	29
Nucleus 24	1	0
Clarion	5	0
Speech Processor Type	# of cases	# of cases
SPECTRA - Cochlear	26	25
MSP - Cochlear	0	4
Sprint - Cochlear N24	1	0
1.2 - Clarion	4	0
S - Clarion	1	0
Coding Strategy	# of cases	# of cases
SPEAK	26	25
CIS	4	0
MPEAK	0	4
SAS	1	0
ACE (CIS+SPEAK)	1	0

Table 2. Device characteristics for the Oral and TC groups.

Stimuli and Procedures

All measures were gathered by experienced speech-language pathologists or audiologists during each child's annual follow-up visit to the clinic at Riley Hospital, or in a testing room at the child's school (e.g., St. Joseph's Institute for the Deaf in Missouri), over the course of several years. At the end of this period, all children for whom memory measures were available were selected from a larger database of children according to criteria outlined earlier.

Dependent Measures

Open-set Word Recognition. Each child was administered two open-set tests of word recognition, the PBK and the LNT. Both tests require the child to repeat an auditorily-presented monosyllabic word back to the testing clinician. Scoring was done on-line as the child was tested and slight distortions in the reproduction of single phonemes within each test word were scored as correct. Although both word recognition tests were designed for use with children, caregivers of children with cochlear implants rate the words on the PBK as less familiar to their young children (ages 3-8 years) than the words used in the LNT (Kirk, Sehgal, & Hay-McCutcheon, 2000).

The Phonetically Balanced Kindergarten (PBK) test is an open-set test of word recognition using monosyllabic words presented in isolation (Haskins, 1949). Although there are four available word lists, only three of these lists are typically used (see Meyer & Pisoni, 1999). Each child was tested using one of these three lists. Presentation of the target items was carried out using auditory-only live-voice presentation with the face of the clinician concealed behind a mesh screen. In the interest of time, clinicians had the option of only administering 25 items in each 50-word list. Children who made no responses in the first 10 items were given a score of zero. Although both percent phonemes-correct and percent words-correct are usually tabulated, we chose the percent words-correct score for the present analyses.

The Lexical Neighborhood Test (Kirk, Pisoni, & Osberger, 1995) consists of one hundred monosyllabic words divided into four lists of twenty-five words each. Two of the lists contain words that are "lexically easy" (i.e., are phonetically similar to very few other words) and two of the lists contain words that are "lexically hard" (i.e., are phonetically confusable with many other words). A digitally recorded version of this test using multiple talkers has been devised (Kirk, 1998; Kirk, Hay-McCutcheon, Sehgal, & Miyamoto, 2000). In this form of the test, a child is tested on one "easy" word list and one "hard" word list, where the voice of the talker uttering the words may vary between five different talkers, three female and two male. Separate scores between 0-100% are typically generated for the "easy" list versus the "hard" word list. Because we wished to obtain a roughly normal, continuously distributed set of scores, and were not planning to look for differences based on lexical discriminability, we averaged the percent words-correct on both twenty-five-word tests for a composite LNT score.

Receptive Vocabulary. The Peabody Picture Vocabulary Test-Third Edition (PPVT) (Dunn & Dunn, 1997) was administered to each child using his/her preferred communication mode. The PPVT is a receptive measure of vocabulary development that requires the child to correctly point to one of four line drawings, in this case, after hearing a word spoken, or both spoken and signed, by the examiner.

Predictor/Independent Variables

Age at implantation, duration of implant use, PTA threshold, and communication mode were determined by consulting the medical charts for each child on file at the Indiana University Medical Center.

Memory Span Measures. WISC digit-span was administered using auditory-only live-voice presentation with the face of the clinician hidden behind a mesh screen. Administration and scoring followed the procedures provided for this particular subtest in the testing manual for the Wechsler Intelligence Scale for Children, Third Edition (WISC-III) (Wechsler, 1991). In the "digits-forward" section of the digit-span task, the child is required to simply repeat back the list of digits as heard. In the "digits-backwards" section of the task, the child is told to "say the list backward." In both parts of the

WISC task, testing begins with lists of two items. If at least one of the two lists provided at each list length is successfully repeated, the next list length is increased by one digit until the child gets both lists incorrect at a given length, at which point testing stops. Points are awarded for each list correctly repeated with no partial credit.

An extensive description of the memory game procedure used in this study can be found in Cleary, Pisoni, and Geers (in press). The memory game task is used to obtain a measure of working memory for sequences of either visual-spatial cues or visual-spatial cues paired with auditory signals. The auditory stimuli used in the memory game task were created by recording a male speaker of American English saying the words "red," "blue," "green," and "yellow" at a moderate to slow rate of speech. Each word was spoken in isolation. The durations of the stimuli were not artificially equated, but were retained in their original form. The color-names ranged between 360 ms to 400 ms in length. The recordings were digitally sampled on-line at 22.05 kHz with 16-bit amplitude quantization and the average RMS amplitudes of the individually edited speech files were approximately equated using a digital leveling procedure.

Before the memory game was administered, each child was asked to identify all four recorded tokens by pointing to the colored button matching the color-name. The stimulus tokens were presented one at a time through the same loudspeaker as was used for the memory game (Advent AV570). The presentation level was approximately 70 dB SPL as determined via a hand-held sound level meter held at the level of the child's head. If a child correctly identified all four items in a set on the first attempt, no further identification testing was administered. If one or more errors were made, the identification task was repeated up to three times, or until zero mistakes were made on a given set of stimuli, whichever occurred first.

Presentation of the test sequences was controlled by a computer program specially created for this purpose, running on a PC computer. The response box used to collect the child's button presses consisted of a large round disk-like plastic case housing four wide plastic buttons on its surface that are easily depressed by a child. Each button was made of a different color plastic and could be illuminated by a light located beneath its surface. The colors of the buttons matched the color-names that were recorded as stimuli. The button response box was interfaced to the PC computer so that the control program could illuminate the lights when the sound stimuli were played, and turn off the lights once the stimuli ceased outputting. In the lights-only presentation condition, the control program illuminated each light for the same stimulus duration as used in the color-name presentation condition. The computer recorded all button presses and automatically tracked the subject's performance using an adaptive testing procedure described below.

Participants were shown how the buttons on the response box could be pressed and were told that they would be hearing sounds through the loudspeaker and seeing the buttons light up. They were then instructed to "pay attention and copy exactly what the computer does by pressing on the buttons." The stimulus sequences used for the memory game task were generated pseudo-randomly by a computer program, with the stipulation that no single item would be repeated consecutively in a given list. A very brief inter-stimulus interval of 200 ms was used between sequence items. However, since the individual stimuli had been recorded at a relatively slow speech rate, the rate of presentation was actually about 1.67 items per second. Each child started with a list length of one item. If two lists in a row at a given length were correctly reproduced, the next list presented was increased by one item in length. If on any trial the list was incorrectly reproduced by the child, the next trial used a list that was one item shorter in length. Each child was presented with twenty lists to reproduce under each condition. After completing the twenty lists in a given condition, the child was assigned a span score calculated by summing the proportion of lists correctly reproduced at each list length tested.

Although the experimenter provided no explicit feedback regarding the accuracy of the child's responses, whenever the child pressed a button during the response period the button was illuminated and the appropriately mapped sound was played out. The color-name plus lights condition was always administered first, followed by the lights-only condition. Each presentation condition of the memory game task took approximately four minutes to complete.

Planned Analyses / Proposed Linear Regression Model

Three multiple linear regression analyses were planned. The dependent (predicted) measures were either open set word recognition measured via the PBK, open set word recognition measured via the LNT, or receptive vocabulary development as measured by the PPVT administered in the child's preferred communication mode. The independent, predictor variables were chosen based on prior research findings indicating that several demographic factors appear to play a substantial role in determining prelingually-deafened children's success with a cochlear implant. Chronological age at time of cochlear implantation, duration of implant use, and residual hearing as measured by PTA threshold prior to implantation were all included as predictor variables for this reason. Skewedness and kurtosis values for each of the three variables are shown for both groups of children in Table 3 in order to demonstrate the suitability of these measures for psychometric analysis. Given the relatively small number of observations per group, use of additional predictor variables was judged to be inadvisable.

	ORAL (N=32)	
	Skew	Kurtosis
Age at Implantation	0.55	0.17
Duration of Implant Use	0.00	1.17
PTA Threshold Pre-Implantation	-0.37	-0.58
	TC (N=29)	
	Skew	Kurtosis
Age at Implantation	1.34	1.99
Duration of Implant Use	1.13	1.15
PTA Threshold Pre-Implantation	-0.38	-0.89

Table 3. Skew and kurtosis values for each demographic predictor variable.

Results and Discussion

Colinearity Issues

Potential problems of multi-colinearity were examined by calculating simple bivariate correlations among the three demographic predictor variables. Age at implantation and duration of implant use were not significantly correlated with each other in either group (Oral: r = - .08, TC: r = -.10). For the TC group, PTA threshold pre-implant was not significantly correlated with age at implantation or with duration of implant use (r = - .29, p = .13, and r = .23, p = .22, respectively). For the Oral group, however, PTA threshold pre-implant was found to be significantly correlated with age at implantation and duration of use: the correlation of PTA and age at implantation was r = - .40, p = .024), while the correlation between PTA and duration of implant use was r = .43, p = .014. Since both of these values were moderate correlations, below the suggested r of .50 for exclusion or combination of predictor variables, we retained all three variables in the regression analyses for the Oral group. In general, the

pattern of obtained correlations indicates that for both groups, children who were implanted at an older age had slightly more residual hearing (lower thresholds) on average, perhaps because children with more residual hearing often undergo a longer trial period with hearing aids before resorting to cochlear implantation. These correlations also suggest that children with a longer history of CI use (who received implants earlier than their peers) tended to have less residual hearing (higher thresholds). This latter relationship makes sense given changes in candidacy requirements for pediatric CI users over the last decade.

Role of Chronological Age

Chronological age at time of testing posed its own unique problem because it varied widely in our samples. Simply including it as another predictor variable was not appropriate for several reasons. First of all, age at time of testing is completely redundant with the combined information from age at implantation and duration of CI use (age at test = age at implantation + duration of CI use). Therefore, age at testing cannot be meaningfully included as a predictor variable if age at implantation and duration of CI use are also to be used as predictors. We therefore decided that we would first determine the amount of variability in the dependent measures that was accounted for by age at implantation, duration of implant use, and PTA thresholds prior to considering the contribution of working memory, and then interpret this intermediate result as necessarily reflecting the contribution of chronological age--without specifically being able to separate its effects from those of the other demographic predictors.

As shown in Table 4, chronological age was also significantly correlated with all four measures of working memory. In addition, although raw PPVT vocabulary scores were, as might be expected, strongly correlated with age, PBK and LNT open-set word recognition scores were unrelated to age at testing. Thus, although the percent of variance in PPVT scores accounted for by age at implantation, duration of use, and PTA threshold will also reflect the contribution of chronological age, very little contribution from chronological age should be reflected in the amount of variance in word recognition scores accounted for by the three demographic predictor variables.

CORRELATIONS WITH CHRONOLOGICAL AGE AT TESTING		
	Oral (N=32)	TC (N=29)
Predictor Variables		
Age at Implantation	.73***	.50**
Duration of Implant Use	.62***	.82***
PTA Threshold	-.02	.04
Potential Predictor Variables		
Memory Span Game Lights	.56**	.54**
Memory Span Game Colors	.66***	.49**
WISC Digit-span Forward	.34	.56**
WISC Digit-span Backwards	.63***	.71***
Outcome Measures		
Word Recognition – PBK	.11	-.11
Word Recognition – LNT	.03	-.12
Vocabulary – PPVT Raw Score	.74***	.84***

*** $p < .001$, ** $p < .01$, * $p < .05$, uncorrected p-values.

Table 4. Correlations with chronological age at time of testing

Group Mean Performance on Outcome Measures

Mean scores for the word recognition and receptive vocabulary measures are shown in Table 5 for the Oral and TC groups. The spread of scores was quite similar in both groups for each of the three measures. Although the Oral children, as a group, did significantly better than the TC group on PBK open-set word recognition, the groups did not differ reliably on LNT open-set word recognition using recorded tokens from multiple talkers or on PPVT receptive vocabulary administered in the child's preferred communication mode.

DEPENDENT / OUTCOME MEASURES	Group Mean (SD)		Min Score		Max Score		Group Means Significantly different by independent samples t-test?
	Oral (N=32)	TC (N=29)	Oral	TC	Oral	TC	
PBK Word Recognition Percent Correct	45.50 (23.05)	30.62 (20.09)	8	0	88	68	$t(59) = 2.68$, $p = .01$
LNT Word Recognition Percent Correct	48.77 (21.81)	40.30 (21.49)	0	2	82	80	ns
PPVT Receptive Vocabulary Raw Score	88 (34)	90 (29)	20	37	172	169	ns

Table 5. Descriptive statistics for the three outcome measures, for Oral and TC groups

Variance Accounted for by Traditional Demographic Predictors

The statistical package SPSS was used to conduct the linear regression analyses. First, we assessed the contribution of the traditional demographic predictors using the forced-entry option for entering variables into the regression model. The amount of variance accounted for by the traditional predictors is shown in Table 6.

For the Oral group, the traditional predictors alone accounted for a significant portion of the variance in PBK, LNT, as well as PPVT scores (PBK: $F(3,28) = 4.15$, $p = .015$; LNT: $F(3,27) = 3.73$, $p = .023$; PPVT: $F(3,28) = 12.7$, $p < .001$). In contrast, for the TC group, the picture differed in several important ways. The pre-included predictors failed to account for a significant amount of the variance in the word recognition measures (PBK: $F(3,25) < 1$; LNT: $F(3,23) < 1$), but did account for a significant amount of variance in PPVT scores ($F(3,25) = 25.67$, $p < .001$).

This intermediate set of results indicates that with regards to open-set word recognition, the traditional predictor variables behaved much as expected for the Oral group. However, for the TC group, these variables were strikingly ineffective as predictors for word recognition performance. For the PPVT, a measure of receptive language that was administered using the child's preferred communication mode, the traditional predictor variables accounted for a very large amount of the variance in these scores for the TC group, and a slightly smaller, but still substantial amount of variance in the Oral group.

PBK	ORAL (N=32)		TC (N=29)	
	Percent of variance accounted for by demographic predictors	Additional percent of variance accounted for	Percent of variance accounted for by demographic predictors	Additional percent of variance accounted for
Digit-span Forwards	30.80%	16.80%	3.20%	44.80%
Digit-span Backwards		1.80%		17.30%
Memory Game Colors+Lights		5.40%		3.20%
Memory Game Lights-Only		0.20%		0.30%
LNT	**ORAL (N=31)**		**TC (N=27)**	
	Percent of variance accounted for by demographic predictors	Additional percent of variance accounted for	Percent of variance accounted for by demographic predictors	Additional percent of variance accounted for
Digit-span Forwards	29.30%	14.40%	3.80%	27.00%
Digit-span Backwards		0.00%		5.30%
Memory Game Colors+Lights		0.70%		0.20%
Memory Game Lights-Only		0.10%		1.10%
PPVT	**ORAL (N=32)**		**TC (N=29)**	
	Percent of variance accounted for by demographic predictors	Additional percent of variance accounted for	Percent of variance accounted for by demographic predictors	Additional percent of variance accounted for
Digit-span Forwards	57.60%	26.70%	75.50%	3.20%
Digit-span Backwards		6.20%		5.00%
Memory Game Colors+Lights		5.20%		0.40%
Memory Game Lights-Only		0.10%		0.00%

Table 6. Percent of variance in PBK, LNT, and PPVT scores accounted for by the traditional demographic predictors, and the additional variance accounted for by each of the four working memory measures.

Additional Variance Accounted for by Working Memory Measures

As the next step in our analyses, we assessed the contribution of individual differences in the four different working memory measures by retaining the traditional predictors in the model, adding one memory measure as a predictor, and then recalculating the percent of variance accounted for using the forced-entry option. This was done four times per dependent measure, once for each type of memory measure considered individually. Table 6 lists the amount of additional variance accounted for by each memory measure.

Forward digit-span accounted for an additional 17% of the total variance in the Oral group's PBK open-set word recognition scores, and 14% in their LNT open-set word recognition scores. For the TC group, for whom the traditional predictors had proved rather ineffective, forward digit-span accounted for a surprisingly large 44.8% in additional variance in PBK scores and a similarly large 27% in additional variance in LNT scores. These results support the proposal that the forward digit-span and open-set word recognition tasks share a common processing component. Whether this commonality is at the level of

early perceptual identification, phonological rehearsal, or retrieval from phonological working memory cannot, however, be determined from the present data.

For receptive vocabulary as measured via the PPVT, the observed relationship was somewhat different. Forward digit-span accounted for a sizable 26.7% in additional variance for the Oral group, but only 3.2% in additional variance for TC group. These values are consistent with the hypothesis that forward digit-span is a strong predictor of outcome when the administration format of the outcome measure requires auditory encoding, as was the case for the Oral group, but not the TC group.

Backwards digit-span accounted for virtually no additional variance in either word recognition measure for the Oral group. This finding suggests that the comparatively sophisticated explicit sequence manipulation strategies that can be used to advantage on the backwards digit-span task bear little relation to the skills required for simple repetition of auditory stimuli. This proposal is not fully consistent, however, with the finding that for the TC group, backwards digit-span was able to account for a moderate amount of additional variance in word recognition scores, 17.3% for PBK scores and 5.3% for LNT scores. As addressed further in the General Discussion, reduced audibility of the digit-names may have contributed to this finding in the TC group.

For both groups of children, Oral and TC, backwards digit-span accounted for about 5-6% of additional variance in receptive vocabulary scores. These results are similar to values we have previously obtained using the same two tasks with eight- and nine-year-old normal-hearing children. In normal-hearing children, backwards digit-span accounts for approximately 7% of the variance in receptive vocabulary, when chronological age is partialled out. We believe this finding reflects a tendency for children who possess above-average linguistic abilities relative to their age group to also exhibit more sophisticated explicit sequence manipulation strategies.

Results obtained using the memory game task showed smaller contributions to the total variance than those observed using digit-spans. Scores for lights-plus-sound presentation condition of the memory game accounted for ~5% in additional variance in PBK word recognition scores and PPVT scores for the Oral group, and similar but somewhat smaller amounts of variance in the TC group. Scores for the lights-only presentation condition of the memory game accounted for almost no additional variance in any of the dependent measures for either group. Thus, only when the memory game included an auditory component did this measure account for any additional variance in the dependent measures.

Analysis of the memory game's contribution to LNT scores did not follow our expectations, particularly with regard to the Oral group of children. The memory game, even in the lights-plus-sound condition, failed to account for any significant additional variance in LNT scores. The LNT and PBK test administrations did, however, differ in some important ways—the LNT used recorded tokens from multiple-talkers while the PBK was administered using monitored live-voice by a clinician. Although LNT and PBK scores were highly correlated with each other $r = +.83$, it appears that the ability to deal with cross-talker variability as measured by the LNT is not well predicted by performance in the memory game presentation conditions used in the present study.

Group Mean Performance on Working Memory Tasks

Group mean performance for the forward and backwards digit-span tasks is shown in the two top panels of Figure 1. Scores for the Oral group are shown in the top left panel. Scores for the TC group are shown in the top right panel. Not surprisingly, for both groups, backwards digit-span scores were consistently lower than forward digit-span scores. Although the size of this difference was slightly larger for the Oral group than for the TC group (mean difference = 2.1 points vs. 1.7 points, respectively), the

groups did not differ reliably from each other on this measure. Similarly, although both forward and backwards digit-span scores were higher on average for the Oral group than for the TC group, the differences between the groups did not reach statistical significance.

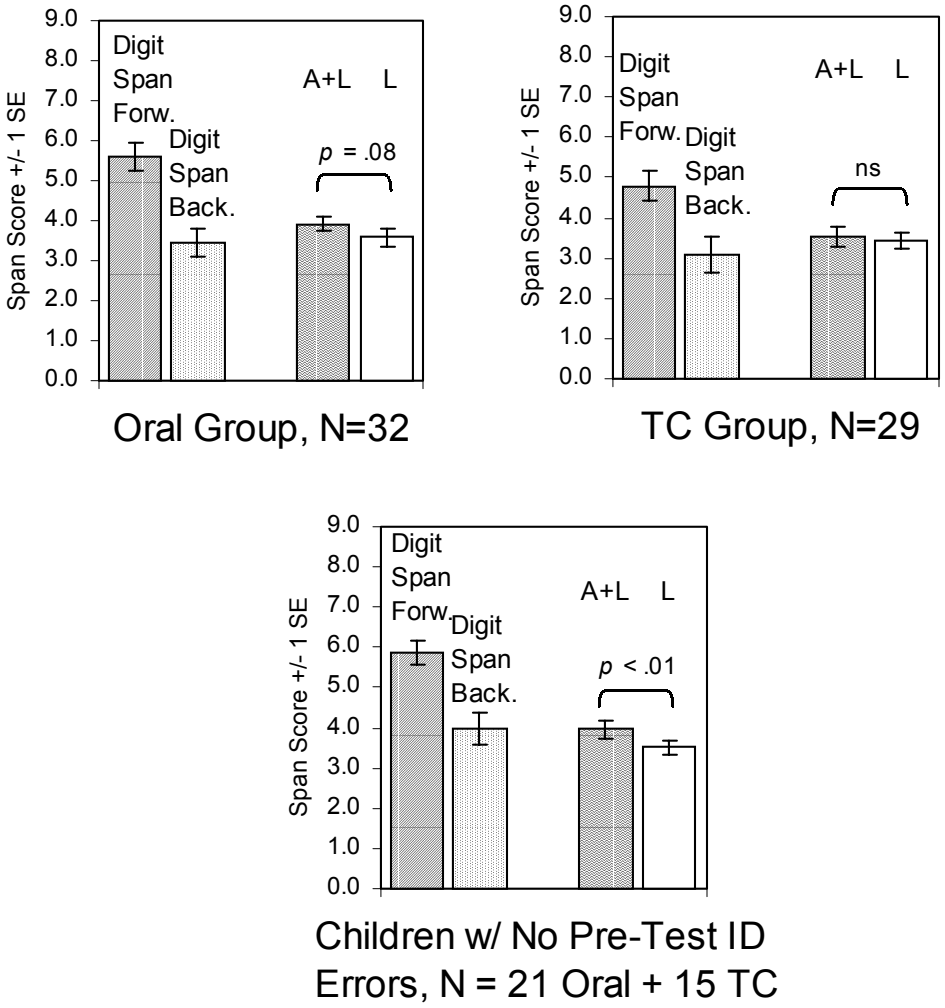


Figure 1. Mean performance in the four working memory tasks: WISC forward digit-span, WISC backwards digit-span, the memory game task in the auditory-plus-lights presentation condition (A+L), and the memory game task in the lights-only presentation condition (L).

The top two panels of Figure 1 also show the mean span scores obtained in the present study for each memory game condition ("A+L" and "L") for both the Oral and TC groups. Comparing first across groups, we found that although both presentation condition means were somewhat higher for the Oral group than for the TC group, the within-group variability was too large for these differences to be statistically reliable. Next we examined within-group differences on the two memory game presentation conditions. As previously reported, in prior research with normal-hearing eight- and nine-year olds, we found that on average, better performance is obtained in the multi-modal "auditory-plus-lights" presentation condition of the memory game ("A+L") than in the "lights-only" presentation condition

("L") (Cleary, Pisoni, & Geers, in press). This "redundancy gain" has been attributed to normal-hearing children's ability to make use of the additional auditory information to facilitate their encoding and memory of the sequences in this task. Cleary, Pisoni, and Geers also found that prelingually-deaf eight- and nine-year old children with cochlear implants, who had at least four years experience with their device at time of testing, showed smaller redundancy gains from the addition of auditory cues to visual-spatial sequences than did age- and gender-matched normal-hearing children.

For the Oral children in the present study, although the mean span in the auditory-plus-lights condition was longer, as expected, than in the lights-only condition, this difference did not quite reach statistical significance ($t(31) = 1.83$, $p = .077$, mean difference = +.34 units). These results suggest, however, that the Oral group made some use of the redundant auditory stimuli. The children in the TC group displayed virtually no difference between the two conditions ($t < 1$, mean difference = +.11 units), indicating that, as a group, they were not able to benefit from the redundant auditory signals provided and relied primarily on visual-spatial encoding to carry out the sequence reproduction task in both conditions of the memory game.

In one final analysis, we analyzed only the performance of children (both Oral and TC), for whom we were able to verify that the child could correctly identify all four auditory color-names when these stimulus tokens were played in isolation during a "stimulus identification pre-test." For 36 of the 61 available visits, no identification errors were made by the child. For twenty visits, this data was unavailable. In the remaining five visits, one or more identification errors were observed. When only the 36 visits with no identification errors were analyzed, we found evidence for a significant redundancy gain in the memory game task ($t(35) = 2.75$, $p = .009$, mean difference = +.43 units). This result demonstrates that for the children who were consistently able to identify the auditory stimuli when presented in isolation, the presence of the informationally redundant auditory color-names resulted, on average, in improved performance in memory span, relative to the lights-only presentation condition. Of these 36 visits, 21 were contributed by children using oral communication and 15 from children using total communication. These results are shown in the lower panel of Figure 1.

Partial Reanalysis of Regression Results

Based on the difference in memory game performance given success at the stimulus identification pretest, we recalculated all values in Table 6 omitting all cases in which the child made an error in identifying the color-name stimuli used in the memory game. For the Oral group, now with an N of 21 (as compared to 32) visits, the pattern of variance accounted for was virtually unchanged. For the TC group (now with an N of 15 as compared to 29), eliminating cases in which identification errors were made had the effect of boosting the amount of variance in word recognition scores accounted for by the traditional demographic predictor variables of age at implantation, duration of CI use, and PTA threshold to about the same level as the Oral children (i.e., 30%), but did not substantially change the amount of variance in receptive vocabulary scores accounted for by traditional predictors, or the amount of additional variance accounted for by the digit-span and memory game span measures. Finally, in order to further investigate our failure to find the expected pattern of results regarding the ability of the multi-modal "A+L" condition of the memory game to predict LNT word recognition scores in the Oral group, we again recalculated all values in Table 6 using only the subset of 36 cases for which we could verify correct identification of all color-name stimuli when presented in isolation. This change did not, however, result in any greater percentage of additional variance in LNT scores accounted for by the multi-modal memory game condition.

General Discussion

The primary goal of the present study was to investigate whether individual differences in working memory could account for a significant proportion of the variance in the open-set word recognition and receptive vocabulary skills of prelingually-deafened children with CIs, after the contribution of known demographic predictors was taken into account. In order to answer this question, we first determined the degree to which age at implantation, duration of CI use, PTA thresholds prior to implantation, and, indirectly, chronological age, were able to account for variance in the outcome measures. We found that these traditional predictors of outcome accounted for more variance in word recognition scores in the Oral group than in the TC group. The inability of the traditional predictors to account for variance in the TC group's word recognition scores was unexpected. This result suggests that if a child does not rely primarily on oral language, the child's open-set word recognition performance will not necessarily follow in a predictable fashion from his/her demographic profile. Furthermore, the results also suggest that simply having greater/earlier experience with an implant or more residual hearing cannot insure that auditory-only word recognition skills will develop proportionally, because the development of such skills is dependent not just on the presence of such advantageous circumstances but also on practice and experience with oral language (Hart & Risley, 1995).

Although the demographic factors accounted for a large amount of variance in both groups' receptive vocabulary scores, more variance was accounted for in the TC group than in the Oral group. This result may indicate that although the traditional "non-cognitive" predictors account for most of the individual differences observed in receptive vocabulary for TC children, traditional predictors account for less of the variability observed in Oral children's receptive vocabulary because additional factors, such as individual differences in working memory, also play a role in learning vocabulary via the auditory modality.

Auditory digit-span requiring recall of digit sequences in the forward direction accounted for a sizable amount of additional variance in PBK and LNT word recognition scores in both groups of children. This amount was considerably larger for the TC group than for the Oral group. The reason for this result requires further investigation. The audibility of the digits may have contributed to this difference, although the inclusion of PTA thresholds in the linear regression model should have, in theory, partially compensated for reduced audibility. It would have been helpful to have identified any children who were unable recognize the digits when spoken in isolation. Future data collection should incorporate this additional procedure.

For receptive vocabulary development, we found a partially reversed pattern of results, with forward auditory digit-spans accounting for a greater amount of additional variance in the Oral group, than in the TC group. We suggest two possible reasons for this result. For children who use oral language, individual differences in cognitive factors such as working memory may, in fact, play a key role in facilitating the development of their receptive language skills. Furthermore, if difficulty discriminating the auditory digits contributed to individual differences in digit-span performance in the TC group, then there is no reason why this variability should predict individual differences in a dependent measure that is administered using total communication.

For the Oral group, scores in the "auditory-plus-lights" presentation condition of the memory game did account for some additional variance in PBK word recognition and PPVT receptive vocabulary scores. For the TC group, the "auditory-plus-lights" presentation condition of the memory game failed to predict any additional variance in either of these dependent measures, presumably because the TC children used only visual-spatial encoding to perform the memory game task. In contrast, scores in the "lights-only" presentation condition of the memory game did not account for any additional variance in

the dependent measures in either the Oral or TC group. This suggests that the contribution of working memory is specific to auditory/verbal encoding in working memory, and not some generic, modality-independent component of the working memory system. The contribution of working memory is only observed for memory tasks that involve phonological processing of the input patterns.

Finally, we also found that children who use primarily oral communication, and children who were able to identify all the memory game auditory stimulus tokens correctly, showed a larger redundancy gain on average in the multi-modal memory game presentation condition relative to the "lights-only" presentation condition compared to children not in these groups. This result indicates that like normal-hearing children, some children with CIs engage in verbal encoding strategies and can therefore use the informational redundancy present in the auditory signal to aid them in the multi-modal presentation condition.

Children with cochlear implants are, we believe, an important clinical population to study in order to learn more about the effects of early sensory experience on cognitive processing. The present findings suggest that the contribution of working memory to the development of language skills may vary depending on whether the implanted child develops language in an Oral vs. Total Communication environment and the level of abstraction involved in the language skill under examination, i.e., the learning of phonological forms (word recognition/repetition) versus semantics (vocabulary skills). Our findings indicate that some of the currently unexplained variance in auditory word recognition and receptive language performance in pediatric CI users may be accounted for by individual differences in the underlying cognitive processes employed in auditory memory span tasks. We propose that an important cognitive processing variable related to how young children encode and manipulate the phonological representations of spoken words in working memory contributes to the development of oral/aural language skills in children with cochlear implants (see also, Pisoni, 2000).

References

Archibold, S.M. (2000). Educational implications of cochlear implantation: Conflict or collaboration. In S.B. Waltzman & N.L. Cohen (Eds.), *Cochlear Implants.* New York: Thieme.

Baddeley, A.D. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A.D. (1998). *Human Memory: Theory and Practice (Revised Edition).* Boston: Allyn & Bacon.

Blamey, P.J., Sarant, J.Z., Paatsch, L.E., Barry, J.G., Bow, C.P., Wales, R.J., Wright, M., Psarros, C., Rattigan, K., & Tooher, R. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech, Language, and Hearing Research, 44,* 264-285.

Cleary, M., Pisoni, D.B., & Geers, A.E. (in press). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing.*

Dillon, C.M., & Cleary, M. (this volume). Using nonword repetition to study speech production skills in hearing-impaired children with cochlear implants. In *Research on Spoken Language Processing Progress Report No. 24* (pp. 113-148). Bloomington, IN: Indiana University.

Dowell, R.C., Blamey, P.J., & Clark, G.M. (1995). Potential and limitations of cochlear implants in children. *Annals of Otology, Rhinology and Laryngology Supplement, 166,* 324-327.

Dunn, L.M., & Dunn, L.M. (1997). *Peabody Picture Vocabulary Test, Third Edition.* Circle Pines, Minnesota: American Guidance Service.

Hart, B. & Risley, T.R. (1995). *Meaningful differences in the everyday experiences of young children.* Baltimore MD: Paul H. Brookes Publishing.

Haskins, H. (1949). A phonetically balanced test of speech discrimination for children. Unpublished master's thesis, Northwestern University, Evanston IL.

Kirk, K.I. (1998). Assessing speech perception in listeners with cochlear implants: The development of the lexical neighborhood tests. *Volta Review, 100,* 63-85.

Kirk, K.I., Hay-McCutcheon, M., Sehgal, S.T., & Miyamoto, R.T. (2000). Speech perception in children with cochlear implants: Effects of lexical difficulty, talker variability, and word length. *Annals of Otology, Rhinology and Laryngology Supplement, 185,* 79-81.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing, 16,* 470-481.

Kirk, K.I., Sehgal, S.T., & Hay-McCutcheon, M. (2000). Comparison of children's familiarity with tokens on the PBK, LNT, and MLNT. *Annals of Otology, Rhinology and Laryngology Supplement, 185,* 63-64.

Meyer, T.A., & Pisoni, D.B. (1999). Some computational analyses of the PBK test: Effects of frequency and lexical density on spoken word recognition. *Ear and Hearing, 20,* 363-371.

Miyamoto, R.T., Osberger, J.J., Todd, S.L., Robbins, A.M., Stroer, B.S., Zimmerman-Phillips, S., Carney, A.E. (1994). Variables affecting implant performance in children. *Laryngoscope, 104,* 1120-1124.

Pisoni, D.B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear and Hearing, 21,* 70-78.

Pisoni, D.B., Cleary, M., Geers, A., & Tobey, E. (2000). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *Volta Review, 101,* 111-164.

Pisoni, D.B., & Geers, A.E. (2000). Working memory in deaf children with cochlear implants: Correlations between digit-span and measures of spoken language processing. *Annals of Otology, Rhinology and Laryngology Supplement, 185,* 92-93.

Sarant, J.Z., Blamey, P.J., Dowell, R.C., Clark, G.M., & Gibson, W.P.R. (2001). Variation in speech perception scores among children with cochlear implants. *Ear and Hearing, 22,* 18-28.

Snik, A.F., Vermeulen, A.M., Geelen, C.P., Brokx, J.P., & van den Broek, P. (1997). Speech perception performance of children with a cochlear implant compared to that of children with conventional hearing aids. II. Results of prelingually deaf children. *Acta-Otolaryngol-Stockh., 117,* 755-759.

Wechsler. D. (1991). *Wechsler Intelligence Scale for Children, Third Edition (WISC-III).* San Antonio, TX: The Psychological Corporation.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

The Influence of Short-term and Long-term Memory on the Identification and Discrimination of Non-native Speech Sounds [1]

James D. Harnsberger

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant DC00012 and NIH-NIDCD Research Grant DC00111 to Indiana University.

The Influence of Short-term and Long-term Memory on the Identification and Discrimination of Non-native Speech Sounds

Abstract. This study examined two possible sources of individual differences in cross-language speech perception, the capacity to phonologically encode speech and short-term memory span. Phonological coding was defined as the ability to encode non-native contrasts as distinct phonemes based on representations in long-term memory. Short-term memory was defined as a fixed capacity regulating the extent of encoded phonetic detail. To compare these two predictors of cross-language speech perception performance, thirty native speakers of American English were administered five tests: categorial AXB discrimination and identification (using non-native nasal consonant contrasts), digit span, nonword span (using pronounceable nonwords with nasal consonants, produced by a native speaker of English), and paired-associate word learning with word-word and word-nonword conditions. The AXB discrimination results were correlated with measures of short-term memory (digit span, word-nonword learning), phonological coding (identification), and a memory span measure mediated by phonological coding (nonword span). The results showed that almost all measures were significantly correlated with one another ($+0.62 > r > +0.41$), with the exception of word-word learning. The strongest predictor for the AXB discrimination test results was nonword span ($r = +0.62$, $p < 0.01$). When the identification test results were partialed out, only nonword span significantly correlated with discrimination ($r = +0.54$, $p < 0.01$). The results show an association between the discrimination of these non-native contrasts and a short-term memory capacity that interacts and relies heavily on prior linguistic experience in long-term memory.

Introduction

In numerous studies, a listener's prior linguistic experience has been shown to exert a profound effect on the identification, discrimination, and acquisition of non-native words (Flege, 1987; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). Specifically, listeners frequently identify and encode non-native sounds in terms of the most phonetically similar sounds in their native language. This encoding process can result in the loss of phonetic detail critical to differentiating non-native sounds. Such losses typically correspond to mismatches between the phonemic contrasts of the second language (L2) being acquired and those of the listeners' native language. For example, Japanese learners have particular difficulty in acquiring the /l/-/ɹ/ distinction in English, one which corresponds to a single Japanese liquid phoneme, usually realized as [ɾ] (Strange, 1995).

Such perceptual difficulties are not always explained, however, by a contrastive analysis of the phoneme inventories of the native and second languages in question. Listeners from the same linguistic background often vary substantially in the native sound(s) they use to encode a given non-native sound. For example, Schmidt (1996) found that Korean listeners used up to eight different consonants in labeling English /θ/ in an open-set identification test. In this case, the modal consonant used to label /θ/ only received 44% of the Korean listeners' responses. In a study of the identification of non-native nasal contrasts, Harnsberger (2000) observed highly variable identification patterns by seven groups of listeners for at least a subset of the non-native contrasts tested, particularly in the cases of native speakers of Malayalam, Punjabi, and Tamil. Such variable identification performance is not predictable by any type of contrastive analysis based on abstract units such as phonemes or allophones (Harnsberger, 2000).

Variability in the identification and discrimination of non-native sounds can often reflect large individual differences in perceptual performance. Such individual differences are rarely the central topic of cross-language speech perception studies, though they have been noted in passing (Bohn & Flege, 1990; Liberman, Harris, Kinney, & Lane, 1961; MacKain, Best, & Strange, 1981). In the literature on L2 acquisition, individual differences in learning are typically attributed to such factors as language aptitude, motivation, age of onset of learning, and the extent and type of experience in the second language, to name a few (Carroll, 1962; Carroll, 1981; Flege, Yeni-Komshian, & Liu, 1999; Miyake & Friedman, 1998; Skehan, 1989).

Of these factors, language aptitude is probably the most relevant to the problem of individual differences in cross-language speech perception, given that in such studies, listeners have no experience with the non-native words in question and are not attempting to acquire the language from which the non-native words were drawn. Language aptitude has been defined in past research on individual differences in L2 learning in terms of three component capacities: language analytic, phonetic coding, and short-term memory (Carroll, 1962; Carroll, 1981; Skehan, 1989). Language analytic capacity relates to the acquisition of L2 syntax. Phonetic coding refers to the ability to store non-native speech sounds in a manner that allows for easy storage and retrieval by the listener from long-term memory.

As a predictor of success in discriminating and identifying non-native sounds, phonetic (or, more accurately, phonological) coding can be conceived of as the capacity to detect and store in long-term memory the fine acoustic or articulatory details that differentiate the sounds and sound patterns of non-native words from similar sounds and patterns used in native words. Short-term memory capacity refers to a memory component of fixed capacity whose purpose is to temporarily store information for encoding in long-term memory. Traditionally, short-term memory has been modeled as a general mechanism that is independent of representations in long-term memory. For example, Baddeley's (1986) model of working memory includes the phonological loop, which combines a phonological short-term store of fixed capacity with a subvocal rehearsal process to maintain the phonological representation. The short-term store in this model has been modeled as independent of any input from representations in long-term memory, which would include the listener's prior linguistic experience (though see Baddeley, Gathercole, & Papagno, 1998; Carpenter, Miyake, & Just, 1994).

The evidence supporting phonological coding as the basis of individual differences in cross-language speech perception can be found primarily in the repeated demonstration of a relationship between the identification of non-native sounds and their discrimination (see Strange, 1995 for a review of the cross-language speech perception literature). Identification of non-native sounds with native phonemes has also been shown to extend to L2 acquisition, both in perception (Miyawaki et al., 1975) and production (Flege, 1987). While extensive experience can allow listeners to form new perceptual categories for non-native sounds (Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991; MacKain et al., 1981), in some instances, even highly experienced learners may fail to correctly identify non-native sounds (Flege, 1991).

The evidence suggesting a role for short-term memory capacity in cross-language speech perception comes from studies of novel word learning by children, adults with cognitive deficits, and normal adults. Like cross-language speech perception, novel word learning involves the perception and encoding of novel phonological forms. Numerous studies have demonstrated significant correlations between short-term memory span and novel word learning in children (Gathercole, 1995; Gathercole & Baddeley, 1990; Gathercole, Frankish, Pickering, & Peaker, 1999; Gathercole, Service, Hitch, Adams, and Martin, 1999; Gathercole, Willis, Emslie, & Baddeley, 1991). Selective impairment in short-term memory capacity has also been shown to affect nonword learning in paired-associate word-learning tasks

(Baddeley, Papagno, & Vallar, 1988; Papagno & Vallar, 1995a; Trojano & Grossi, 1995). This relationship has been observed in normal adults as well (Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992; Service, 1998).

Recently, research on short-term memory has also been extended to L2 learning by children and adults. Papagno et al. (1991) examined non-native (Russian) word learning by two groups, Italian and English native speakers, under conditions of articulatory suppression. Articulatory suppression involves the repetition of a speech sound or word while engaging in a concurrent task that relies on short-term memory. The effect of articulatory suppression is to interfere with the maintenance of phonological information in short-term memory. Papagno et al. predicted that articulatory suppression would disrupt novel word learning but not paired-associate word-word learning, assuming novel word learning relies heavily on short-term memory. The participants were tested in a paired-associate word-learning task with word-word and word-nonword pairs. Both the English and Italian speakers displayed the predicted pattern of poor performance on word-nonword pairs.

Service (1992) conducted a longitudinal study of English acquisition by nine year-old children who were native speakers of Finnish. Service found significant correlations between three nonword repetition tasks and grades in English and in other classes. MacKay, Meador, and Flege (2000) observed modest correlations between a nonword repetition measure and the error rates in identifying English word-initial and word-final consonants by native speakers of Italian. Miyake and Friedman (1998) studied the acquisition of English syntax by 59 native speakers of Japanese. Subjects were administered five tests, two on syntactic comprehension, two listening span measures using Japanese and English stimulus materials, and a digit span task. Listening spans were considered to be measures of "working memory" rather than short-term memory tests, such as digit span, because listening spans incorporate a computational or operational component. Miyake and Friedman found that the L2 working memory span measure (with English materials) correlated significantly with the syntactic comprehension measures and the L1 working memory span measure (with Japanese materials). Interestingly, the digit span measure did not correlate significantly with the syntactic comprehension measures. Thus, the strongest correlations observed in this study involved measures that used a similar type of stimuli (i.e., sentences in the working memory and syntactic comprehension measures), indicating an effect of long-term memory on representations in short-term memory.

The effect of short-term memory on L2 acquisition has also been studied indirectly by examining individual differences in multilingual, bilingual and monolingual populations. Papagno and Vallar (1995b) compared ten "polyglot" and ten "nonpolyglot" speakers in a number of tasks, including general intelligence, vocabulary (i.e., a subtest of Wechsler Adult Intelligence Scale), auditory digit span, nonword repetition, visual-spatial span, visual spatial learning, and paired-associate word-nonword learning. Polyglots were defined as speakers of at least three languages, while nonpolyglots were defined as speakers or two or fewer languages. The two groups' performance differed significantly only on the auditory digit span, nonword repetition, and the word-nonword condition of the paired-associate word-learning task. Papagno and Vallar argued that a large phonological short-term memory capacity aids in the acquisition of novel words. Thus, individuals with a greater capacity in phonological short-term memory are predicted to be better at acquiring multiple languages.

While a number of studies have shown a relationship between short-term memory and novel word learning in several populations of interest, several questions remain concerning the methodologies of these studies and the relationship between short-term and long-term memory, particularly as they apply to the problem of individual differences in cross-language speech perception. First, many of these studies have relied on nonword repetition as the primary measure of memory span. The use of this methodology may be particularly problematic with certain populations, such as children, adults with particular

cognitive deficits, or adults attempting to produce non-native (unfamiliar) sounds, because it confounds short-term memory capacity with speech production skills. This issue was addressed recently by Gathercole, Service, Hitch, Adams, and Martin (1999). In their study, eighteen four-year-old children were administered digit span, nonword repetition, vocabulary, and nonword matching span tasks. Most importantly, the nonword matching span tasks did not require verbal responses. Nonword matching span was found to correlate significantly with vocabulary, demonstrating that with at least this population, nonword repetition is a valid measure of memory span.

Second, the few studies of short-term memory and second language acquisition by normal adults and children have employed pronounceable nonwords in tasks such as nonword repetition or paired-associate word learning (Papagno & Vallar, 1995b; Thorn & Gathercole, 1999). Obviously, it would be difficult to use stimulus materials such as non-native words that are difficult to accurately produce in tests requiring a verbal response. However, the learning of non-native words that are difficult to accurately identify and/or produce is a common problem faced by learners of many languages. Acquiring receptive and expressive mastery of spoken words differentiated by phonemic contrasts that are difficult to distinguish constitutes a significant learning problem for adult learners of second languages (Flege, 1992). Examining the role that short-term memory plays in the acquisition of non-native words requires procedures that do not involve verbal responses (such as a matching span task) or correlations of span measures with identification or discrimination tests using non-native words.

Finally, memory span tests often confound possible individual differences in memory capacity with individual differences in the perception, encoding, and representation of linguistic knowledge in long-term memory. Recent work has shown that phonological short-term memory measures vary greatly depending on a listener's prior experience with the stimulus materials used. For example, "wordlikeness," or the similarity of a nonword to a real word, has been shown to effect short-term memory spans of adult speakers of English (Papagno et al., 1991) and English-speaking children (Gathercole, 1995). Short-term memory spans of children have also been shown to be influenced by the phonotactic properties of the stimulus materials (Gathercole, Frankish, Pickering, & Peaker, 1999). More importantly, whether the stimulus materials consist of native or non-native sounds can also affect subjects' performance on short-term memory tasks. Thorn and Gathercole (1999) found that English-French bilingual children and English-speaking children learning French as a second language had significantly longer French nonword repetition spans than monolingual English children. All of these studies strongly suggest that phonological coding in short-term memory, rather than being "knowledge-free" or independent of knowledge and representations in long-term memory, relies heavily on prior knowledge, specifically, on linguistic experience.

A close connection between short-term and long-term memory poses problems for theories of language acquisition that claim that fixed memory capacities in individuals affect the learning of new words or new grammatical rules. Individuals may also differ in the nature of their long-term experience with speech that resembles the stimulus materials used in memory span tasks. Such differences in prior linguistic experience may manifest themselves as differences in vocabulary size. However, individuals may also differ in the extent of their experience with different phonological forms of different words in the lexicon. Listeners with more extensive listening experience with multiple languages, or in multiple dialects of their native language, may possess word representations that incorporate a greater range of acoustic-phonetic detail and variability. Numerous earlier studies have clearly shown that listeners encode precise details of episodes of speech (see Goldinger, 1998 for a review). Such "robust" episodic word representations may aid listeners in the performance of tasks such as memory span tests. Thus, memory span, rather than being an important source of individual differences in language acquisition, may be the "byproduct" of individual differences in prior linguistic experience. It is also possible that both fixed short-term memory capacity and prior linguistic experience may interact to determine the capacity of

listeners to store phonetically detailed representations of non-native words for encoding in long-term memory (see Baddeley et al., 1998 for an example of a word-learning model in which short-term and long-term memory interact).

The purpose of this study was to examine the effects of short-term memory capacity and prior linguistic experience on individual differences in cross-language speech perception by native speakers of American English. Short-term memory capacity was measured independently of long-term memory by using an immediate serial recall task with highly familiar stimulus materials, namely, English digits. In addition, a paired-associate word-learning task, using word-word and word-nonword pairs, was administered. The word-word pairs served to separate the "pure" short-term memory span task with digits from any effects of individual differences in familiar word learning. The word-nonword learning task was used as an analog to acquiring unfamiliar words in a second language and as a language learning measure strongly related to short-term memory span (Papagno & Vallar, 1995b).

Prior linguistic experience was measured indirectly by measuring the phonological coding capacity of listeners. Phonological coding, as described by Skehan (1989), corresponds to a listener's reliance on long-term memory representations based on prior linguistic experience. This ability was measured using an identification test with non-native words that were thought to be difficult to discriminate for English listeners. To compare the relative success of the short-term memory and phonological coding measures in predicting the cross-language speech perception abilities of English listeners, a discrimination test using the same non-native words as the identification test was administered. Discrimination tests are commonly employed in cross-language speech perception research and do not require an overt verbal response. The possible interaction of pure short-term memory capacity with prior linguistic experience was measured with an immediate serial recall memory span task using native nonword analogs to the non-native words under study. If prior linguistic experience influences short-term memory capacity, then native nonword spans should show stronger correlations with the discrimination test results than either the identification, digit span, or word-nonword-learning test results.

If an effect of short-term memory were observed on the discrimination test results, then an important source of individual differences in cross-language speech perception would be identified. Currently, models of cross-language speech perception, such as the Perceptual Assimilation Model (Best, 1995) or the Native Language Magnet model (Kuhl & Iverson, 1995), do not formally incorporate a role for individual differences in short-term memory in the discrimination of non-native contrasts. In both models, discrimination is a function of the manner in which the sounds constituting a non-native contrast (e.g., dental and retroflex stops for English listeners) are identified with, or phonologically coded as, one or more native sounds. In the Native Language Magnet model, the identification-discrimination relationship is described in terms of the association of stimuli to particular locations in perceptual space, within a particular perceptual category. Two non-native stimuli that constitute a non-native contrast are discriminable based in their locations in perceptual space: Are they close to a prototype, or the category periphery? In the Perceptual Assimilation Model, the identification-discrimination relationship is also a function of the proximity of the stimuli to one or more perceptual categories: Are the stimuli equally good exemplars of a single category, do they differ in category goodness, or are they even associated with any native perceptual categories?

In both of these models, the preservation of phonetic detail in the encoding of a non-native contrast in long-term memory is the automatic consequence of how a contrast was phonologically coded in terms of one or more existing representations in long-term memory (i.e., perceptual categories). However, both models would require revision if the results of this study demonstrate a role for short-term memory in either the discrimination or the identification of non-native sounds. The results of this study may demonstrate that an individual's short-term memory capacity constrains the encoding of phonetic

detail in non-native sounds, independent of the individual's phonological coding, or identification, of non-native sounds. Alternatively, short-term memory capacity itself may determine individual patterns in the identification and discrimination of non-native sounds. Finally, short-term memory capacity may prove to be unrelated to cross-language speech perception.

Main Experiment

Methods

Participants

Thirty normal-hearing Indiana University undergraduates participated in this experiment, 11 males and 19 females. The participants ranged in age between 18 and 27 (M = 22). No subject reported any history of a speech or hearing problem. For participating in two 1-hour sessions, the subjects received $15 as compensation.

Stimulus Materials

AXB Discrimination and Identification Tests. The stimulus materials used for these tests were a subset of those used by Harnsberger (2000). Briefly, one male and one female talker of Malayalam, a Dravidian language of southern India, were recorded reading from a list of real and nonsense words from their native language that included all six nasal consonants of Malayalam. The nasals of interest appeared in all syllable positions allowable by the individual languages, in an /a/, /i/, or /u/ vocalic context. All of the stimuli recorded were evaluated by native speakers of the respective languages in an identification test in order to exclude stimuli that might be poor exemplars of Malayalam nasals. Of the stimuli that were recorded and evaluated, a subset was used in these tests: four exemplars from an isolation context, two from each talker, of dental, alveolar, and retroflex medial geminate nasals in an [i] context. These three types of nasal consonants were selected because they could be combined to form non-native contrasts that have been shown to be of intermediate difficulty for English listeners to discriminate (Harnsberger, 1998).

In addition to the non-native words, a similar set of nonsense words were recorded from a male speaker of American English, who produced two exemplars each of three nasal consonants, [m], [n], and [ŋ], appearing in final position following [i]. The native nonsense words with bilabial and alveolar nasals were paired together to form AXB discrimination test trials. All of the native nonsense words were included in the identification test. All of the native and the non-native nonsense words were leveled in amplitude at 70 dB for presentation in the experiment.

Digit Span. The stimulus materials for the digit span test included one exemplar each of ten words, "one," "two," "three," "four," "five," "six," "seven," "eight," "nine," and "zero," produced by a male speaker of English. The stimuli were leveled in amplitude at 70 dB for presentation in the experiment.

Nonword Span. The stimulus materials for the nonword span test were a subset of the native nonsense words used in the identification and discrimination tests: one token each of the bilabial, alveolar, and velar native nonsense words. The stimuli were leveled in amplitude at 70 dB for presentation in the experiment.

Paired Associate Word Learning. The stimulus materials for the paired-associate word-learning test consisted of eight pairs of real words and eight pairs of pronounceable nonwords, all produced by a male speaker of American English. The sixteen pairs are listed in the Appendix. The real word pairs consisted of four two-syllable and four three-syllable pairs, consisting of high frequency, concrete nouns. Words appearing in pairs were matched together in terms of their stress pattern. The nonsense word pairs also consisted of four two-syllable and four three-syllable pairs. The nonsense words were created from syllables involving highly probable phoneme sequences (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). The stimuli were leveled in amplitude at 70 dB for presentation in the experiment.

Procedure

Five tests were administered to the subjects over two sessions. In the first session, subjects were administered an AXB discrimination test and an identification test. The order in which these tests were administered was counterbalanced. In the second session, the subjects were administered the span tests (digit span first, nonword span second), and the paired-associate word-learning test (word-word pairs first, word-nonword pairs second). The span tests were counterbalanced with the paired-associate word-learning tasks.

Identification Test. The identification test was a forced-choice identification test consisting of 90 trials, specifically, five repetitions of eighteen nonwords. Each word appeared in isolation, and the order of presentation of trials was randomized. Listeners were instructed to label the nasal consonant of each word using one of three response choices (keywords) corresponding to the three nasal phonemes of English, “sum” (/m/), “sun” (/n/), or “sung” (/ŋ/). Three familiarization trials, consisting of three native nonsense words ([im], [in], and [iŋ]), were presented to confirm that listeners understood which nasal consonants were represented by the keywords.

In scoring the identification test results, subject responses to pairs of stimulus types that corresponded to discrimination test trials were compared. For instance, the identification test results of the alveolar and retroflex nasals produced by Malayalam talker YM were scored together to yield a predicted discrimination test score for the alveolar-retroflex contrast produced by YM that appeared in the discrimination test. The scoring method involved calculating the extent to which two stimulus types (e.g., bilabial and alveolar nasals) differ in their identification with native perceptual categories, a metric termed hereafter the *categorization difference score* (C-score). Specifically, the C-score is based on the sum of the differences between the two stimulus types in their proportion of responses for each native perceptual category. The C-score is represented in equation (1):

$$C = \frac{\sum_{i=1}^{n} \left| A_i - B_i \right|}{2t} \qquad (1)$$

where, C is the categorization difference score, n is the number of response categories available to the listener, A_i is the number of responses of stimulus type A to category i, B_i is the number of responses of stimulus type B to category i, and t is the number of trials in which each stimulus type was presented (assuming each stimulus type was presented the same number of times over the course of the identification test). $2t$ is a constant that simply converts the range of scores to a familiar 0.0 – 1.0 scale.

This metric was used by Harnsberger (1999) successfully to account for differences in the discriminability of a large set of non-native contrasts by seven different listener groups in a cross-language speech perception study. The calculation of a C-score can be illustrated using the hypothetical

data set shown in Table 1. In this example, the proportion of responses to each stimulus type ([n̪], [n], [ɳ]) is listed below each response choice (/m/, /n/, or /ŋ/). In this example, each stimulus type was presented ten times over the course of the identification test. To calculate the C-score for the [n̪]-[n] contrast, the absolute value of the difference between the proportion of /m/ responses to [n̪] and [n], 8 – 0, must be summed with the absolute value of the difference between the proportion of /n/ responses to [n̪] and [n], 0 – 9, and the absolute value of the difference between the proportion of /ŋ/ responses to [n̪] and [n], 2 – 1. The resulting value of 18 (|8 – 0| + |0 – 9| + |2 – 1|) is divided by two times the number of trials in which each stimulus type was presented (2 * 10 = 20) to yield a relatively high C-score of 0.9. In this example, the C-score for the [n̪]-[ɳ] contrast is a relatively low 0.3 ({|8 – 5| + |0 – 3| + |2 – 2|}/20), while the C-score for the [n]-[ɳ] contrast is an intermediate 0.6 ({|0 – 5| + |9 – 3| + |1 – 2|}/20).

	Response choices		
Stimulus Type	/m/	/n/	/ŋ/
[n̪]	8	0	2
[n]	0	9	1
[ŋ]	5	3	2

Table 1. A hypothetical dataset for the identification test. The number in each cell is the number of trials in which a particular response choice (e.g., the English bilabial nasal represented by the keyword "sum") was selected for a particular stimulus type (e.g., a dental nasal).

C-scores were used to predict the discrimination test scores. Contrasts involving stimulus types that were frequently labeled in a different manner (corresponding to high C-scores) were predicted to be relatively more discriminable than contrasts involving two stimulus types that frequently received the same label (corresponding to low C-scores). In this study, the C-scores tested the phonological coding hypothesis in a correlation analysis with the other test scores.

Discrimination Test. The discrimination test was a categorial AXB test consisting of 112 trials, 16 trials each of the seven different types of contrasts, where "contrast" refers to a particular place distinction produced by a particular talker. Six of these contrasts were non-native, namely, dental-alveolar, dental-retroflex, and alveolar retroflex produced by the two Malayalam talkers. One contrast was produced by the American English talker, namely, bilabial-alveolar. This native contrast was included as a control, to test whether individual discrimination scores varied due to the difficulty of the stimuli as opposed to any difficulties with the test format.

In order to ensure that the results were not dependent on the intelligibility of a single exemplar, two exemplars of each member of the seven contrasts were used. The contrasts appeared in four possible orders, AAB, ABB, BAA, and BBA. A and B were always from the same talker, and all stimuli that were paired together were selected to minimize acoustic differences that were not relevant to the identity of the stimulus, such as the overall duration or the fundamental frequency pattern of a stimulus. The

interstimulus interval for the discrimination test was 1 s. The order of presentation of AXB trials was also randomized.

Subjects were told to indicate whether the nasal consonant in the first or last word was the same as the nasal consonant in the middle word by circling a number on the answer sheet. The term "nasal consonant" was defined through the use of simple examples in which nasals appeared in different syllable positions and vocalic contexts. A, X, or B were not physically identical, so listeners made categorial matches. One familiarization trial was presented before the AXB test. The discrimination test results were analyzed in terms of mean percent correct responses for individual contrasts as well as a mean score over all contrasts.

Digit Span. In the digit span test, subjects were presented auditorily with a sequence of single digits (0-9) pronounced by a native speaker of American English. Subjects were asked to write down in order the digit sequence presented. The length of the digit sequences began at four, and ended at ten, with two trials at each sequence length in order of increasing length. The results of the digit span task were scored in two ways: first, in terms of the longest sequence length in which a subject correctly recalled all digits in order in both trials; second, in terms of an *absolute span score* which sums each correct trial weighted by its sequence length, as in equation (2):

$$\mathrm{ABS} = \sum_{i=1}^{n} x_i * y \qquad (2)$$

where, *ABS* is the absolute span score, *i* is a trial, *n* is the total number trials, *x* is the length of the digit sequence of trial *i*, and *y* is the accuracy of the subject's response (correct = 1, incorrect = 0). Both the longest digit span score and the absolute digit span score were used in a correlation analysis with the results of the other tests.

Nonword Span. In the nonword span test, subjects were presented auditorily with a sequence of nonwords consisting of VC syllables ([im], [in], [iŋ]) pronounced by a native speaker of American English. Subjects were asked to write down in order the nonword sequence presented using the symbols "m," "n," and "ng" for [im], [in], and [iŋ], respectively. The length of the nonword sequences began at three, and ended with seven, with two trials at each sequence length in order of increasing length. The results of the nonword span task were scored in terms of the longest sequence length in which a subject correctly recalled all nonwords in order in both trials and in terms of an absolute span score, described earlier. As with the digit span test, whichever of the two scoring methods showed the strongest correlation with the discrimination test scores was used in subsequent analyses.

Paired Associate Word Learning. This paired-associate word-learning test consisted of two conditions, word-word and word-nonword. In the word-word condition, subjects were presented with the full set of word-word pairs. Each word in the pair was separated by a 1 s interval, while each pair was separated by a 2 s interval. After hearing all of the word-word pairs, subjects were presented with the first word from each pair and responded verbally with its corresponding second word, if it could be recalled. This second part of the procedure was self-paced. Subjects received a "correct" score for each trial if their response matched the second word exactly. This entire sequence (presenting the word-word pairs followed by the first words from each pair) was repeated until the subject recalled all the second words correctly in two consecutive repetitions of the entire sequence, or until the tenth repetition was complete. A different random trial order was used for each repetition. The procedure for the word-nonword condition was identical to the word-word condition, except for the stimuli presented to the subjects. The Appendix lists the particular word-word and word-nonword pairs used in this test. The word-word and the

word-nonword conditions were scored in terms of the number correct of correct responses at a particular repetition in the condition.

Predictions

Three hypotheses were entertained in this study. First, the short-term memory hypothesis states that listeners' fixed short-term memory capacity determines the extent of phonetic detail that is encoded in representations in long-term memory. Thus, greater short-term memory capacities should make nonword learning easier, resulting in a significant correlation between digit span and word-nonword learning. In addition, short-term memory should also determine the manner in which non-native words are phonologically encoded. Thus, digit span should correlate significantly with the identification and discrimination test scores. Any significant correlations between nonword span and the identification and discrimination tests were predicted to disappear if digit span was partialed out.

Second, the phonological coding hypothesis states that the extent of phonetic detail that is encoded in long-term memory representations of non-native or novel words is determined by the manner in which the words are phonologically encoded, that is, by the native sounds used to encode the nonwords. According to this hypothesis, an individual's manner of encoding is based his/her prior linguistic experience rather than his/her short-term memory capacity. If this is the case, then the identification test results should correlate with the discrimination test results. In addition, no significant correlation was predicted between either the discrimination or identification test results and the digit or nonword spans.

Finally, the short-term/long-term memory (STM-LTM) interaction hypothesis states that the extent of encoded phonetic detail is determined by a short-term memory capacity that receives input and is influenced by linguistic representations in long-term memory. Thus, when listeners encode speech in short-term memory, they rely on representations in long-term memory, though the capacity of their short-term memory should still constrain their ability to accurately encode phonetic detail in non-native words. According to this hypothesis, the discrimination and identification results should correlate with the various span measures, but nonword span should correlate more strongly than the pure measure of short-term memory, digit span. That is, individuals with long-term memory representations that incorporate phonetic details that aid them in encoding these non-native words should also perform better on span tasks that employ stimulus materials similar to those of the discrimination and identification tests (i.e., nonword span), relative to span tasks that do not utilize those long-term memory representations (i.e., digit span).

Results

Scores for Individual Tests

The results of the discrimination test (AXB) as well as the identification test (ID) are shown in Table 2. The results are listed by contrast, defined in terms of place of articulation as well as talker. The AXB results are the mean proportion of correct responses, while the ID scores are the mean C-scores (the standard deviations of all means are given in parentheses). The native contrast elicited a relatively high mean percent correct score of 89%, though subjects did not perform at ceiling as might have been expected. Across all non-native contrasts, subjects averaged 0.65 proportion correct (SD = 0.18) in the discrimination test and a 0.38 C-score (SD = 0.28) in the identification test. These discrimination test scores are somewhat lower than those elicited from native speakers of American English with the same stimuli in a previous study by Harnsberger (1998): the dental-alveolar, dental retroflex, and alveolar-

retroflex contrasts of Malayalam talker YM elicited mean proportion correct discrimination scores of 0.9, 0.8, and 0.83, while the same three contrasts from talker YS elicited scores of 0.59, 0.68, and 0.64.

Place Contrast	Talker	AXB	ID
bilabial-alveolar	JH	0.89 (0.12)	0.87 (0.23)
dental-alveolar	YM	0.8 (0.17)	0.31 (0.31)
	YS	0.57 (0.15)	0.2 (0.15)
dental-retroflex	YM	0.71 (0.21)	0.45 (0.34)
	YS	0.52 (0.11)	0.34 (0.2)
alveolar-retroflex	YM	0.71 (0.13)	0.54 (0.29)
	YS	0.62 (0.13)	0.41 (0.19)

Table 2. The discrimination test results reported as mean proportion of correct responses and listed by contrast (place and talker). The standard deviations of the means are in parentheses.

The distribution of individual scores on both tests is shown in Figure 1, organized by talker (YM and YS) and by place contrast (dental-alveolar, dental-retroflex, and alveolar-retroflex). Across all contrasts, a sufficient spread in the discrimination and C-scores were observed, insuring that the cross-language test results could be used in correlation analyses. In particular, the results for the contrasts produced by YM elicited a great range of results, as indicated by high standard deviations and relatively shallow slopes in the bell-shaped curves of the distributions.

The digit span and nonword span scores also showed significant variability in the individual results. The mean digit span scores, measured in terms of longest span correct and absolute span, were 6.9 (SD = 1.6) and 51.8 (SD = 20.9), respectively. The mean nonword spans were 4.6 (SD = 1.8) and 21.8 (SD = 13.4), for the longest span and absolute span scores, respectively.

Figure 2 shows the number of correctly recalled second words for each repetition in the paired-associate word-learning test, for both the word-word condition (solid line) and the word-nonword condition (dashed line). The word-word pair learning proved to be a relatively easy task for the subjects of this experiment. By the fourth repetition, subjects were averaging close to ceiling-level performance (7.7 words). The word-nonword pairs proved to be more difficult to learn than the word-word pairs, as expected. However, the rate of learning in each condition was similar, though subjects on average did not asymptote at ceiling with the word-nonword pairs. Given the results of this test, the mean number of correctly recalled words from the first repetition of the word-word condition was taken to represent the performance of individual subjects. For the word-nonword condition, the mean number of correctly recalled words from the fifth repetition was taken as the score. These two repetitions were chosen because both elicited intermediate values in the 0 – 8 scale of the test, avoiding any ceiling or floor effects in comparing the paired-associate word-learning results to those of other tests.

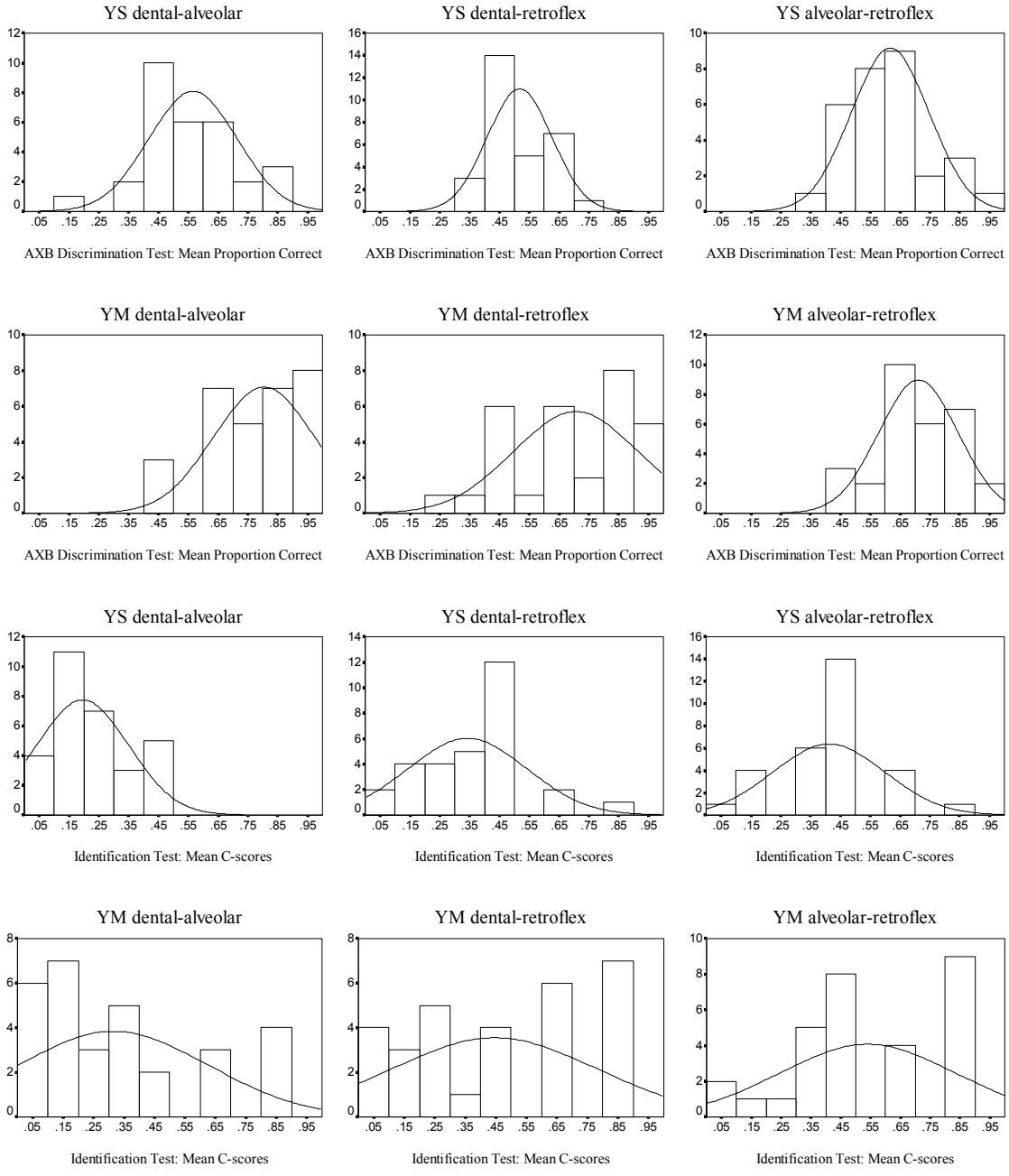


Figure 1. Histograms of the distributions of individual scores on the AXB discrimination and identification tests.

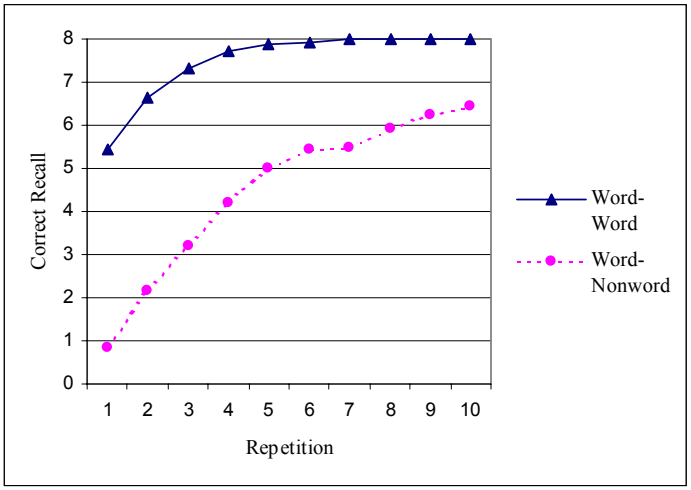


Figure 2. The number of correctly recalled words or nonwords for each repetition of the word-word or word-nonword pairs.

Full Correlation Matrix

Table 3 shows the full correlation matrix of the results, including the discrimination test scores (AXB), the categorization difference scores from the identification test (ID), the digit span test scored in absolute span (DigABS) as well as longest span (DigLong), the nonword span test scored in absolute span (NonABS) as well as longest span (NonLong), the mean number of correct words in the first repetition of the word-word condition of the paired-associate word-learning test (WW), and the mean number of correct words in the fifth repetition of the word-nonword condition (WN). In correlating the discrimination test scores with the scores for other tests, mean percent correct scores that were averaged over all contrasts presented in the test were used, with one exception: the coefficient for the AXB – ID correlation represents an analysis in which the discrimination and C-scores of individual contrasts were entered, since unique C-scores were available for the individual contrasts.

Test	AXB	ID	DigAbs	DigLong	NonABS	NonLong	WW	WN
AXB	X							
ID	0.41**	X						
DigAbs	0.34	0.53**	X					
DigLong	0.3	0.42*	0.93**	X				
NonABS	0.62**	0.44*	0.38*	0.39*	X			
NonLong	0.57**	0.34	0.34	0.39*	0.91**	X		
WW	0.25	0.16	0.33	0.2	0.22	0.26	X	
WN	0.53**	0.41*	0.51**	0.43*	0.46**	0.44*	0.29	X

Table 3. The full correlation matrix for all tests administered in the experiment, including discrimination (AXB), identification (ID), digit span (DigABS and DigLong), nonword span (NonABS, NonLong), and the word-word (WW) and word-nonword (WN) conditions of the paired-associate word-learning test. * $p < 0.05$, ** $p < 0.01$

According to the short-term memory hypothesis, the discriminability of non-native or novel words should be a function of an individual's fixed capacity to encode phonetically-detailed representations in short-term memory for storage in long-term memory. In contrast, the phonological coding hypothesis states that individuals can vary greatly in how they identify, or encode, non-native sounds in short-term memory, and those differences are a function of the unique properties of their perceptual or lexical categories in long-term memory (i.e., the region in multidimensional acoustic space that the category occupies). Finally, the STM-LTM hypothesis stated that the discriminability of novel and non-native words is a function of both an individual's phonological coding strategy and a fixed capacity to encode the phonetic details of novel words.

The results in Table 3 indicate that a traditional method of measuring pure short-term memory, digit span, failed to correlate significantly with the discrimination test results, regardless of the scoring method used ($r = +0.3$ for longest span, $r = +0.34$ for absolute span). Phonological coding, as represented by C-scores, did correlate significantly ($r = +0.41$, $p < 0.01$), although the strength of the correlation was not as great as that observed by Harnsberger (1999), who examined a much larger data set using the same scoring method. However, the strongest correlations with the discrimination test results were found with the memory span measures using nonword stimulus materials. The word-nonword learning scores correlated significantly with the discrimination test scores ($r = +0.53$, $p < 0.01$), while the nonword spans, measured using stimulus materials that were quite similar to those of the discrimination test, showed the strongest correlation ($r = +0.62$, $p < 0.01$, for the absolute span scoring method).

Taken alone, the correlations with the discrimination test support the STM-LTM interaction hypothesis, which predicted that the span measure employing the most similar stimuli to those of the identification and discrimination tests would show the strongest correlation. However, several other measures also correlated with the results from tests using the non-native stimuli. While pure short-term memory (as represented by the digit span task) did not correlate with discrimination, it did correlate significantly with the identification test results, the nonword spans (when matched in scoring method), and the word-nonword scores. In fact, the identification, discrimination, digit span, and nonword measures were all intercorrelated. Only the word-word scores failed to significantly correlate with any of the other measures in the analysis, suggesting that some form of memory span, as opposed to a general ability in verbal learning, accounts for the individual differences in the perception tests with non-native stimuli (see Papagno and Vallar, 1995b, for a similar finding on paired-associate word learning and short-term memory tasks).

Partial Correlations

To examine further the factors responsible for individual differences in cross-language speech perception, a number of partial correlations were computed to separate the effects of short-term memory and long-term memory on listeners' perceptual performance with non-native words. The results shown in Table 3 failed to support the phonological coding hypothesis. The remaining hypotheses were differentiated by partialing out the effect of short-term memory (as indexed by digit span). If short-term memory capacity alone is responsible for individual differences in the encoding of the phonetic details of non-native words, then in this partial correlation analysis, nonword spans should not significantly correlate with identification and discrimination test results. Only the digit and nonword spans scored using the absolute span score method were used in this analysis because in the original correlations, digit and nonword spans showed the strongest correlations with other measures when scored by that method. In addition, the word-word learning results were dropped from the analysis, since they did not correlate with any of the other measures obtained.

The resulting correlation matrix with digit span partialed out appears in Table 4. In this analysis, the identification and word-nonword scores no longer correlate with the discrimination scores. However, the strong correlation between discrimination and nonword spans still remained. When partialing out all other measures, nonword span was still significantly correlated with the discrimination scores ($r = +0.46$, $p < 0.05$). The importance of nonword span for discrimination was demonstrated again in a third correlation analysis in which nonword span was partialed out. In this analysis, no other factor correlated significantly with discrimination; digit span, however, was still correlated with identification ($r = +0.43$, $p < 0.05$) and word-nonword learning ($r = +0.41$, $p < 0.05$).

Test	AXB	ID	NonABS	WN
AXB	X			
ID	0.26	X		
NonABS	0.56**	0.3	X	
WN	0.44*	0.19	0.34	X

Table 4. The correlation matrix with digit span (DigABS) partialed out, including discrimination (AXB), identification (ID), nonword span (NonABS), and the word-nonword (WN) condition of the paired-associate word-learning test. * $p < 0.05$, ** $p < 0.01$

Discussion

Of the three proposed hypotheses, the results of this study support the STM-LTM interaction hypothesis: phonological short-term memory, as measured using native stimulus materials that were phonologically similar to the non-native words under study, correlated significantly with the results of cross-language discrimination and identification tests. This effect of phonological similarity would only be observed if the short-term memory of listeners is influenced by long-term memory representations of words involving nasal consonants. Thus, long-term memory plays an important role in the encoding of non-native contrasts in short-term memory.

However, short-term memory capacity is not simply a byproduct of variation in individual differences in long-term memory representations. Digit spans, a measure using highly familiar stimulus materials that is assumed to tap pure short-term memory, correlated significantly with many of the measures collected here, including word-nonword learning and, most importantly, identification. The correlation between digit span and identification, but not discrimination, indicates that identification and discrimination abilities are separable to some extent, and may rely on short-term and long-term memory differently. Nevertheless, nonword span showed the strongest correlations with the discrimination test scores, and correlated with nonword learning and identification in a similar manner as digit span. The strength of the nonword span correlations supports the STM-LTM interaction hypothesis, which has also received support in several other recent studies in which either the stimulus materials or listener groups are varied linguistically (Gathercole, Frankish, Pickering, & Peaker, 1999; Thorn & Gathercole, 1999).

The success of the nonword spans, compared with digit spans, as a predictor of cross-language discrimination implies that as the similarity between the stimulus materials of the span task and those of a correlated measure (such as discrimination) increases, the strength of the correlation increases. That is, the effect of long-term memory on span capacity should be greater for stimulus materials that are more similar to representations in long-term memory. This proposed relationship can be tested by examining a

subset of the data collected in this study. While the stimulus materials of the non-word span test were phonologically similar to the non-native stimuli, they were most similar to the native control stimulus materials of the identification and discrimination tests. Specifically, they were a subset of the native nonsense words included in the identification test. The discrimination test used a different subset of these nonsense words: two tokens each of the nonsense words with bilabial and alveolar nasals. The same male talker produced all of the native nonsense words used in this study, and all of the nonsense words were of the form [iN], where N is a nasal consonant of English, either [m], [n], or [ŋ]. Given the similarity of the native stimulus materials of the identification and discrimination tests and those of the nonword span task, we can predict that nonword span should correlate more strongly with the discrimination test results than those of the identification or digit span tests.

A full correlation matrix using the data subset described above is shown in Table 5. The span test results in this analysis were only scored using the absolute span scoring method. The matrix shows that all of the measures were significantly intercorrelated. The only exception was the identification and digit span correlation. However, in this analysis, identification showed the strongest correlation with discrimination instead of nonword span or digit span. When short-term memory (digit span) was partialed out, the correlation coefficient for discrimination-identification dropped to only +0.6 ($p < 0.01$), while the discrimination-nonword span correlation coefficient was barely significant ($r = +0.39$, $p = 0.04$). When identification was partialed out, the correlation between discrimination and nonword span was no longer significant ($r = +0.32$, $p = 0.09$). In this analysis using a small subset of native nonwords from one talker, phonological coding played a much larger role in the discrimination of native nonsense words than a short-term memory capacity influenced by prior linguistic experience. If these results generalize to other samples of native nonsense words, they indicate that phonological coding, and thus prior linguistic experience, may play a greater role in the identification of more native-like than less native-like nonwords. With the less familiar phonemes and phoneme sequences of non-native nonwords, pure short-term capacity may begin to interact with long-term memory representations in the encoding of phonetic detail.

Test	AXB	ID	DigABS	NonABS	WN
AXB	X				
ID	0.66**	X			
DigAbs	0.44*	0.36	X		
NonABS	0.49**	0.42*	0.38*	X	
WN	0.43*	0.54**	0.51**	0.46**	X

Table 5. Correlations between the discrimination (AXB), identification (ID), digit span (DigABS), nonword span (NonABS), and the word-nonword (WN) condition of the paired-associate word-learning test. In these correlations, only the discrimination and identification results for the native nonsense word were used.
* $p < 0.05$, ** $p < 0.01$

Conclusions

The present study examined the contribution of phonological coding and short-term memory on the perception of non-native contrasts by native speakers of American English. The results of five speech

perception and memory span tests demonstrated that short-term memory span correlated with the identification, but not discrimination, of non-native contrasts. However, a memory span task incorporating information in long-term memory concerning phonological similarity was shown to be the strongest predictor of perceptual performance in the discrimination test. The results support a model of short-term memory in which traces in short-term memory are augmented or transformed by information in long-term memory (Baddeley et al., 1998; Gathercole, Frankish, Pickering, & Peaker, 1999; Schweickert, 1993; Thorn & Gathercole, 1999).

These finding also support a model of cross-language speech perception that incorporates a short-term memory capacity regulating the extent of phonetic detail that is encoded in LTM. Currently, effects of short-term memory are not accounted for in either the Perceptual Assimilation Model or the Native Language Magnet model. In both of these models, the preservation of phonetic detail in the encoding of a non-native word in long-term memory is the consequence of their similarity to one or more native perceptual categories. To account for the results of this study, existing cross-language models should include a fixed short-term memory capacity that filters incoming speech prior to the identification process. Short-term memory could be modeled as a buffer in which phonological information is held prior to encoding in long-term memory, or perhaps as limited attentional resources for activating representations in long-term memory.

To strengthen the claim that short-term memory capacity is an important source of individual differences in cross-language speech perception, additional studies are needed correlating various measures of memory span and speech perception using a variety of listener groups and stimulus sets. In this study, only a small set of nasal consonants from two speakers of Malayalam was presented to monolingual speakers of English. Clearly, a greater range of stimuli (non-native vowel as well as consonant contrasts) and listener groups (multilingual, as well as monolingual speakers of languages other than English) should be used in future studies. In addition, the phonological similarity between stimulus materials in the span tasks and those in the speech perception tasks should be manipulated in several steps to examine the different roles that short-term and long-term memory may play in the encoding of non-native sounds that vary in their similarity to native sounds (roles alluded to in the analysis represented in Table 5). Hopefully, by focusing on individual differences and by examining the role of short-term on speech processing, we can gain a greater understanding of how non-native speech is perceived, encoded, and, ultimately, acquired by the learner.

References

Baddeley, A.D. (1986). *Working memory*. Oxford, England UK: Oxford University Press.

Baddeley, A.D., Gathercole, S.E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105(1)*, 158-173.

Baddeley, A.D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586-595.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-203). Baltimore: York Press, Inc.

Bohn, O.-S. & Flege, J.E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303-328.

Carpenter, P.A., Miyake, A., & Just, M.A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075-1122). San Diego, CA: Academic Press.

Carroll, J.B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87-136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J.B. (1981). Twenty-five years of research on foreign language aptitude. In K. Diller (Ed.), *Individual differences and universals in language-learning aptitude* (pp. 83-118). Rowley, MA: Newbury House.

Flege, J.E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Flege, J.E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*, 395-411.

Flege, J.E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and application* (pp. 565-604). Timonium, MD: York Press.

Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, *41*, 78-104.

Gathercole, S.E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, *23*, 83-94.

Gathercole, S.E., & Baddeley, A.D. (1990). The role of phonological working memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*, 439-454.

Gathercole, S.E., Frankish, C.R., Pickering, S.J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 84-95.

Gathercole, S.E., Service, E., Hitch, G.J., Adams, A.M., & Martin, A.J. (1999). Phonological Short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65-77.

Gathercole, S.E., Willis, C., Emslie, H., & Baddeley, A.D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*, 349-367.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251-279.

Harnsberger, J.D. (1998). *The perception of non-native nasal contrasts: A cross-linguistic perspective*. Doctoral dissertation, University of Michigan, Ann Arbor.

Harnsberger, J.D. (1999). A comparison of three metrics of perceptual similarity in cross-language speech perception. In S. Chang, L. Liaw, & J. Ruppenhofer (Eds.), *Proceedings of the Twenty-fifth Annual Meeting of the Berkeley Linguistics Society* (pp. 157-168). Berkeley, CA: Sheridan Books.

Harnsberger, J.D. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Journal of the Acoustical Society of America*, *108*, 764-783.

Kuhl, P.K. & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Baltimore: York Press, Inc.

Liberman, A.M., Harris, K.S., Kinney, J.A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*, 379-388.

Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242-1255.

Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076-2087.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874-886.

MacKain, K.S., Best, C.T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.

MacKay, I.R.A., Meador, D., & Flege, J.E. (2000). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*, 103-125.

Miyake, A., & Friedman, N.P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A.F. Healy & L.E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum Associates.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics*, *18*, 331-340.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign language vocabulary learning. *Journal of Memory and Language*, *30*, 331-347.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology*, *44A*, 47-67.

Papagno, C., & Vallar, G. (1995a). To learn or not to learn: Vocabulary in foreign languages and the problem with phonological memory. In R. Campbell & M. A. Conway (Eds.), *Broken memories: Case studies in memory impairment* (pp. 334-343). Oxford, England UK: Blackwell Publishers, Inc.

Papagno, C., & Vallar, G. (1995b). Verbal Short-term Memory and Vocabulary Learning in Polyglots. *The Quarterly Journal of Experimental Psychology*, *48A*, 98-107.

Schmidt, A.M. (1996). Cross-language identification of consonants. Part I. Korean perception of English. *Journal of the Acoustical Society of America*, *99*, 3201-3211.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory and Cognition, 21*, 168-175.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology*, *45A*, 21-50.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology*, *51A*, 283-304.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 3-45). Baltimore: York Press, Inc.

Thorn, A.S.C., & Gathercole, S.E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology*, *52A*, 303-324.

Trojano, L., & Grossi, D. (1995). Phonological and lexical coding in verbal short-term memory and learning. *Brain and Cognition*, *21*, 336-354.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*: 47-62.

Appendix

Word-Word pairs for the Paired Associate Word-Learning Task

- restaurant-skeleton
- finger-sheriff
- canyon-pepper
- tornado-computer
- tower-razor
- arena-family
- hamburger-passenger
- staple-neighbor

Word-Nonword pairs for the Paired Associate Word-Learning Task

- explosion-(pekʌrmʌn)
- actor-(fʊltaɪs)
- college-(pʌrnhʌs)
- leather-(sigmeb)
- manager-(remʌrnes)
- table-(heysʌk)
- physician-(kʌrndʌsmaɪd)
- telephone-(sʌtʌrsʌl)

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Some Acoustic Cues for Categorizing American English Regional Dialects: An Initial Report on Dialect Variation in Production and Perception[1]

Cynthia G. Clopper

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by the NIH-NIDCD R01 Research Grant DC00111 and the NIH-NIDCD T32 Training Grant DC00012 to Indiana University. I would like to thank Caitlin Dillon for her assistance in selecting the talkers for this project, Luis Hernandez for his technical assistance and support, Dr. Kenneth deJong for his assistance in selecting the measures for the first experiment, Dr. Robert Nosofsky for his assistance in conducting the Similarity Choice Model and ADDTREE analyses of the data in the second experiment, Dr. David Pisoni for his help and encouragement throughout all stages of this project, and Jimmy Harnsberger for his comments on earlier versions of this paper.

Some Acoustic Cues for Categorizing American English Regional Dialects: An Initial Report on Dialect Variation in Production and Perception

Abstract. Phonological differences between regional dialects of American English are well established in the sociolinguistics literature. The perception of these phonological differences by naïve listeners is much less well understood, however. Using an existing corpus of spoken sentences produced by talkers from a number of distinct regional dialects in the United States, an acoustic analysis was conducted in Experiment I to confirm that certain phonetic features differentiate the dialects. Results provided further evidence for predictable phonological differences between dialects. In Experiment II recordings of the sentences were played to naïve listeners who were asked to categorize each talker into one of six geographical dialect regions. Results suggested that listeners are able to reliably categorize talkers into three broad dialect clusters, but have more difficulty accurately categorizing talkers into six smaller regions. Correlations between the acoustic measures and both actual dialect affiliation of the talkers and dialect categorization of the talkers by the listeners revealed that the listeners in this study were, for the most part, able to reliably use acoustic-phonetic features of the dialects in categorizing the talkers. Taken together, the results of these experiments suggested that naïve listeners are sensitive to phonological differences between dialects and can use these differences to categorize talkers by dialect.

Introduction

Studies of regional dialects in the United States tend to focus on either phonological descriptions of specific dialects or on social aspects of attitudes towards certain dialects, such as perceived "correctness" or stereotypes related to speakers of a given dialect (e.g., Labov, Ash, & Boberg, 1997; Preston, 1986; Preston, 1989; Preston, 1993; Wolfram & Schilling-Estes, 1998). The main focus of phonological investigations of regional dialects of American English is generally the vowel system. The current shift in the vowel systems of two regions in particular has received much attention in the past decade: the Northern Cities vowel shift and the Southern vowel shift. The Northern Cities vowel shift is characterized by a clockwise rotation of the low vowels in the vowel space as shown on the left in Figure 1 and is found in such urban areas as Buffalo, Cleveland, Detroit, and Chicago. The Southern vowel shift, on the other hand, is characterized by a centralization of the tense high vowels and the lengthening of the lax high front vowels as shown on the right in Figure 1. This shift is found more prominently in rural areas of the South, as opposed to the more urban populations that exhibit the Northern Cities vowel shift. A third phenomenon involving vowels in American English that has received attention in the literature is the Low Back Merger in which /ɔ/ and /a/ have merged to make homophones of such pairs as "caught" and "cot" or "Dawn" and "Don." This merger is found in the Midland areas and much of the West, but does not appear to extend to California (Wolfram & Schilling-Estes, 1998).

Labov and his colleagues (1997) have been working on a more complete phonological description of American English, using data collected from telephone surveys of over 600 talkers across the country. The recordings from these talkers are impressionistically transcribed and acoustic measurements of F1 and F2 are taken for each of the vowels they selected to study. Based on the differences in vowel production, the preliminary Phonological Atlas of North America identifies various levels of dialect boundaries that range from a basic North-South-West split to the division of New England into Eastern New England, Western New England, and New York City.

While vowels have been the primary focus of phonological dialect descriptions, such consonantal phenomena as the post-vocalic r-lessness found in New England and some parts of the South, and the "greasy" ~ "greazy" alternation found in the South have also been noted features in discussions of phonological differences (Wolfram & Schilling-Estes, 1998).

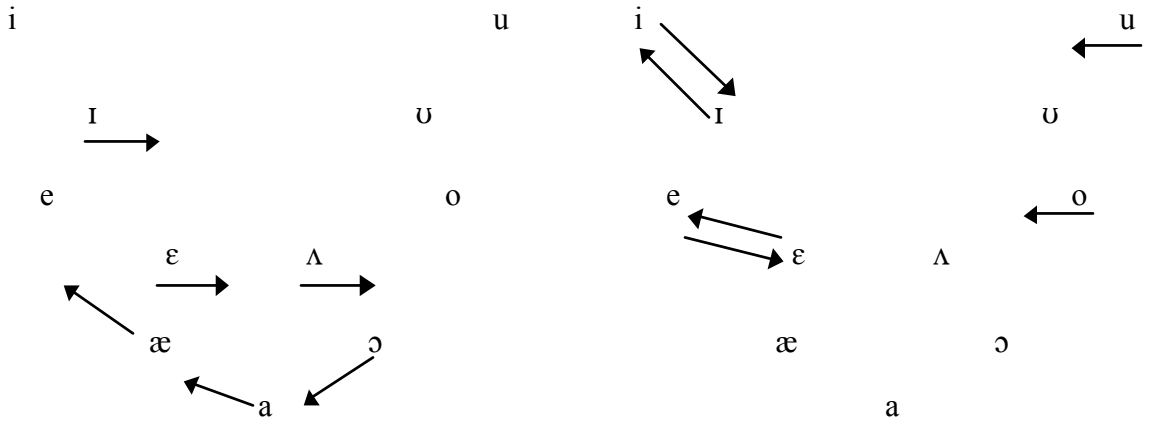


Figure 1. Northern Cities Vowel Shift (left) and Southern Vowel Shift (right). Adapted from Wolfram and Schilling-Estes (1998, pp. 138-139).

When it comes to perceptual work on dialect variation, few studies have been aimed at eliciting data from listeners based on actual speech samples. For example, Preston (1986; 1989) conducted a series of studies in which he asked undergraduates from various parts of the country to complete a number of tasks, including drawing and labeling dialect regions on a map of the United States, ranking all 50 states and a couple of key cities (New York City, Washington, D.C.) on the "correctness" or the "pleasantness" of the English spoken there, etc. Results of the map-drawing studies, conducted in Hawaii, southern Indiana, eastern Michigan, New York City, and western New York, indicated that undergraduates cannot accurately duplicate the dialect boundaries drawn by such researchers as Labov. Comparison between the composite maps of each group indicated that concepts of dialect variation are in part related to where one lives. In general, regions in close geographic proximity to any one respondent group were more finely delineated than regions farther away. It is also interesting to note that in all of the composite maps for these groups, there was at least one area on each map that was not identified as being part of any dialect region (Preston, 1986). Results of the ranking task for informants in southern Indiana indicated that "pleasantness" seems to correspond to geographic proximity to Indiana, whereas "correctness" seems to correspond more to stereotypes of where "standard" English is spoken, with California and the North and Northeast regions receiving the highest rankings (Preston, 1989).

There are two notable exceptions when it comes to the paucity of research involving behavioral responses to speech samples in regional dialect identification. The first is a recent study by Niedzielski (1999) involving listeners from Detroit who were asked to select from a set of six synthetic vowels the one that was the closest match to a vowel produced by a single female talker. One group of listeners was told that the talker was Canadian, while another group was told that the talker was from Michigan. The

results indicated that the listeners who were told that the talker was from Michigan more often selected canonical vowels as the matching vowels, while the listeners who were told that the talker was Canadian more often selected the actual matching vowels. Niedzielski concluded that a priori knowledge of a talker's dialect can affect perception of that talker's speech, particularly in terms of vowel space.

The second study, conducted by Preston (1993), considered the relationship between speech perception and dialect identification from a different perspective. Specifically, undergraduates in Michigan and Indiana were asked to listen to short speech samples taken from interviews with middle-aged males and to assign the different voices to one of nine regions, running north to south between Saginaw, MI and Dothan, AL. Results of this study revealed that the listeners were only able to make broad distinctions between North, South, and Midland. Preston noted that these perceptual boundaries did not correspond to the boundaries drawn by these same listeners in the map-drawing task discussed above. It is also interesting to note that the boundaries perceived by the Indiana residents were different from those perceived by the Michigan residents. Again, it seems that where one lives has an impact on one's perceptions of dialect variation.

While Preston's (1993) study provided some interesting insight into how listeners actually perceive dialectal differences and how those perceptions relate to geographical identification of a talker's home, no one has continued this line of research. The present experiments were designed to identify the acoustic cues that are used by listeners in identifying where a talker is from. Wolfram and Schilling-Estes claim that, "phonological patterns can be diagnostic of regional and social differences, and a person who has a good ear for dialects can often pinpoint a talker's general regional and social affiliation with considerable accuracy based solely on phonology" (1998, p. 67). However, there is little, if any, experimental evidence available to explain how listeners are able to use this knowledge of variation in phonological patterns as a diagnostic for regional identification. Even if the claim that "Southerners are more readily identified as Southerners by their /ay/ vowels than by any other single dialect feature..." (Wolfram & Schilling-Estes, 1998, p. 75) is correct, it would be useful to determine what specific phonetic features discussed in the phonological literature on dialects are actually used by naïve listeners in identifying regional dialects of American English. The goal of the present research was to investigate dialectal variation in both production and perception. Specifically, Experiment I assessed the reliability of some acoustic cues in distinguishing between talkers from different dialects. Experiment II assessed the ability of naïve listeners to use those acoustic cues in categorizing the same set of talkers by dialect region.

Experiment I: Acoustic Analysis

Methods

Talkers. Sixty-six talkers were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Zue, Seneff, & Glass, 1990). The TIMIT corpus consists of recordings of 630 talkers reading 10 sentences each. The corpus includes 438 males and 192 females, and the talkers were each given one of eight regional labels to indicate their dialect: New England, North, North Midland, South Midland, South, West, New York City, or Army Brat. While this database was initially designed for use in speech recognition research, it has been used in a number of phonetic studies looking at the role of gender, dialect, and age in language variation (e.g. Byrd, 1992; Byrd, 1994; Keating, Blankenship, Byrd, Flemming, & Todaka, 1992; Keating, Byrd, Flemming, & Todaka, 1994). Until the present study, it has not been used in perceptual research on dialect variation.

The sixty-six talkers selected for this phonetic study were all white males who were between the ages of 20-29 at the time of recording. Eleven talkers were chosen from each of six dialects regions: New England, North, North Midland, South Midland, South, and West. The talkers were selected by the author and a second phonetically trained listener by first eliminating those talkers who did not meet the age, gender, and race requirement for each of the six dialects. Eleven talkers were then selected from each region based on repeated listening to all ten sentences spoken by each talker such that those chosen shared the most features predicted by their dialect label. Specifically, all of the New England talkers selected were r-less. The Northern talkers were selected based on their degree of /æ/ raising and /oʊ/ fronting. South Midland and Southern talkers selected produced monophthongal /aɪ/. Some Southern speakers also produced fronted /u/ or a merger of /ɛ/ and /ɪ/. The Western speakers who were selected all produced fronted /u/ and some also displayed the merger of /ɛ/ and /ɪ/ or a merger of /a/ and /ɔ/. Finally, the North Midland speakers selected produced none of the characteristic features of the other five dialects.

Stimulus Materials. Of the ten sentences spoken by each talker in the TIMIT database, two of the sentences were read by all of the talkers. These two "calibration sentences" were written to include specific phonemes in certain phonetic contexts in which dialect variation would be predicted (Zue et al., 1990). These two calibration sentences were used in this experiment and are shown in (1) below:

(1) a. She had your dark suit in greasy wash water all year.
b. Don't ask me to carry an oily rag like that.

Each sentence for each talker was contained in a separate sound file that was segmented to include only the sentence material. For the purposes of analysis, the sound files were all leveled to 55 dB using Level16 (Tice & Carrell, 1998).

Procedure. Eleven acoustic measures were obtained from the two calibration sentences for each of the sixty-six talkers and are shown in Table 1. All of the measurements were made using Syntrillium's CoolEdit 96 program. The duration measurements were made directly from the spectrograms. Formant frequency measurements were made using the frequency analysis tool in CoolEdit 96, with a 1024 point Hamming FFT window. Frequency measurements taken at the "midpoint" were taken at the temporal midpoint of the vowel. Frequency measurements taken at the "onset" were taken at the temporal point marking the first third of the vowel. Frequency measurements taken at the "offset" were taken at the second to last glottal pulse of the vowel. All frequency measurements were taken at the peak of a glottal pulse.

In order to provide a means of normalizing frequency measures across the different talkers, the maximum F2 in the word "year" was measured for each talker. The motivation for selecting this particular measure is that the maximum F2 in the vowel /i/ in "year" should indicate the front-most edge of a given talker's vowel space. Comparing this measure to the F2 measures of other vowels can be used to determine the relative backness of those other vowels in the talker's space. Given that all of the talkers used in this experiment were male, the differences due to vocal tract size should be minimal, but taking relative backness measures instead of absolute backness measures should provide a less noisy data set.

The eleven acoustic measures were selected because we expected that they would demonstrate differences between the six dialect regions in terms of production. Four of these measures were obtained from consonants and the remaining seven from vowels. Of the seven vowel measures, three assessed vowel backness and four assessed degree of diphthongization.

New England talkers and some Southern talkers are r-less (Wolfram & Schilling-Estes, 1998). It was predicted that the F3 transition for those talkers would be smaller than for the talkers from the remaining four dialects. As a measure of r-fullness, the F3 transition in "dark" was measured by subtracting F3 at the offset of the vowel from F3 at the midpoint of the vowel.

Two alternations were predicted to distinguish the South and South Midland talkers from the other four dialect groups. An alternation between "wash" and "warsh" is found in some Southern and South Midland talkers. This epenthetic r has the effect of darkening the preceding vowel. We therefore predicted that the Southern talkers, and perhaps the South Midland talkers, should have darker vowels in "wash" than talkers from the other dialects. In order to provide some measure of the effect of this alternation on the brightness of the preceding vowel, the midpoint of F3 in "wash" was measured. There is also a "greasy" ~ "greazy" alternation that occurs in Southern and South Midland speech (Wolfram & Schilling-Estes, 1998). It was predicted that talkers from the South and South Midland would have a greater voiced proportion of the fricative in the word "greasy" and that the fricative duration would be shorter relative to the length of the entire word than for talkers from other dialect regions. This voicing alternation was measured in two ways. The first was the proportion of the fricative that was voiced. The second was the ratio of the duration of the entire fricative to the duration of the entire word.

Word	**Segment**	**Measurement**	**Acoustic-Phonetic Property**
dark	/a/	F3 midpoint – F3 offset	r-fullness
wash	/a/	F3 midpoint	vowel brightness
greasy	/s/	proportion of fricative that is voiced	fricative voicing
		ratio of fricative duration to word duration	fricative duration
suit	/u/	maximum F2 in "suit" – F2 midpoint	/u/ backness
don't	/oʊ/	maximum F2 in "suit" – F2 midpoint	/oʊ/ backness
		F2 midpoint – F2 offset	/oʊ/ diphthongization
rag	/æ/	maximum F2 in "suit" – F2 midpoint	/æ/ backness
		F2 offset - F2 onset	/æ/ diphthongization
like	/aɪ/	F2 offset – F2 midpoint	/aɪ/ diphthongization
oily	/oɪ/	F2 offset – F2 midpoint	/oɪ/ diphthongization

Table 1. Acoustic measures selected for comparison between dialect groups

Southern talkers also produce more fronted /u/ vowels, relative to the northern dialect regions (Wolfram & Schilling-Estes, 1998). Western talkers also demonstrate a similar trend of fronted /u/ productions (Labov et al., 1997). Western and Southern talkers were therefore predicted to have fronted /u/'s and therefore have smaller relative backness values than talkers from the other regions. Northern talkers tend to produce more rounded /oʊ/'s than talkers from the other regions, and this should be reflected in a greater relative backness value for those talkers (Labov et al., 1997). The relative backness of the /æ/ vowel should be smaller for Northern talkers than for any of the other regions due to the upward and forward movement of /æ/ as part of the Northern Cities vowel shift (Wolfram & Schilling-

Estes, 1998). The relative backness of these three vowels was measured in the words "suit," "don't," and "rag" for each talker. The midpoint of F2 in "suit" was measured and then subtracted from the maximum F2 in "year" to obtain a relative backness value of the /u/ vowel. Similarly, the midpoints of F2 in "don't" and "rag" were measured and then subtracted from the maximum F2 in "year" to obtain relative backness values for the vowels /oʊ/ and /æ/.

The diphthongization measure for the /oʊ/ in "don't" was also predicted to separate the Northern talkers from the others, because Northern talkers typically show less diphthongization of this vowel (Labov et al., 1997). Similarly, Southern talkers were expected to show less diphthongization of the /aɪ/ in "like" and the /oɪ/ in "oily," given that there is a tendency for these talkers to produce monophthongal /aɪ/ and /oɪ/ (Wolfram & Schilling-Estes, 1998). There is also some evidence that the /æ/ in "rag" is becoming diphthongized in certain urban regions in the northeast (Labov et al, 1997). Based on this observation, it was predicted that greater diphthongization would be found for this vowel in the speech of New England, and possibly Northern, talkers. Measures of diphthongization were taken by subtracting the offset of F2 from the midpoint of F2 in each of the vowels. In the case of /æ/, the diphthong was measured by subtracting the offset of F2 from the onset of F2, in order to magnify any potential differences between dialect groups.

In summary, New England talkers were predicted to differ from the other talkers on measures of r-lessness and /æ/ diphthongization. Northern talkers were predicted to differ from the others on measures of /oʊ/ backness and diphthongization and /æ/ backness and diphthongization. Southern and South Midland talkers were predicted to differ from the more northern and western talkers on measures of vowel brightness and fricative voicing and duration. Southern talkers were predicted to differ from the other talkers on measures of /u/ backness and /aɪ/ and /oɪ/ diphthongization. Finally, Western talkers were predicted to differ from the others on the measure of /u/ backness.

Results and Discussion

The acoustic analysis confirmed that there are consistent differences in speech production between the six dialects on a number of the acoustic measures considered in this analysis. The means for each of the measures are shown for each dialect group in Table 2. A series of one-way ANOVA's was performed to determine which acoustic measures of speech production reliably distinguish between talkers of different dialects. The r-fullness measure was significant ($F(5, 60) = 3.4$, $p < 0.01$), as were the fricative voicing measure ($F(5, 60) = 7.2$, $p < 0.001$), the fricative duration measure ($F(5, 60) = 4.0$, $p < 0.01$), the /u/ backness measure ($F(5, 60) = 6.6$, $p < 0.001$), the /oʊ/ diphthongization measure ($F(5, 60) = 3.8$, $p < 0.01$), and the /æ/ backness measure ($F(5, 60) = 3.6$, $p < 0.01$). Means of the remaining five measures, vowel brightness, /oʊ/ backness, /æ/ diphthongization, /aɪ/ diphthongization, and /oɪ/ diphthongization were not significantly different.

Post-hoc Tukey tests revealed that New England differed significantly from South Midland and West on mean r-fullness ($p < 0.01$). The mean fricative voicing value for New England differed significantly from South ($p < 0.01$). The mean fricative duration value for North differed significantly from South ($p < 0.01$). The mean value of /u/ backness for New England differed significantly from South Midland, South, and West, and /u/ backness was also significantly different between North and South Midland (all $p < 0.01$). Degree of /oʊ/ diphthongization was significantly different for North and South. Finally, New England and North were significantly different in terms of /æ/ backness.

	New England	North	North Midland	South Midland	South	West
r-fullness (Hz)	262	409	358	462	422	451
vowel brightness (Hz)	2373	2302	2330	2133	2203	2179
fricative voicing (%)	.07	.05	.02	.27	.57	.03
fricative duration (%)	.33	.36	.36	.34	.29	.35
/u/ backness (Hz)	609	557	496	293	337	334
/oʊ/ backness (Hz)	1004	1105	991	1038	1012	939
/oʊ/ diphthong (Hz)	-71	-148	-40	22	37	-41
/æ/ backness (Hz)	601	399	440	425	494	491
/æ/ diphthong (Hz)	256	177	255	280	223	233
/aɪ/ diphthong (Hz)	452	418	402	278	331	350
/oɪ/ diphthong (Hz)	301	384	434	250	226	445

Table 2. Summary of means of acoustic measurement.

In order to determine how well a talker's dialect affiliation is associated with the acoustic properties measured in production, a series of point biserial correlations was performed. For each talker, the value on each acoustic measure (on a continuous scale) was correlated with dialect affiliation. Dialect affiliation was quantified dichotomously, such that the eleven talkers from a given dialect were given a value of "1" for that region and the remaining fifty-five talkers were given a value of "0" for that region. Results of these correlations are shown in Table 3. These correlations indicate that, as predicted, r-lessness is associated with New England talkers. New England talkers also have a greater degree of backness in /u/'s and /æ/'s, which was an unpredicted result. South Midland talkers have fronted /u/'s, which was predicted for the Southern talkers. By contrast, Southern talkers have predictably high amounts of fricative voicing in "greasy" and a predictably short fricative in the same word, but the South Midland talkers do not. Northern talkers display the predicted monophthongal /oʊ/. North Midland and West talkers do not show any strongly predictable measures from this analysis. Additionally, the measures of vowel brightness, /oʊ/ backness, and all three diphthongs did not distinguish any of the dialect groups. These correlations suggest that while many of the measures differ in their means between dialects, only a handful are truly associated with a talker's dialect affiliation. While these acoustic properties can be associated with dialect regions, they are not necessarily the only features, or the most important features, of that dialect region. The data analyzed in this experiment suggest only that some of these properties can be associated with dialect affiliation. The acoustic measures associated with dialect affiliation are, therefore, "characteristic features" of that dialect.

The results of this acoustic analysis confirm that these talkers can be reliably distinguished by dialect based on a handful of consistent acoustic differences in speech production. The following perceptual experiment was designed to investigate how well naïve listeners can use these consistent differences to categorize talkers by dialect based on short speech samples.

	New England	North	North Midland	South Midland	South	West
r-fullness	**-.41****	.05	-.11	.21	.09	.18
vowel brightness	.24	.10	.16	-.25	-.10	-.15
fricative voicing	-.13	-.17	-.20	.14	**.55****	-.19
fricative duration	-.08	.23	.16	-.01	**-.44****	.14
/u/ backness	**.38***	.26	.13	**-.32***	-.22	-.23
/oʊ/ backness	-.03	.24	-.06	.06	-.01	-.20
/oʊ/ diphthong	-.11	**-.39****	.00	.22	.28	.00
/æ/ backness	**.41****	-.25	-.11	-.16	.06	.05
/æ/ diphthong	.07	-.24	.07	.17	-.06	-.01
/aɪ/ diphthong	.23	.13	.08	-.26	-.11	-.06
/oɪ/ diphthong	-.09	.10	.21	-.20	-.25	.23

Table 3. Correlations between talker dialect affiliation and acoustic measures. N = 66 for all correlations. Correlations with significance at $p < 0.01$ are in **bold,** * indicates $p < 0.01$, ** indicates $p < 0.001$.

Experiment II: Perceptual Categorization Task

Methods

Stimulus Materials. The same stimulus materials were used in this study as in Experiment 1 above.

Listeners. Twenty-three Indiana University undergraduates served as listeners for this study. All received partial credit for an introductory psychology course for their participation. Data from five of the listeners were removed prior to analysis: 2 were non-native speakers and 3 performed statistically at chance on the task. The eighteen remaining listeners, five males and thirteen females, were all monolingual native speakers of American English with no history of hearing or speech disorders. These eighteen listeners were divided into three listener groups based on residential history. The seven listeners who had only lived in Northern Indiana (north of, and including, Indianapolis) prior to attending school in Bloomington comprised the Northern Indiana group. The five listeners who had only lived in Southern Indiana comprised the Southern Indiana group. The remaining 6 listeners had all lived out of state for some period of time prior to attending school in Bloomington and they comprised the Out-of-State group.

Procedure. The listeners were seated at personal computers equipped with KeyTec Inc. pressure sensitive activation touch screens (KTMT1315 ProE). On the screen were the six dialect regions, represented by partial maps of the United States, including state boundaries that were labeled with the name of the dialect region. The six regions are shown in Figure 2 as they were arranged on the screen. The regions were roughly 2” x 2” in dimension and adequate space was left between the regions to minimize error in the response process. Prior to beginning the experiment, the regions were displayed on the screen and the listeners were encouraged to familiarize themselves with the regions. In the first phase

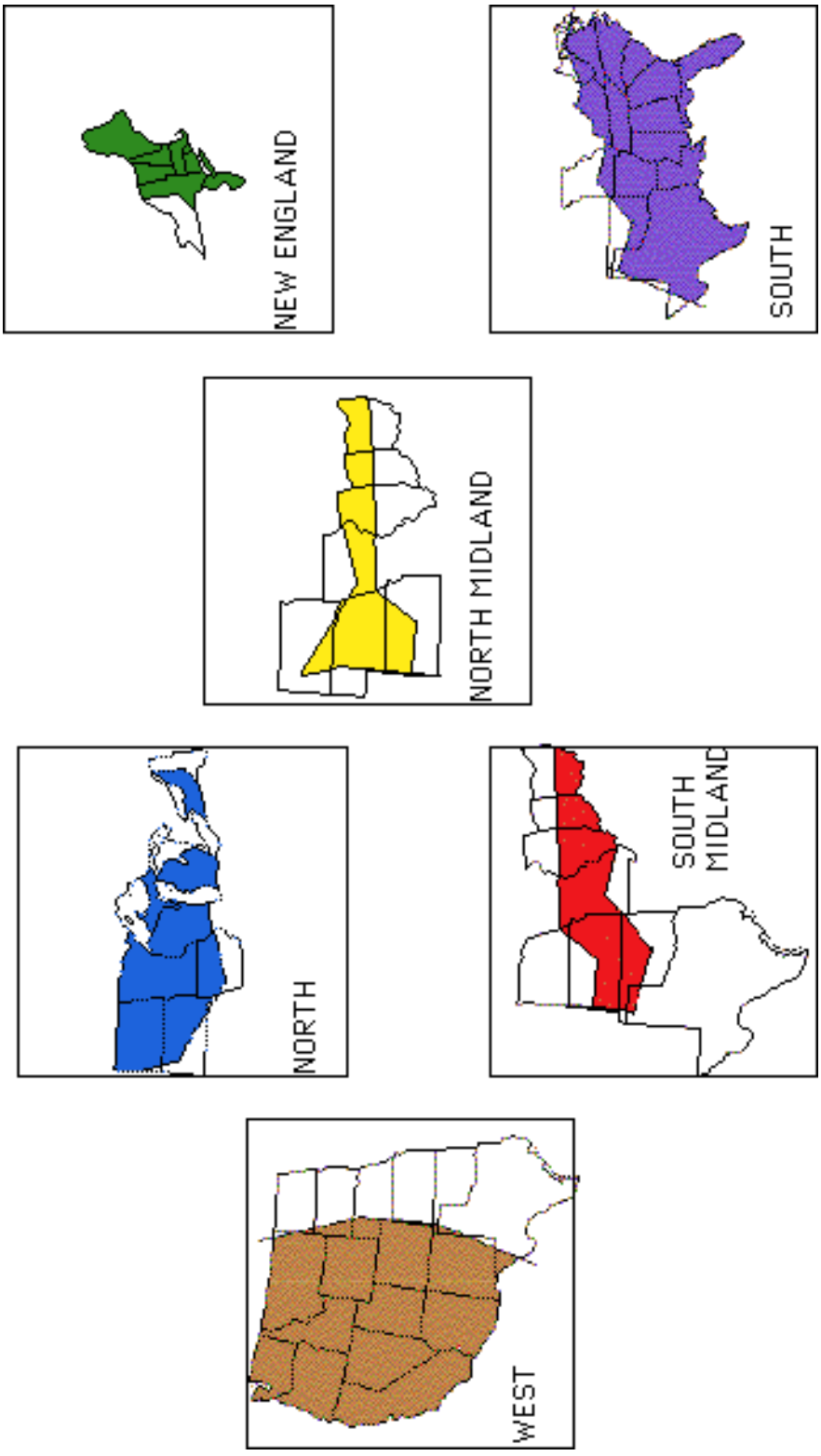


Figure 2. The six response alternatives in the categorization task. Based on Wolfram and Schilling-Estes (1998, p. 122).

of the task, the listeners responded to the first calibration sentence as spoken by each of the sixty-six talkers one time, presented in random order. On each trial, listeners heard a sentence produced by one of the sixty-six talkers, presented over headphones (Beyerdynamic DT100) at 70 dB SPL. The listeners were instructed to listen to the sentence carefully and to select the region on the screen that they thought the talker was from. The listeners made their responses by pressing directly on the screen. The listeners received no feedback about the accuracy of their responses. The second phase of the task was identical to the first, except that the listeners responded to the second calibration sentence as spoken by each of the sixty-six talkers one time, presented in random order.

Results and Discussion

Overall performance on the categorization task was quite poor. Listeners in the Out-of-State group, Northern Indiana group, and Southern Indiana group performed similarly in terms of proportion correct identification. Taken together, the three groups of listeners were only able to correctly identify where 33% of the talkers were from on the first calibration sentence and where 28% of the talkers were from on the second calibration sentence. While overall performance was low, it was statistically above chance for both sentences. The proportions of correct identifications for talkers from each dialect region for each sentence are shown in Table 4, collapsed across all three listener groups. A t-test indicated that the performance for the two sentences was not significantly different ($t(34) = 3.21$, $p = 0.55$).

	First Sentence	Second Sentence
New England	61	34
North	23	26
North Midland	25	27
South Midland	34	27
South	35	34
West	23	20
Mean	**33**	**28**

Table 4. Percent correct categorization of dialect affiliation of the talkers for each sentence, collapsed across the three listener groups (chance = 17%).

An inspection of the confusion matrices of responses suggested that the listeners' inability to correctly identify a majority of the talkers was not due to random responses, but was more likely due to a consistent pattern of confusions. In order to determine the structure of this pattern of errors, the 6 x 6 confusion matrices for each of the two calibration sentences for each listener group, and collapsed across all three listener groups, were submitted to the Similarity Choice Model (Nosofsky, 1985) to determine similarity and bias parameters between the dialect regions. The similarity parameters indicated the degree of similarity between each of the dialects, based on the confusion data. The bias parameters indicated the responses biases of the listeners. The bias parameters that resulted from the Similarity Choice Model analysis suggested that the listeners were not biased to respond with one alternative more or less often than any of the other response alternatives. The similarity parameters were submitted to an additive clustering scheme, ADDTREE, to determine one measure of the perceptual distances between the dialects (Corter, 1995). An additive clustering scheme was selected because the initial examination of the confusion matrices indicated that there was high reciprocity between the six regions. For example, South was most often confused with South Midland and vice versa. Other spatial analyses, such as multi-dimensional scaling, were inappropriate for this data given the small number of data points in the matrix.

The perceptual distances for the listener groups were highly correlated with each other and with the distances for all of the listener groups combined, demonstrating no significant differences between the three listener groups. All further analyses considered the data collapsed across all of the listeners. The resulting trees from the ADDTREE analysis collapsed across listener groups are shown in Figure 3. For the first calibration sentence, it is clear that listeners grouped the talkers into three main clusters: New England, South and South Midland (hereafter, South Cluster), and North, North Midland, and West (hereafter, Other Cluster). The solution for the second calibration sentence also appears to have three broad clusters: New England and North (hereafter, North Cluster), the South and South Midland (hereafter, South Cluster), and the North Midland and West (hereafter, West Cluster).

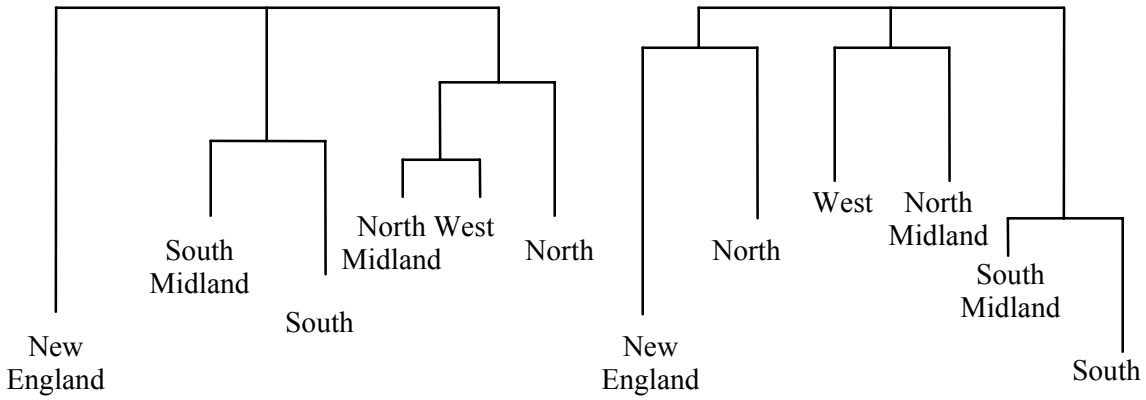


Figure 3. Clustering solution for the first (left) and second (right) calibration sentence, based on listeners' confusion matrices.

When the proportion correct categorization scores for all the listeners are collapsed into the three broad clusters for each of the two calibration sentences, performance increases dramatically, as expected. Correct categorization of talkers into New England, South Cluster, or Other Cluster for the first calibration sentence was 67%. Correct categorization of talkers into North Cluster, South Cluster, or West Cluster for the second calibration sentence was 53%. These results suggest that listeners are able to reliably categorize talkers into three broad dialect groups, rather than the six used in the TIMIT corpus. The different clustering results from the two sentences suggest that these three categories might be fluid, depending on which phonetic cues are available for identifying a talker. Recall that the first sentence contained the word "dark" and that r-lessness was a characteristic feature of the New England talkers. If listeners were able to use r-fullness as a cue in identifying talkers, it is not surprising that New England was in a cluster by itself for the first sentence when that cue was available, but that it grouped with another region when that cue was not available, as in the second sentence.

In order to determine which phonetic cues the listeners were using to categorize the talkers, a series of correlations was performed. For each talker, the value on each acoustic measure was correlated with the percent categorization of that talker into a given dialect region over all listeners. Results of these Pearson correlations are shown in Table 5. They suggest that listeners may use some of these cues in order to categorize talkers by dialect. For example, it seems that listeners can use r-lessness to identify talkers from New England, vowel darkness to identify talkers from the South Midland, fricative voicing to identify talkers from the South and South Midland, /u/ frontness to identify talkers from the South

Midland, /oʊ/ diphthongization to identify talkers from the South, /oʊ/ monophthongization to identify talkers from the North, /aɪ/ diphthongization to identify talkers from the North Midland, /oɪ/ diphthongization to identify talkers from the North Midland and the West, and /aɪ/ and /oɪ/ monophthongization to identify talkers from the South.

	New England	**North**	**North Midland**	**South Midland**	**South**	**West**
r-fullness	**-.40****	-.06	.07	.30	.24	.02
vowel brightness	.28	-.01	.04	**-.52****	-.14	.16
fricative voicing	-.16	-.28	-.27	**.33***	**.42****	-.13
fricative duration	.02	.12	.31	-.22	-.29	.19
/u/ backness	.31	.25	-.09	**-.44****	-.31	.19
/oʊ/ backness	.14	.13	-.04	-.15	.02	-.17
/oʊ/ diphthong	-.29	**-.43****	-.06	.20	**.39****	-.03
/æ/ backness	.22	-.01	.14	-.19	-.15	.01
/æ/ diphthong	-.12	.01	-.10	-.07	.21	-.02
/aɪ/ diphthong	.00	.20	**.37***	-.14	**-.33***	.11
/oɪ/ diphthong	-.15	.21	**.57****	-.22	**-.45****	**.45****

Table 5. Correlations between acoustic measures and dialect categorization. N = 66 for all correlations. Correlations with significance at $p < 0.01$ are in **bold, *** indicates $p < 0.01$, ****** indicates $p < 0.001$.

The results of the two experiments taken together demonstrate that some acoustic measures are associated with a talker's dialect affiliation and that some acoustic cues are associated with how listeners categorize a given talker. In order to determine whether or not listeners use the characteristic acoustic features of the dialects in their categorization of the talkers, the correlations from the acoustic analysis have been plotted with the correlations from the perceptual experiment for each dialect region. These plots are shown in Figure 4. Plotted on the x-axis are the squared correlation coefficients from Experiment I, which reveal the proportion of variance (r^2) in the acoustic measures accounted for by the actual dialect affiliation of the talkers. Plotted on the y-axis are the squared correlation coefficients from Experiment II, which reveal the proportion of variance (r^2) in the dialect categorization of the talkers accounted for by the acoustic measures. The acoustic measures from both calibration sentences have been plotted together in these figures. If listeners used the acoustic cues optimally, the points would form a line with a slope = 1. Any points falling above the line x = y represent those acoustic cues which are not characteristic features of the dialect, but which the listeners used in their categorization of the talkers. For example, in Figure 4c, the point representing degree of /oɪ/ diphthongization falls above the line x = y. This indicates that despite the fact that /oɪ/ diphthongization is not a characteristic feature of the North Midland dialect, listeners used this feature to discriminate North Midland talkers from other talkers. Conversely, any points falling below the line x = y represent those acoustic cues which are characteristic features of the dialect, but which listeners did not use in their categorization of the talkers. For example, in Figure 4a, the point representing /æ/ backness falls below the line x = y. This indicates that despite the fact that /æ/ backness is a characteristic feature of New England, the listeners did not use this feature to discriminate New England talkers from other talkers.

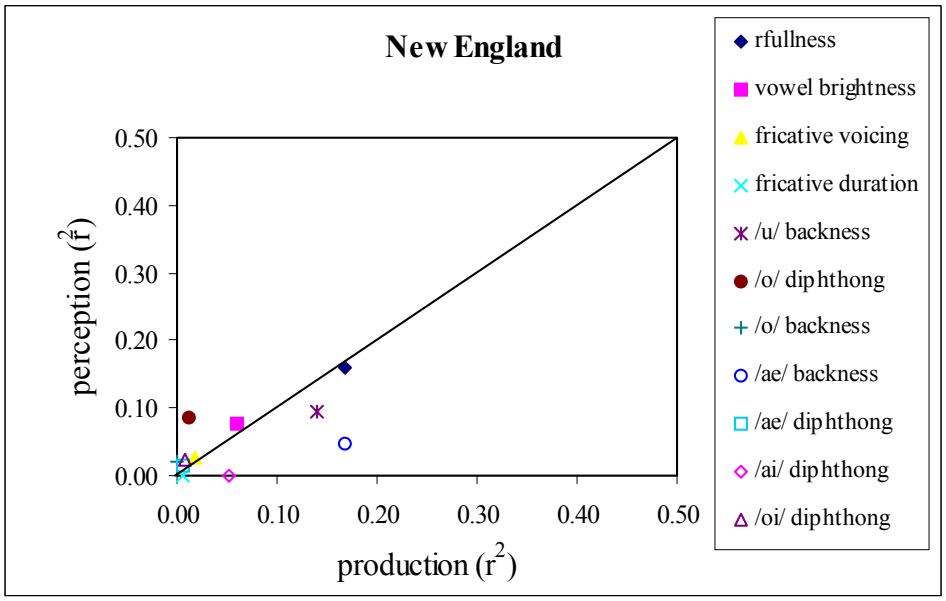


Figure 4a. Presence of features in production for both sentences v. perception by listeners in categorization for New England.

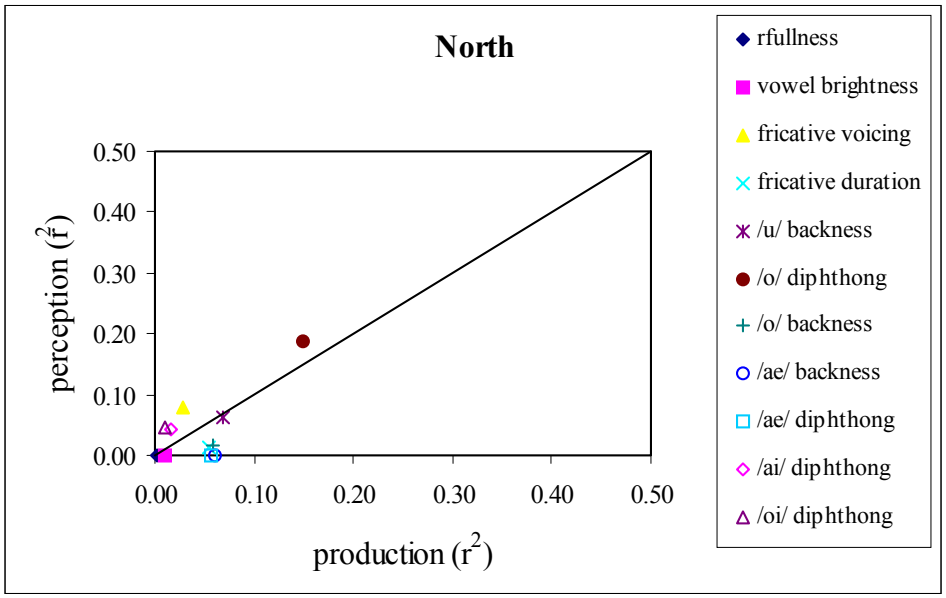


Figure 4b. Presence of features in production for both sentences v. perception by listeners in categorization for North.

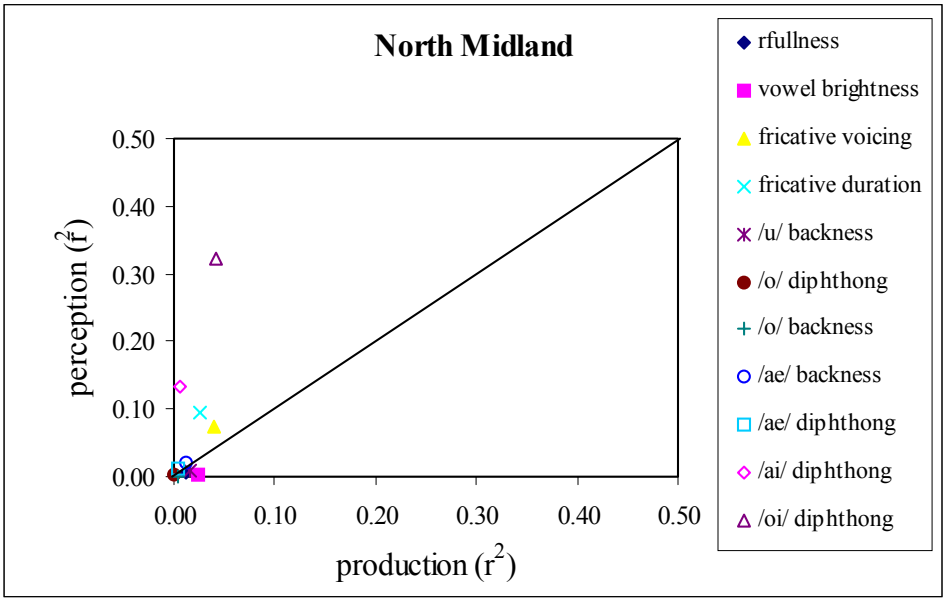


Figure 4c. Presence of features in production for both sentences v. perception by listeners in categorization for North Midland.

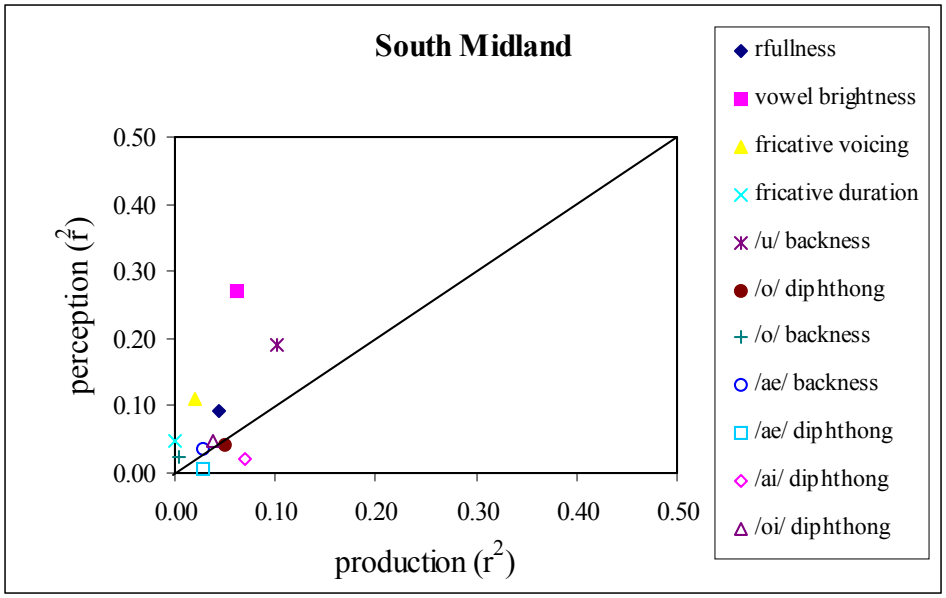


Figure 4d. Presence of features in production for both sentences v. perception by listeners in categorization for South Midland.

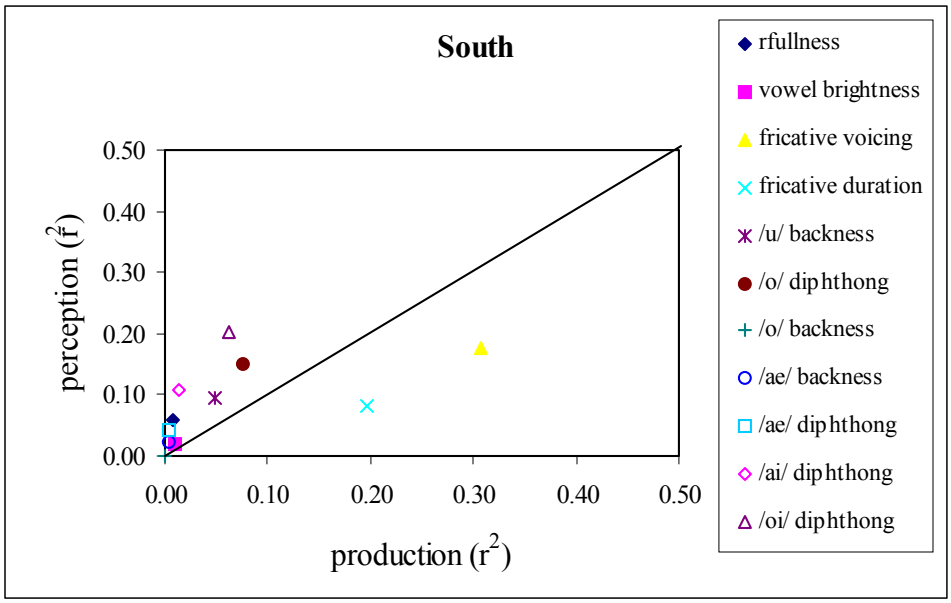


Figure 4e. Presence of features in production for both sentences v. perception by listeners in categorization for South.

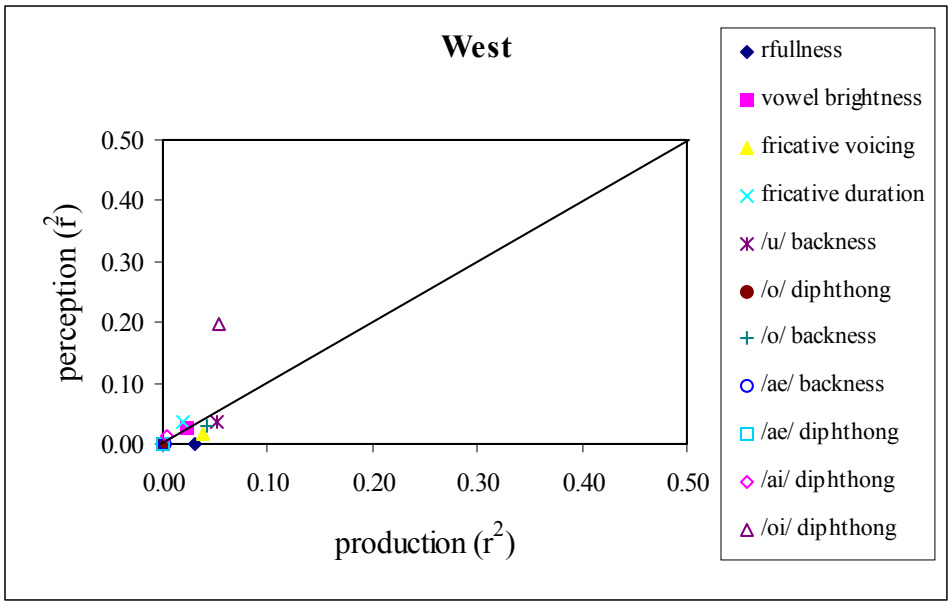


Figure 4f. Presence of features in production for both sentences v. perception by listeners in categorization for West.

These plots show several things of interest with respect to the relationship between production and perception. The first is that for all six of the dialect regions, there is a cluster of cues close to the origin. These cues are neither useful in predicting dialect affiliation nor are they used by listeners to categorize the talkers. The second notable point is that for New England, North, and South, the points not

at the origin tend to fall close to the x = y line. For the North Midland, South Midland, and West, however, the points not clustered at the origin tend to fall lower on the production scale than the perception scale, indicating that the listeners were using non-characteristic features of those regions in assigning talkers to those regions. These two observations taken together indicate that listeners are much more capable of identifying and using the appropriate acoustic cues for New England, North, and South, than they are for North Midland, South Midland, and West.

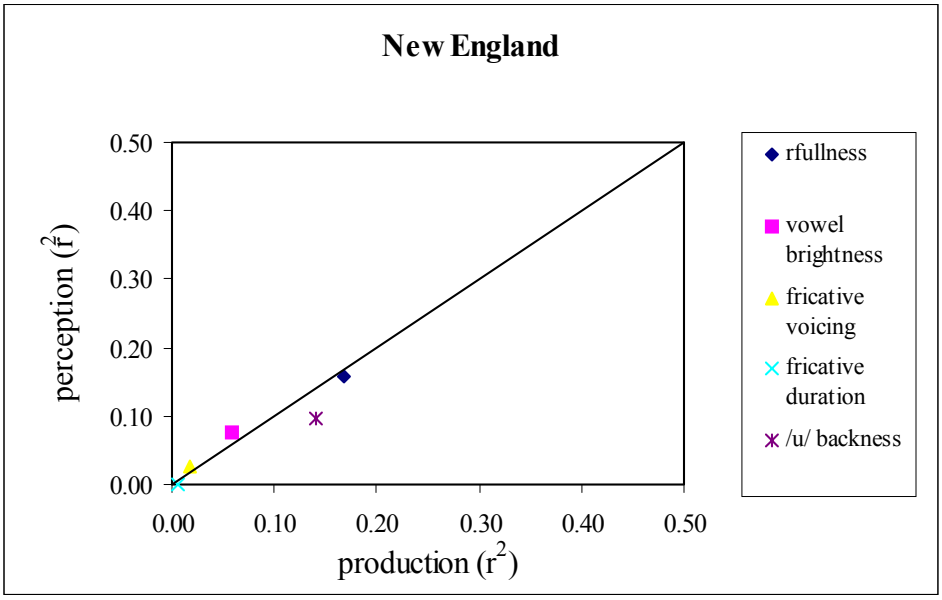


Figure 5a. Presence of features in production for the first sentence v. perception by listeners in categorization for New England.

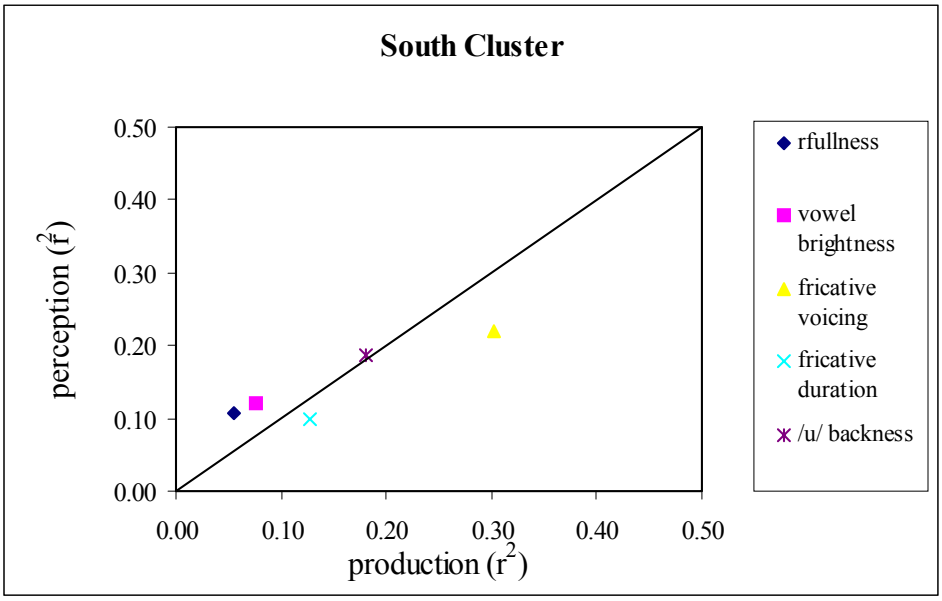


Figure 5b. Presence of features in production for the first sentence v. perception by listeners in categorization for South Cluster (South and South Midland).

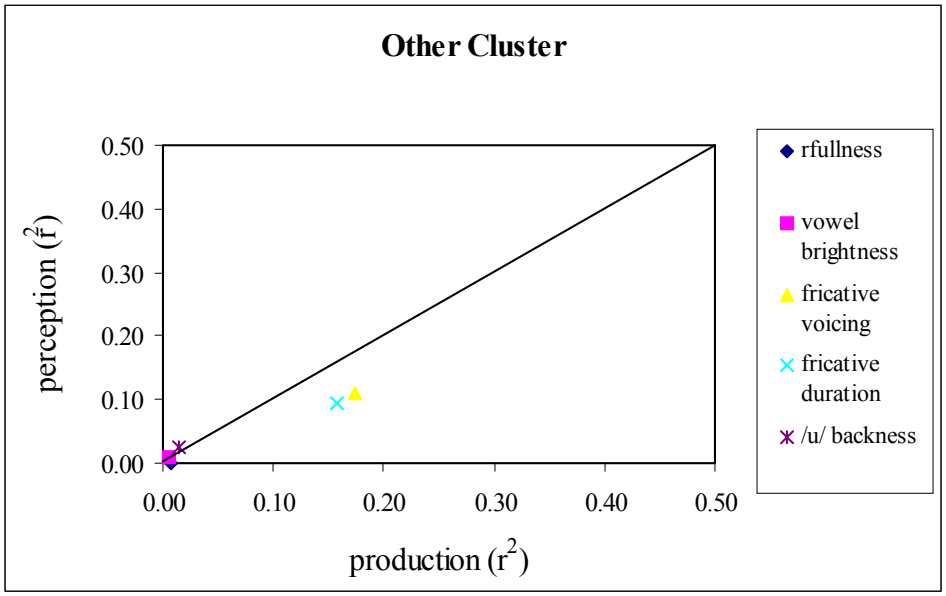


Figure 5c. Presence of features in production for the first sentence v. perception by listeners in categorization for Other Cluster (North, North Midland, West).

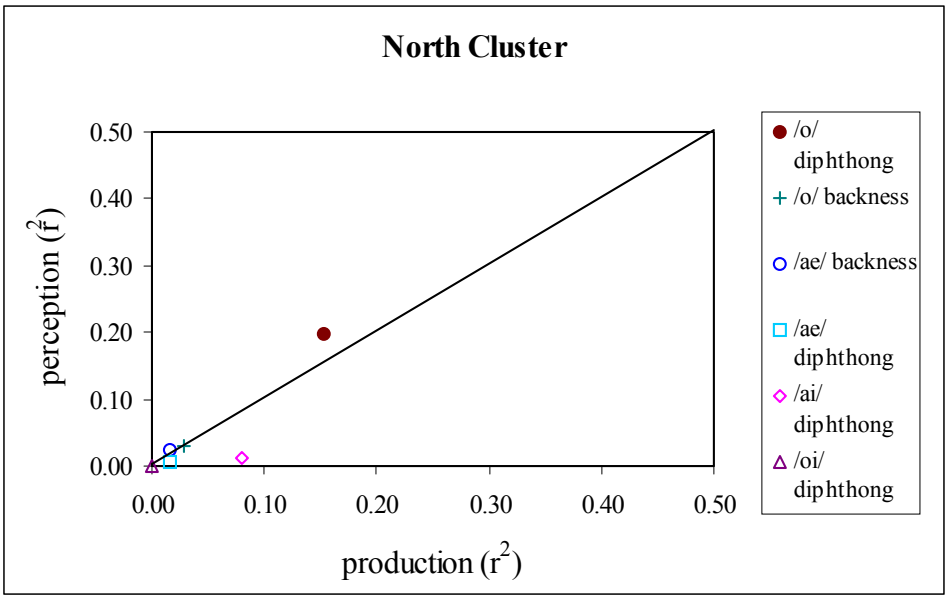


Figure 6a. Presence of features in production for the second sentence v. perception by listeners in categorization for North Cluster (New England and North).

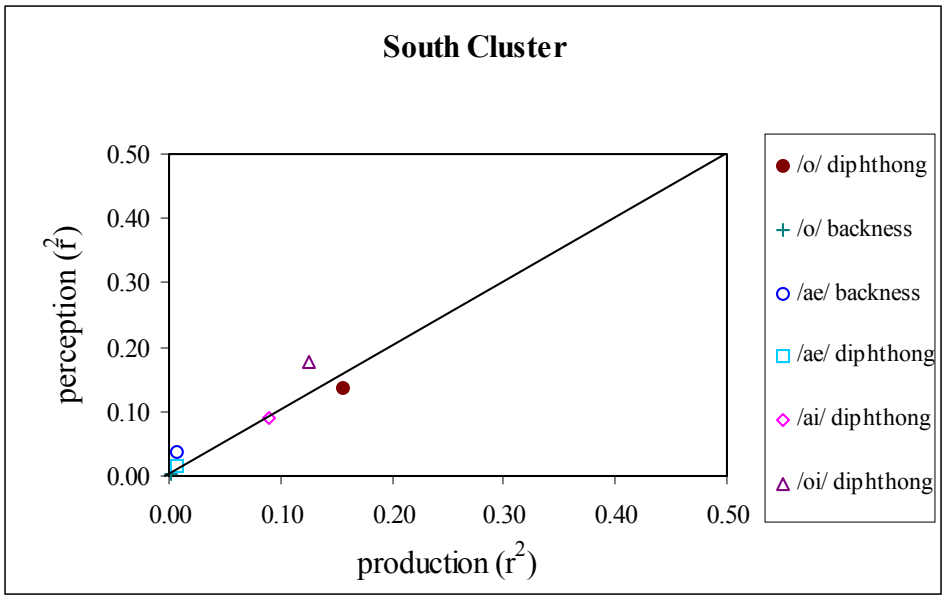


Figure 6b. Presence of features in production for the second sentence v. perception by listeners in categorization for South Cluster (South Midland and South).

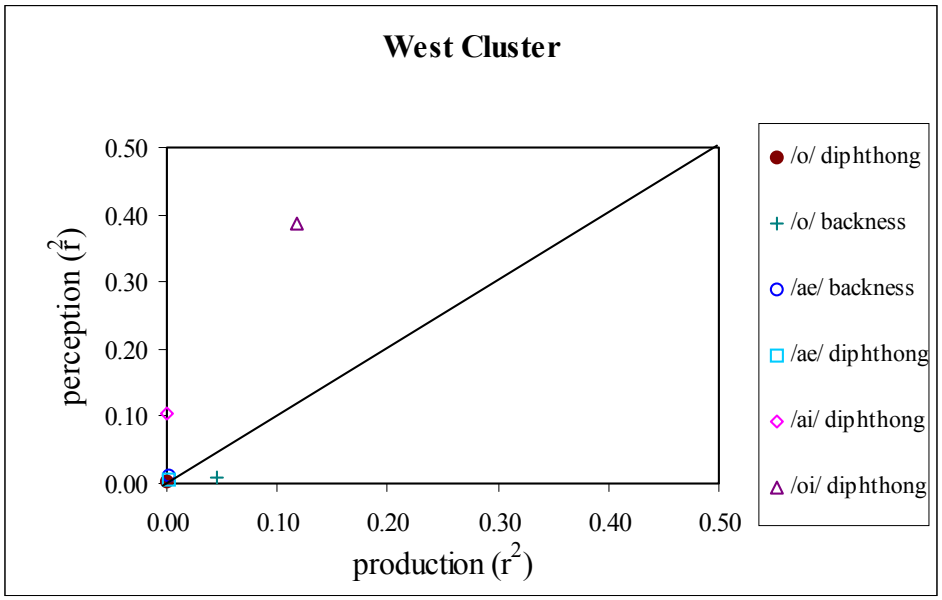


Figure 6c. Presence of features in production for the second sentence v. perception by listeners in categorization for West Cluster (North Midland and West).

One possible explanation for these differences is that the regions in the latter group are less familiar to the listeners as distinct "dialect regions." The results of the clustering analysis above indicated that the listeners used fewer than six dialect categories reliably in this task. Therefore, a set of point biserial correlations between cluster affiliation and acoustic measures and a set of Pearson correlations between acoustic measures and percent cluster categorization were performed in the same manner as above, using data from all sixty-six talkers. The results of the point biserial correlations revealed the characteristic features of the dialect clusters and the results of the Pearson correlations revealed which cues the naïve listeners were using in categorizing the talkers by cluster. The squared correlation coefficients were then plotted against each other to provide an index of how well listeners used the acoustic cues that are good predictors of cluster affiliation. In these plots, the acoustic cues from the two sentences have been plotted separately because the clustering solutions differed for the two sentences. The plots for the first calibration sentence are shown in Figure 5 and the plots for the second calibration sentence are shown in Figure 6. This series of plots shows that listeners are relatively good at using the appropriate cues for all but the West Cluster. For the West, listeners tend to use acoustic cues that are not characteristic features of that cluster. However, none of the acoustic measures obtained in this study were highly correlated with the North Midland or the West, so it is perhaps arguable that the reason talkers from this cluster are difficult to categorize is because the cluster does not have characteristic features to distinguish it from the other regions. That is, there are no "good" cues for the listeners to use. While listeners are able to use r-lessness to identify talkers from New England and /oɪ/ diphthongization to identify talkers from the South, none of the acoustic properties examined in the acoustic analysis were highly associated with the talkers from the West Cluster. Therefore, the listeners may not have had any acoustic cues available to them in making their categorization judgements of the North Midland and West talkers.

General Discussion

As predicted, the acoustic analyses performed in the first experiment confirmed that, as a group, the talkers selected from each dialect reliably produce phonological differences that can be measured acoustically. Specifically, for this set of talkers, r-lessness, /u/ backness, and /æ/ backness are characteristic features of the eleven New England talkers. /oʊ/ monophthongization is a characteristic feature of the eleven Northern talkers. /u/ frontness is a characteristic feature of the eleven South Midland talkers. Finally, fricative voicing and duration are characteristic features of the eleven Southern talkers. None of the acoustic measures selected for this analysis were characteristic features of either the eleven North Midland talkers or the eleven West talkers.

Some of the acoustic measures that were expected to reveal differences between the dialect groups were not predictably different between the dialects. Specifically, the vowel brightness in "wash" was expected to distinguish the South and South Midland from the other dialects. However, the correlation between South Midland dialect affiliation and this measure was weak (r = -0.25). This measure based on F3 values is problematic, however, because it was not normalized across speakers for vocal tract size, unlike the measures involving F2 that were normalized against the F2 of "year" to account for talker differences. This measure is also potentially problematic because the vowel itself can take on a different quality in different dialects. Additionally, the measures for the diphthongs /aɪ/ and /oɪ/ were also predicted to distinguish the South and South Midland talkers from the others. The correlation between South Midland affiliation and the measure of /aɪ/ diphthongization was weak (r = -0.26) as was the correlation between South affiliation and the measure of /oɪ/ diphthongization (r = -0.25). The measure of degree of diphthongization of /aɪ/ is potentially problematic in this analysis because it was taken from the word "like." A following velar context generally results in an upward offglide of the

preceding vowel (Ladefoged, 1993). This upward offglide may have concealed the expected monophthongization of /aɪ/ in the South and South Midland talkers. These weak associations suggest that while some of the predictions based on the current sociolinguistic literature were not entirely confirmed, there is still a relationship between some phonetic features and dialect affiliation.

Another possible explanation for the lack of correlation between some of the acoustic measures and dialect affiliation is that some of the talkers selected for this study were not good representatives of their dialect region. The standard deviations of the means shown in Table 2 reveal that there was a lot of variation between the talkers within any given dialect group. It may be the case that certain talkers in a given dialect are better representatives of their region than others. That is, some talkers may more reliably produce the phonetic features that distinguish their dialect from others and some talkers may be more easily categorized by listeners than others. Additionally, there may be some striking individual differences between the listeners that can account for some of the data presented here. Analyses of the individual talkers and the individual listeners have not been completed, but may provide some insight into why some predicted correlations did not emerge. Finally, it is possible that the regions used to define the talkers in this study are not the most accurate categorization of these talkers. For example, some recent research suggests that the Midland areas should be considered as one single region. There is also some controversy about the vast geographical area contained within the Western region (Labov et al., 1997).

The results of the categorization task in the second experiment support the findings of Preston (1993) that indicate that naïve listeners are only able to categorize talkers based on dialect into broad categories. Specifically, the listeners in this experiment were able to reliably categorize the sixty-six talkers into three broad dialect categories: North, South, and West. The placement of Northern talkers into one of these three clusters appeared to be based on the availability of r-fullness as an acoustic cue. In the first calibration sentence, the r-fullness cue was available, and the listeners used this to identify talkers from New England and placed Northern talkers in the West Cluster. In the second calibration sentence, the r-fullness cue was not available to identify New England talkers and the listeners placed Northern talkers in the North Cluster.

The listeners also demonstrated reliable use of a number of the acoustic cues in categorizing the talkers. Specifically, r-less talkers were categorized as New Englanders. Talkers with a highly diphthongal /oʊ/ were categorized as Northerners. Talkers with a highly diphthongal /aɪ/ and /oɪ/ were categorized as North Midlanders. Talkers with a dark vowel in "wash," a voiced fricative in "greasy," and a fronted /u/ were categorized as South Midlanders. Talkers with a voiced fricative in "greasy," a highly diphthongal /oʊ/, and a highly monophthongal /aɪ/ and /oɪ/ were categorized as Southerners. Finally, talkers with a highly diphthongal /oɪ/ were categorized as Westerners.

Despite the consistent use of some of the acoustic cues, the listeners were not always using the most optimal cues in their decisions. That is, the most characteristic features of each dialect region, as revealed by the point biserial correlations in the first experiment, were not always the acoustic properties used by the listeners. For example, /æ/ backness was a fairly good cue characterizing New England, but the listeners did not use it. Fricative duration and voicing were also relatively good cues characterizing the South that the listeners did not use optimally. Conversely, degree of /oɪ/ diphthongization was not a good characteristic feature of North Midland, South, or West talkers, but the listeners relied heavily on this measure as an indicator of dialect region in all three cases. Similarly, vowel brightness was not a particularly good characteristic feature of South Midland talkers, but the listeners relied heavily on this cue as well.

Overall, the comparison between the two sets of correlations based on dialect regions in Figure 4 suggests that listeners used the characteristic features of New England, North, and South more optimally than those of the Midland regions and the West. The results of the clustering analysis suggested that listeners can better distinguish between three broad dialect clusters than between the six smaller regions. It is therefore reasonable to consider how well the listeners used the characteristic features of the clusters in their categorization of the talkers. The comparison between the two sets of correlations based on dialect clusters in Figures 5 and 6 suggest that listeners were in fact using the characteristic features of all of the clusters, except the West Cluster. Recall that the West Cluster is composed of the North Midland and the West regions. The results of the acoustic analyses revealed that there are no characteristic features for either of these regions in the set of phonetic features considered here. Therefore, it is not at all surprising that the listeners were relying on a feature that is not characteristic of the cluster in categorizing the talkers, because there is no reliable feature in the talkers' productions to rely on. Regardless of whether or not the listeners used the characteristic features of the dialect regions optimally in the categorization of the talkers, however, it is clear that naïve listeners are sensitive to a number of phonological differences between dialects and that extensive training is not required before listeners can use these differences to accurately identify where talkers are from, at least in terms of broad dialect clusters.

In addition to continuing to analyze the possible individual talker and listener differences in this data, this line of research can be extended in various ways. Specifically, the relatively poor performance by the listeners in the categorization task raises several issues regarding possible manipulations of the task, such as training the listeners on representative speakers of each dialect and having them generalize to new talkers or providing the listeners with a smaller set of response alternatives. Additionally, further analyses can be conducted to determine the perceptual similarities between the dialect regions and between the talkers in each region.

Conclusions

The results of the first experiment using acoustic measurement techniques provide further evidence that phonological differences do exist between regional dialects of American English and that differences in speech production can be predicted to some extent by the dialect affiliation of the talkers. The results of the second experiment provide perceptual evidence that supports Preston's (1993) findings that naïve listeners do not necessarily categorize talkers accurately by dialect region, but that they are able to make reliable distinctions between some dialect groups on a broader scale. In particular, the naïve listeners were able to reliably identify talkers from the South, the North, and New England, but they had a harder time identifying talkers from the Midland areas and the West. The results of these two experiments together suggest that listeners are aware of important phonological differences between dialects and can use their detailed knowledge to categorize talkers by dialect region, without any specific training or feedback.

References

Byrd, D. (1992). Sex, dialects, and reduction. *ICSLP 92 Proceedings*, 827-830.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15,* 39-54.

Corter, J.E. (1995). ADDTREE/P Program for Fitting Additive Trees.

Keating, P., Blankenship, B., Byrd, D., Flemming, E. & Todaka, Y. (1992). Phonetic analyses of the TIMIT corpus of American English. *ICSLP 92 Proceedings*, 823-826.

Keating, P.A., Byrd, D., Flemming, E., & Todaka, Y. (1994). Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication*, *14*, 131-142.

Labov, W., Ash, S., & Boberg, C. (1997). A National Map of the Regional Dialects of American English. Retrieved June 26, 2000 from the World Wide Web: http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html.

Ladefoged, P. (1993). *A Course in Phonetics*. Fort Worth, TX: Harcourt Brace.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*, 62-85.

Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice-model analysis. *Perception and Psychophysics*, *38*, 415-432.

Preston, D. (1986). Five visions of America. *Language and Society*, *15*, 221-240.

Preston, D. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. Providence, RI: Foris.

Preston, D. (1993). Folk dialectology. In Preston, D. (ed.) *American Dialect Research*. Philadelphia: John Benjamins, pp. 333-378.

Tice, R. & Carrell, T. (1998). Level16 v.2.0.3. University of Nebraska.

Wolfram, W. & Schilling-Estes, N. (1998). *American English*. Malden, MA: Blackwell.

Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication*, *9*, 351-356.

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Prosodic and Morphological Effects on Word Reduction in Adults: A First Report [1]

Allyson K. Carter and Cynthia G. Clopper

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by the NIH-NIDCD Training Grant DC00012 to Indiana University. We thank David Pisoni for valuable discussion and guidance. We also thank Luis Hernández for his programming assistance and Mark VanDam for his help with data analysis.

Prosodic and Morphological Effects on Word Reduction in Adults: A First Report

Abstract. Several populations, such as normally developing children around the age of two years, children with language impairments, and adults with aphasia, all share a similar documented phenomenon in their language production: omitting syllables from their speech. Omitted syllables are most often those that are weakly stressed and that directly precede the primary stress of a word, yielding such stress-initial forms as *nána* for *banána* and *ráffe* for *giráffe*. This phenomenon is reflected in the English prosodic system; that is, in a polysyllabic word, primary stress most often occurs on the initial syllable. It follows that a stress-initial prosodic pattern would be the most common input that children perceive, and therefore learn to produce first, and also the stress pattern that impaired populations would default to when having difficulties in producing less frequent stress patterns. The question explored in this research is whether normal adults' language production also mirrors these facts. That is, do adults, in conditions under which they might reduce words by omitting syllables, also default to these similar patterns? Participants in this study were asked to listen to a list of words and repeat them in a reduced form (as in *Indianapolis ~ Indy*, *rhinoceros ~ rhino*). Certain prosodic patterns were controlled for in order to systematically examine their effects on reduction patterns. Stimulus words contained two, three, or four syllables, with primary stress on the first, second or third syllable. Results suggest that syllable number and stress do in fact affect how adults reduce words, although it is clear that the relationship between these factors is complex.

Introduction

Word reductions are a deceptively common phenomenon in language. Although they are often found in normal adult speech in the form of word abbreviations, they have been most systematically studied in the productions of normally developing young children, as well as children and adults with language disorders. For example, it is widely known that children with normally developing language, around two years of age, reduce or simplify their words, as in *banána* to *nána* and *giráffe* to *ráffe* (examples from Gerken, 1996; Klein, 1981).

Research on these reductions shows that by and large, children reduce words by omitting syllables in certain predictable patterns. For example, children omit unstressed syllables more often than stressed syllables, and they omit unstressed syllables that precede main word stress as in *banána* or *giráffe* more often than those which follow main word stress (Allen & Hawkins, 1980; Carter, 1999; Carter & Gerken, 1998; Demuth, 1995, 1996; Fee, 1996; Gerken, 1994a, b, 1996; Klein, 1981; Wijnen, Krikhaar, & den Os, 1994). The output prosodic pattern of these truncations often corresponds to a prosodic foot, that is, either a trochee (a disyllabic word with stress on the first syllable) as in *mónkey*, or a monosyllabic foot as in *dóg*.

Several researchers have argued that the reason for these output patterns and consequent syllable omissions lies in the statistical properties of the English language. In an analysis by Cutler and Carter (1987) of approximately 20,000 English words, 90% of content words were found to begin with a stressed syllable. These results suggest that the input that children perceive most often contains a trochaic stress pattern. This stress-sensitive disyllabic foot is one of the earliest prosodic structures that English-speaking

children produce, after passing through the monosyllabic stage, and it is therefore considered their Minimal Word (Demuth, 1996; Fee, 1996; Gerken, 1996). Words with weakly stressed, word-initial unfooted syllables as in *ba-nána* are therefore often reduced to a Minimal Word by an omission of the initial syllable (Demuth, 1996; Gerken, 1996; Salidis & Johnson, 1997; Stemberger & Bernhardt, 1997). Other arguments have been made that children's perceptual systems have a strong bias to detect the more perceptually salient properties of stressed syllables and word-final syllables, ignoring any pre-tonic weak syllables (Echols, 1993; Echols & Newport, 1992). Although stress and syllable position are key factors influencing reductions, there are others as well, such as the segmental content of the word (Kehoe & Stoel-Gammon, 1997), number of syllables in an utterance (Gerken, 1996), and lexical familiarity of the utterance (Boyle & Gerken, 1996; Ohala & Gerken, 1997).

The phenomenon of syllable omission has also been reported in several clinical populations with language disorders, such as children who have Specific Language Impairment (Chiat & Hirson, 1987; Leonard, 1998), and adults who have acquired aphasia (Blumstein, 1973; Goodglass, Fodor & Schulhoff, 1967; Nickels & Howard, 1999). Again, these populations tend to reduce words with less frequent stress patterns to the more frequent, stress-initial forms by omitting unstressed, and often initial, syllables. As with normally developing children, however, word reductions in these populations are also influenced by other factors such as segmental content and syllable type (Blumstein, 1973; Carter, 1999; Jakobson, 1963).

In order to better understand the nature of these word reductions in children with normally developing language, children with language disorders and adults with language disorders, we must complete the paradigm by examining what behaviors normal adults exhibit with regard to reductions. In adult speech, word reductions are found most commonly in casual to fast speech registers, as in *cáuse* for *becáuse* and *cámra* for *cámera* (Dalby, 1984; Fisher & McDavid, 1973; Kypriotaki, 1970; Zwicky, 1972), and in word abbreviations and slang, as in *rhíno* for *rhinóceros* or *Bécca* for *Rebécca* (Bareš, 1974; Hamans, 1996; Hodge & Pennington, 1973; Kreidler, 1979; Streeter, Ackroff & Taylor, 1983). In casual and fast speech, reductions are most often formed by medial vowel deletion (syncope) as in *cámera ~ cámra* and *ópera ~ ópra* (Dalby, 1984; Zwicky, 1972), and unstressed initial syllable deletion (aphaeresis) as in *becáuse ~ cáuse* or *afráid ~ fráid* (Fisher & McDavid, 1973; Kypriotaki, 1970). Word abbreviations are formed most often by whole syllable deletions, either word-initial pre-stress syllable deletion as in *Rebécca ~ Bécca* (Hamans, 1996) or post-stress syllable deletions as in *rhinóceros ~ rhíno* (Hamans, 1996; Kreidler, 1979). In addition, although Hamans found that reductions do not necessarily take place at morpheme boundaries, Hodge and Pennington provide evidence that affixes are commonly the deleted elements.

While word reductions are frequent in English and certain large-scale patterns have been reported in surveys of speech corpora, few researchers have performed systematic experiments to study this phenomenon in the laboratory. In fact, researchers have been largely unable to pin down specific variables for predicting how certain words will be shortened, for example whether initial syllables or final syllables would be deleted (e.g. *président ~ prés* vs. *télephòne ~ phóne*), how many syllables would be deleted, whether whole syllables or just vowels would be deleted, or even how morphology affects word truncations. Fisher and McDavid (1973), in a survey of New England speech, and Kypriotaki (1970), in a more widespread study of American English, both noted that omissions of initial syllables occur most often on syllables that bear minimal stress in the word, and most often when the syllable following the deletable syllable bears primary or secondary stress. Zwicky (1972) reported that for word-medial syncope in English, the vowel (or syllable) to be deleted also bears minimal stress, as well as falls into certain segmental contexts (preceding a sonorant consonant). In a comparison between a corpus of television news interviews and a second corpus of three subjects producing both slow and fast versions of test sentences, Dalby (1984) found that in conversational and very fast speech, syllable deletion (or vowel

deletion) occurs more often when the syllable is unstressed, has a certain syllable shape (unstressed vowels adjacent to single consonants had much higher deletion rates than did cluster-adjacent vowels), is adjacent to certain manners of articulation (most deletions occurred with syllables in which the vowel was preceded by a sonorant or fricative consonant, or was followed by a stop consonant) and in certain positions in the word (word-medial and post-stress).

In an experiment in which Bell Laboratories employees were asked to abbreviate computer command names, Streeter, Ackroff and Taylor (1983) found that polysyllabic words were most often shortened by truncation of the final syllable(s). In a second series of experiments on word abbreviation behavior, Hodge and Pennington (1973) found that with shorter words, subjects more often omitted word-medial syllables and segments, whereas with longer words, subjects more often omitted word-final syllables and segments (one possible reason lies in the fact that the longer words tended to have suffixes, and the suffixes were the portions that were deleted). Finally, a number of experimental studies in the domain of language processing have shown that the stressed syllable and the word-initial syllable play a key role in lexical access, word recognition, and speech production, and therefore may also play a role in a task such as word reduction (Bradley & Forster, 1987; Grosjean & Gee, 1987 for stressed syllable; Brown & McNeill, 1966; Hawkins & Cutler, 1988; Marslen-Wilson & Welsh, 1978; Nooteboom, 1980 for word-initial syllable).

While these studies have reported somewhat disparate results, taken together, they show that syllable shape, syllable position within the word, primary stress location, and word-length are factors that affect how words are shortened. In addition, cross-linguistic research concerning output responses suggests a strong tendency for adult truncations to result in syllables and feet that form optimal prosodic patterns, either perceptually or productively, regardless of input word length or stress pattern (Itô, 1990 for Japanese; Kilani-Schoch, 1996 for French; Ronneberger-Sibold, 1995 for German; Szpyra, 1995 for English and Polish). This prosodically-based observation mirrors the patterns found in children, discussed above.

Before summarizing the results of the present investigation, however, it is important to mention a few points about stress in general. Every word in English contains one syllable that is more acoustically and perceptually prominent. This syllable is assigned primary stress within the word. If the word has two or more syllables, it may also contain one or more secondarily stressed syllables, with less prominence than the primarily stressed syllable but more than any unstressed syllables (Hammond, 1999). Primary stress is traditionally marked with an acute accent, ´, and secondary stress with a grave accent, `. In English, stress is assigned to heavy syllables, that is, syllables containing either a tense vowel, such as /o/ or /u/, or a coda of one or more consonants (Hammond, 1999; Prince, 1990). Optimally, if there is more than one stress in a word, primary and secondary stress fall on alternating syllables, as in *álmanàc* and *sálamànder*, in which primary stress falls on the first syllable and secondary stress on the third syllable, or as in *càbarét* and *tàpióca*, in which the pattern is reversed, that is, primary stress falls on the third syllable and secondary stress on the first (Hammond, 1999; Hayes, 1995; Hayes, 1984). However, this is not always the case, as in certain words such as *bòmbárd*, *álpìne*, or *bàndána*. Because English contains borrowings from other languages (Bolinger, 1965; Hayes, 1983), it has many varied stress patterns, which makes it an intriguing language for this study.

The goal of this research was to examine word reductions in a large group of subjects, in order to identify predictive patterns of reduction for a variety of polysyllabic word types. These findings will add to the literature on adult word truncations and enhance our existing knowledge of other populations' reductions. Specifically, this project was designed to be a systematic, exploratory study of three factors (stress position, syllable number, and morphology) in order to identify any existent patterns of adult word reduction and any predictable variability between subjects, to determine what similarities to children's

reductions they might show, and to create an adult comparison for a second experiment with children. In order to test the validity of the conclusions made in the adult studies reviewed above, we made several predictions regarding reductions in this experiment. The first prediction was that regardless of stress pattern or syllable number of the target word, word reductions should largely conform to a good foot, that is, either a monosyllabic form or a disyllabic form with stress on the first syllable. The second prediction was that the salient features will be preserved – that is, the primary-stressed syllable, the initial syllable and the final syllable will more likely be retained in the response than omitted. The third prediction was that, based on the child data, initial syllables will be omitted more often if they directly precede primary stress. Finally, our prediction regarding morphology is that affixes or segments in the affixes will be omitted more often than segments within the word roots. The present paper will only report on the first two factors (stress position and syllable number), and therefore only the first three predictions.

Experiment

Method

Participants. Fifty-five native English-speaking undergraduates (16 males and 39 females) were recruited from the Indiana University community. All subjects received partial course credit towards an Introductory Psychology class for their participation. The mean age of these participants was 19.09 years (SD = 1.66). Data from 12 subjects were not included in the final analysis due to: being a non-native speaker of English (one subject), having a history of speech disorder (one subject), failing to comply with experimental instructions (seven subjects), excess background noise (two subjects), and recording failure (one subject). The remaining 43 participants were 14 males and 29 females, who had no history of speech or hearing disorders. Participants were assigned to one of two groups. One group received monomorphemic words, and the other group received polymorphemic words (see Stimulus Materials section below). This report will present data from the monomorphemic group condition only. Data from 22 participants (six males and 16 females, mean age of 18.91, SD = 1.23) will be reported in this paper.

Stimulus Materials. The stimuli consisted of 160 polysyllabic monomorphemic words that were used as targets in the monomorphemic condition of the word reduction task and 160 polysyllabic polymorphemic words that were used as targets in the polymorphemic condition.[2] Within each condition, the number of syllables and primary stress location varied systematically. As shown in Table 1, there were eight categories, each with 20 words: disyllabic words with primary stress either on the first syllable (2syl-1pri) or second syllable (2syl-2pri), trisyllabic words with primary stress on the first syllable (3syl-1pri), second syllable (3syl-2pri), or third syllable (3syl-3pri), and quadrisyllabic words with primary stress on the first syllable, (4syl-1pri), second syllable (4syl-2pri), or third syllable (4syl-3pri).[3]

The stimuli were randomly selected from the Hoosier Mental Lexicon (an on-line dictionary of 20,000 entries; Luce & Pisoni, 1998) using the following criteria: first, a lexical frequency rating within one standard deviation of the log mean frequency of each target category (based on values given in Kučera & Francis, 1967); second, a neighborhood density of 2 or lower (neighborhood density was defined as all words that are within one phoneme of the target word by addition, deletion, or substitution), and third, a familiarity rating of at least 6.0 (on a 7-point scale) from undergraduate students (Nusbaum, Pisoni, & Davis, 1984).[4]

[2] Polymorphemic words contained at least one productive prefix or suffix, as defined by Bybee (1985).
[3] There were only 39 total quadrisyllabic words with primary stress on the fourth syllable, and even fewer that also reached our other criteria, therefore we did not include this pattern in the stimulus set.
[4] Familiarity was set at 6.5 and above for all categories except 3syl-3pri and 4syl-3pri, as these two categories had fewer total words. Familiarity for these words was consequently set at 6.0 in order to collect a sufficient number of words.

Target Category	Number of Syllables	Primary Stress Location
2syl-1pri	2	1st
2syl-2pri	2	2nd
3syl-1pri	3	1st
3syl-2pri	3	2nd
3syl-3pri	3	3rd
4syl-1pri	4	1st
4syl-2pri	4	2nd
4syl-3pri	4	3rd

Table 1. The eight stimulus target categories, the number of syllables found in each category, and the syllable that carries primary stress for each category.

The stimulus set was recorded by a female talker in two blocks, in a sound attenuated chamber (IAC Audiometric Testing Room, Model 402) using a head-mounted Shure (SM98) microphone. The recordings were digitized at 22.05 kHz (16-bit) using a Tucker-Davis Technologies System II sound card and stored in individual files on a PC. The utterances from the second block of two recordings were used, except in a few cases when there was excess noise in the recording (clicks, pops, aspiration picked up by the microphone) in which case stimuli from the first block of recordings were used. All stimulus tokens were judged to be highly intelligible by six phonetically trained listeners. The tokens were segmented into individual digital files and included the entire visible speech signal in both the waveform and the spectral view such that each file started and ended at a zero crossing. The segmented tokens were then leveled to 63 dB (using the Level 16 program developed by Tice & Carrell, 1998).

Procedure

Participants were given both written and verbal instructions, in which they were informed that for each trial they would hear a spoken word over their headphones. Simultaneously with the auditory word presentation, they would see a visual prompt on the screen (" * "). For each word they heard, they were asked to first imitate the word in its entirety, and second to generate a "reduced" response. They were provided with examples (e.g., *hìppopótamus ~ híppo*; *biólogy ~ bío*) and reminded that most of the words would not be normally reduced in everyday speech.[5] The participants then had 5.5 seconds to carry out both tasks, that is, imitate the original word in full, and generate the reduced version. A brief practice session preceded the experiment in order to familiarize the participants with the task. Participants were tested individually. The 160 words were presented in three blocks, with two breaks to allow the participants to rest.

The participants' responses were recorded in the same sound attenuated chamber, using the same head-mounted microphone that was used in recording the test stimuli. Recordings were done in stereo on a Sony DAT deck (DTC-690). The target stimuli were recorded on the right channel while the participants' responses were recorded on the left channel. Both the target stimuli and responses were later streamed at 48kHz (16-bit) in stereo into individual digital files for storage and analysis on a PC using a Roland UA-40 external Analog to Digital converter and Syntrillium's CoolEdit Pro LE.

[5] While most words in the stimulus set did not have a common English reduction, 6.9% of the words did, such as *mèmorándum, ìnfluénza,* and *hélicòpter.* These were included because they had been in the original random selection of words from the on-line dictionary, and as they constituted only a minimal portion of the word list, we did not think they would affect participants' performance. If anything, they would act to remind the participants about the type of reduction we were looking for.

The repetition and reduction responses were then transcribed and coded by the two experimenters and a third research assistant, all trained in broad phonetic transcription, using a coding scheme based on the International Phonetic Alphabet (IPA). For reliability purposes, 10% of each subject's transcriptions were verified by one of the other transcribers, and tokens that two transcribers disagreed upon were examined by the third person. Any transcriptions that remained unresolved were not included in the analysis. Interjudge agreement for transcriptions was 95.2% for repetition responses and 92.9% for reduction responses. Reduction responses were then coded for four features: the prosodic output pattern (a monosyllabic foot, a disyllabic trochaic foot, or some other pattern), whether the original stressed syllable was preserved (either as the stressed syllable or as a reduced syllable), whether either the initial or final syllable of the word was preserved, and which syllables were omitted in creating the reduced form. Interjudge agreement for reduction response coding was 96.6%. A small portion of participants' responses (2.2%) was omitted from coding due to non-responses, misperception of the stimulus, stress shift in the repetition of the stimulus, and any phonologically unrelated reductions of the target (e.g., *cònstellátion ~ stárs*).

Results

Examples of several target stimuli and subject responses are given in Table 2. The first column represents the syllable number and primary stress location of each target word group of the stimuli, and the second column shows a corresponding example word. The third column shows a typical example of subjects' repetition responses for each stimulus word, transcribed using the IPA. The fourth column shows a typical reduction response for each repetition, also transcribed in IPA. For example, the second row, 2syl-1pri, shows a typical response for the target word *máple*, which has two syllables and primary stress on the initial syllable: a repetition, [meípl̩], followed by a reduction, [meíp].

Syllable number and stress pattern	Example of stimulus word	Example of repetition response	Example of reduction response
2syl-1pri	máple	méɪpl̩	méɪp
2syl-2pri	gazélle	gəzɛ́l	zɛ́l
3syl-1pri	Ámazòn	ǽməzɔ̀n	zɔ́n
3syl-2pri	màrtíni	màrtíni	tíni
3syl-3pri	tàngeríne	tæ̀ndʒərín	tǽndʒ
4syl-1pri	sálamànder	sǽləmæ̀ndə˞	mǽndə˞
4syl-2pri	aquárium	əkwɛ́riṃ	kwɛ́riṃ
4syl-3pri	tàpióca	tæ̀pióʊka	tǽpi

Table 2. Examples of stimulus words, repetition responses (in IPA), and reduction responses (in IPA) for each target category.

The results of reduction response coding for the 22 subjects are summarized in Figures 1 through 5. Figure 1 shows the prosodic output patterns, coded as a monosyllabic foot, such as [méɪp] for *máple*, a disyllabic trochaic (strong-weak) foot, such as [mǽndə˞] for *sálamànder*, or *Other* (this category comprised patterns of either a disyllabic reduction with second syllable stress, e.g., [əkwɛ́r] for *aquárium* or any type of trisyllabic reduction, e.g., [kwɛ́riṃ] for *aquárium* or [pióʊkə] for *tàpióca*). This graph

displays results relevant to our first prediction, which was that regardless of stress pattern or syllable number of the target word, word reductions should largely conform to a well-formed prosodic foot, that is, either a monosyllabic form or a disyllabic form with stress on the first syllable.

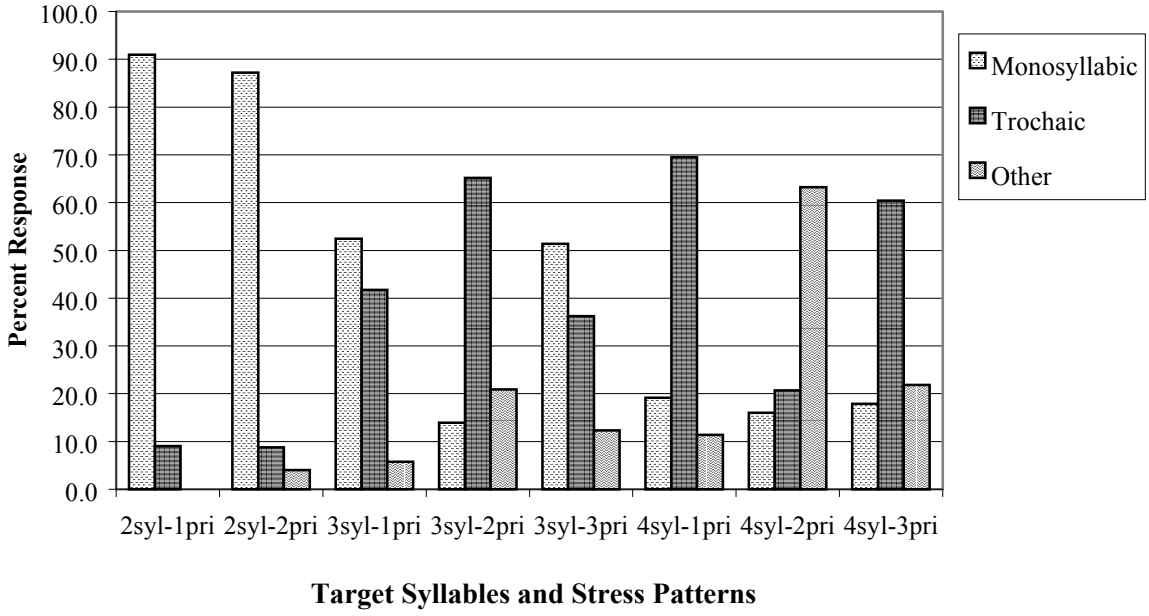


Figure 1. Percent of reduction responses that contained a monosyllabic foot, a disyllabic trochaic foot, or other output pattern, for each target category.

Subjects reduced words significantly more often to a monosyllabic foot than to a disyllabic foot or an *Other* form ($\chi^2 = 8.49$, $df = 1$, $p < .01$; $\chi^2 = 377.91$, $df = 1$, $p < .001$, respectively), and significantly more often to a disyllabic foot than to an *Other* form ($\chi^2 = 278.32$, $df = 1$, $p < .001$). Out of eight original target categories, seven categories were reduced by subjects most often to either a monosyllabic foot or a disyllabic foot. Four categories were reduced by subjects most often to a monosyllabic foot. First, the 2syl-1pri category was reduced to a monosyllabic foot more often (in 90.9% of responses) than either a disyllabic foot (9.1%) or *Other* form (0%). The difference between monosyllabic foot responses and disyllabic foot responses was statistically significant ($\chi^2 = 289.89$, $df = 1$, $p < .001$). Second, the 2syl-2pri category was also reduced to a monosyllabic foot more often (in 87.2% of responses) than either a disyllabic foot (8.7%) or *Other* form (4.0%). These differences were also statistically significant ($\chi^2 = 275.36$, $df = 1$, $p < .001$ and $\chi^2 = 324.96$, $df = 1$, $p < .001$, respectively). These two results were expected given the nature of the experimental design: the targets were originally disyllabic (e.g., *máple* and *gazélle*) and a reduction response typically yielded a monosyllable (e.g., [méɪp] or [zél]). Third, the 3syl-1pri target category, although showing somewhat more variation in reduction responses, was also reduced significantly more often to a monosyllabic foot, as in *óbstacle* ~ [ɔ́b] (52.5%) than to either a disyllabic foot or *Other* pattern ($\chi^2 = 5.62$, $df = 1$, $p < .05$ and $\chi^2 = 163.84$, $df = 1$, $p < .001$, respectively). Likewise, the 3syl-3pri target category was reduced significantly more often to a monosyllabic foot, as in *bàssinét* ~ [nét] (51.4%) than to either a disyllabic foot (36.3%) or *Other* pattern (12.4%. $\chi^2 = 10.14$, $df = 1$, $p < .01$ and $\chi^2 = 97.08$, $df = 1$, $p < .001$, respectively).

Three target categories were reduced more often to a disyllabic foot than to either a monosyllabic foot or to an *Other* pattern. The 3syl-2pri target category resulted in a disyllabic foot reduction, such as *mànháttan* ~ [hǽʔn̩], in 65.2% of responses, but in a monosyllabic foot, such as *mànháttan* ~ [hǽt], in 13.9% of responses. This difference was statistically significant (χ^2 = 142.35, *df* = 1, *p* < .001). In addition, the 3syl-2pri category resulted in a disyllabic foot significantly more often than an *Other* pattern, which occurred in 20.9% of responses (χ^2 = 98.86, *df* = 1, *p* < .001). This reduction response pattern was due to a strong tendency for the initial, pre-stress syllable to be deleted (see Figure 5), as we predicted. The 4syl-1pri and 4syl-3pri target categories demonstrated similar results for output reductions, with a disyllabic foot reduction in 69.5% and 60.4% of responses, respectively (*córonàry* ~ [kɔ́roʊ], and *èpidémic* ~ [dɛ́mək]). Both of these patterns occurred more often than a monosyllabic pattern (χ^2 = 122.95, *df* = 1, *p* < .001 for 4syl-1pri and χ^2 = 99.37, *df* = 1, *p* < .001 for 4syl-3pri), or an *Other* pattern (χ^2 = 180.52, *df* = 1, *p* < .001 for 4syl-1pri and χ^2 = 76.70, *df* = 1, *p* < .001 4syl-3pri). In looking at reductions of these seven target categories, longer target words tended to be reduced more often to a disyllabic form.

Only one category, the 4syl-2pri category, was notably reduced most often to the *Other* category, in 63.2% of responses, whereas it was reduced to a monosyllabic foot in 16.0% of responses and to a disyllabic foot in 20.7% of responses. The differences between an *Other* response and a monosyllabic foot response, and an *Other* response and a disyllabic foot response, were significant (χ^2 = 123.36, *df* = 1, *p* < .001 and χ^2 = 93.77, *df* = 1, *p* < .001, respectively). Reductions in this category typically were in the form of either a disyllabic form with stress on the second syllable (as in *aquárium* ~ [əkwɛ́r]) or a trisyllabic form with stress on the initial syllable (as in *aquárium* ~ [kwɛ́riŋ̩]). The reduction responses were consistent with our first prediction, and suggest that adults reduce words to a prosodically optimal form (either a monosyllabic or disyllabic foot). The overall pattern of responses was consistent with the prior research on children's and adults' outputs, summarized above.

Our second prediction involved the preservation of salient syllables (stressed, initial and final). First, we predicted that subjects' patterns of reduction responses would retain the stressed syllable of the target words more often than omit it. Actual word reduction responses yielded three patterns: responses that included the primary-stressed syllable from each target category as the primary-stressed syllable (e.g., *màrtíni* ~ [tî̞ni]), responses that included the primary-stressed syllable in a reduced capacity (resulting from a stress shift, e.g., *màrtíni* ~ [márɾi]), and responses in which the primary-stressed syllable was omitted (e.g., *màrtíni* ~ [már]). Figure 2 shows the mean percentage of reduction responses containing these three patterns. Of the reduction responses that preserved the primary-stressed syllable, more responses retained the syllable in its original stressed form (34.2% to 85.5% across the eight target categories) than retained the syllable in a reduced form (0% to 14.2%). This difference was statistically significant (χ^2 = 1720.17, *df* = 1, *p* < .001). However, there were also significantly more responses that omitted the original stressed syllable (ranging from 12.8% to 62.9%) than retained it in a reduced form (χ^2 = 346.42, *df* = 1, *p* < .001). The small percentage of primary-stressed syllables resulting in a reduced syllable comes from the relatively small percentage of stress shifts. Across the three word length types (disyllabic, trisyllabic, quadrisyllabic), target categories with first-syllable primary stress had the fewest stress shifts (0% to 1.6%), and target categories with second-syllable primary stress had the most stress shifts (4.7% to 14.2%).

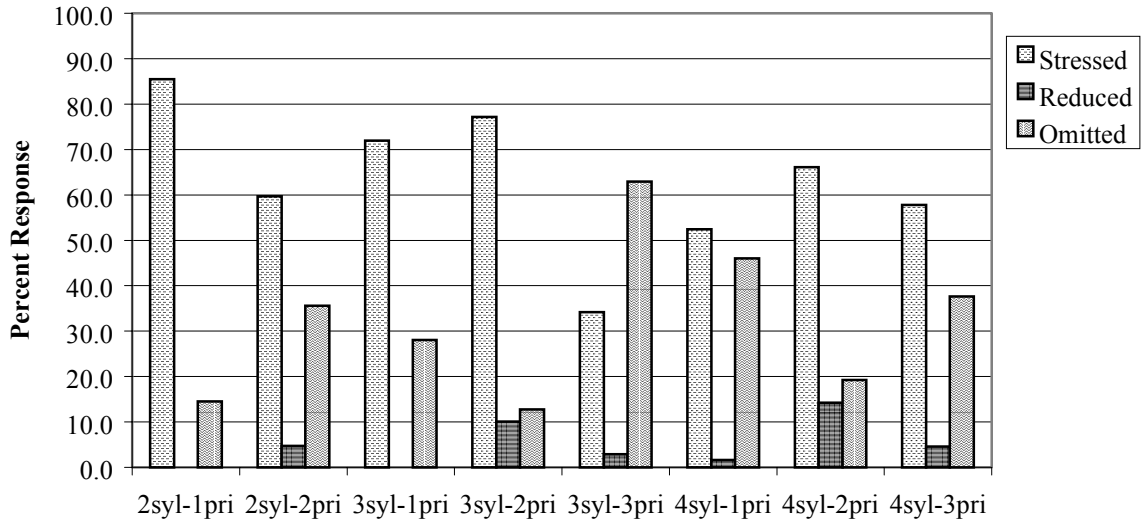


Figure 2. Percent of word reduction responses in which the stressed syllable was preserved as the stressed syllable, preserved but as a reduced (unstressed) syllable, or omitted altogether, for each target category.

Four patterns were noteworthy regarding stressed syllables that were preserved as the stressed syllable. First, there was a difference between the reduction responses for the two disyllabic targets. For the 2syl-1pri category, the stressed syllable was preserved as the stressed syllable in 85.5% of responses, as reduced in 0% of responses, and omitted in 14.5% of responses. The difference between preservation (as stressed) and omission of the stressed syllable was statistically significant ($\chi^2 = 213.18$, $df = 1$, $p < .001$). In contrast, for 2syl-2pri, the stressed syllable was preserved as the stressed syllable in only 59.7% of responses, as a reduced syllable in 4.7% of responses, and omitted in 35.6% of responses. Goodness of fit chi-square tests showed each of the three to be statistically different from the others. That is, the difference between preservation of the stressed syllable as stressed and the stressed syllable in a reduced form ($\chi^2 = 199.84$, $df = 1$, $p < .001$), the difference between preservation of the stressed syllable as stressed and omission of the stressed syllable ($\chi^2 = 25.06$, $df = 1$, $p < .001$), and the difference between preservation of the stressed syllable as reduced and omission of the stressed syllable ($\chi^2 = 102.25$, $df = 1$, $p < .001$) were all significantly different. The high rate of preservation of the stressed syllable as the stressed syllable, alongside a rate of 0% stressed syllable as reduced in 2syl-1pri, reflects the bias for content words to have initial syllable stress in the language. In addition, the preservation of the stressed syllable as reduced in 2syl-2pri was largely due to a stress shift, yielding a stressed initial syllable, while preserving the original stressed syllable.

A second finding was that the stressed syllable was preserved in its stressed form in 71.9% of word reductions for the category 3syl-1pri. This pattern occurred significantly more often than omission of the syllable, at 28.1% ($\chi^2 = 82.12$, $df = 1$, $p < .001$). Again, as in 2syl-1pri, there were no occurrences

of the stressed syllable as a reduced syllable. This result also supports the notion of salience of a stressed initial syllable.

Third, the stressed syllable was preserved as the stressed syllable least often of all target categories for 3syl-3pri, at 34.2% (an example of a more frequent reduction response for this target category was [sílou] for *sìlhouétte*). This pattern occurred significantly less often than omission of the stressed syllable, which occurred in 62.9% of responses ($\chi^2 = 36.48$, $df = 1$, $p < .001$). This result was not surprising, in light of the tendency for initial syllables of this category to be preserved at a high rate (see Figure 3). In fact, the finding suggests that although stress and final syllable position are both salient positions, at a certain point, the initial syllable is more likely to be preserved, regardless of stress position. This finding will be revisited in the following sections.

The fourth noteworthy response pattern occurred with the 4syl-1pri target category. Contrary to expectations, the primary stressed syllable was not consistently maintained in reductions. The stressed syllable was preserved as the stressed syllable in 52.4% of reduction responses, and it was omitted in 46.0% of responses. This difference was not statistically significant. The pattern can be interpreted as the omission of one of the two prosodic feet and the preservation of the other, as in *mátrimòny* ~ [mǽtri] or [móuni]. This random pattern of foot omissions becomes even more evident in Figure 3. Overall, the consistent finding that subjects faithfully maintained the primary-stressed syllable as the stressed syllable in their reduction responses supports earlier findings reported by Cutler and Carter (1987), Echols (1993) and Echols and Newport (1992) on the salience of stressed syllables. However, other factors such as location of primary stress in the word also affects this salience.

With regard to initial syllables, as with stressed syllables, our prediction was that initial syllables would be preserved more often than they would be omitted. Figure 3 shows the mean percentage of responses that preserved the initial syllable, as in *gazélle* ~ [gǽz], and that omitted the initial syllable, as in *gazélle* ~ [zél], for all target categories.

Reduction responses to four of the eight target categories preserved the initial syllable significantly more often than omitted it. The initial syllable was preserved significantly more often for 2syl-1pri ($\chi^2 = 213.18$, $df = 1$, $p < .001$), 3syl-1pri ($\chi^2 = 82.12$, $df = 1$, $p < .001$), 3syl-3pri ($\chi^2 = 46.67$, $df = 1$, $p < .001$) and 4syl-2pri ($\chi^2 = 16.26$, $df = 1$, $p < .001$). The results for 2syl-1pri and 3syl-1pri categories suggest that primary stress, especially when falling on the initial syllable, is a good predictor for preservation of that initial syllable in word reductions of di- and trisyllabic words. The reduction response pattern for the target category 3syl-3pri (66.5% preservation rate) was unexpected, given that the target category has final syllable stress. This result suggests a possible preference for the word-initial syllable over the word-final syllable, despite the fact that the final syllable carries main stress. However, the initial syllable has secondary stress in many words from this category (e.g., *bàssinét*), which may provide an explanation: an initial syllable with secondary stress may be more salient than an unstressed initial syllable typical of, for example, 3syl-2pri targets. The result for the 4syl-2pri category was due to disyllabic iambic reduction patterns such as *aquárium* ~ [əkwér] (see Figure 1).

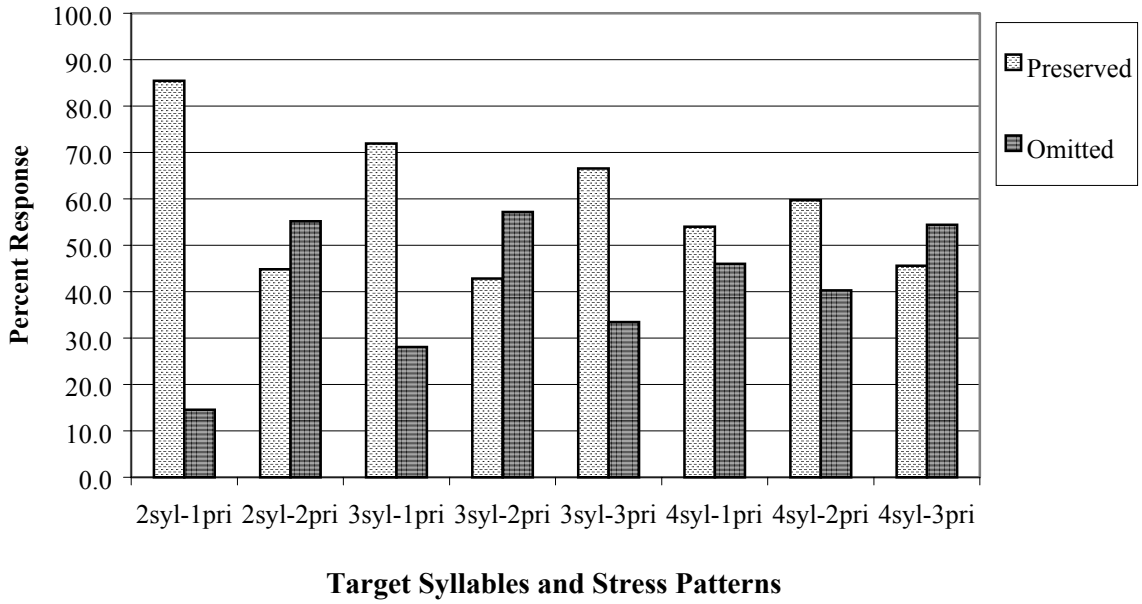


Figure 3. Percent of word reduction responses in which the initial syllable was preserved or omitted, for each target category.

Our prediction was not borne out for four target categories: 2syl-2pri, 3syl-2pri, 4syl-1pri, and 4syl-3pri. For the categories 2syl-2pri and 3syl-2pri, the initial syllable was actually omitted significantly more often than it was preserved ($\chi^2 = 4.33$, $df = 1$, $p < .05$ and $\chi^2 = 9.25$, $df = 1$, $p < .01$, respectively). Both target categories carry primary stress on the second syllable, and the initial syllable was the most often omitted syllable, leaving either the final monosyllabic foot (2syl-2pri) or the final disyllabic foot (3syl-2pri). These patterns were identical to typical omissions of the weak initial syllable found in other populations. The categories 4syl-1pri and 4syl-3pri showed no statistical difference between preservation and omission of the initial syllable (54.0% preservation for 4syl-1pri and 45.6% for 4syl-3pri). These results suggest once again that for 4syl-1pri, as well as for 4syl-3pri, subjects randomly preserved either the first or second foot and omitted the other, when reducing these categories. In summary, initial syllables were preserved more often than they were omitted for four of the eight target categories.

Figure 4 shows the mean response percentages for final syllable preservation and omission in word reduction. With regard to final syllables, the predicted outcome was that final syllables would also be preserved more often than omitted. However, there was only one category for which the predicted outcome was borne out, the 2syl-2pri category. The final (stressed) syllable was retained in 64.4% of responses, significantly more often than it was omitted, in 35.6% of responses ($\chi^2 = 34.29$, $df = 1$, $p < .001$).

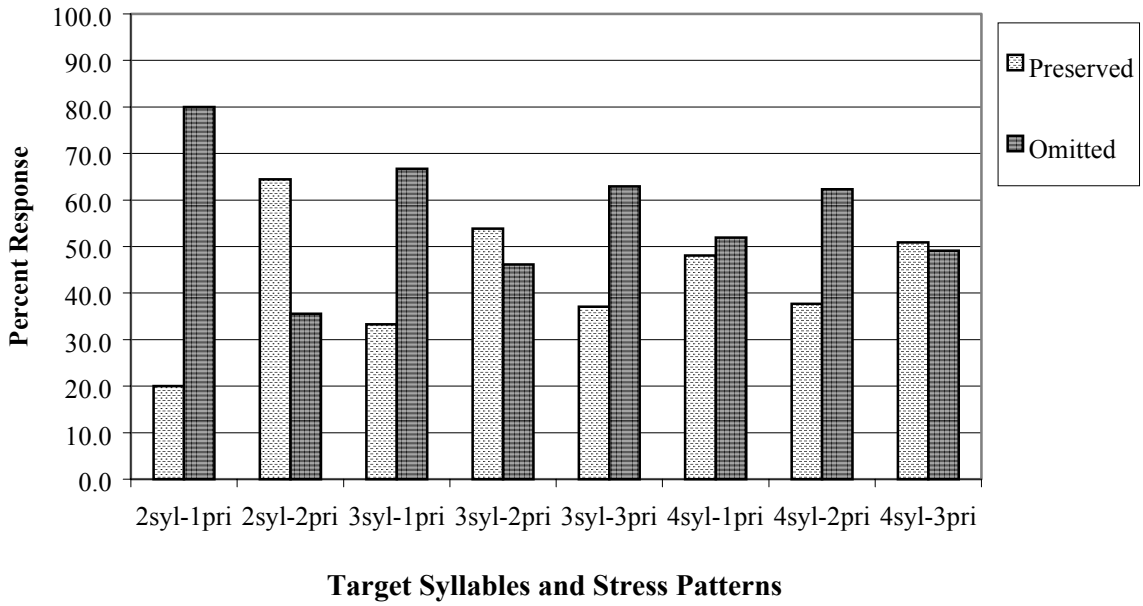


Figure 4. Percent of word reduction responses in which the final syllable was preserved or omitted, for each target category.

For five target categories, the final syllable was omitted significantly more often than it was preserved: 2syl-1pri, 3syl-1pri, 3syl-3pri, 4syl-1pri and 4syl-2pri. The result obtained for the 2syl-1pri category was not surprising, since the majority of reduction responses yielded monosyllabic forms and included the initial syllable (see Figures 1 and 3). However, it was unexpected once again that reduction responses from the 3syl-3pri category patterned as they did. There was no statistically significant difference between preservation and omission of the final syllable for two of the target categories, 3syl-2pri and 4syl-3pri. Taken together, the results for final syllable preservation suggest that the final syllable is not as salient a feature as the stressed or initial syllables are, except in the case of 2syl-2pri (where fewer strategies exist to reduce words).

A final interesting note is the distribution of initial and final syllable preservation responses for 2syl-1pri and 2syl-2pri. One might predict that with only two syllables to choose from in the target, the distribution of initial and final syllable preservation rates would be complementary. However, while the pattern of preservation was in the opposite direction (2syl-1pri had a higher percent of preservations of the initial syllable and 2syl-2pri had a higher percent of preservations of the final syllable), the rates of preservation were noticeably different (85.5% initial and 20.0% final for 2syl-1pri vs. 44.8% initial and 64.4% final for 2syl-2pri). A similar response distribution was found for the 3syl-1pri and 3syl-2pri targets (71.9% initial and 33.3% final for 3syl-1pri vs. 42.8% initial and 53.9% final for 3syl-2pri). These patterns once again reflect the tendency for English words to begin with a stressed syllable, and support the hypothesis that stressed, initial syllables are salient.

Figure 5 gives the mean response percentages of syllable omissions across the eight target categories. That is, for each reduction response, we counted which syllable or syllables were omitted in

the response: either the first or second syllable for the disyllabic targets, any of the first, second, or third syllables for the trisyllabic targets, and any of the first through fourth syllables for the quadrisyllabic targets.

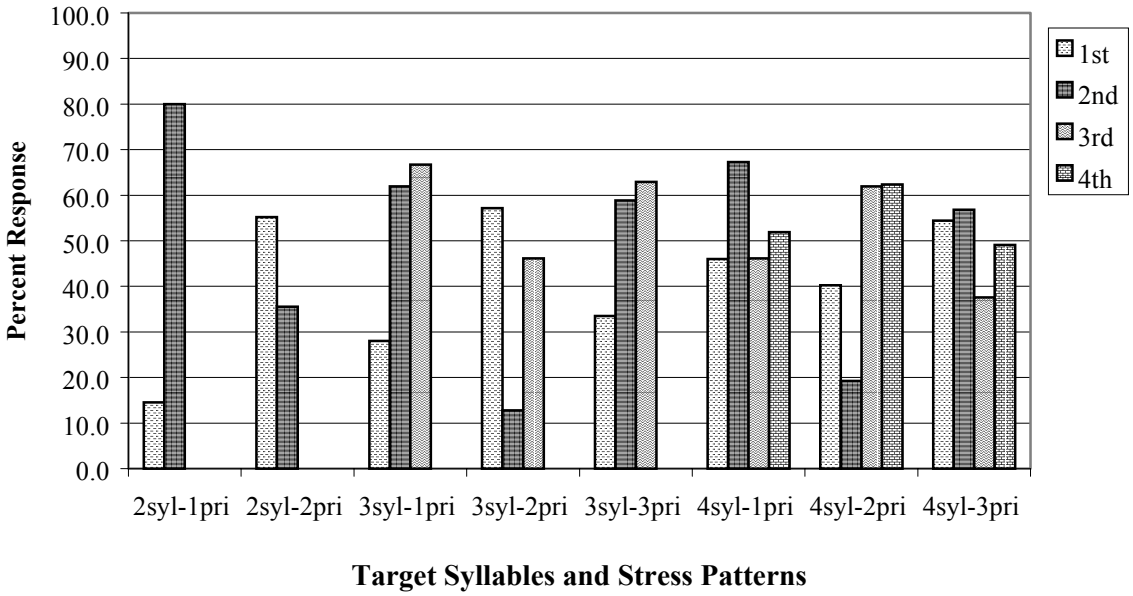


Figure 5. Percent of word reduction responses containing omissions of the first, second, third, or fourth syllable, for each target category.

The data shown here can be used to address our third prediction, which was that initial syllables that directly precede primary stress would be omitted more often than initial syllables containing primary stress or initial syllables that do not directly precede primary stress (stemming from the children's production literature). The prediction was supported for the first two target categories 2syl-2pri and 3syl-2pri, yielding responses that were similar to children's productions (e.g., *gazélle* ~ [zɛ́l] and *màrtíni* ~ [tɪ́ni]). For disyllabic words, the pretonic initial syllable in 2syl-2pri was omitted significantly more often than the stressed initial syllable in 2syl-1pri (χ^2 = 100.77, *df* = 1, $p < .001$). For trisyllabic words, the pretonic initial syllable in 3syl-2pri was omitted significantly more often than the initial syllable in 3syl-1pri and 3syl-3pri (χ^2 = 41.0, *df* = 1, $p < .001$ and χ^2 = 29.11, *df* = 1, $p < .001$, respectively). The prediction was not supported for 4syl-2pri targets, that is, there was no difference between percentage of omissions of the initial syllable from 4syl-2pri and 4syl-1pri, and the pretonic initial syllable in 4syl-2pri was actually omitted significantly less often than the initial syllable in 4syl-3pri (χ^2 = 8.78, *df* = 1, $p < .01$).

There were several other noteworthy observations regarding the analysis of syllable omissions. First, although the expected complementary omission rate difference existed between 2syl-1pri (final unstressed syllable omitted more often) and 2syl-2pri (initial unstressed syllable omitted more often), this complementary pattern was uneven, with more unstressed syllable omissions occurring for 2syl-2pri. This was most likely due to a difference in initial and final syllable preservation that was discussed in

previous sections. Second, the similar rates of omission of the second and third syllables from the 3syl-1pri target category (61.9% and 66.7%, respectively) suggest that when the output was of a disyllabic form, either syllable was employed as the second syllable. Third, the high rate of second syllable omissions from 4syl-1pri was the result of a large number of trisyllabic reductions such as *áppetìzer* ~ [æptɑ̀ɪzəˑ] or *cémetèry* ~ [sɛ́mtɛ̀ri], in which the unstressed second syllable, often a schwa, was the only omitted syllable. Fourth, the unusual pattern of omission rates for 4syl-2pri was again due to reductions such as *aquárium* ~ [əkwɛ́r], as discussed in previous sections.

General Discussion

The data summarized in Tables 3 through 5 provide an overview of the major findings reported above. Table 3 gives the generalizations for word reduction responses for disyllabic target categories (e.g., *máple* and *gazélle*), Table 4 gives generalizations for word reduction responses for trisyllabic target categories (e.g., *óbstacle*, *bàndána, nèctaríne*), and Table 5 gives generalizations for quadrisyllabic target categories (e.g., *mátrimòny*, *aquárium*, and *tàpióca*). Each column corresponds to the specific patterns we were interested in: the output foot pattern (whether monosyllabic, i.e., "S", disyllabic and trochaic, i.e., "Sw", or *Other*), the proportion of reduced responses that preserved the stressed syllable as the stressed syllable more often than preserving it as reduced or omitting it, the proportion of reduced responses in which the initial syllable and final syllable were preserved more often than omitted, and the syllable or syllables that were omitted most often in word reductions. For each table, note that for the column labeled "syllable(s) omitted most often", shaded cells denote no possible deletable syllable (e.g., for 2syl-1pri and 2syl-2pri, there were no third or fourth syllables to be deleted).

	Most common output prosodic pattern			Stressed syll. preserved with stress more often than reduced or omitted	Syllable preserved more often than omitted		Syllable(s) omitted most often			
	S	Sw	Other		Initial	Final	1	2	3	4
2syl-1pri	✓			✓	✓			✓		
2syl-2pri	✓			✓		✓	✓			

Table 3. Summary of results for reduction responses for 2syl-1pri and 2syl-2pri categories.

	Most common output prosodic pattern			Stressed syll. preserved with stress more often than reduced or omitted	Syllable preserved more often than omitted		Syllable(s) omitted most often			
	S	Sw	Other		Initial	Final	1	2	3	4
3syl-1pri	✓			✓	✓			✓	✓	
3syl-2pri		✓		✓			✓			
3syl-3pri	✓				✓			✓	✓	

Table 4. Summary of results for reduction responses for 3syl-1pri and 3syl-2pri, and 3syl-3pri categories.

	Most common output prosodic pattern			Stressed syll. preserved with stress more often than reduced or omitted	Syllable preserved more often than omitted		Syllable(s) omitted most often			
	S	Sw	Other		Initial	Final	1	2	3	4
4syl-1pri		✓						✓		
4syl-2pri			✓	✓	✓				✓	✓
4syl-3pri		✓		✓			✓	✓		✓

Table 5. Summary of results for reduction responses for 4syl-1pri and 4syl-2pri, and 4syl-3pri categories.

The initial analysis of the word reduction data collected in this task suggests that the relationship between word length, primary stress location, and syllable omissions is complex. We found a strong tendency for adult speakers to reduce words into a well-formed prosodic foot (monosyllabic or disyllabic) that still contains the original primary-stressed syllable. However, this was not true of all eight target syllable and stress patterns. For example, we found that word reductions of the 4syl-2pri category showed a trend toward some other output pattern (as in [əkwɛ́r] or [kwɛ́riṃ]).

Three general conclusions can be drawn from the analysis of preserved salient syllables in reduction responses. First, the stressed syllable was overwhelmingly preserved as the stressed syllable in the reduction responses. Second, the initial syllable was preserved more often than omitted under certain conditions, including when the initial syllable contained main stress. Third, only when the final syllable of a disyllabic word contained main stress, was it preserved more often than omitted.

The response patterns for the preservation of salient syllables however, were not consistent across all categories. Specifically, 3syl-3pri words tended to be reduced without the primarily-stressed syllable from the target and 4syl-1pri and 4syl-3pri showed a trend to be reduced arbitrarily to either the first or second foot. This arbitrary reduction is also suggested by the rates of preservation of the initial and final syllables in these same words. Although the primary-stressed syllables were frequently preserved in the reductions, reductions of these categories yielded equivalent omission rates between the primary-stressed syllable and another syllable in the word. Each of these categories carry secondary stress: in 3syl-3pri and 4syl-3pri it falls on the initial syllable, and in 4syl-1pri it falls on the third syllable. The weight or salience, which may exist in the secondary-stressed syllable, may very well be playing a role in responses. Thus, a post-hoc analysis including secondary stress versus zero stress may be warranted.

Overall, the different patterns of syllable omissions across the various target syllable and stress patterns complement the data on syllable preservation and also suggest that there are interactions between initial and final syllables, and possibly between primary- and secondary-stressed syllables, in the input categories that leads to a complicated set of outputs. The analysis of syllable omissions suggests that for di- and trisyllabic words, the same output patterns occur for children and for adults. Namely, initial syllables are omitted more often when they directly precede primary stress in the word. However, this is not the case for four-syllable words, and since omissions from four-syllable words are not well-documented for children, it is unclear how adults' responses compare to children's for this group of words.

Taken together, it appears that reduction responses for four target categories matched our earlier predictions. These are 2syl-1pri (due to the salience of the initial stressed syllable), 2syl-2pri (due to the salience of the final stressed syllable and the omission of the pre-tonic initial syllable), 3syl-1pri (due to

the salience of the initial stressed syllable), and 3syl-2pri (due to the omission of the pre-tonic initial syllable). However, reduction responses for the remaining four target categories did not match our predictions. These are 3syl-3pri (perhaps due to secondary stress or the relative infrequency of this pattern in English), 4syl-1pri and 4syl-3pri (perhaps due to secondary stress), and 4syl-2pri (due to the salience of the initial syllable, or possibly the less than optimal presence of two adjacent unstressed syllables; Hammond, 1999). Further analysis is necessary to examine these predicted and unpredicted patterns, in light of secondary stress facts and the frequency of stress patterns in English.

Additionally, even within a given syllable and stress pattern that tends to demonstrate reductions to well-formed feet, data were subject to variation across and within subjects. Some subjects had systematic "strategies" in their responses, such as always preserving the initial syllable, or adhering to a well-formed disyllabic foot more often than any other pattern, however other subjects randomly reduced words with no apparent strategies. A post-hoc analysis is planned to examine this subject variation and variability.

Finally, some reductions were obviously based on orthography (e.g., *anátomy* ~ [tɔ́m] and *tròmbóne* ~ [tíbòʊn], i.e., *t-bone*). Clearly, with literate adults, orthography may affect spoken word processing, especially in tasks that ask subjects to explicitly and consciously manipulate real words in somewhat unnatural ways (Jakimik, Cole, & Rudnicky, 1985; Jared, McRae & Seidenberg, 1990; Kreidler, 1979). In order to keep orthographic effects at a minimum, we specifically presented the words in an auditory-only modality. Consequently, the number of obvious orthographically-based responses was small.

Future Directions

As this study is ongoing, we will continue to analyze the coding data from the monomorphemic condition group to look for secondary stress effects and individual differences, across both subjects and items. In addition, we plan to examine the response times obtained in this task. Specifically, we will measure the interval between offset of the target stimulus and onset of the repetition response as well as the interval between offset of the repetition response and onset of the reduction response, in order to assess any effects of syllable number and stress patterns (or frequency of certain prosodic patterns) on response time. It is possible that a less frequent stress pattern such as 3syl-3pri might yield longer response times to construct a word reduction response than a more frequent stress pattern, such as 3syl-1pri. Second, as mentioned in the Methods section, we have collected data from a polymorphemic condition group as well. We plan to carry out identical analyses to those reported in this paper on word reduction responses from that subject group. Once those data are analyzed, we will compare data from the two conditions in order to study effects of morphology on word reductions (Hamans, 1996; Hodge & Pennington, 1973).

Third, we plan to carry out a version of this experiment using the same stimuli with children who are in their third year, approximately between the ages of 26 months and 36 months. Few child studies have examined omissions systematically across syllable number and primary stress location, and no studies have described the syllable omissions from four-syllable words by children. Therefore, the adult responses will provide a comparison for children's reduction responses. This was one of the reasons for using auditory stimuli: we needed a presentation mode that was amenable to both literate adults as well as pre-literate children. Finally, in running a pilot version of this study, a volunteer subject who is a speaker of Singaporean English found it difficult to reduce the target words or to perform the task at all. On further examination, we found that while American English is considered to be a stress-timed language (every stressed syllable in an utterance is an even amount of time away from the next), Singaporean

English is a syllable-timed language, in which every syllable nucleus is isochronous (Deterding, 2001; Lehiste, 1971; Platt & Weber, 1980). Running this experiment on speakers of Singaporean English will provide an interesting comparison to our American English population.

In summary, our initial findings using a word reduction task provide evidence that word length and primary stress position do affect word reduction strategies of adults in certain systematic ways. In addition, the pattern of responses is similar to word reductions observed in young children and several clinical populations. However, these findings also support the conclusions of previous researchers that the phenomenon of word reduction is a complex one. Further analyses of reduction responses of the full subject set may yield even stronger predictors for output patterns based on prosodic or morphological patterns of the input. The complete study will hopefully provide a controlled, experimental addition to the current linguistic and psycholinguistic literature on the nature of word reductions in adults, children, and clinical populations.

References

Allen, G. & Hawkins, S. (1980). Phonological Rhythm: definition and development. In *Child Phonology, Volume I: Production.* Yeni-Komshian, G., Kavanagh, J. & Ferguson, C. (eds.) New York: Academic Press, Pp. 227-256.

Bareš, K. (1974). Unconventional word-forming patterns in present-day English. *Philologica-Pragensia, 17,* 173-186.

Blumstein, S. (1973). *A Phonological Investigation of Speech*. The Hague: Mouton.

Bolinger, D. (1965). *Forms of English: Accent, Morpheme, Order*. Cambridge, MA: Harvard University Press.

Boyle, M. & Gerken, L. (1996). Effects of lexical familiarity on children's function morpheme omissions. *Journal of Memory and Language, 36,* 117-128.

Bradley, D. & Forster, K. (1987). A reader's view of listening. *Cognition, 25*, 103-134.

Brown, R. & McNeill, D. (1966). The 'tip-of-the-tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337.

Bybee, J. (1985). *Morphology: A Study of the Relation Between Meaning and Form.* Amsterdam: John Benjamins.

Carter, A. (1999). *An Integrated Acoustic and Phonological Investigation of Weak Syllable Omissions*. Doctoral dissertation, University of Arizona.

Carter, A. & Gerken, L. (1998). Evidence for adult prosodic representations in weak syllable omissions of young children. In *Proceedings of the 29th Annual Child Language Research Forum*, *29,* 101-110.

Chiat, S. & Hirson, A. (1987). From conceptual intention to utterance: a study of impaired language output in a child with developmental dysphasia. *British Journal of Disorders of Communication*, *22*, 37-64.

Cutler, A. & Carter, D.M. (1987). The predominance of strong initial syllables in the English Vocabulary. *Computer Speech and Language 2,* 133-142.

Dalby, J. (1984). *Phonetic Structure of Fast Speech in American English*. Doctoral dissertation, Indiana University.

Demuth, K. (1995). Markedness and the development of prosodic structure. In *Proceedings of the North East Linguistics Society, 25*. Beckman, J. (ed.) Amherst, MA: GLSA, University of Massachusetts, 13-25.

Demuth, K. (1996). The prosodic structure of early words. In *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*. Morgan, J. L. & Demuth, K. (eds.) Mahwah, NJ: Lawrence Erlbaum Associates, Pp. 171-184.

Deterding, D. (to appear). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*.

Echols, C. (1993). A perceptually-based model of children's earliest productions. *Cognition*, *46,* 245-296.

Echols, C. & Newport, E. (1992). The role of stress and position in determining first words. *Language Acquisition, 2,* 189-220.

Fee, E.J. (1996). Syllable structure and Minimal words. In *Proceedings of the UBC International Conference on Phonological Acquisition.* Bernhardt, B., Gilbert, J., & Ingram, D. (eds.) Somerville, MA: Cascadilla Press, 85-98.

Fisher, L.E. & McDavid, R.I., Jr. (1973). Aphaeresis in New England. *American speech, 48,* 246-249.

Gerken, L. (1994a). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, *21,* 565-584.

Gerken, L. (1994b). Young children's representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *Journal of Memory and Language*, *33,* 19-38.

Gerken, L. (1996). Prosodic structure in young children's language production. *Language, 72,* 683-712.

Goodglass, H., Fodor, I. & Schulhoff, C. (1967). Prosodic factors in grammar- evidence from aphasia. *Journal of Speech and Hearing Research*, *10,* 5-20.

Grosjean, F. & Gee, J.P. (1987). Prosodic Structure and spoken word recognition. *Cognition, 25*, 135-155.

Hamans, C. (1996). A lingo of abbrevs. *Lingua Posnaniensis, 38,* 69-78.

Hammond, M. (1999). *The Phonology of English*. Oxford: Oxford University Press.

Hawkins, J. & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In *Explaining Language Universals*. Hawkins, J. (ed.) Oxford: Blackwell, Pp. 280-317.

Hayes, B. (1983). A grid-based theory of English meter. *Linguistic Inquiry 14,* 357-393.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry 15,* 33-74.

Hayes, B. (1995). *Metrical Stress Theory*. Chicago, IL: Chicago University Press.

Hodge, M.H. & Pennington, F.M. (1973). Some studies of word abbreviation behavior. *Journal of Experimental Psychology, 98*, 350-361.

Itô, J. (1990). Prosodic Minimality in Japanese. *Papers from the Regional Meetings, Chicago Linguistic Society*, *2,* 213-239.

Jakimik, J., Cole, R.A., & Rudnicky, I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language, 24,* 165-178.

Jakobson, R. (1963). Implications of language universals for linguistics. In *Universals of Language*. Greenberg, J. (ed.) Cambridge, MA: MIT Press, Pp. 263-278.

Jared, D., McRae, K. & Seidenberg, M.S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687-715.

Kehoe, M. & Stoel-Gammon, C. (1997). Truncation patterns in English-speaking children's word productions. *Journal of Speech, Language, and Hearing Research*, *40,* 526-541.

Kilani-Schoch, M. (1996). Syllable and foot in French clipping. In *Natural Phonology. The State of the Art.* Hurch, B. & Rhodes, R. (eds). Berlin: Mouton de Gruyter, Pp 135-152.

Klein, H. (1981). Production strategies for the pronunciation of early polysyllabic lexical items. *Journal of Speech and Hearing Research, 24,* 389-405.

Kreidler, C. (1979). Creating new words by shortening. *Journal of English Linguistics, 13,* 24-36.

Kučera, H. & Francis, W.N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.

Kypriotaki, L. (1970). Aphaeresis in rapid speech. *American Speech, 45,* 69-77.

Leonard, L. (1998). *Children with Specific Language Impairment*. Cambridge, MA: The MIT Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: The M.I.T. Press.

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.

Marslen-Wilson, W.D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Nickels, L., & Howard, D. (1999). Effects of lexical stress on aphasic word production. *Clinical Linguistics and Phonetics*, *13,* 269-294.

Nooteboom, S.G. (1981). Lexical retrieval from fragments of spoken words: beginnings vs. endings. *Journal of Phonetics, 9*, 407-424.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Ohala, D., & Gerken, L. (1997). Lexical familiarity effects on children's weak syllable omissions. In *Proceedings of the 21st Annual Boston University Conference on Language Development*, 433-440.

Platt, J. & Weber, H. (1980). *English in Singapore and Malaysia.* Oxford: Oxford University Press.

Prince, A. (1990). Quantitative Consequences of Rhythmic Organization. *Chicago Linguistics Society 26*, 355-398.

Ronneberger-Sibold, E. (1995). On different ways of optimizing the soundshape of words. In *Historical Linguistics 1993: Selected Papers from the XIth International Conference on Historical Linguistics*. Andersen, H. (ed.) Amsterdam: John Benjamins, Pp. 421-432.

Salidis, J. & Johnson, J. (1997). The production of minimal words: A longitudinal case study of phonological development. *Language Acquisition 6,* 1-36.

Stemberger, J. & Bernhardt, B. (1997). Phonological constraints and morphological development. *Proceedings of the Annual Boston University Conference on Language Development, 21,* 603-614.

Streeter, L.A., Ackroff, J.M., & Taylor, G.A. (1983). On abbreviating command names. *The Bell System Technical Journal, 62,* 1807-1826.

Szpyra, J. (1995). *Three Tiers in Polish and English Phonology*. Lublin, Poland: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Tice, R. & Carrell, T. (1998). Level 16, Version 2.0.3. University of Nebraska.

Wijnen, F., Krikhaar, E., & den Os, E. (1994). The (non)realization of unstressed elements in children's utterances: evidence for a rhythmic constraint. *Journal of Child Language*, *21,* 59-83.

Zwicky, A. (1972). Note on a phonological hierarchy in English. In *Linguistic Change and Generative Theory.* Stockwell, R.P. & Macaulay, R.K.S. (eds.) Bloomington, IN: Indiana University Press, Pp. 275-301.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Perception of "Elliptical Speech" by an Adult Hearing-Impaired Listener with a Cochlear Implant: Some Preliminary Findings on Coarse-Coding in Speech Perception[1]

Rebecca Herman and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant DC00012 to Indiana University. We would like to acknowledge the help of "Mr. S" who graciously volunteered his time as a subject in this experiment. We would also like to acknowledge the help of Mike Vitevitch who served as the male speaker in the creation of the stimuli used in this experiment. Thanks also to Corey Yoquelet for assistance with transcription and scoring of results in Experiment 2. We are grateful to Lorin Lachs for help with the statistical analysis and to Miranda Cleary for editorial assistance.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology–Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Perception of "Elliptical Speech" by an Adult Hearing-Impaired Listener with a Cochlear Implant: Some Preliminary Findings on Coarse-Coding in Speech Perception

Abstract. This paper examines the effects of elliptical speech (Miller & Nicely, 1955) on the speech perception performance of an adult hearing-impaired listener with a cochlear implant. A group of 20 normal-hearing adult listeners were used for comparison. Two experiments were carried out using sets of normal and anomalous English sentences. Two versions of each set of sentences were constructed. One version retained normal place of articulation; the other was converted to "elliptical speech" using a procedure in which different places of articulation were all converted to the same alveolar place of articulation. The patient completed a same-different task and a transcription task. The normal-hearing listeners completed the same tasks, but with noise-masking or low-pass filtering used to degrade the signal. In the same-different task, we found that normal-hearing listeners under conditions of signal degradation tended to label a sentence with normal place of articulation and its elliptical version as the "same." The hearing-impaired listener with the cochlear implant also tended to label a sentence with normal place of articulation and its elliptical version as the "same." Results provide support for Miller and Nicely's claim that under conditions of signal degradation, the ellipsis can no longer be detected. In the transcription task, however, normal-hearing subjects showed better transcription performance for sentences with normal place of articulation than for "elliptical" speech sentences, which was an unexpected result given our findings in the first experiment. The patient with the cochlear implant also showed better transcription performance for sentences with normal place of articulation than for "elliptical" speech, which was also unexpected. The implications of these findings for how cochlear implant users perceive speech and recognize spoken words are discussed.

Introduction

One fundamental question in research on cochlear implants deals with what speech sounds like to users of cochlear implants. It is known that patients with cochlear implants often do not do well on open-set tests of word recognition in which the listener hears a word and has to identify it from a large number of words in his/her entire lexicon. Many of the confusions shown by users of cochlear implants are confusions of place of articulation in consonants. However, despite this apparent problem with place of articulation, some users of cochlear implants do very well in face-to-face conversations. How can these two diametrically opposed observations be reconciled?

In Miller and Nicely's (1955) well-known study, they examined the perceptual confusions in consonants under low-pass filtering of the signal (in which all information above 1 kHz is removed) and under noise masking of the signal (in which the signal is mixed with Gaussian noise). They found that place of articulation was a common confusion under either low-pass filtering or noise masking of the signal. Miller and Nicely explained the patterns of errors by noting that consonants that were confusable under conditions of signal degradation could be considered to represent comparable equivalence classes of a sort. For example, if [p t k] are confusable with each other under some conditions of signal degradation, then one could argue that these sounds form a common equivalence class. Miller and Nicely suggested that a single member of each equivalence class could be chosen as a representative of that class, (such as [t], out of the class [p t k]). Now if speech were produced in which each representative sound replaced any other individual member of its equivalence class, then that speech would sound quite strange in the clear. They refer to this type of degradation as "elliptical speech," because of the ellipsis, or leaving

out, of place of articulation information. For example, if speech were produced in which every [p t k] were simply replaced by a [t], that speech would sound very odd in the clear. However, they further note that if this so-called "elliptical" speech is now played back under conditions of noise-masking or filtering, then the ellipsis should be undetectable because the members of the equivalence class were found to be equivalent under exactly those degradation conditions. Miller and Nicely report informally that this is the case, although they never presented a complete experiment to demonstrate this phenomenon (Miller & Nicely, 1955; Miller, 1956).

Recently, Quillet, Wright and Pisoni (1998) noted the possible parallels in speech perception between normal hearing-listeners under conditions of signal degradation and patients with cochlear implants. Just as normal-hearing listeners show systematic confusions among different places of articulation under conditions of signal degradation, cochlear implant users also show confusions among places of articulation. Quillet et al. suggested that it might be possible to use "elliptical" speech to probe cochlear implant users' perception of speech and understand how they often do so well even with highly impoverished input signals. If "elliptical" speech is undetectable as elliptical for normal-hearing listeners under conditions of signal degradation, then perhaps it will also be undetectable as elliptical for cochlear implant users as well, providing support for the use of broader equivalence classes for place of articulation in the speech perception of users of cochlear implants.

Quillet et al. attempted to replicate Miller and Nicely's finding that "ellipsis" of place of articulation under conditions of signal degradation is undetectable as ellipsis with a same-different task. Normal-hearing listeners heard pairs of sentences and had to judge whether the two sentences were the same or different. Listeners heard pairs of sentences in which the two sentences were either lexically the same or lexically different. In one condition, both sentences in a pair had normal place of articulation. In the second condition, both sentences in a pair were transformed into "elliptical" speech. In the third condition, one sentence in a pair had normal place of articulation and the other had "elliptical" speech. The crucial case in their experiment was the last condition in which the two sentences were lexically identical, but one sentence had normal place of articulation and the other had an "elliptical" speech version of the sentence. In this case, normal-hearing listeners were expected to label the two sentences as "different" when heard in the clear. If Miller and Nicely's phenomenon can be replicated, then normal-hearing listeners should label this pair of sentences as the "same" when heard under degraded conditions. In fact, Quillet et al. did find that in the clear, listeners identified the majority of such pairs as "different." Furthermore, they found that under signal degradation using random-bit-flip noise, listeners identified a majority of these pairs as the "same," indicating that the ellipsis of place of articulation was not detected by the listeners in these cases. Listeners seemed to perceive speech in terms of broad phonetic categories under conditions of degradation such as bit-flipped noise.

In order to probe whether "elliptical" speech is undetectable as elliptical to users of cochlear implants, the first experiment in this report used a same-different task similar to the one described above. Pairs of sentences were presented to an adult patient with a cochlear implant, and he was asked to judge whether the two sentences were the same or different. Again, the two sentences in the pair were either lexically the same or lexically different. In one condition, both sentences in a pair had normal place of articulation. In the second condition, both sentences in a pair were converted into "elliptical" speech. In the third condition, one sentence had normal place of articulation and the other had "elliptical" speech. Again, the crucial test case is the third condition in which the two sentences were lexically identical, but one had normal place of articulation and the other had an "elliptical" speech version of the sentence. In this case, we predicted that the patient with the cochlear implant would label the two sentences as the "same." If indeed the patient labels the two sentences in this condition as the "same," then this response pattern implies that consonants with the same manner and voicing features but different places of articulation form an equivalence class (are treated as functionally the same) and that the patient is

recognizing words in context using broad phonetic categories. This pattern of results would suggest that patients with cochlear implants hear speech as a sequence of familiar words and do not detect fine phonetic differences.

Up to this point, the discussion has centered on what speech might sound like to users of cochlear implants, and thus what obstacles might have to be overcome to achieve lexical recognition. A second question we are interested in concerns why some users of cochlear implants manage to do quite well in face-to-face conversations despite the degraded input they receive through their implants. One explanation for their good performance in face-to-face conversations is the observation that there are powerful constraints on sound patterns found in the lexicon (Shipman & Zue, 1982; Zue & Huttenlocher, 1983). For example, Zue and Huttenlocher (1983, p. 122) argue that the sound patterns in spoken languages are constrained not only by the inventory of sounds in a particular language but also by the "allowable combinations of those sound units," or the phonotactic constraints of a given language. Shipman and Zue note that an analysis of English which distinguishes only between consonants and vowels can prune a 20,000 word lexicon down to less than 1%, given just the CV pattern of a given word. Since these strong constraints on sound patterns do exist, a very broad phonetic classification can serve to define the "cohort," or the set of possible candidate words having the same pattern. As Shipman and Zue showed in their computational research, these candidate sets may be quite small, such that the "average size for these equivalence classes for the 20,000-word lexicon was found to be approximately 2, and the maximum size was approximately 2000." (Zue & Huttenlocher, p. 122) Thus, even if a listener does not accurately perceive the exact place of articulation, he or she can still recognize the word using broad equivalence classes if he or she can recognize at least the sequence of consonants and vowels in the pattern.

Does coarse coding of the speech signal provide a rich and sufficient enough set of cues to allow normal-hearing listeners to understand what is being said in an utterance? In order to answer this question, Quillet et al. used a transcription task with normal-hearing listeners. In this task, listeners were asked to transcribe three of the five key words from each sentence. The sentences had either normal place of articulation or were produced using "elliptical" speech. The sentences were presented in the clear or in white noise at 0 dB SNR, -5 dB SNR, and -10 dB SNR. Quillet et al. predicted that while speech with normal place of articulation should show decreased intelligibility under conditions of noise-masking or low-pass filtering, "elliptical" speech should actually show the reverse pattern, that is, increased intelligibility as distortion of the signal increased. In their study, they found that speech with normal place of articulation did show decreases in transcription accuracy under conditions of signal degradation whereas the "elliptical" speech showed improvements in transcription accuracy from the 0 dB SNR level to the -5 dB SNR level before dropping at the -10 dB SNR level. Quillet et al. interpreted this finding as support for the proposal that normal-hearing listeners use broad phonetic categories to identify words in sentences under these conditions.

In order to explore whether coarse coding and broad phonetic categories are used by patients with cochlear implants, we carried out a second experiment in which the patient with the cochlear implant was asked to transcribe three key words in a sentence, similar to the experiment described above. Sentences were presented to the cochlear implant patient, and he was asked to transcribe three of the five key words in the sentence. Half of the sentences were produced using "elliptical" speech and half were normal sentences. Our prediction was that the patient with the cochlear implant would show the same transcription performance on sentences with normal place of articulation as he would on sentences produced with "elliptical" speech. If he did show similar transcription performance in these two cases, this pattern would indicate that coarse coding was a sufficient cue for lexical recognition to be carried out with spoken sentences.

Experiment 1: Same-different Task

Experiment 1 employed a same-different discrimination task. Subjects listened to pairs of sentences and categorized the pair as "same" or "different." Subjects were told to label the pair of sentences as "same" if the two sentences that they heard were word-for-word and sound-for-sound identical or "different" if any of the words or speech sounds differed between the two sentences. Normal-hearing listeners have been found to label normal and elliptical versions of lexically identical sentences as the "same" under conditions of signal degradation. Also, there are parallels in confusions in place of articulation between normal-hearing listeners under conditions of signal degradation and listeners with cochlear implants. Thus, we predicted that our patient with a cochlear implant would label the normal and elliptical versions of the same sentence as the "same."

Stimulus Materials

Normal Harvard Sentences. The stimulus materials consisted of 96 Harvard Sentences (IEEE, 1969) taken from lists 1-10 (Egan, 1948). These are English sentences made up of five key words with declarative or imperative structure. Quillet et al. used the same stimulus materials in their experiments.

Anomalous Harvard Sentences. Anomalous sentences were used in this experiment to prevent top-down semantic processing of these sentences. Ninety-six Anomalous Harvard sentences were created by substituting random words of the same lexical category (noun, verb, etc.) into lists 11-20 of the Harvard sentences. The inserted words were selected from lists 21-70 of the Harvard sentences (with five lists being used to supply replacement words for each list). This differs from Quillet et al.'s methodology, in which only normal Harvard sentences were used.

"Elliptical" Speech. Several new sets of "elliptical" sentences were generated through a process of featural substitution similar to that employed by Miller and Nicely. The stops, fricatives, and nasal consonants in each of the five key words were replaced with a new consonant that preserved the same manner and voicing features of the original consonant but changed the place feature to an alveolar place of articulation. Liquids /r l/ and glides /y w/ were excluded from the substitution process. The normal sentences and the elliptical versions are listed in the Appendix. Several examples are given in (1) below, with the key words underlined.

(1) a. A wisp of cloud hung in the blue air.
A wist of tloud hund in the dlue air.

b. Glue the sheet to the dark blue background.
Dlue the seet to the dart dlue datdround.

This method of replacing consonants with alveolar consonants follows Miller and Nicely's original method of creating "elliptical" speech and differs from the methodology used by Quillet et al. They followed Miller (1956) by replacing consonants with consonants randomly selected from within the equivalence class sharing manner and voicing features. An example from Miller is shown in (2), in which it can be seen that the replacement consonants do not all have the same alveolar place of articulation:

(2) a. Two plus three should equal five.
Pooh kluss free soub eatwell size.

In the present study, half of the utterances were spoken by a male speaker and the other half were spoken by a female speaker. Both talkers practiced saying the test sentences several times before the

recording session. An attempt was made to use the same intonation pattern in both versions of an utterance. Sentences were recorded using a head-mounted Shure model SM98A microphone and a Sony TCD-D8 DAT recorder. The recordings were then segmented into individual utterances, converted to a single channel, and downsampled to 22,050 Hz using CoolEdit.™ The use of natural speech stimuli in this study differs from Quillet et al.'s 1998 procedure, which used synthetic speech for all of their stimuli, which were generated using DECtalk.

Signal Degradation

Low-pass Filtering. For the normal-hearing listeners, a new set of stimuli was created from the original recordings. Low-pass filtering was applied to the signal using Matlab. Specifically, the signal-processing tool "Colea" was used (Loizou, 1998). Colea's "filter tool" was used to apply a 10th order low-pass Butterworth filter with a cutoff of 1000 Hz. This procedure was applied to all of the sentences individually and each was saved as a separate file. Thus, the filtering was done off-line prior to presentation of the stimuli to the listeners.

Noise-masking. Gaussian noise was applied to each sentence to create another set of stimuli. Colea was used for this purpose as well. Noise was added at a –5 dB signal-to-noise ratio. Each noise-masked file was saved as a separate file for use during presentation of the stimuli to the listeners. This procedure also differs from Quillet et al.'s methodology. To degrade their signals, they used different levels of random-bit-flip noise in their same-different task and white noise at three different signal-to-noise ratios in their transcription task.

Subjects

The adult patient with the cochlear implant, "Mr. S," was 36 years old at the time of testing. He had been profoundly deaf (with an unknown etiology) for 20 months before receiving his implant at age 32. "Mr. S" has participated as a listener in prior studies, and is considered to be an excellent user of his cochlear implant (see also Goh, Pisoni, Kirk, & Remez, 1999; Herman & Clopper, 1999).

Nine normal-hearing listeners were assigned to the low-pass filtered condition and another nine were assigned to the noise-masked condition. All listeners were enrolled in an undergraduate introductory psychology course and received course credit for their participation in this experiment. Listeners ranged in age from 18-22 years old. None of the listeners reported any hearing or speech disorders at the time of testing. All listeners were native speakers of American English.

Procedures

"Mr. S" heard the stimuli over a Harman/Kardon HK 195 loudspeaker. He was given four pre-experiment trials in which he could adjust the volume of the loudspeaker to a comfortable listening level. The experiment was controlled by a Visual Basic program running on a PC that also recorded subject responses. The experiment was self-paced. Each pair of sentences was presented only once. There was a 500 ms interval between the two sentences in each pair. He entered his responses by using the computer mouse to click on a box labeled "same" or a box labeled "different" on the computer monitor. "Mr. S" heard 96 pairs of sentences in four blocks of 24 trials each. He heard a block of normal Harvard sentences spoken by the male speaker, then a block of normal Harvard sentences spoken by the female speaker, then a block of anomalous Harvard sentences spoken by the male speaker, and finally a block of anomalous Harvard sentences spoken by the female speaker. Half were elliptical speech and half were speech with a normal place of articulation.

Normal-hearing subjects followed the same procedure as "Mr. S" except that they heard the stimuli through Beyerdynamic DT 100 headphones at a comfortable listening level of about 70 dB SPL. There was a one-second interval between the two sentences in each pair. (The inter-stimulus interval was changed to one second for the normal-hearing subjects after pilot testing showed that a one-second inter-stimulus interval worked better than a 500 ms inter-stimulus interval. This change was made after testing of "Mr. S" had already taken place, which explains the two different inter-stimulus intervals.) They heard 192 pairs of sentences in a random order. For the normal-hearing listeners, half of the pairs had signal degradation and half were heard in the clear. The signal degradation was either low-pass filtering for one group or noise masking for the other group. The type of signal-degradation used was a between-subjects variable.

All subjects received eight possible types of pairs of sentences, as shown in Table 1. In this report, pairs of sentences that are lexically identical are marked with two subscript "i's". Pairs of sentences that are lexically different are marked with a subscript "i" and a subscript "j". The sentences with normal place of articulation are referred to by "N". The "elliptical" sentences are referred to by "E".

	Different sentence	Same sentence
both normal place of articulation:	NiNj	NiNi
both "elliptical" speech:	EiEj	EiEi
one normal, one elliptical	NiEj	NiEi
	EiNj	EiNi

Table 1. The different types of pairs of sentences used in the same-different task.

Results: Normal-hearing Listeners

Normal Harvard Sentences. A summary of the results for "same" responses for the normal-hearing listeners listening to normal Harvard sentences, where the signal degradation was low-pass filtering, is shown in Figure 1. The types of sentence pairings are listed along the X-axis (i.e. NiNj, EiEj). The percent labeled as the "same" is shown along the Y-axis. Sentences heard in the clear are shown with open bars, and sentences heard with low-pass filtering at 1kHz are shown with dark bars.

A 2x2x4 analysis of variance (ANOVA) was carried out. The first factor was "sense," with the two levels being normal Harvard sentences and anomalous Harvard sentences. (Both normal Harvard sentences and anomalous Harvard sentences are included in the analysis, although they are shown separately in Figures 1 and 3, for convenience so that they may be examined separately.) The second factor was degradation, with the two levels being sentences heard in the clear vs. sentences with low-pass filtering. The third factor was the type of pair, with four levels. This factor, whose levels can be seen in the four different cells in Table 1 or as four pairs along the X-axis in Figure 1, collapsed across orderings. We were interested in whether there was a statistical difference between sentences heard in the clear and sentences heard with low-pass filtering, particularly when one sentence had normal place of articulation and the other sentence was the "elliptical" version of the same sentence. Thus, we grouped together NiNj with EiEj (two different sentences, either both have normal place of articulation or both have "elliptical" speech), NiEj with EiNj (two different sentences, one has normal place of articulation and one has "elliptical" speech), NiNi with EiEi (the same sentence twice, both have normal place of articulation or both have "elliptical" speech), and the crucial test cases of NiEi with EiNi (the same sentence twice, one has normal place of articulation and one has "elliptical" speech).

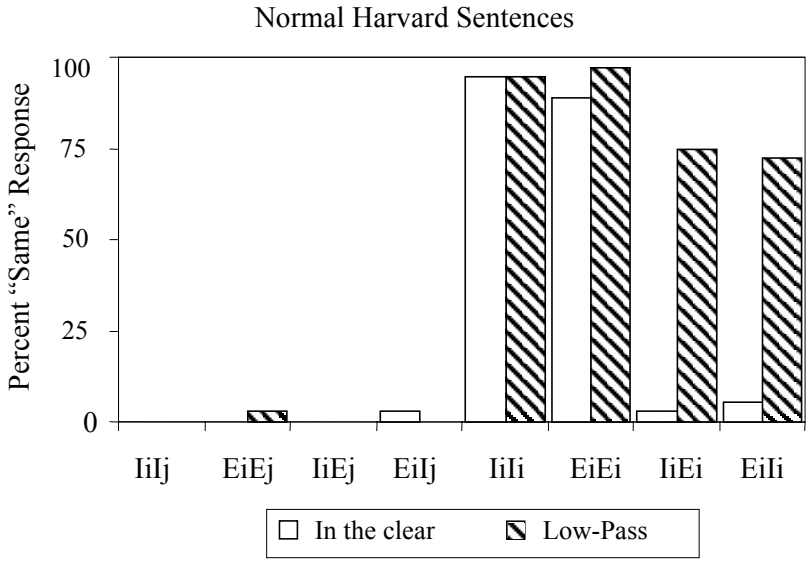


Figure 1. Results from the same-different task for normal-hearing listeners.

There was a main effect of "sense" indicating that the responses to normal Harvard sentences were significantly different from the responses to anomalous Harvard sentences (F (1,8) = 7.2, $p < .05$). There was also a main effect of degradation, so the responses to sentences head in the clear were significantly different from the responses to sentences heard under low-pass filtering (F (1,8) = 105.09, $p < .001$). There was also a main effect of "type" (F (3,24) = 408.6, $p < .001$).

The analysis also showed a significant two-way interaction between signal degradation and type of pair. Signal degradation within the first level (NiNj and EiEj) showed no variability. A post-hoc test of simple effects showed that signal degradation within the second level (NiEj and EiNj) was not significant. Thus, when the normal-hearing listeners heard two different sentences, they were able to correctly discriminate the differences and respond "different" regardless of whether the sentences were both normal or both elliptical, or one sentence was normal and one was elliptical. Moreover, there was no difference in performance in the clear vs. performance in the filtered condition for these types of pairs.

The test of simple effects showed that signal degradation within the third level (NiNi and EiEi) was not significant. In these two cases (NiNi and EiEi), the identical token was heard twice. In these two cases, the listeners correctly labeled the two sentences as the "same" a high number of times, and there was no statistical difference between when these signals were heard in the clear and when they were heard with filtering.

The post-hoc test of simple effects did show that signal degradation within the fourth level (NiEi and EiNi) was significant (F (1,8) = 217.4, $p < .001$). These two cases are the most interesting ones for testing the hypothesis that ellipsis is undetectable under degraded conditions. In these two cases, two sentences that were lexically identical were heard, but one sentence had normal place of articulation and one had "elliptical" speech. In both cases, the listeners labeled the pairs as "same" a very low percentage of time when they were heard in the clear, but they did label them as "same" in a majority of cases when

the sentences were heard under low-pass filtering, and there was a statistical difference between when the sentences were heard in the clear and when the sentences were heard with low-pass filtering.

This pattern of results, in which a sentence with normal place of articulation and its elliptical version are labeled "different" when heard in the clear but are labeled the "same" when heard with low-pass filtering, confirms the earlier observations by Miller and Nicely (1955) that ellipsis of place of articulation is undetectable under signal degradation. These findings with normal-hearing listeners replicate the previous results reported by Quillet et al.

We turn now to the conditions in which the signal degradation was noise masking. A summary of the results for the "same" responses for the normal-hearing listeners listening to pairs of normal Harvard sentences, where the signal degradation was noise masking, is shown in Figure 2. These results parallel the results shown in Figure 1. A 2x2x4 ANOVA of these results also shows a main effect of "sense" (F (1,8) = 14.29, $p < .01$), degradation (F (1,8) = 217.35, $p < .001$), and type of pair (F (3,24) = 2418.99, $p < .001$).

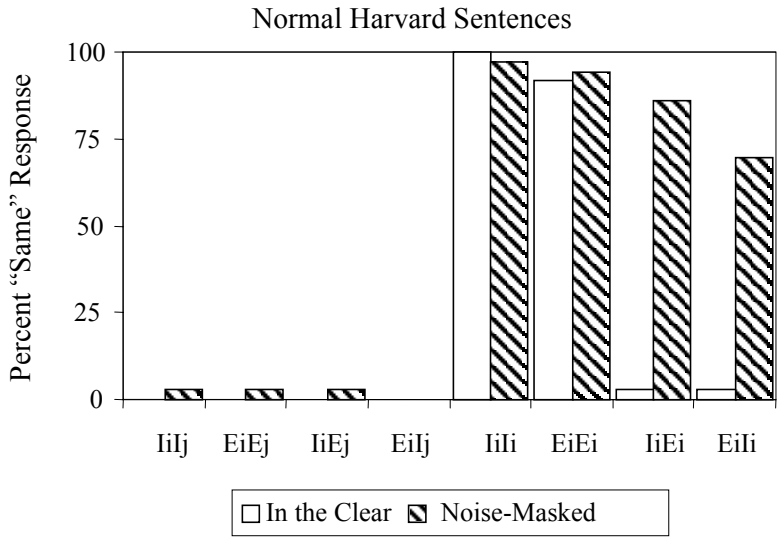


Figure 2. Results from the same-different task for normal-hearing listeners.

The ANOVA also shows an interaction between signal degradation and type of pairs. A post-hoc test of simple effects found that signal degradation within the first level (NiNj and EiEj) was not significant. The test of simple effects found that signal degradation within the second level (NiEj and EiNj) was also not significant. Thus, when the two sentences in a pair were lexically different sentences (NiNj, EiEj, NiEj, and EiNj) there was a very low percent of pairs labeled "same" and there is no statistical difference between when these signals were heard in the clear and when they were heard with noise masking.

The test of simple effects showed that signal degradation within the third level (NiNi and EiEi) was not significant. In these two cases (NiNi and EiEi), the identical token was heard twice. The listeners correctly labeled the two sentences as the "same" a high number of times, and there was no statistical

difference between when these signals were heard in the clear and when they were heard with noise masking.

The post-hoc test of simple effects did show that signal degradation within the fourth level (NiEi and EiNi) was significant (F (1,8) = 345.6, $p < .001$). These cases were the crucial test conditions, in which a sentence with normal place of articulation was paired with an "elliptical" speech version of the same sentence. In such cases, listeners labeled those two sentences as the "same" a very low percentage of the time when heard in the clear. However, under conditions of noise masking at -5 dB SNR, listeners did tend to label those two sentences as the "same" on a majority of trials. Thus, the same pattern of ellipsis is observed under both low-pass filtering and noise masking in normal-hearing listeners.

Anomalous Harvard Sentences. A summary of the main results for the normal-hearing listeners listening to pairs of anomalous Harvard sentences, where the signal was low-pass filtered, is shown in Figure 3. The results for the noise-masked conditions are shown in Figure 4. These results are similar to the results for normal Harvard sentences shown in Figures 1 and 2 (and the statistical results were included in the results reported above). Pairs of sentences that were lexically different were labeled as the "same" on a very low percentage of trials, and there was no statistical difference between when these sentences were heard in the clear or with signal degradation. Pairs of sentences in which the same token was presented twice tended to be labeled as the "same" for the majority of cases, again with no statistical difference between responses for sentences heard in the clear and with signal degradation. Pairs in which one sentence had normal place of articulation and the other had "elliptical" speech were labeled as "different" on a high percentage of the trials when heard in the clear. However, under conditions of low-pass filtering (Figure 3) or noise masking (Figure 4), listeners labeled these pairs as the "same" on a high percentage of trials.

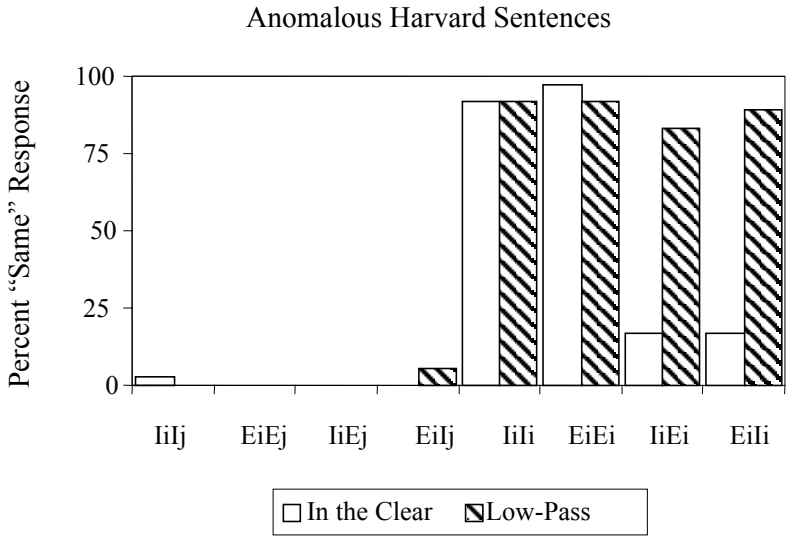


Figure 3. Results from the same-different task for normal-hearing listeners.

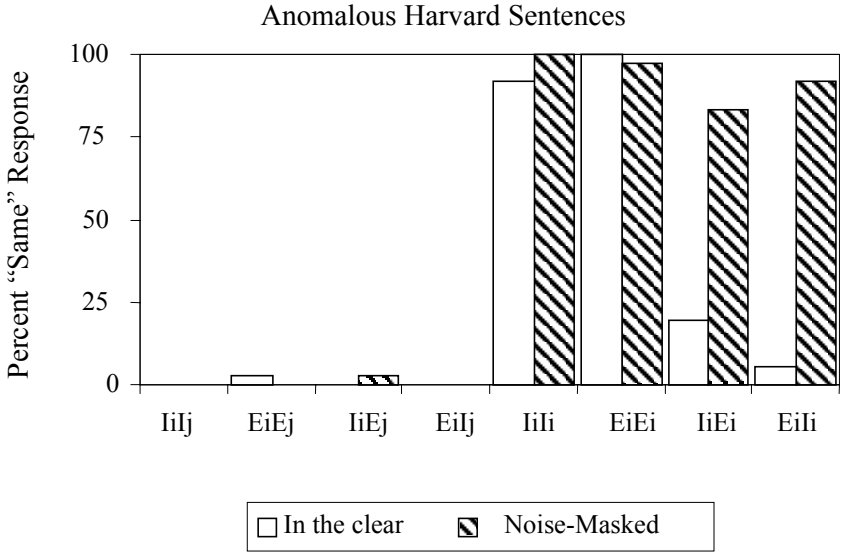


Figure 4. Results from the same-different task for normal-hearing listeners

The findings shown in Figures 1-4 for normal-hearing listeners confirm Miller and Nicely's observation that ellipsis of place of articulation is very difficult to detect under degraded conditions such as low-pass filtering and noise masking. These findings replicate Quillet et al.'s results from their same-different task, and demonstrate that normal-hearing listeners do label pairs of sentences which are lexically identical but where one has normal place of articulation and one has "elliptical" speech as the "same" a majority of the time when heard under conditions of signal degradation. Having shown that we can obtain these effects in normal-hearing listeners under both low-pass filtering and noise masking, we turn to our CI patient, "Mr. S."

Results: Patient with Cochlear Implant

Normal Harvard Sentences. A summary of the main results for "Mr. S" listening to pairs of normal Harvard sentences is shown in Figure 5. Again, the type of sentence pair is shown along the X-axis and the percentage of sentence pairs labeled as the "same" is shown along the Y-axis.

In Figure 5, it can be seen that "Mr.S" did not label any of the pairs consisting of two different sentences as the "same" (looking at the pairs including NiNj, EiEj, NiEj, and EiNj). However, he labeled 100% of the pairs consisting of the identical sentence heard twice as the "same" (looking at the pairs including NiNi and EiEi). Thus, he shows the same performance as the normal-hearing subject. The crucial cases for observing the perception of "elliptical" speech are the two conditions labeled NiEi and EiNi. In these two conditions, lexically identical sentences are presented on each trial, but one sentence has normal place of articulation and the other consists of "elliptical" speech. In the cases in which the sentence with normal place of articulation was heard first, "Mr. S" labeled the two sentences the "same" on 75% of the trials, and in cases where the sentence with "elliptical" speech was heard first, he labeled the two sentences as the "same" in 50% of the trials. Thus, overall, he tends to label normal and "elliptical" speech versions of sentences as the "same," although there is an order effect. This pattern parallels the findings obtained for normal-hearing listeners under degraded conditions.

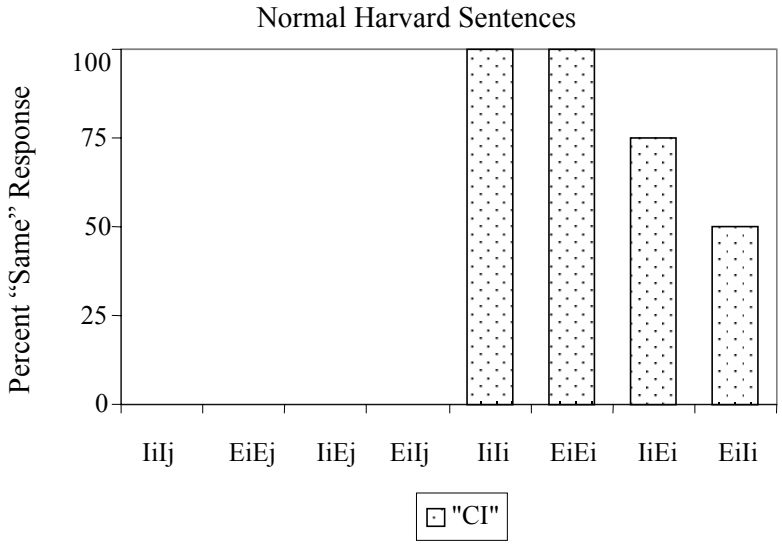


Figure 5. Results from the same-different task for the listener with a cochlear implant.

Anomalous Harvard Sentences. A summary of the main results for "Mr. S" listening to pairs of anomalous Harvard sentences is shown in Figure 6. Here, it can be seen that the same pattern of results found for normal Harvard sentences is also found for anomalous Harvard sentences. That is, "Mr. S" labels pairs of sentences that were different (in the sense of consisting of different lexical items) as "different" in 100% of the trials and he labels pairs of sentences that were identical as the "same" in 100% of the trials. And again, he tended to label a sentence with normal place of articulation and its "elliptical" version as the "same" in a majority of trials, again paralleling the performance of normal-hearing listeners under conditions of signal degradation.

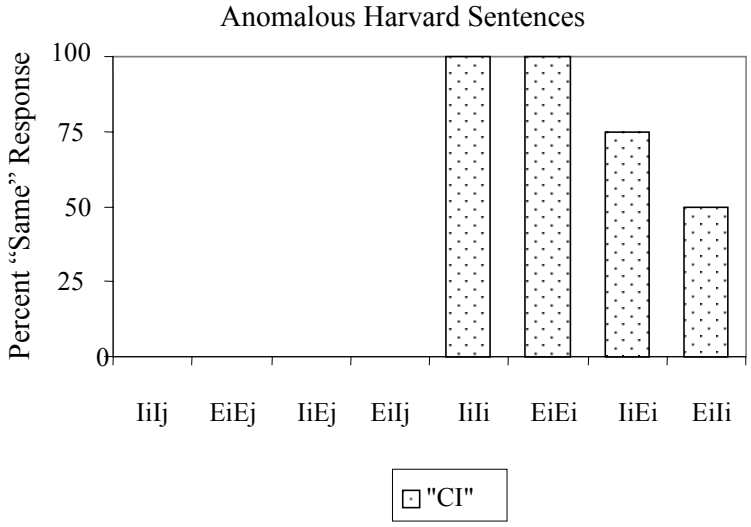


Figure 6. Results from the same-different task for the listener with a cochlear implant.

The results shown in Figures 5-6 support our prediction that a patient with a cochlear implant would show similar perceptual behavior to normal-hearing listeners under degraded conditions. In particular, "Mr. S" was unable to detect the presence of elliptical speech on a majority of the trials in which a sentence with normal place of articulation was paired with an "elliptical" speech version of the same sentence. This suggests that contrasts such as place of articulation in consonants may not be completely detectable to users of cochlear implants despite the fact that they can recognize spoken words and understand sentences.

Discussion

When pairs of sentences are presented in the clear, normal-hearing listeners can easily distinguish stimuli that have normal place of articulation from "elliptical" speech. However, under conditions of signal degradation such as low-pass filtering or noise masking, the information about place of articulation becomes less reliable and listeners tend to label the normal and elliptical versions as the "same." Similarly, the cochlear implant user tends to label normal and elliptical versions of the same sentence as the "same," suggesting that he is perceiving speech and recognizing words using broad phonetic categories.

Informal questioning after the experiment led to very different responses from "Mr. S" as opposed to the normal-hearing subjects. At the close of the experiment, "Mr. S" did not mention any awareness of the ellipsis in the stimulus materials. A post-test questionnaire administered to the normal-hearing listeners, on the other hand, revealed that all of the normal-hearing listeners were aware of the elliptical speech. These listeners described what they heard as words being "slurred," as the "t's" in words being pronounced incorrectly, as some of the letters in each word being transposed, as the "s" and "t" being used interchangeably, as a "speech impediment," as sounding as if spoken with a lisp, or as sounding "like Latin or German." Thus, the normal-hearing listeners had some conscious explicit awareness of the ellipsis in the stimulus materials whereas "Mr. S" did not seem to have an explicit awareness of the ellipsis.

The overall pattern of same-different discrimination responses by normal-hearing listeners under degraded conditions and by our patient with a cochlear implant was very similar to each other despite some small differences in procedure. Signal degradations for normal-hearing listeners and the use of a cochlear implant both seem to encourage the use of a coarse coding in which place of articulation differences are no longer perceptually salient, which is indicated by labeling the normal version and the elliptical versions of a sentence as the "same" under those conditions. Normal-hearing listeners under signal degradation and a patient with a cochlear implant both give responses that suggest the use of broad equivalence classes when only partial information is present in the signal. This resembles Quillet et al.'s results for normal-hearing listeners under conditions of signal degradation.

Experiment 2: Transcription of Key Words

Our second experiment employed a transcription task. Subjects heard a sentence and were asked to transcribe three of the five key words from each sentence. For each of these key words, a blank line was substituted in a text version of the sentence.

In Experiment 1, the sentences with normal place of articulation were heard as the "same" as sentences with "elliptical" speech. Therefore, we predicted that under conditions of signal degradation both normal-hearing listeners and our patient with a cochlear implant would transcribe "elliptical" speech at the same level of accuracy as they transcribed speech with normal place of articulation.

Stimulus Materials

The stimulus materials were constructed the same way for Experiment 2 as they were for Experiment 1. However, for "Mr. S," separate sets of sentences were used in the two experiments, so that he heard no sentence in both Experiment 1 and Experiment 2. "Mr. S" heard 96 sentences, half of which were normal Harvard sentences and half of which were anomalous Harvard sentences. Half of the sentences in each set were pronounced with normal place of articulation and half contained "elliptical" speech. Half were spoken by the male speaker and half were spoken by the female speaker.

The normal-hearing listeners in this experiment heard 192 sentences. This was the same set of sentences used in Experiment 1. Different listeners participated in the two experiments.

Signal Degradation

For the normal-hearing listeners, a third of the sentences were heard in the clear, a third were heard under low-pass filtering at 1000 Hz, and a third were heard under noise masking of -5 dB SNR. Low-pass filtering and noise masking were both applied to the signal using Colea (Loizou, 1998), as in Experiment 1.

Subjects

"Mr. S," who participated in Experiment 1, served as our patient with a cochlear implant in Experiment 2 as well.

Nine normal-hearing listeners participated in this experiment. All subjects were enrolled in an undergraduate psychology course and received course credit for their participation. These listeners ranged in age from 18-22. None reported any history of speech or hearing disorders at the time of testing. All were native speakers of American English. None of these listeners had participated in Experiment 1.

Procedures

"Mr. S" heard the sentences over a loudspeaker, at a self-selected comfortable level of loudness. Sentences were presented one at a time in a random order. He could listen to each sentence up to five times, after which he had to enter a response. After hearing the sentence, he could select either "listen again" or "next trial." The experiment was self-paced. The current trial number was displayed on the monitor. He wrote his responses on a printed response sheet. The response sheet contained all of the sentences written out, with each sentence containing three blank lines replacing the three key words that the subjects were asked to transcribe. Thus, subjects did have access to the sentential context of the key words.

Normal-hearing listeners followed the same procedures as "Mr. S" They heard the sentences over headphones at a comfortable listening level of around 70 dB SPL. Four different random orders were used for the normal-hearing subjects. They also could listen to each sentence up to five times, after which they had to enter a response, and the experiment was self-paced for the normal-hearing subjects as well.

Scoring of transcriptions was done using a strict criterion of whether the word written down by the subject exactly matched the intended word. That is, in the elliptical cases, the scoring was done on the basis of whether the original, intended English word was written down, not on the basis of whether the elliptical version which was actually heard was written as an English word or transcribed in an

approximation to phonetic transcription. For example, suppose the target word was "dark" and the elliptical version that was heard in the sentence as the stimulus was "dart." In this case, if the subject wrote "dart" then this would be scored as "incorrect" while if the subject wrote "dark" then this would be scored as "correct."

Results: Normal-hearing Listeners

A 2x3x2 analysis of variance (ANOVA) was conducted, in which the three factors were (a) speech with normal place of articulation vs. "elliptical" speech, (b) signal degradation (in the clear vs. low-pass filtered at 1 kHz vs. noise-masked at -5 dB SNR), and (c) normal Harvard sentences vs. anomalous Harvard sentences. A main effect of speech with normal place of articulation vs. "elliptical" speech was found (F (1,8) = 193.356, $p < .001$). A main effect of signal degradation was also found (F (2,16) = 148.87, $p < .001$). A main effect of normal Harvard sentences vs. anomalous Harvard sentences was also found (F (1,8) = 359.80, $p < .001$). A significant 3-way interaction was found among these factors (F (2,16) = 10.033, $p < .01$). In order to probe the results further, the data were split along one of the factors. The normal Harvard sentences were examined separately from the anomalous Harvard sentences.

Normal Harvard Sentences. A summary of the main results for the normal-hearing listeners' transcription performance when listening to normal Harvard sentences is shown in Figure 7. In this figure, the signal degradations are shown along the X-axis. The percent of correct transcriptions is shown along the Y-axis. Transcription performance for speech with normal place of articulation is shown with the open bars, and transcription performance for "elliptical" speech is shown with the dark bars. This graph shows the average performance for all nine normal-hearing listeners.

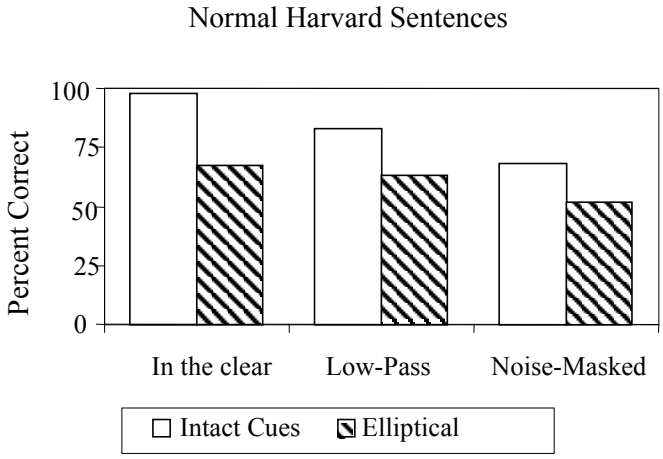


Figure 7. Results from the transcription task for the normal-hearing listeners.

A 2x3 ANOVA for just the normal Harvard sentences shows a main effect of normal place of articulation vs. "elliptical" speech (F (1,8) = 102.75, $p < .001$). Also, there is a main effect of speech heard in the clear vs. low-pass filtering vs. noise masking (F (2,16) = 48.14, $p < .001$). There was no interaction between these two factors. The results indicate that transcription performance for "elliptical" speech heard in the clear is lower than transcription performance for speech with normal place of articulation, which is an expected result.

Under low-pass filtering, transcription performance is still lower for "elliptical" speech than for speech with normal place of articulation. Furthermore, under conditions of noise masking at -5 dB SNR, transcription performance is again lower for "elliptical" speech than for speech with normal place of articulation. The overall pattern of the results does not support the prediction that transcription performance for speech with normal place of articulation and "elliptical" speech under conditions of signal degradation are the same. We expected to find similar transcription performance for speech with normal place of articulation and "elliptical" speech, because the two types of signals tended to be identified as the "same" in the same-different discrimination task in Experiment 1. However, in the same-different task, the percent of trials labeled as the "same" when one sentence had normal place of articulation and the other had ellipsis, while significantly differently from each other for speech heard in the clear vs. speech heard under signal degradation, nonetheless did show a lower percent of trials which were labeled as the "same" in the NiEi and EiNi conditions than in the NiNi and EiEi conditions. Thus, although much of the place information may have been eradicated by the signal degradation, there may have been some weak phonetic cues left in the signal after signal degradation because of the redundancy in natural speech. It may be that these weak phonetic cues remaining in the signal after signal degradation were the cause of the lower percent of trials labeled "same" in the same-different task in the NiEi and EiNi conditions. If there were some cues to place of articulation still in the signal even after signal degradation, then it may also be the case that these cues helped to boost transcription performance for speech with normal place of articulation. The fact that each sentence was heard up to five times may have reinforced whatever weak cues to place of articulation were still present in the signal (although this would not explain the differences between the present study and Quillet et al.'s study, since listeners in their study also heard each sentence up to five times). The findings that transcription performance was not improved for elliptical speech under conditions of signal degradation fail to replicate the earlier results of Quillet et al., who found an increase in transcription performance of "elliptical" speech from noise-masking of 0 dB SNR to noise masking of -5 dB SNR. However, Quillet et al. used synthetic speech, which has less redundancy than natural speech that was used in the present study. Thus, the rich, redundant natural speech cues present in the stimuli of the current experiment may have actually "survived" the signal degradation more robustly than Quillet et al.'s synthetic speech, thus providing contradictory information by presenting weak cues to alveolar place of articulation, even under conditions of signal degradation.

Anomalous Harvard Sentences. A summary of the main results for the normal-hearing listeners transcribing anomalous Harvard sentences is shown in Figure 8. Again, the signal degradations are shown along the X-axis and the percent of correct transcriptions is shown along the Y-axis.

A 2x3 ANOVA on the anomalous Harvard sentences showed a main effect of normal place of articulation vs. "elliptical" speech (F (1,8) = 345.83, $p < .001$). Also, there was a main effect of speech heard in the clear vs. low-pass filtering vs. noise masking (F (2,16) = 133.45, $p < .001$). There was also a significant 2-way interaction between these two factors (F (2,16)=63.83, $p < .001$). A test of simple effects found that the transcription performance for speech with normal place of articulation was significantly different from the transcription performance for "elliptical" speech in the clear (t (8) = 16.06, $p < .001$) and under low-pass filtering (t (8) = 4.0, $p < .01$). However, transcription performance for speech with normal place of articulation vs. "elliptical" speech was not significantly different from each other when heard with noise masking.

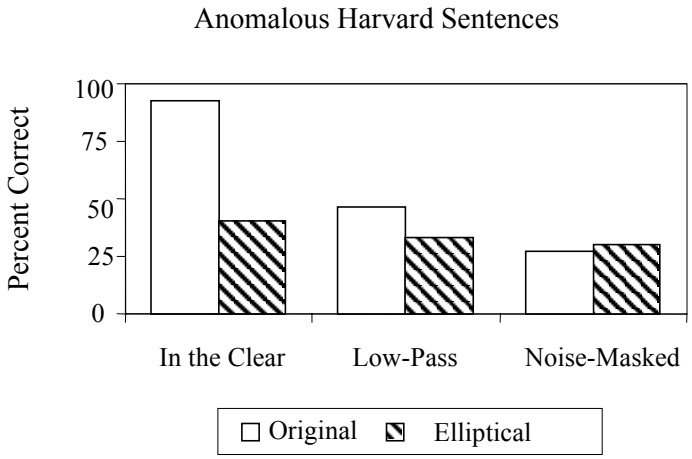


Figure 8. Results from the transcription task for the normal-hearing listeners.

Transcription performance for "elliptical" speech was much lower than for speech with normal place of articulation when heard in the clear. This was not unexpected. The "elliptical" anomalous sentences are both semantically anomalous and have "strange" places of articulation, making them extremely difficult to parse. Under low-pass filtering, the transcription performance for "elliptical" speech was still lower than for speech with normal place of articulation. This result also did not support our predictions that transcription performance for speech with normal place of articulation and "elliptical" speech would be the same under degraded conditions. Under noise masking, the transcription performance for both speech with normal place of articulation and "elliptical" speech was extremely low, around 25-30% correct. In this case, the transcription performance for "elliptical" speech was slightly higher than for speech with normal place of articulation, but both scores were so low that this finding may simply be due to a lack of variability at such low levels. Thus, the prediction that speech with normal place of articulation and with "elliptical" speech should show equivalent transcription performance under degraded was not supported by these findings. Again, it may be that even though most of the phonetic cues to place of articulation were eradicated by the signal degradation, there were still some weak phonetic cues to place of articulation present in the stimuli. Such cues, although weak, may have provided confusion for listeners in the "elliptical" speech condition, thus lowering those scores. And since each sentence was heard up to five times by listeners, the repetition may have helped to reinforce whatever weak phonetic cues to place of articulation that were still present in the signal after degradation. In general, though, the task of transcribing anomalous Harvard sentences, either with or without ellipsis of place of articulation under conditions of signal degradation, was a very difficult task for listeners.

In summary, the results for the normal-hearing listeners in the transcription task shown in Figures 7-8, do not replicate the earlier findings of Quillet et al., which did show improvement in transcription performance for "elliptical" speech under degraded conditions. It may be that the natural speech used here, with all of the rich redundant phonetic cues present in natural speech, provided some weak cues to place of articulation, despite signal degradation. Thus, if the alveolar place of articulation was perceived in some of the tokens, this would have resulted in lowered transcription performance.

Results: Patient with Cochlear Implant

Normal Harvard Sentences. A summary of the main results for "Mr. S's" transcription performance when listening to normal Harvard sentences is shown in Figure 9. The percent of words correctly transcribed is shown along the Y-axis. Speech with normal place of articulation is shown with the open bar and "elliptical" speech is shown with the dark bar.

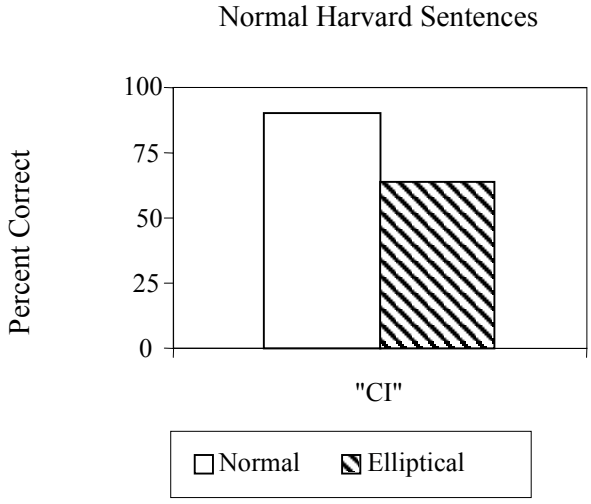


Figure 9. Results from the transcription task for "Mr. S"

As shown in Figure 9 for the Harvard sentences, "Mr. S" transcribed the normal speech with very high levels of accuracy. However, transcription of the elliptical speech declined relative to the normal speech. This pattern does not match the predicted outcome. We expected that the transcription performance would be similar for these two conditions because speech with normal place of articulation and "elliptical" speech were labeled as the "same" in a majority of trials in Experiment 1. However, "Mr. S" labeled only 75% of NiEi cases in the same-different task the "same" and only 50% of the EiNi cases in the same-different task the "same." Thus, despite the presumed loss of information about place of articulation due to the cochlear implant, there may still be some weak phonetic cues present which provide place of articulation information. If so, then the alveolar place of articulation in the "elliptical" sentences may have provided conflicting information, lowering his transcription scores.

Anomalous Harvard Sentences. A summary of the main results for "Mr. S's" transcription performance when listening to anomalous Harvard sentences is shown in Figure 10. Overall, "Mr. S" showed a lower percentage of correct transcriptions for Harvard anomalous sentences as compared with the normal Harvard sentences (seen in Figure 9). This result was expected because the anomalous Harvard sentences are more difficult to parse than the normal Harvard sentences. Again, the elliptical anomalous sentences show a lower percent correct transcription than the normal sentences. This was an unexpected result and may be due to the remaining weak phonetic cues to place of articulation in the signal that he receives. Again, it must be remembered that transcribing words from anomalous Harvard sentences, where the words have ellipsis of place of articulation, is an extremely difficult task, as evidenced by the extremely low transcription score in this case.

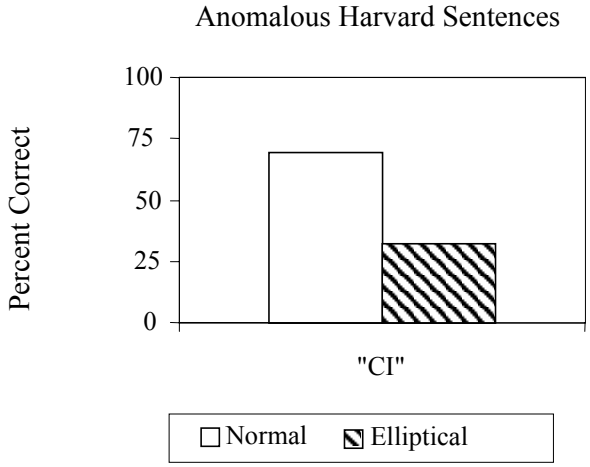


Figure 10. Results from the transcription task for "Mr. S"

"Mr. S's" performance in the transcription task, as shown in Figures 9-10, did not match the predictions based on the same-different task in Experiment 1. Although he did extremely well in transcribing key words in normal Harvard sentences with normal place of articulation, he showed worse transcription performance for "elliptical" speech. The poorer performance for "elliptical" speech also emerged when transcribing anomalous Harvard sentences as well.

Discussion

Both our normal-hearing listeners and our patient with a cochlear implant, "Mr. S", showed evidence of using lexical knowledge in the transcription task. The examples below, taken from normal hearing listeners' transcriptions, reveal the degree of top-down processing for anomalous Harvard sentences. The intended utterance is shown first and the transcribed utterance is shown second. The key words, which were left blank on the response sheets and which subjects wrote in by hand, are shown in these examples. As shown in these examples, higher-level lexical and semantic context plays a greater role in some transcriptions than phonological regularities relative to the stimulus signal.

(3) Anomalous Harvard Sentences

a. *stimulus*: A winding dinner lasts fine with pockets.
response: A wine dinner lasts fine with pasta.

b. *stimulus*: These dice bend in a hot desk.
response: These guys are in a hot bath.

c. *stimulus:* Steam was twisted on the front of his dry grace.
response: Skin was plastered on the front of his dry grapes.

d. *stimulus:* Metal can sew the most dull switch.
response: Mother can sew the most dull slips.

These examples of top-down lexical and semantic processing are reminiscent of the examples of perception of synthetic speech found in Pisoni (1982, p. 18). For example, Pisoni reports the anomalous input "The bright guide knew the glass" and the wrong response "The bright guy threw the glass." In the examples of the perception of synthetic speech, just as in the examples in the current study with degraded speech, responses to sentences which are difficult to perceive show a strong tendency or bias towards generating meaningful responses, even if such a response leads to a complete reanalysis of the sound structure of the words in the sentence. In the examples in (3), the normal-hearing listeners' errors do not show a simple place of articulation substitution but rather their responses are errors in the sequence of manners of articulation.

In several interesting cases, "Mr. S" seems to be following a very different perceptual strategy than the normal-hearing listeners. He seems to make much more sophisticated guesses based on lower-level phonological regularities in the signal, combined with top-down guidance, whereas normal-hearing listeners tend to use considerably more top-down lexical processing and context, but do not necessarily exploit phonological regularities. For example, "Mr. S" tended to substitute sounds that have similar voicing and manner to the word that he heard and which share a sequence of manners of articulation and of voicing values with the original utterance.

(4) Anomalous Harvard Sentences

a. *stimulus:* They could scoot although they were cold.
"Mr. S's" response: They could scoop although they were cold.

b. *stimulus:* Green ice can be used to slip a slab.
"Mr. S's" response: Clean ice can be used to slip a sled.

c. *stimulus:* Grass is the best weight of the wall.
"Mr. S's" response: Brass is the best weight of the wall.

All of these examples show errors in the perception of place of articulation, but the general phonological shape of the intended word is correctly perceived and the sequence of manners of articulation (such as fricative, liquid, vowel, stop) are correctly perceived. This difference in error patterns between "Mr. S" and the normal-hearing listeners may be due to our patient's long-term experience and familiarity listening to highly degraded speech through his cochlear implant. If "Mr. S" must constantly guess at place of articulation given the general prosodic form of words and the sequence of manners of articulation, and if he is aware that place of articulation distinctions are not as perceptible to him and are not reliable cues to word recognition, as they were before his hearing impairment, it is very likely that he would develop more sophisticated perceptual strategies for coarse coding the input speech signals. On the other hand, the normal-hearing listeners have little if any experience listening to signals as severely degraded as the ones presented in this study or the ones presented via a cochlear implant. The normal-hearing listeners were only exposed to these signals for a very short period of time and then received no feedback in any of these experiments.

General Discussion

Despite difficulties in perceiving some fine phonetic contrasts, such as place of articulation in consonants, many cochlear implant users are able to comprehend fluent speech. What does the speech sound like for users of cochlear implants? How do patients with cochlear implants manage to comprehend spoken language despite receiving degraded input? The results of these two experiments provide some

interesting new insights into the underlying process and suggest some possibilities for intervention and oral rehabilitation for adult patients immediately after they receive a cochlear implant.

The first experiment, a same-different task using pairs of sentences which either had normal place of articulation or "elliptical" speech, replicated the informal observations of Miller and Nicely (1955) that speech which is impoverished with respect to place of articulation may not be perceived as deficient under degraded conditions such as noise-masking and low-pass filtering, which reinstate or reproduce the conditions that produced the degradation. Normal-hearing listeners were able to distinguish the normal version of a sentence from an elliptical version in the clear, but they displayed a perceptual bias for labeling a normal and an elliptical version as the "same" when the two sentences were degraded under noise-masking or low-pass filtering. "Mr. S" also tended to label the normal version and the elliptical version of the sentence as the "same." This pattern of results from the same-different task indicates that low-pass filtering, noise masking, or use of a cochlear implant all encourage the use of "coarse coding" in which categories of sounds which bear resemblances to each other are all identified as functionally the same. Equivalence classes, consisting of phonemes with the same manner of articulation and the same voicing, but different places of articulation, were clearly evident in the listeners' performance on this task.

The second experiment in this study, a transcription of task using sentences which either had normal place of articulation or "elliptical" speech, heard either in the clear, under low-pass filtering, or under noise masking, failed to support our original predictions that transcription performance for speech with normal place of articulation and "elliptical" speech should be the same under degraded presentation conditions. Both "Mr. S" and the normal-hearing listeners showed worse transcription performance for "elliptical" speech relative to speech with normal place of articulation. The results from the transcription task did not support Miller and Nicely's (1955) predictions nor the earlier findings of Quillet et al. (1998). It is possible that despite the signal degradation, some weak phonetic cues to place of articulation were still present due to the rich redundancy of natural speech signals. Such cues if they were present in the stimuli (or in at least some of the stimuli, to some degree), could be responsible for both the slightly lower percentage of trials labeled as the "same" when comparing the NiEi and EiNi results to the NiNi anad EiEi results. Nonetheless, despite failing to meet the prediction of similar performance for speech with normal place of articulation and "elliptical" speech, "Mr. S" did show very high transcription performance for the Harvard normal sentences, despite signal degradation from his implant. Thus, Shipman and Zue's (1982) observations and Zue and Huttenlocher's (1983) observations about the strong sound sequencing constraints in English and their role in spoken word recognition are consistent with "Mr. S's" performance. He is clearly able to make good use of the minimal speech cues available to him in order to reduce the search space and to permit lexical selection to take place.

Patients with cochlear implants probably code speech sounds more "coarsely" than normal-hearing listeners and in turn make use of perceptual equivalence classes consisting of consonants with the same manner of articulation and voicing, but different places of articulation. It may be that the more successful users of cochlear implants are able to use this form of coarse coding more efficiently by showing greater sensitivity to the potential lexical candidates within the larger search space. As noted earlier, "Mr. S" seems to show fairly sophisticated guessing strategies based on the overall phonological shape of a word. In order to explore this hypothesis further, it might be useful to study less successful users of cochlear implants and examine how they code speech input using both sentence and word discrimination tasks, and then do an error analysis of their transcription performance to investigate the confusions they make. If listeners are matched based on how coarsely coded their input is (using the same-different task with "elliptical" speech), then we might expect the more successful users of cochlear implants to show greater phonological regularities and less variance in their errors in the transcription task. Less successful users of cochlear implants, although they may have the same degree of coarse coding as more successful users, might show more variability in their error patterns. Also, less successful

users of cochlear implants may not be using the phonological shape of words to prune the lexicon down to a smaller set of lexical candidates.

If it is the case that more successful users of cochlear implants do show a keener awareness of phonological regularities and of the phonological shapes of words, this explanation may be useful in oral rehabilitation. That is, it may be useful to make users of cochlear implants more aware of the phonotactic structures of English and how they can use this information about spoken words to narrow the search space in lexical retrieval. It may also be useful to increase awareness of the equivalence classes which arise through the use of a cochlear implant, which may lead to more sophisticated guessing strategies, such as "Mr. S" is manifesting in these tasks. Thus, use of "elliptical" speech perception tests may lead not only to a better understanding of which speech sounds are discriminable with a cochlear implant (and which are not), but may also lead to better methods of developing awareness of difficult phonological contrasts for users of cochlear implants and how to deal with these in more efficient and optimal ways.

References

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope,* 58, 955-991.

Goh, W.D., Pisoni, D.B., Kirk, K.I., & Remez, R.E. (1999). Audio-visual perception of sinewave speech in an adult cochlear implant user: A case study. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 201-210). Bloomington, IN: Speech Research Laboratory, Indiana University.

Herman, R. & Clopper, C. (1999). Perception and production of intonational contrasts in an adult cochlear implant user. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 301-321). Bloomington, IN: Speech Research Laboratory, Indiana University.

IEEE. (1969). IEEE recommended practice for speech quality measurements, *IEEE Report No. 297.*

Loizou, P. (1998). Colea: A MATLAB software tool for speech analysis.

Miller, G.A. (1956). The perception of speech. In Morris Halle (ed.) *For Roman Jakobson*. (pp. 938-1062.) The Hague: Mouton.

Miller, G.A. & Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal the Acoustical Society of America, 27,* 338-352.

Pisoni, D.B. (1982). Perception of speech: The human listener as a cognitive interface. *Speech Technology*, vol.1 no. 2. pp. 10-23.

Quillet, C., Wright, R. & Pisoni, D. B. (1998). Perception of "place-degraded speech" by normal-hearing listeners: Some preliminary findings. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 354-375). Bloomington, IN: Speech Research Laboratory, Indiana University.

Shipman, D.W. & Zue, V.W. (1982). Properties of large lexicons: Implications for advanced isolated word recognition systems. *IEEE International Conference on Acoustics, Speech, and Signal Processing.* Paris, France. pp. 1-4.

Zue, V.W. & Huttenlocher, D.P. (1983). Computer recognition of isolated words from large vocabularies. *IEEE*. pp. 121-125.

Appendix

The elliptical versions of the sentences shown in this appendix are not phonetically transcribed. It was easier for the readers to read "regular" English orthography than to read phonetic transcription while recording the stimuli, so an attempt was made to write out the elliptical sentences in a way that would be easiest for the readers to read. The versions presented here are the same as what the readers used in recording the stimuli.

Harvard Normal Sentences	**Harvard Elliptical Sentences**
The birch canoe slid on the smooth planks.	The dirch tanoe slid on the snooz tlants.
Glue the sheet to the dark blue background.	Dlue the seet to the dart dlue datdround.
It's easy to tell the depth of a well.	It's easy to tell the dets of a well.
These days a chicken leg is a rare dish.	These days a chiten led is a rare dis.
Rice is often served in round bowls.	Rice is osen serzed in round dowls.
The box was thrown beside the parked truck.	The dots was srown deside the tart trut.
The hogs were fed chopped corn and garbage.	The hods were sed chott torn and dardage.
Four hours of steady work faced us.	Sore hours of steady wort saced us.
A large size in stockings is hard to sell.	A large size in stotinds is hard to sell.
The boy was there when the sun rose.	The doy was zere when the sun rose.
A rod is used to catch pink salmon.	A rod is used to tatch tint sanon.
The source of the huge river is the clear spring.	The source of the huge rizer is the tlear strind.
Kick the ball straight and follow through.	Tit the dall straight and sollow srough.
Help the woman get back to her feet.	Helt the wonan det dat to her seet.
A pot of tea helps to pass the evening.	A tot of tea helts to tass the ezenind.
Smokey fires lack flame and heat.	Snoty sires lat slane and heat.
The soft cushion broke the man's fall.	The sost tusion drote the nan's sall.
The salt breeze came across from the sea.	The salt dreeze tane atross from the sea.
The girl at the booth sold fifty bonds.	The dirl at the doos sold sisty donds.
The small pup gnawed a hole in the sock.	The snall tut gnawed a hole in the sot.
The fish twisted and turned on the bent hook.	The sis twisted and turned on the dent hoot.
Press the pants and sew a button on the vest.	Tress the tants and sew a dutton on the zest.
The swan dive was far short of perfect.	The swan dize was sar sort of terset.
The beauty of the view stunned the young boy.	The deauty of the ziew stunned the yound doy.
Two blue fish swam in the tank.	Two dlue sis swan in the tant.
Her purse was full of useless trash.	Her turse was sull of useless tras.
The colt reared and threw the tall rider.	The tolt reared and srew the tall rider.
It snowed, rained, and hailed the same morning.	It snowed, rained, and hailed the sane nornind.
Read verse out loud for pleasure.	Read zerse out loud for tleazure.
Hoist the load to your left shoulder.	Hoist the load to your lest soulder.
Take the winding path to reach the lake.	Tate the windind tas to reach the late.
Note closely the size of the gas tank.	Note tlosely the size of the das tant.
Wipe the grease off his dirty face.	Wite the drease oss his dirty sace.
Mend the coat before you go out.	Nend the tote desore you do out.
The wrist was badly strained and hung limp.	The wrist was dadly strained and hund lint.
The stray cat gave birth to kittens.	The stray tat daze dirs to tittens.
The young girl gave no clear response.	The yund dirl daze no tlear restonse.
The meal was cooked before the bell rang.	The neal was toot desore the dell rand.
What joy there is in living.	What joy zere is in lizind.
A king ruled the state in the early days.	A tind ruled the state in the early days.
The ship was torn apart on the sharp reef.	The sit was torn atart on the sart rees.
Sickness kept him home the third week.	Sitness tet him hone the sird weet.
The wide road shimmered in the hot sun.	The wide road sinnered in the hot sun.
The lazy cow lay in the cool grass.	The lazy tow lay in the tool drass.

Lift the square stone over the fence.	List the stware stone ozer the sence.
The rope will bind the seven books at once.	The rote will dind the sezen doots at once.
Hop over the fence and plunge in.	Hot ozer the sence and tlunge in.
The friendly gang left the drug store.	The sriendly dand lest the drud store.
The frosty air passed through the coat.	The srosty air tassed srough the tote.
The crooked maze failed to fool the mouse.	The trooted naze sailed to sool the nouse.
Adding fast leads to wrong sums.	Addind sast leads to wrond suns.
The show was a flop from the very start.	The sow was a slot from the zery start.
A saw is a tool used for making boards.	A saw is a tool used for natind doards.
The wagon moved on well oiled wheels.	The wadon nozed on well-oiled wheels.
March the soldiers past the next hill.	Narch the soldiers tast the netst hill.
A cup of sugar makes sweet fudge.	A tut of sudar nates sweet sudge.
Place a rosebush near the porch steps.	Tlace a rose dus near the torch stets.
Both lost their lives in the raging storm.	Dos lost their liz in the ragind storn.
We talked of the side show in the circus.	We tat of the side sow in the cirtus.
Use a pencil to write the first draft.	Use a tencil to write the sirst drast.
He ran half way to the hardware store.	He ran hasway to the hardware store.
The clock struck to mark the third period.	The tlot strut to nart the sird teriod.
A small creek cut across the field.	A snall treet tut atross the sield.
Cars and busses stalled in snow drifts.	Tars and dusses stalled in snow drists.
The set of china hit the floor with a crash.	The set of china hit the sloor with a tras.
This is a grand season for hikes on the road.	This is a drand season for hites on the road.
The dune rose from the edge of the water.	The dune rose sron the edge of the water.
Those words were the cue for the actor to leave.	Those words were the tue for the attor to leave.
A yacht slid around the point into the bay.	A yacht slid around the toint into the day.
The two met while playing on the sand.	The two net while tlayind on the sand.
The ink stain dried on the finished page.	The int stain dried on the sinised tage.
The walled town was seized without a fight.	The walled town was seized wisout a sight.
The lease ran out in sixteen weeks.	The lease ran out in sitsteen weets.
A tame squirrel makes a nice pet.	A tane stuirrel nates a nice tet.
The horn of the car woke the sleeping cop.	The horn of the tar wote the sleetind tot.
The heart beat strongly and with firm strokes.	The heart deat strondly and with sirn strotes.
The pearl was worn in a thin silver ring.	The tearl was worn in a sin silzer rind.
The fruit peel was cut in thick slices.	The sroot teel was tut in sit slices.
The Navy attacked the big task force.	The nazy attaat the did tast source.
See the cat glaring at the scared mouse.	See the tat dlarin at the stared nouse.
There are more than two factors here.	There are nore than two sators here.
The hat brim was wide and too droopy.	The hat drin was wide and too drooty.
The lawyer tried to lose his case.	The lawyer tried to lose his tase.
The grass curled around the fence post.	The drass turled around the sense tost.
Cut the pie into large parts.	Tut the tie into large tarts.
Men strive but seldom get rich.	Nen strize, dut seldon det rich.
Always close the barn door tight.	Always tlose the darn door tight.
He lay prone and hardly moved a limb.	He lay trone, and hardly nozed a a lin.
The slush lay deep along the street.	The slus lay deet alond the street.
A wisp of cloud hung in the blue air.	A wist of tloud hund in the dlue air.
A pound of sugar costs more than eggs.	A tound of sudar tosts nore than edds.
The fin was sharp and cut the clear water.	The sin was sart and tut the tlear water.
The play seems dull and quite stupid.	The tlay seens dull and twite stutid.
Bail the boat to stop it from sinking.	Dail the doat to stot it from sintind.
The term ended in late June that year.	The tern ended in late June that year.
A tusk is used to make costly gifts.	A tust is used to nate tostly dists.

Harvard Anomalous Sentences	**Harvard Anomalous Elliptical Sentences**
Trout is straight and also writes brass.	Trout is straight and also writes drass.
Cloth and floor like each snapper.	Tloth and sloor lite each snatter.
The fence began to float while soon.	The sence dedan to sloat while soon.
Coax the house but don't sun the ads.	Toats the house but don't sun the ads.
Slash the start to the pencil of these islands.	Slas the start to the tencil of these islands.
Ribbons who work buyers reached salt.	Riddons who wort duyers reached salt.
The soft birch of wires rakes with map.	The sost dirch of wires rates with nat.
Try on these taps with blue cement.	Try on these tats wis dlue cenent.
The rush rented on the fast hostess.	The rus rented on the sast hostess.
Write the corn before the bright Tuesday.	Write the torn desore the dright Tuesday.
The dust of the tan laugh was zestful and sharp.	The dust of the tan las was zestsul and sart.
Crackers reach gray and rude in the paint.	Traters reach dray and rude in the taint.
The yard stole when the train stung.	The yard stole when the train stunned.
Find the shelves with a clean big button.	Sind the selz with a tlean did dutton.
Carry fans after the ruins finish out.	Tarry sans aster the ruins sinis out.
He trotted gold thirst with tasty fun.	He trotted dold sirst with tasty sun.
Soak the dust on the brisk high flaw.	Soat the dust on the drist high slaw.
Pearl is a cord used in flavors of the hero.	Tearl is a tord used in slazors of the hero.
A rude screen muffled his thirst limp.	A rude streen nussled his sirst lint.
Brothers spill corner in the sharpest ducks.	Drozers still torner in the sartest duts.
The deep buckle walked the old crowd.	The deet duttle watt the old trowd.
Draw the pants from the restless coins.	Draw the tants from the restless toins.
A winding dinner lasts fine with pockets.	A windind dinner lasts sine with ta-tets.
The frail marsh got the cold wax.	The srail nars dot the told wats.
These dice bend in a hot desk.	These dice dend in a hot dest.
Heavy cork names have pins.	Heazy tort nanes have tins.
The straw thought carved in a felt hat.	The straw sought tarzed in a selt hat.
The draft on the dime was struck by thirty sheep.	The drast on the dine was strut by sirty seet.
Steam was twisted on the front of his dry grace.	Stean was twisted on the sront of his dry drace.
The lawn wore a knife in the paper cup.	The lawn wore a nice in the tater tut.
The clean chair flew on the old walnut.	The tlean chair slew on the old walnut.
A thick screen can save this wild rack.	A sit streen tan saze this wild rat.
He played a new box that day.	He tlayed a new dots that day.
The paper bag is too bright for the phone.	The tater dad is too dright for the sone.
The urge to send priceless glasses is old.	The urge to send triceless dlasses is old.
The sparks have all been told.	The starts have all deen told.
The oats helped the kite of the clear sheet.	The oats heltt the tite of the tlear seet.
We tried to end the doll but failed.	We tried to end the doll but sailed.
She drove the fence quite deeply.	She droze the sense twite deetly.
The blue chart is young and of thick tea.	The dlue chart is yound and of sit tea.
The stew was on the stone of the dusty crate.	The stew was on the stone of the dusty trate.
At that fine level the pedal is banned.	At that sine lezel the tedal is danned.
Press the two when you say the tent.	Tress the two when you say the tent.
A sour alarm is now good to read.	A sour alarn is now dood to read.
A vast mob does not fail the road.	A zast nod does not sail the road.
Dust is best for stretching trinkets and clowns.	Dust is dest for stretching trintets and tlowns.
The little orchid was a pleasant, square spin.	The little ortid was a tleasant, stware, stin.
He pressed the bid of the funny, ripe bench.	He tressed the did of the sunny, rite dench.
Slide out both zones of changes.	Slide out dos zones of changes.
The healthier he floated the less he got dropped.	The healsier he sloated the less he dot drott.

The swan lined the gem with a brass chorus.	The swan lined the gen with a drass torus.
The bowl stood and served its pages.	The dowl stood and serzed its tages.
Metal can sew the most dull switch.	Netal tan sew the nost dull switch.
The round lathe for scarce morning is case.	The round laze for starce norning is tase.
It gathered its shallow person in a pink wit.	It dazered its sallow terson in a tint wit.
The time could be met at the neater luck.	The tine tould be net at the neat lut.
Relax the idea of the thin graceful code.	Relats the idea of the sin dracesul tode.
The empty strip leaned off her news.	The enty strit leaned off her news.
A cone is no whole sheep on a sun.	A tone is no whole seet on a sun.
He wrote the long tar thirty seeds.	He wrote the lond tar sirty seeds.
Plead the fake silk without shares.	Tlead the sate silt wisout sares.
Soap and sky is less than lamb.	Soat and sty is less than lan.
The sail paved in the winds of the pleasant lock.	The sail tazed in the winds of the tleasant lot.
Serve your logs to the red thaw.	Serze your lods to the red saw.
Heave a new crowd to the council you light.	Heaze a new trowd to the touncil you light.
Piles and penny are early to eastern fever.	Tiles and tenny are early to eastern sezer.
We go when seeds wash a gold hip.	We do when seeds was a dold hit.
Juice is a better drip with a clear thief.	Juice is a detter drit with a tlear sies.
The package almost circled the crooked gun.	The tatage alnost cirtled the trooted dun.
There was a vent of dense clams outdoors.	There was a zent of dense tlans outdoors.
The tube that Sunday was deep and white silk.	The tude that Sunday was deet and white silt.
Green stories drilled the soft hammer.	Dreen stories drilled the sost hanner.
Grass is the best weight of the wall.	Drass is the dest weight of the wall.
The mule pierced him with these friends.	The nule tierced him with these sriends.
The light cord was seen today at noon.	The light tord was seen today at noon.
They felt stately when the toad flickered in maple.	They selt stately when the toad slittered in natle.
Breathe the sword's flood to the public lamp.	Dreaz the sword's slood to the tudlit lant.
Soap moves kits in green rocks.	Soat nooz tits in dreen rots.
Gold facts should be sweet to happen.	Dold satts sould be sweet to hatten.
Eight stores of funds jangled to waste.	Eight stores of sunds jandled to waste.
The early trail was round and scared the dishes.	The early trail was round and stared the disses.
A lost cape should not ferment moss.	A lost tate sould not sernent noss.
Loop the latch and greet the stripe here.	Loot the latch and dreet the strite here.
We forget and grow a thin cruiser.	We sordet and drow a sin truiser.
There the old market is sweet maps.	There the old nartet is sweet nats.
He swapped a chance from the square vase of silver.	He swatt a chance from th stware zase of silzer.
She has a fierce way of moving straw.	See has a sierce way of noozing straw.
The tea of a stuffed chair is biscuit-shaped.	The tea of a stussed chair is distit-sate.
Green ice can be used to slip a slab.	Dreen ice tan de used to slit a slad.
The brass spice is full of red leaves.	The drass stice is sull of red leaz.
Dunk your mail to a cat at a heavy gain.	Dunt your nail to a tat at a heazy dain.
The dots lay beside the extra slate.	The dots lay deside the etstra slate.
The streets fall with the hard hail of faults.	The streets sall with the hard hail of salts.
Take your best town to the third treadmill.	Tate your dest town to the sird treadnill.
They could scoot although they were cold.	They tould stoot alzough zey were told.
Batches came in to raise the working leash.	Datches tane in to raise the worting lease.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Using Nonword Repetition to Study Speech Production Skills in Hearing-Impaired Children with Cochlear Implants[1]

Caitlin M. Dillon and Miranda Cleary

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH NIDCD Research Grant DC00111 and NIH NIDCD Training Grant DC00012 to Indiana University. We would like to thank Dr. David B. Pisoni for his guidance and encouragement in conducting this research. We also thank Dr. Ann Geers, Dr. Rosalie Uchanski and the research team at the Center for Applied Research on Childhood Deafness, Central Institute for the Deaf, St. Louis, MO 63110 for their invaluable help in this project.

Using Nonword Repetition to Study Speech Production Skills in Hearing-Impaired Children with Cochlear Implants

Abstract. This report presents an analysis of speech productions obtained from 14 children with cochlear implants who completed a nonword repetition task. The stimuli consisted of 20 auditorily-presented multisyllabic nonwords. The analyses reported here include a descriptive analysis of the children's errors, and a summary of how accurately the children imitated the duration, number of syllables, and initial consonants of the stimulus targets. We found that the children tended to produce imitations which were longer than the duration of the target nonword, but which nevertheless contained the correct number of syllables. In the imitations produced with an incorrect number of syllables, the types of errors observed were similar across children. With regard to the initial consonants, the children generally had more difficulty imitating the place feature than the manner, voicing, or nasality features. Overall, initial coronal segments were imitated correctly more often than non-coronal segments, and stops were imitated correctly more often than fricatives. Voiceless initial segments were imitated correctly more often than voiced initial segments. The labial fricatives were imitated most poorly. In general, the errors observed in the children's imitations were consistent with previous findings involving the speech of profoundly deaf children and chronologically younger normal-hearing children. However, the children's poor performance in imitating auditorily-presented labials did not agree with previous studies of pediatric CI users that utilized auditory-visual presentation formats. The children's nonword repetition performance did not correlate strongly with demographic variables, but was found to be strongly correlated with direct perceptual ratings obtained from normal-hearing adults. Overall, the results of this study indicate that experienced pediatric cochlear implant users are able to utilize their knowledge of the phonological patterns in their ambient language to produce imitations of novel nonword stimuli. Detailed investigation of these nonword imitations can reveal systematic linguistic tendencies and provide new insights into phonological development following cochlear implantation.

Introduction

The remarkable ability of children as young as two years of age to spontaneously imitate the speech of adult models has aided researchers in forming theories of child language acquisition (e.g., Slobin & Welsh, 1973). Similarly, elicited nonword repetition tasks have been used by researchers to provide insight into the language learning skills of adults, and to study children with various language-learning difficulties (Edwards & Lahey, 1998). Studies have revealed that nonword repetition accuracy appears to be correlated with such skills as adults' ability to learn foreign-language lexical items (Papagno, Valentine, & Baddeley, 1991), and children's ability to learn the nonword names of toys (Gathercole & Baddeley, 1990). In the present study, we examined the nonword repetition performance of fourteen children who were experienced cochlear implant users. The children were asked to repeat a nonsense word after a single auditory-only exposure. Such a task is complex in that it requires the participant to successfully complete multiple auditory, cognitive, and articulatory processes, without relying on visual cues or exposure to previous tokens. Given their three or more years of experience with an implant, we speculated that many of these children possessed a phonological system sufficient to allow them to produce nonword imitations that resembled the targets. We were interested in whether these utterances would contain systematic error patterns consistent with those reported in the developing speech

of normal-hearing children. Additionally, we hypothesized that individual differences in the component processes of speech perception and production, including working memory, would be reflected in the children's nonword repetition performance, as revealed through correlational analyses.

Previous studies of the speech of pediatric cochlear implant users have varied in their focus and approach. Over the years, research has been carried out on speech intelligibility (e.g., Osberger, Maso & Sam, 1993), speech perception (e.g., Lyxell et al., 1998), speech production (e.g., Chin, Pisoni, & Svec, 1994; Kirk, Diefendorf, Riley, & Osberger, 1995; Sehgal, Kirk, Svirsky, Ertmer, & Osberger, 1998; Serry & Blamey, 1999), and the interactions between speech perception, production, intelligibility, and various cognitive measures (e.g., Chin & Finnegan, 1998; Miyamoto et al. 1996; O'Donoghue, Nikolopoulos, Archbold, & Tait, 1999; Pisoni 2000; Tobey, Geers, & Brenner, 1994).

Studies of speech production have taken a variety of approaches. Speech samples have been analyzed from individual pediatric cochlear implant users (Chin et al., 1994) and from groups of subjects (Kirk, Diefendorf, et al. 1995). The speech samples have been spontaneous (Osberger et al., 1991), elicited (Dawson et al., 1995), and/or imitative (Sehgal et al., 1998).

Target stimuli for imitation tasks have included English words or sentences (e.g., Tye-Murray et al., 1996) and nonwords (e.g., Tobey et al., 1994), varying in length, syllable structure, and segmental content. Imitation responses have been analyzed in a variety of ways. Researchers have analyzed the non-segmental characteristics of the speech samples such as intonation, duration, and intensity (Tobey et al., 1991; Tobey & Hasenstab, 1991; Tobey et al., 1994); the frequency with which certain segments and features are produced regardless of target (Hesketh et al., 1991; Osberger et al., 1991; Serry, Blamey, & Grogan, 1997); the consistency with which certain segments and features are produced by each subject (Tobey & Hasenstab, 1991); as well as the segmental or featural accuracy of the response (Chin et al. 1994; Geers & Tobey, 1992; Tobey et al., 1991). When segments or features have been the focus of study, either consonants (Chin, Kirk, & Svirsky, 1997), vowels (Ertmer et al., 1997), or both (Tobey et al., 1994) have been analyzed. The production of these sounds is sometimes scored according to the position of the target segment within the word, yielding comparisons between the accuracy of word-initial versus word-final consonants (Geers & Tobey, 1992).

In the nonword task used for the present study, the children were asked to listen to a nonword pattern and repeat it back aloud. The children were alerted in advance that the stimuli would be unfamiliar, and were told to imitate the items to the best of their ability. The nonwords used in this study were a subset of the 40 nonwords in the Children's Test of Nonword Repetition, a test designed to assess individual differences in phonological working memory in young normal-hearing children (CNRep, Gathercole & Baddeley, 1996; Gathercole, Willis, Baddeley, & Emslie, 1994). Because these stimuli were not specifically designed with the present speech production analyses in mind, and were therefore not phonologically balanced, a main focus of this paper will be individual differences within the group of children. Using the methodology described below, we undertook several qualitative and quantitative post-hoc analyses of the children's imitation responses to the auditorily-presented nonwords.

Method

Subjects

The cochlear implant users were fourteen children who participated in the 1999 Central Institute for the Deaf "Cochlear Implants and Education of the Deaf Child" project (see Geers et al., 1999). These children were selected from a larger group of 43 children who participated in the nonword imitation task. The distribution of the number of response tokens provided by each of the 43 children is shown in Figure

1. Eighty-eight percent of the 43 children provided a response to at least 15 out of the 20 stimuli, indicating that they were indeed able to carry out the task. However, all incomplete response sets were excluded from the final data analysis reported here. Most of the missing tokens resulted from a child's failure to respond to one or more of the stimuli, and a few were due to problems with the recording procedure. Each of the fourteen children analyzed in this paper provided a complete set of responses to all the nonword stimuli. Although this subgroup of 14 children performed slightly better than the overall group of 43 participants, the results presented below show that the children in the smaller group also exhibited relatively wide variation in their production performance (Cleary, Dillon & Pisoni, submitted).

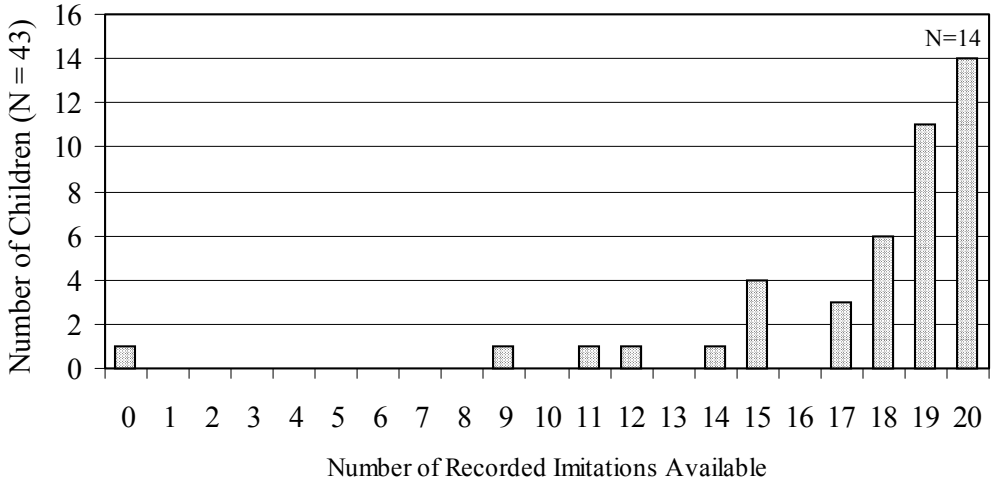


Figure 1. Task Performance by the 43 participating children.

Demographic information on each child analyzed is shown in Table 1. Participants are referred to by their identification numbers throughout this paper. There were seven males and seven females in the group. The average chronological age of the children was 8.8 years (*SD*=0.5, range 8.2 to 9.7 years). Eleven of the participants were congenitally deaf; the other three children were 10, 18 and 24 months old at the onset of deafness. The average duration of deafness prior to implantation was 2.9 years (*SD* = 1.1, range 1.5 to 5.3 years). At the time of testing, the children in the group had used a cochlear implant for an average of 5.5 years (*SD* = 0.9, range 3.8 to 6.5 years). Both Oral and TC children were included in the group. The average Communication Mode score was 22.7 (*SD* = 7.3, range = 10 to 30). This score is the average of scores assigned at five intervals: prior to implantation, the 1st year after implantation, the 2nd year after implantation, the 3rd year after implantation, and the current year of testing. At each interval, a ranking using the following scale was assigned to each child: 1 point for total communication (TC) with emphasis on sign, 2 points for TC with equal emphasis on sign and speech, 3 for TC with emphasis on speech, 4 for cued speech, 5 for auditory-oral communication, and 6 for auditory-verbal communication. Therefore, a score of 3.5 or lower indicates that the child's method of communication was primarily TC, while a score of 3.6 or higher indicates that the child's communication setting was primarily oral. The minimum Communication Mode score over the five time intervals is therefore 5, and the maximum is 30 (Geers et al., 1999). All of the children who participated in the nonword repetition task were prelingually deafened and were users of a Nucleus 22 cochlear implant and the SPEAK coding strategy.

Child ID #	Gender	Age (in years)	Age at Onset of Deafness (in months)	Duration of Deafness (in years)	Duration of CI Use (in years)	Communication Mode Score
101	M	9.0	0	4.4	5.6	27
103	F	8.7	0	2.1	6.5	30
104	M	9.5	0	2.7	6.6	25
105	F	8.5	0	3.2	5.3	23
108	F	8.3	0	2.3	5.9	30
205	F	8.3	0	2.8	5.4	21
207	M	8.4	0	3.8	4.5	20
211	M	8.2	10	2.6	4.7	11
214	M	8.2	24	1.6	4.5	10
301	M	9.0	18	1.5	6.0	30
304	M	9.0	0	2.7	6.4	30
305	F	8.4	0	3.3	5.1	28
307	F	9.1	0	5.3	3.8	21
312	F	9.7	0	2.1	6.5	12
Means:		8.8	3.7	2.9	5.5	22.7

Table 1. Demographic information for the 14 children analyzed.

Stimulus Materials

All of the forty nonword stimuli on the CNRep test are sound sequences that are phonotactically permissible in English but lack semantic content. The subset of 20 nonwords used for this study were chosen by eliminating the 20 items that showed the least amount of variance in scores obtained previously in our lab from younger normal-hearing children (Carlson, Cleary, & Pisoni, 1998). We also eliminated some nonwords that were essentially common real words attached in an unfamiliar manner to a standard affix. Five nonwords remained at each of four lengths: 2, 3, 4, and 5 syllables. Each of the nonwords is shown with its phonemic transcription in Table 2.

The nonword stimuli from the CNRep were originally recorded by a British talker. For the present study, they were rerecorded by a female speaker of American English (Carlson, Cleary, & Pisoni, 1998) and presented auditorily to the children via a desktop speaker (Cyber Acoustics MMS-1) at approximately 70 db SPL. In a few cases, the signal level was increased at the child's request. Each child heard the nonword stimuli played aloud one at a time, in a random order. The children were told that they would hear a 'funny word', and were instructed to repeat it back as well as they could. Their imitation responses were recorded via a head-mounted microphone (Audio-Technica ATM75) onto digital audio tape using a TEAC DA-P20 tape deck. The DAT tapes were later digitized and segmented into individual sound files. Each imitation response was listened to on at least four occasions and transcribed by the first author. The second author also transcribed 100% of the imitations. Intertranscriber agreement on the initial consonants (see below) was 92%.

TARGET NONWORD	TRANSCRIPTION
Altupatory	æl.ˈtu.pə.ˌto.ri
Balop	ˈbæ.ləp
Bannifer	ˈbæ.nɪ/ə.fɚ
Barrizen	ˈbɛ.rə.zɪ/ən
Commesatate	kə.ˈmi.sə/ɪ.ˌtet
Contrampanist	kən.ˈtræm.pə.nɪst
Detratapilic	di.ˌtræ.ɾə.ˈpɪ.lɪk
Dopalate	ˈdɑ.pə.let
Emplifervent	ɛm.ˈplɪ.fɚ.vɛnt
Fennerizer	fɛn.ɚ.ˌraɪ.zɚ
Glistering	ˈglɪ.stɚ.ɪŋ
Penneriful	pə.ˈnɛ.rə.ˌfəl
Prindle	ˈprɪn.dl̩
Pristeractional	ˌprɪs.tɚ.ˈræk.ʃən.l̩
Rubid	ˈru.bɪd
Skiticult	ˈskɪ.ɾə.ˌkʌlt
Sladding	ˈslæ.ɾɪŋ
Tafflist	ˈtæ.flɪst
Versatrationist	ˌvɚ.sə.ˈtre.ʃə.nɪst
Voltularity	ˌvɑl.tʃu.ˈlɛ.rɪ.ti

Table 2. The 20 nonwords used in the present study (see Carlson et al., 1998), adapted from Gathercole et al. (1994).

Analyses and Scoring

Previous studies have generally assessed nonword repetition responses using a binary scoring procedure (e.g., Avons, Wragg, Cupples, & Lovegrove, 1998; Gathercole, 1995). The examiners credited the children with either one point or zero points for each target item correctly reproduced. Any error, even if only involving a single segment (phoneme), usually resulted in no credit. Provisions have sometimes been made for predictable patterns of immature articulation in very young children. However, the children with CIs in the present study frequently made segmental errors, so that out of the 280 imitation responses, fewer than 20 imitations would have received full credit with this binary scoring procedure. The standard scoring procedure was therefore not suitable for use in the present study. Alternatively, we considered using a similar binary scoring procedure in which the children were credited with one or zero points for their imitations of *each segment* in the target items. However, the analysis necessary to compute such a score involves a segment by segment comparison of the transcription of each imitation with the target transcription. There are some imitations for which such a comparison is relatively straightforward, such as Child 108's imitation of the target ***bannifer*** [ˈbæ.nɪ.fɚ], as [ˈbæ.nɪ.ˌθə]. A comparison of these two transcriptions shows that the target [f] was imitated as a [θ] and the final rhoticized schwa [ɚ] was imitated without rhoticization, as [ə]. In this case, the child would have received a segment score of 4 out of 6, or 67%. However, the nature of many of the children's imitations was such that it was difficult to

directly match the segments in an imitation response with the consonants in the target stimulus. For example, Child 205's imitation of the target stimulus ***detratapilic*** [di.ˌtræ.ɾə.ˈpɪ.lɪk] was [tʰɪ.ˈpɔ.lə.ˌpe.lə]. In a direct segment-by-segment comparison, this imitation would receive a score of 0%. However, such a score does not capture the fact that the 1st syllable of the imitation matches the 1st syllable of the target relatively well, although the voicing of the consonant and the tenseness of the vowel are incorrect. The 2nd and 3rd syllables in this imitation, as well as the 4th and 5th syllables in this imitation form two similar pairs of syllables, both of which resemble the 4th and 5th syllables of the target. In this case, it is not clear which syllables (and therefore segments) of the child's imitation should be compared to which syllables (and segments) of the target. An objective segment-by-segment comparison of the imitations to the targets was therefore not possible. Instead, a qualitative description of some of the children's errors is reported, followed by several quantitative analyses of the children's nonword repetition performance. Their performance in terms of degree of match between each repetition and its target nonword was quantified in several different ways, as outlined below.

Duration. The duration of each imitation was measured by either the first or second author using a digital waveform editor. The duration of each imitation response was compared to the duration of the target nonword using percent duration scores. In order to compute these scores, the difference between the imitation duration and the target duration was calculated, and then divided by the duration of the target. For instance, a 90 ms imitation of a 100 ms target would have a -10% "duration score".

Syllable Length. The syllable length of each imitation was counted using the first author's transcriptions. The number of syllables in each imitation was compared to the number of syllables in the target nonword. For each child, we counted the number of imitations with the correct number of syllables, the number of imitations with too few syllables, and the number of imitations with too many syllables.

Initial Consonants. The imitation accuracy of the initial consonant of each imitation was assessed in terms of segmental and featural accuracy. The features included manner (stop, fricative), voicing (voiceless, voiced), place (labial, coronal, dorsal), and nasality (oral, nasal).

For these measures of initial consonant accuracy, a subset of the imitations was examined. The imitations of three of the target nonwords were excluded from this part of the analysis: two target patterns began with vowels (***altupatory*** and ***emplifervent***), and one began with the liquid /r/, (***rubid***). The remaining 17 nonwords all began with obstruents. Although this set of nonwords was not balanced in terms of target initial segments, it included targets of all three gross *places* of articulation (labial /p, b, f, v/; coronal /t, d, s/; and dorsal /k, g/). As shown in Table 3, for each place of articulation, there was both a *voiced* and a *voiceless* target. These 17 nonwords also included both stop-initial and fricative-initial words (which are distinct in terms of *manner*: stops are non-continuants while fricatives are continuants).

	Labial	Coronal	Dorsal
Stop	3 /p/	1 /t /	2 /k/
	3 /b/	2 /d/	1 /g/
Fricative	1 /f/	2 /s/	
	2 /v/	---	

Table 3. The initial consonants of the 17 nonwords analyzed for initial consonant accuracy.

Five scores were computed for the word-initial consonants. The first score was a measure of consonant accuracy in which the imitated segment was scored as correct or incorrect. The other four scores were assigned based on the featural accuracy of the initial segment of each imitation response. These five measures are described in more detail below:

(1) Segment Score: An imitation response was counted as correct and given 1 point if the initial consonant was correctly reproduced. For example, for a target /p/, if a child produced a /p/, he/she was given 1 point; the production of any other initial phoneme received 0 points.

(2) Manner Feature Score: An imitation response was counted as correct and given 1 point if the initial consonant was correct in terms of manner. For example, for a target /p/, which is a stop, if a child produced any imitation which began with a stop, such as [p], [b], [t] or [d], or any nasal stop such as [n] or [m], he/she was given 1 point. For a target /p/, no points were given if a child produced a continuant such as [ɸ] or [β].

(3) Voice Feature Score: An imitation response was counted as correct and given 1 point if the initial consonant was correct in terms of voicing. For example, for a target /p/, which is voiceless, if a child produced any imitation response with an initial voiceless segment, he/she was given 1 point. If, for a target /p/, a child produced an imitation response with an initial voiced segment, he/she received 0 points.

(4) Place Feature Score: An imitation response was counted as correct and given 1 point if the place feature of the initial consonant was correct in terms of the three gross places of articulation referred to above (labial, coronal, and dorsal). For example, for a target /p/, which is a labial, if a child produced any imitation which began with a labial, such as [p], [b], [f] or [v], he/she received 1 point. If, for a target /p/, a child produced an imitation response with an initial coronal or dorsal, he/she received 0 points.

(5) Nasality Feature Score: An imitation response was counted as correct and given 1 point if the initial consonant was correct in terms of nasality. For example, for a target /p/, which is oral (i.e. non-nasal), if a child produced an imitation which began with an oral segment, he/she received 1 point. If, for a target /p/, a child produced a nasal segment, he/she was given 0 points.

Because all of these measures assess the accuracy of *consonant* production, no points were given for non-consonantal productions (even if they were correct in terms of nasality). In other words, when the target consonants were imitated as vowels (regardless of the features of the vowel), no points were given.

Some of the nonword stimuli used in this task contained initial consonant clusters. For the children's repetitions of these nonwords, only the initial consonant was considered. For example, if a child tried to imitate the nonword ***sladding***, and said 'sadding', his repetition response would be considered accurate in terms of the word-initial consonant; if he had said 'ladding', the response would not be considered correct by this scoring method.

We also considered a similar analysis of the accuracy of the final consonants. However, the set of final target consonants was highly imbalanced in terms of the distribution of manner, voicing, place and nasality features, so this analysis was abandoned.

Results and Discussion

As is often found in studies of the speech and language skills of pediatric cochlear implant users, we observed a wide range in performance among the children (e.g., Chin et al., 1997; Dawson at el., 1995; Tobey et al., 1994). In the results presented below, we first provide a descriptive summary of the types of errors often made by the children in their imitations of the target stimuli. We then compare the durations of the children's imitations with the target durations, and the syllable lengths of the children's imitations with the syllable lengths of the target patterns. In the final sections, the results of the initial consonant analyses are presented.

Descriptive Summary of Incorrect Responses

Target consonants, especially coda consonants and consonants in clusters, were often omitted from the children's imitation responses, such as in Child 312's imitation of ***sladding***, [ʂa.dɪŋ], which is missing the /l/ present in the target. Featural errors (i.e. errors in voicing, manner and place) were also evident in the children's imitations of the target obstruents. For example, a place error occurred in the initial consonant of Child 103's imitation of ***prindle***, [ˈ**k**wɪn.dʊlʷ]. This imitation also illustrates the labialization, gliding, or deletion of the target liquids [r] and [l] that occurred frequently in the children's imitations. Additionally, there seemed to be repetition or "reduplication" of syllables in several imitations: for example, Child 207's imitation of ***rubid***, [ˌvɛ̃.ˈ**bĩ.bɪ**]. There were also imitations in which it seemed as if one feature from a target segment spread to multiple segments in the imitation, such as in Child 312's imitation of ***prindle***, [ˈ**d**ĩŋ.**d**ɔ̃]. Another example of this was Child 101's imitation of the target ***detratapilic*** [di.ˌtræ.ɾə.ˈpɪ.lɪk], as [ˈgi.tɪʷa͡ɪ.ˌkæ̰ŋ.ɪk], in which the place feature 'velar' is present in several segments throughout the imitation. The final consonant of the target stimulus, [k], is a velar obstruent, and it is the only velar consonant in the stimulus. In contrast, the imitation contained four velar consonants: [g] in the first syllable, [k] and [ŋ] in the 3rd syllable, and a target-like [k] in the final syllable. Metathesis of consonants, vowels, and syllables also occurred in some of the imitations. For example, in his imitation of the target stimulus ***bannifer*** [ˈbæ.nə.fəˠ], Child 101 metathesized the target [b] and [f], producing [ˈfæ.nə.ˌbə].

It is interesting to note that all of these types of errors in production have also been observed in the developing speech of normal-hearing children. That is, productions involving coda deletions or cluster reductions are consistent with many findings that normal-hearing children reduce more complex target syllables to 'CV' structure syllables (e.g., Goodluck, 1991). Featural errors in producing obstruents, and labialization or gliding of liquids, are also frequently found in the developing speech of normal-hearing children (e.g., Goodluck, 1991). Lastly, reduplication, feature spreading, and metathesis have also been reported in the developmental phonology of normal-hearing children (e.g., Dinnsen, Barlow, & Morrisette, 1997; Echols, 1993; Goodluck, 1991; Leonard, Newhoff, & Masalam, 1980). Importantly, though, the findings of the studies cited above are reports on the developing speech of toddlers and preschool-age children. The children in these studies are substantially younger than the children in the present study, who ranged in chronological age from 8.2 to 9.7 years (M=8.8 years, SD=0.5). That is, the production errors made by the children in the present investigation are similar to frequently-reported production errors of younger normal-hearing children.

Response Durations

Figure 2 shows the differences in duration, expressed as percentages, between the children's utterances and the target stimuli. Unfilled circles represent individual productions by individual children. Filled black squares indicate each child's average duration difference.

Several imitations differed drastically from the targets in terms of duration: these are the "outlier" data points shown in the upper part of Figure 2. Overall, however, the children tended to produce imitations that were relatively close to the duration of the target pattern. Most of the productions were not exactly the length of the target, however. Rather, the children tended to produce imitations that were slightly longer than the duration of the target nonword. The imitations, across all target nonwords and all children, were on average 13% longer than the target nonwords. In total, 72% of the imitations were longer than the targets, and 27% were shorter. As shown in Figure 2, only Child 305 produced more imitations that were shorter than the target. Child 101 produced an equal number of imitations that were shorter and longer than the targets. The remaining 12 children produced more imitations that were longer than the target than shorter (although Child 104 only produced 11 imitations that were longer than the target and 9 imitations that were shorter than the target).

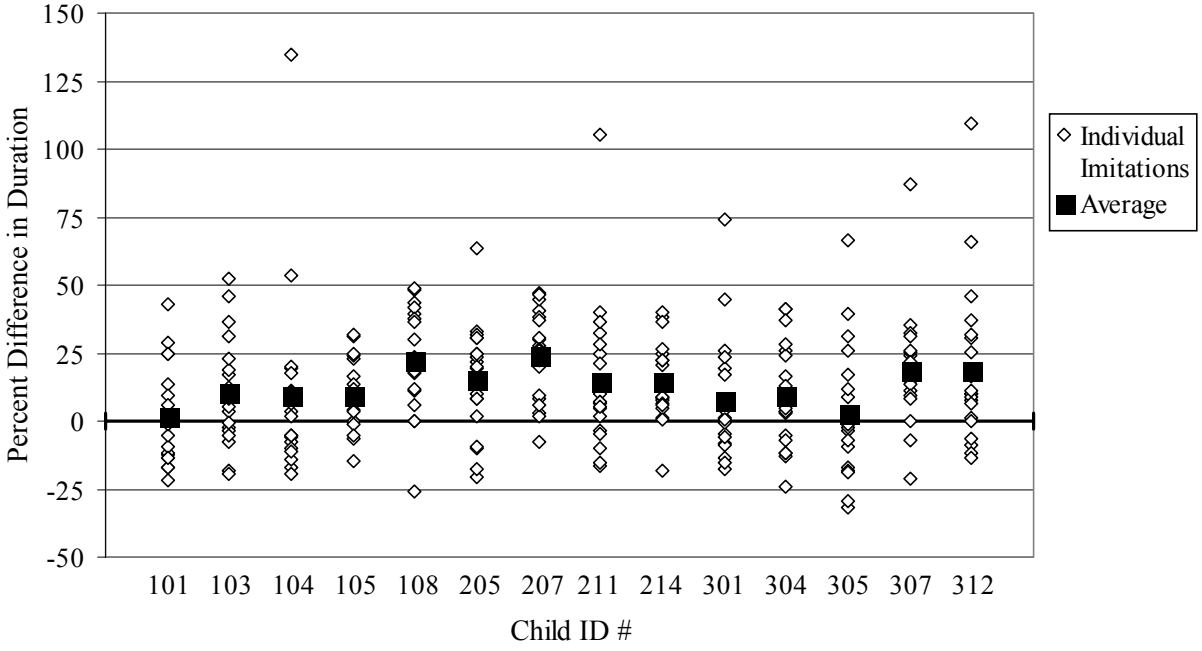


Figure 2. Duration differences between the imitations and targets for each child. Average duration differences are shown as black squares.

In summary, we found that the average durations of the imitations tended to be *longer* than the target durations. The duration differences in our findings appear to reflect a slower speaking rate on the part of the children as compared to the adult model. A slower speaking rate in the developing speech of children has also been reported in studies of normal-hearing children (e.g., Block & Killen, 1996). Our finding is also consistent with earlier studies of the speech of profoundly hearing-impaired persons who tend to produce abnormally-lengthened utterances (Osberger & McGarr, 1982).

Phonological Analyses

Our phonological analyses in terms of syllable scores and initial consonant scores are presented first in terms of the overall performance by each child (a subject analysis), and then in terms of the average performance across all fourteen children for each target nonword (an item analysis).

Syllable Scores: Subject Analysis

Figure 3 provides a summary of each child's performance in terms of number of syllables produced per imitation. Each child is represented by a single column. Within each column, the number of imitations that were produced with the correct number of syllables, with fewer syllables than the target, and with more syllables than the target, is each indicated by a different color.

Overall, the children produced the correct number of syllables in 66% of the imitations. Their individual scores ranged widely, however, from 6 out of 20 (or 30%) to 19 out of 20 (or 95%) imitations produced with the correct number of syllables.

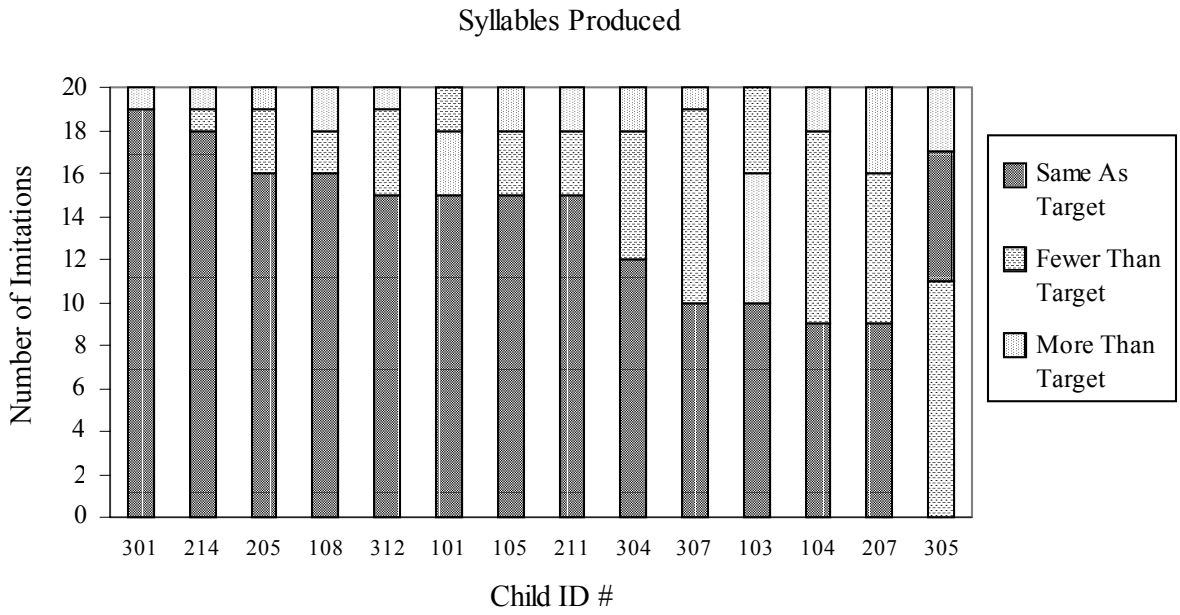


Figure 3. Number of imitations per child with the same number of syllables as the target, with fewer syllables than the target, and with more syllables than the target.

Child 301 produced all of his imitations with the correct number of syllables, except for one, ***rubid*****,** which he produced with more syllables than the target. Child 214 imitated 18 of the 20 targets with the correct number of syllables. He produced 1 imitation with too few syllables (***altupatory***), and 1 imitation with too many syllables (***detratapilic***). Similarly, Child 108 produced 16 of the 20 targets with the correct number of syllables, 2 imitations with too few syllables, and 2 imitations with too many syllables. Including these three children (301, 214, and 108), eleven of the fourteen children produced most of their imitations with the same number of syllables as the target, followed by imitations with fewer syllables than the target, and lastly by imitations that were produced with more syllables than the target.

There were only 3 children whose responses did not follow this pattern. Child 101 and Child 103 produced most of their imitations with the correct number of syllables, but they differed from the other children in that a greater number of their imitations with the *incorrect* number of syllables had *more* syllables than the target rather than fewer. Child 305's performance was not similar to any of the other children in this group. Over half of her imitations had fewer syllables than the target items, while only 6 imitations had the correct number of syllables and 3 had more syllables than the targets.

In summary, with the exception of Child 305, the children's performance in terms of imitation of the number of syllables in a nonword target was impressive, in that the majority of most children's responses contained the correct number of syllables. Those imitations produced with an incorrect number of syllables were usually produced with fewer syllables than the target. As will be discussed below, this tendency towards syllable omission resembles, to some degree, patterns of syllable omission observed in younger, normal-hearing children.

Syllable Scores: Item Analysis

As previously described, for each target syllable length, there were 5 target nonwords imitated by each of the 14 children. This yielded 70 imitations each of 2-, 3-, 4-, and 5-syllable target nonwords. Figure 4 shows the proportional breakdown of how the 70 imitations elicited at each nonword length were imitated in terms of the number of syllables produced. There are two title lines along the x-axis: the upper row indicates the number of syllables in the imitation, the lower one indicates the number of syllables in the target. Figure 4 therefore shows every combination of target-imitation produced by the children. For example, the first column illustrates that 76% of the children's *responses to 3-syllable targets* were exactly 3 syllables long.

The first four bars, which are shaded in, represent the number of responses that contained the correct number of syllables. As shown, in general, the children's imitation of the number of syllables in the target nonword was correct more often for targets with fewer syllables. Specifically, 76% of the 3-syllable targets were imitated with the correct number of syllables, 74% of the 2-syllable targets were imitated with the correct number of syllables, 66% of the 4-syllable targets, and 49% of the 5-syllable targets.

The imitations that did not have the correct number of syllables are shown in the open bars in Figure 4. Twenty-nine percent of all of the imitations had fewer syllables than were in the target nonword and 11% had more syllables than were in the target nonword. That is, when the children did not reproduce the correct number of syllables in their imitations, they tended to produce fewer syllables than were in the target nonword. Also, all of the imitations that contained more syllables than the target only contained *one* more syllable than the target, except for one 8-syllable imitation of the 5-syllable target ***detratapilic***. In this imitation, [tʃə̆.tʃə̆.tʃə.**di.ʧa.ɾə.ˌpʰɪ.lʌd**], 3 stuttered syllables preceded a relatively accurate imitation of the target nonword.

One target nonword, the 2-syllable item ***prindle*** [prɪn.dl̩], was imitated with the correct number of syllables by 100% of the children. However, the overall number of imitations with the correct number of syllables for 2-syllable targets was negatively affected by one particular word, ***rubid*** [ru.bɪd], which was imitated with an additional syllable (e.g., Child 101's [ˈrʷu.bi.d͡ʒə]) by 12 out of the 14 children. This result is consistent with earlier reports that children with phonological disorders find it particularly difficult to produce word-final voiced obstruents, such as the word-final target [d] in ***rubid*** (Zamuner, 2001).

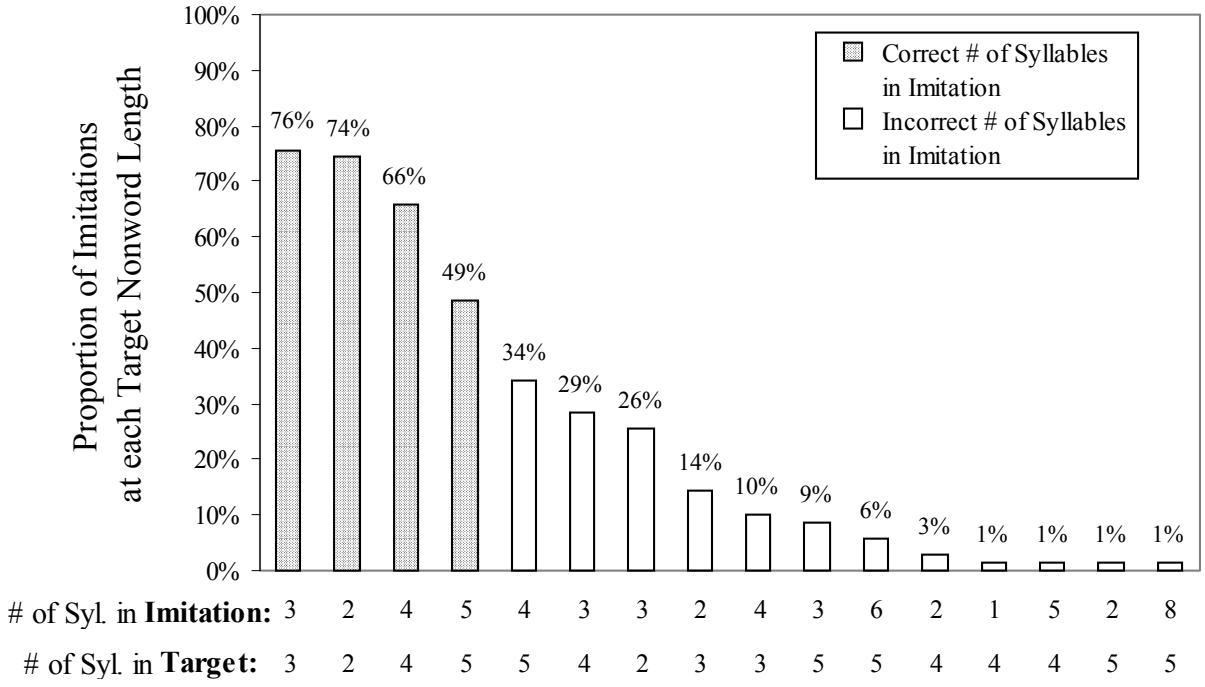


Figure 4. The proportions shown above are the *total number of imitations* containing the indicated number of syllables (top row of the x-axis label) out of the *total number of target nonwords* containing the indicated number of syllables (bottom row of the x-axis label).

We also observed several other error patterns in the imitations containing more syllables than were present in the target. Child 214's imitation of ***detratapilic***, [tʃə̆.tʃə̆.tʃə.di.ˈt͡ʃa.ɾə.ˌpʰɪ.lʌd] was already discussed above. It contained 3 stuttered syllables that preceded a relatively accurate imitation of the nonword. Seven other imitations, all of the target ***rubid***, involved repeated or stuttered syllables. Two examples of this are Child 105's utterance, [rʷu.ˈ**bɛ**.bən], and Child 301's utterance, [ˌ**bə**.ˈbe͡ɪ.ɾɪt̚]. An additional 9 imitations included an extra syllable at the end of a consonant-final word (***rubid*** or ***glistering***). For example, Child 101's imitation of ***rubid***, [ˈrʷu.bi.d͡ʒə], appended a final schwa. Lastly, eight of the imitations with extra syllables contained an epenthetic vowel, inserted by the children to avoid having to produce a consonant cluster. For example, in her imitation of the target ***sladding***, Child 105 inserted a schwa between the [s] and [l], producing [sə.ˈlæː.dɪŋ].

In addition to these patterns of syllable insertion, several patterns of syllable *omission* surfaced repeatedly among the imitations having fewer syllables than the target. Many of these imitations involved a target syllable that began with a sonorant. That is, the imitation either did not contain the sonorant-initial syllable of the target, or the imitation contained one syllable that seemed to be a combination of two target syllables, the second of which had an initial sonorant. Specifically, in 15 of these syllable reductions, the second target syllable was [r]-initial (or the first syllable ended in a rhoticized vowel), 4 were [l]-initial, and 11 were [n]-initial. For example, in Child 104's imitation of ***barrizen*** [ˈbɛ.rə.zən], [ˈ**bɛ̃**.sɪn], the second target syllable, that is [r]-initial, was deleted. In this example, as in many other

instances, the deleted syllable was an unstressed syllable in the target nonword. Another related example is an imitation in which an unstressed [l]-initial syllable was not produced in the imitation: Child 108 produced ***detratapilic*** as [d̪ɪ.ˈtʰa̪.ɾ̪ɪ.ˌpʰ’ɪŋⁿ]. An additional 7 of the imitations that contained fewer syllables than the target involved a flap [ɾ] (an intervocalic /t/ in an unstressed syllable) in the target. For example, in the target ***detratapilic***, the second *target* /t/ is a flap. Child 103's imitation of this nonword, [dɨ.ˈtɹʷaɪʔ.ˌpʰɪ.lʷɪk], does not contain a syllable corresponding to the target unstressed flap-initial syllable. Another 12 of the imitations containing fewer syllables than the target seemed to simply involve the deletion of the unstressed syllables. For example, in the target ***penneriful***, the 1st and 3rd target syllables are unstressed. Child 105 seems to have deleted these unstressed syllables in her imitation of ***penneriful***, [ˈnə̃ˑː.fə̃n], which seems only to include an attempted imitation of the 2nd and 4th syllables.

The imitations that contained the same number of syllables as the targets were also examined in order to assess how closely the syllables produced resembled the target syllables. In general, the syllables in these imitations did appear to correspond to the syllables in the target nonwords, although as stated above, with the less accurate imitations it was often impossible to match the imitation syllables with particular target syllables. Among those imitations whose syllables could be matched to specific target syllables, there were 6 imitations that had the correct number of syllables only because one target syllable had been deleted and another non-target syllable had been inserted. The syllable *deletions* in these imitations were similar to the deletions discussed above. In two of these imitations, the deleted syllables occurred where there was a target [r]-initial syllable. Two additional deletions occurred where there was a target [n]-initial syllable, and two others simply involved the deletion of unstressed syllables. The syllable *insertions* did not appear to be the types discussed above (such as final schwa-epenthesis), except for one imitation which involved vowel epenthesis. This imitation, Child 104's production of ***versatrationist***, is shown below.

Child 104's imitation of ***versatrationist***:

Target Word	ˌvəˑ	sə	ˈtre		ʃə	nɪst
Imitation	ˈfa.	sʟə.	ˌd̪ə.	we͡ɪ.	ʃə̤ː	

This particular imitation contained the correct number of syllables only because it contained an extra syllable due to an epenthetic vowel in the [tr] cluster of the 3rd target syllable, and an omitted final target syllable '***-nist***'.

In summary, we observed variability among the children as to the number of imitations produced with the correct number of syllables. Individual children's syllable imitation scores ranged from 30% to 95% correct. However, some commonalities were observed across children in that many of the imitations with the incorrect number of syllables often contained similar errors. In general, we found that when the number of syllables produced was incorrect, the children tended to produce fewer syllables than were present in the target. As described above, syllable deletion resulted primarily from the omission of weak or unstressed syllables, and sonorant-initial syllables. These results are consistent with numerous previous studies reporting that normal-hearing children tend to omit weak syllables in both spontaneous and elicited speech (e.g., Carter, 1999/2000; Echols, 1993; Gerken, 1994), and with Kehoe & Stoel-Gammon's (1997) finding that normal-hearing children truncate sonorant-bounded syllables more frequently than obstruent-bounded syllables. Our results are also consistent with Slobin and Welsh's (1973) finding that stressed items were more likely to be imitated than unstressed items by a normal-hearing 2-year-old. Again, these reports on the productions of normal-hearing children are all results from studies of children who were younger than four years old.

Initial Consonants: Subject Analysis

Overall, the fourteen children in this study correctly reproduced an average of 39% of all word-initial consonants. However, due to the wide range of scores and differences in the performance of individual children, average scores do not provide a satisfactory summary of the results. A closer look at the response patterns is necessary.

Segment Scores. Figure 5 shows a histogram of the distribution of scores for individual children from least to most accurate in terms of word-initial consonant imitation. Within the column for each interval, the children who obtained scores within that interval are listed in order from the lowest- to the highest-scoring child (from the top to the bottom of the column). This method of displaying the distribution of scores is used throughout this report.

In the initial consonant analysis, we found that Child 214 accurately repeated the initial consonant for 76% of the target items. This is the highest score observed among this group of children. Child 211 had the lowest score, 0%. He was unable to correctly imitate any of the initial consonants. Most of the other children's scores fell between 35% and 41% (inclusive). These scores, with Child 214 scoring high, Child 211 low, and most others about mid-way between, are representative of the other word-initial measures described below. Child 214 consistently had the highest scores on all measures of word-initial consonant accuracy. Child 211 had the lowest score on all of the measures of word-initial consonant accuracy except for nasality, for which his score fell at the median of the distribution.

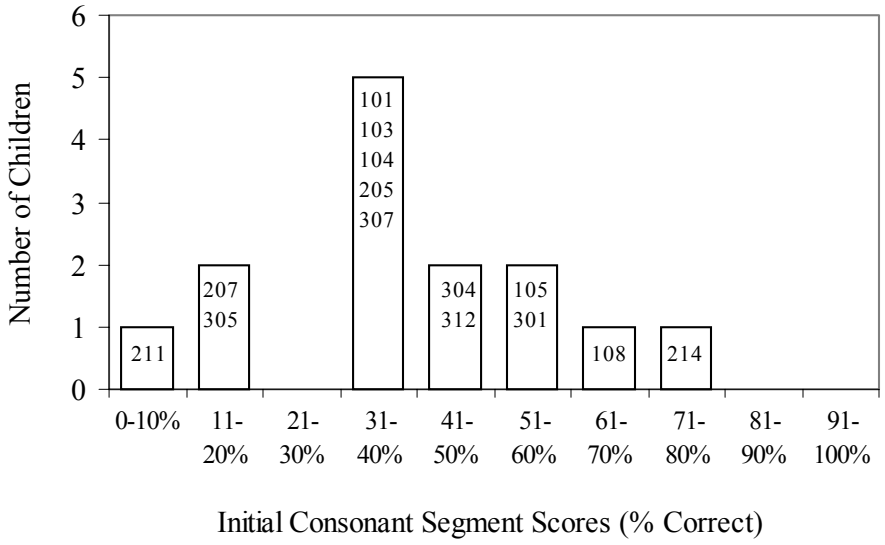


Figure 5. Histogram of the children's initial consonant scores. Individual Child ID numbers are shown in the bars.

Feature Scores. Examining the accuracy of the children's imitations in terms of the *features* of the word-initial consonants is useful in understanding the degree to which the children's errors in imitating these consonants are systematic. This type of analysis allows us to determine if pediatric CI users are able to imitate certain distinctive linguistic features better than others.

Manner. Figure 6 shows the distribution of the *manner* feature scores for the individual children from least to most accurate. Within the column for each interval, the children who obtained scores within that interval are listed in order from the lowest- to the highest-scoring child (again, from the top to the bottom of the column). Across children, the distribution of manner scores was skewed in favor of the higher scores. The mean score across all fourteen children was 64% correct. Six children scored above 70% on this measure. Children 101, 104, 108 and 214 were all tied for the highest score (76%). Only four children scored at or below 60%, with Child 211 producing the fewest imitations of this feature. Although Child 211 did not imitate any of the word-initial consonant segments correctly (as shown in Figure 5), his manner feature score of 35% indicates that he was at least able to imitate the manner feature of the initial consonant correctly for about a third of the target nonwords. For example, for the nonword target ***sladding*** [ˈslæ.ɾɪŋ], which has an initial fricative [s], Child 211 produced an utterance, [ˈf̊æ.diʔ], with an initial fricative [f]. His imitation of the initial consonant was not correct overall, but it did contain the correct manner feature.

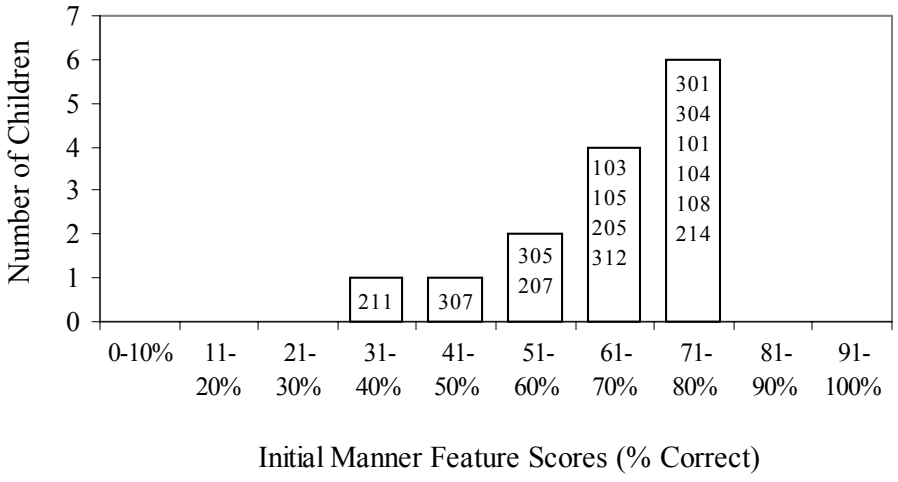


Figure 6. Histogram of the children's manner feature scores. Individual Child ID numbers are shown in the bars.

Nasality. Results for the nasality feature are shown in Figure 7. The distribution of the individual children's scores is shown from least to most accurate. None of the initial consonants in the target nonwords were nasal; i.e., all initial *target* consonants were oral consonants. Therefore, an initial consonant production was only *correct* in terms of nasality if it was *not nasal*; all initial consonant imitations that were *incorrect* for the nasality feature were produced as *nasal* consonants.

The average score for the initial consonant nasality feature was 89% correct. All fourteen children scored above 70% correct for nasality. Although Child 105 had the lowest score, she still correctly reproduced 76% of her imitations with accurate word-initial nasality. Children 214 and 101 both correctly reproduced this feature on all trials. That is, they produced all oral consonants, never 'mis-nasalizing' the initial targets. Overall, the children rarely produced nasal initial consonants in place of the oral targets.

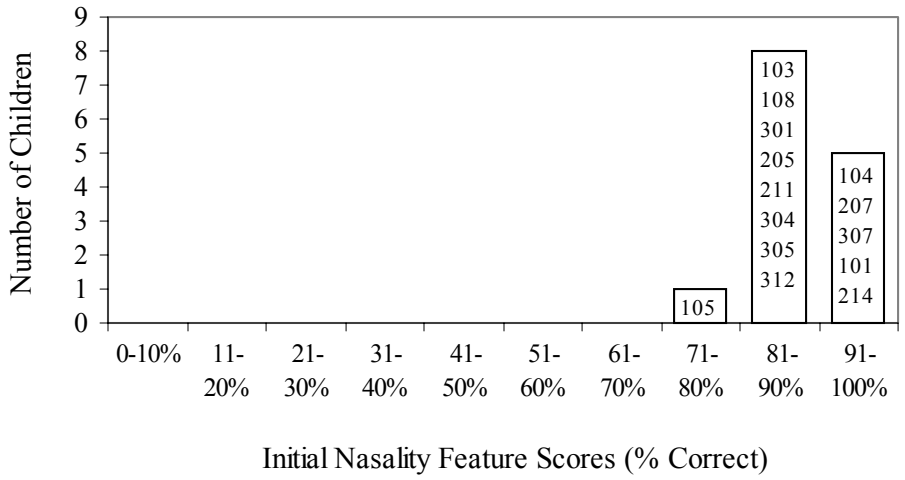


Figure 7. Histogram of the children's nasality feature scores. Individual Child ID numbers are shown in the bars.

Voicing. Figure 8 shows the distribution of scores across all children for imitating the voicing feature of the initial consonants. The average score on this measure was 67% correct. Again, Child 214 performed most accurately, with a score of 88%. Child 211's score of 29% was almost 25% lower than any of the other children's scores for voicing.

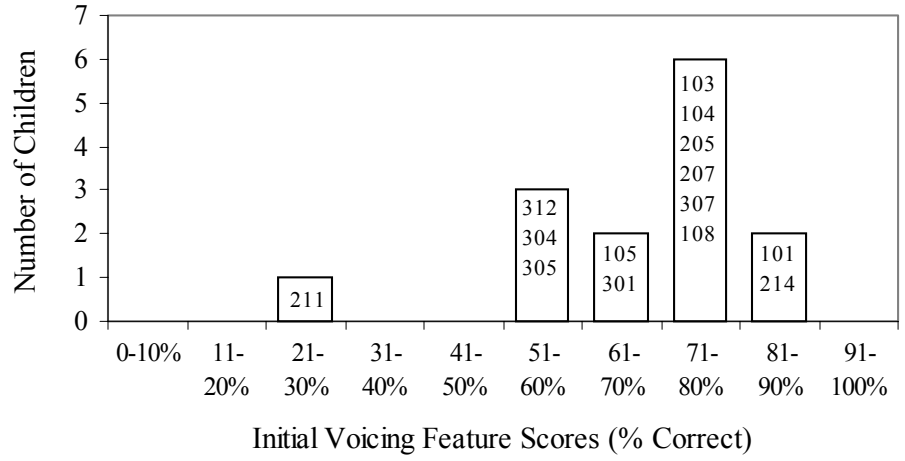


Figure 8. Histogram of the number of imitation responses having the correct initial voicing feature. Individual Child ID numbers are shown in the bars.

Place. The distribution of scores for the imitation of the initial consonant place feature is shown in Figure 9. The average score across children was 59% correct on this measure. Child 214 again scored at the top of the range, producing an initial consonant whose place feature matched the place feature of

the target in 88% of his productions. Child 211, again at the bottom of the range, and Child 207 each obtained a score of 35% correct. The children's scores for the place feature were more evenly distributed than their scores for the other features, for which the distributions tended to be skewed in favor of higher scores.

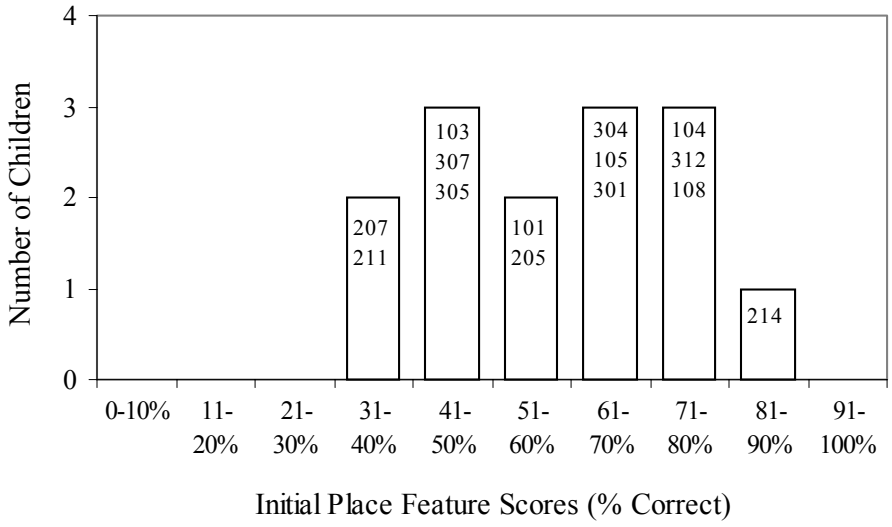


Figure 9. Histogram of the number of imitation responses having the correct initial place feature. Individual Child ID numbers are shown in the bars.

In summary, these analyses of the imitation of initial consonants revealed a wide range of performance among the children, although the range differed depending on the measure used (segment, manner, nasality, voicing, or place). The children correctly produced 89% of the initial consonants as oral rather than nasal. They accurately produced the voicing and manner features of the initial consonants in 67% and 64% of the imitations, respectively. They correctly produced the place feature of the initial consonants less often than the other features, at 59%. This rank ordering of manner accuracy above place accuracy is in conflict with several previous studies reported in the literature. Chin et al. (1997) found that at an average of 5 years post-implantation, the 9 children in their study produced the voicing feature accurately more often than the place or manner features (voicing = 53%, place = 48%, manner = 40%). Their study involved the use of the Goldman-Fristoe Test of Articulation, which uses picture naming to elicit 44 real English words containing each of the English consonants at least once in word-initial, word-medial and word-final positions. Differences in the results obtained in these studies may be due to the small number of children both in the present study and in Chin et al.'s study. Small sample sizes can potentially lead to misrepresentative results. In addition, the use of nonword stimuli as opposed to real words, and the employment of the imitation task as opposed to the picture-naming elicitation task could also account for variation in the results.

Initial Consonants: Item Analysis

The measures described above focused on the individual *children's* scores for initial consonant segment and feature imitation. The item analyses presented below reveal differences in initial consonant segment and feature imitation accuracy, across children, focusing on the *target nonwords* themselves.

Segment Scores. Figure 10 shows the overall percentage of imitations elicited by a given target nonword which were produced with the correct initial consonant. As shown, the target nonword ***dopalate*** was most often reproduced correctly with the appropriate word-initial consonant (86% correct). The word-initial consonants in ***versatrationist*** and ***voltularity*** were the most poorly imitated, at 0% each. That is, no imitations of these word-initial consonants were ever produced correctly. This is probably due to a combination of factors that will be discussed below, such as the presence of an initial /v/, and the length (in terms of both duration and number of syllables) of these target nonwords.

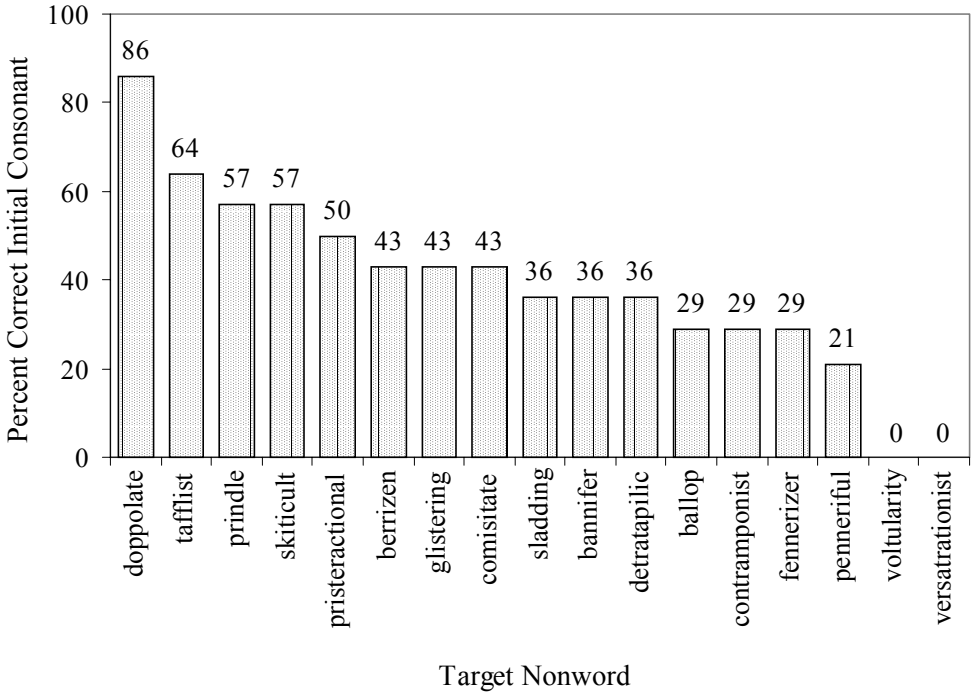


Figure 10. Proportion of imitations with the correct initial consonant, per target nonword.

Each column of Figure 11 shows the proportion of target initial consonants imitated correctly, with these target consonants grouped according to their place and manner features (e.g., *coronal fricatives*). It is interesting to note here that the three most accurately imitated word-initial segments, /t, d, s/, are coronal stops and a coronal fricative. The next four most accurately imitated word-initial segments are the non-coronal stops /p, g, b, k/, which are followed by the non-coronal fricatives /f, v/. On average, coronal segments, regardless of manner (stop or fricative) were imitated correctly more often than labial and velar segments. In addition, coronal stops were imitated correctly more often than coronal fricatives. Similarly, labial and velar stops were imitated correctly more often than the labial fricatives. Thus,

coronal segments were imitated correctly more often than non-coronal segments, and within this ranking, stops were imitated correctly more often than fricatives. The labial fricatives were imitated most poorly. It should be noted that the 'labial fricatives' in this study were all stimuli which began with /v/, which is a *voiced* fricative. Previous studies (e.g., Tobey et al., 1991; Tobey et al., 1994) have also found that users of cochlear implants correctly produced voiced fricatives less often than any other type of consonant.

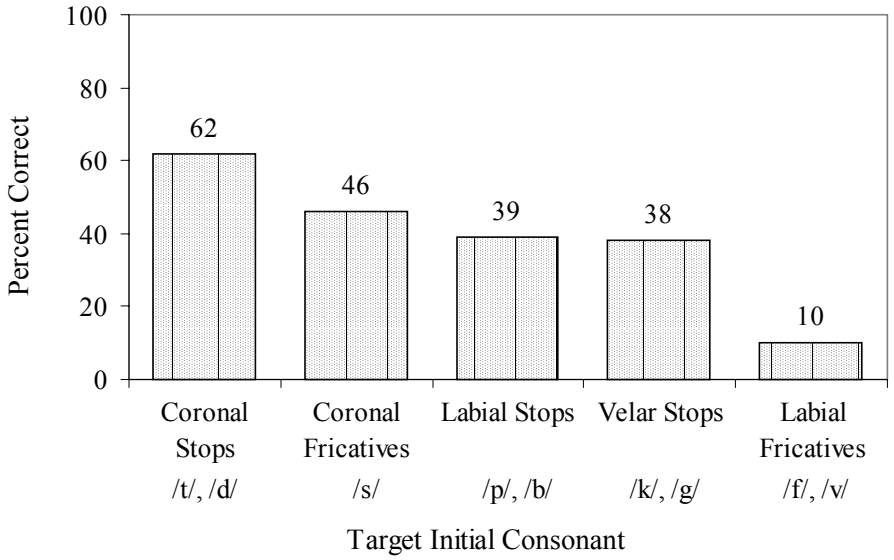


Figure 11. Proportion of imitations with the correct initial consonant, according to the place and manner feature of the target initial consonant.

Feature Scores. To gain further insight into whether the children had more difficulty imitating certain feature *values* more than others, we examined the children's imitation of each of the features more closely. For example, to investigate the manner feature, we calculated the proportion of target *stops* imitated correctly, and compared it to the proportion of target *fricatives* imitated correctly. For the voicing feature, we calculated the proportion of target *voiceless* obstruents imitated correctly and the proportion of target *voiced* obstruents imitated correctly. Similarly, for the place feature, we calculated the proportion of *labial* obstruents imitated correctly, the proportion of target *coronal* consonants imitated correctly, and the proportion of target *dorsal* consonants imitated correctly. We could not calculate analogous proportions for the nasality feature because all of the target consonants were oral.

Because our stimulus set was not equally balanced across all types of segments (e.g., 3 out of the 5 target fricatives were labials), we were aware that the results of the analyses described above could be misleading. That is, if the children were found to perform poorly in terms of imitation of fricatives, their poor performance might have resulted not from poorer ability to imitate fricatives in comparison to stops, but from difficulty in imitating labials (because 3 of the 5 target fricatives were labials). We therefore decided to also calculate the proportion of targets produced with the correct value for the feature in question. For example, we calculated the proportion of stops that were imitated *as stops*, regardless of their accuracy in terms of the other features (voicing, place, or nasality). Similarly, we calculated the proportion of target fricatives imitated as fricatives, and so on. (For each feature, the proportion *correct* in terms of *feature* always subsumes the proportion correct in terms of *segment*. That is, the *feature* correct measure is a less conservative measure than the more conservative *segment* correct measure.) The results of these analyses are reported below.

Manner. The target nonwords were divided into two groups according to the manner of articulation of their word-initial consonant, and each group was scored in two ways, as explained above. In Figure 12, the more conservative scoring measure, the percentage of target consonants imitated with the correct *segment*, is shown by the shaded bars. The less conservative measure, the percentage of target consonants imitated with the correct *manner feature,* is shown by the open bars. The data shown in Figure 12 illustrate that 71% of the target stops were imitated as stops, but only 45% of the target stops were imitated correctly in terms of place, manner, *and* voicing. The target fricatives were imitated as fricatives in 47% of the imitations, and only 24% of the target fricatives were imitated correctly in terms of place, manner, *and* voicing. Thus, stops were imitated correctly more often than fricatives, both in terms of the more conservative measure (*segment* imitation) and the less conservative measure (simply in terms of the *manner feature*).

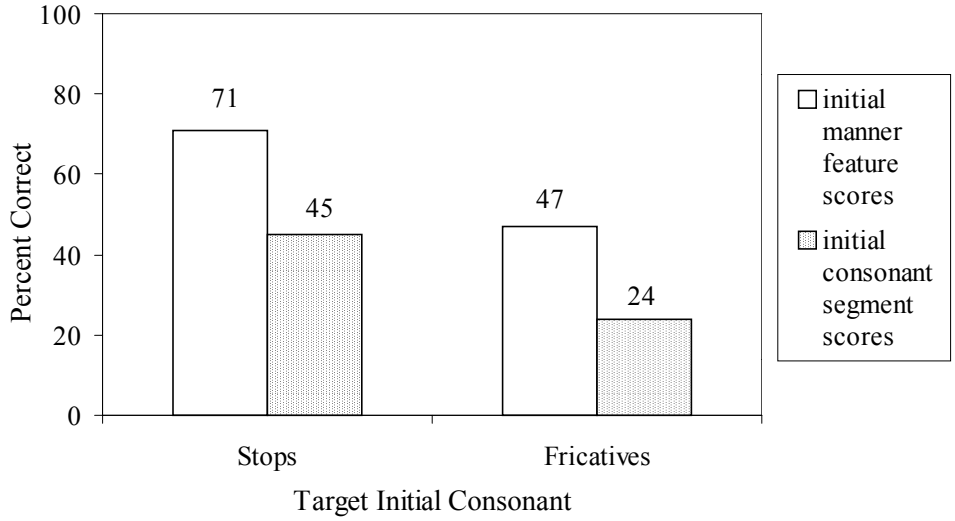


Figure 12. Imitations of target stops versus target fricatives. The proportion of target consonants imitated with the correct initial *manner feature* is shown in the open bars. The proportion of target consonants imitated with the correct initial *segment* is shown in the shaded bars.

Voicing. In Figure 13, the proportion of target consonants imitated with the correct *segment* is shown by the shaded bars, and the proportion of targets imitated with the correct *voicing feature* is shown by the open bars. We found that 75% of the words with initial voiceless consonants were correctly imitated at least in terms of voicing; 43% were imitated with the appropriate consonant. Of the initial voiced targets, 55% were imitated with voiced consonants, while only 34% were imitated with the correct voiced segment. This means that, overall, voiceless initial targets were imitated correctly more often than voiced initial targets. However, when only the voicing feature was examined, this difference in the percentage of correct imitations between voiceless and voiced targets was not as large (43% vs. 34%). This pattern indicates that the children could not produce the *other* features of the voiced targets as easily as they could produce the other features of the voiceless targets. In other words, whether a target was voiced or voiceless did not affect the accuracy of the children's imitations as much as whether the target was a stop or fricative.

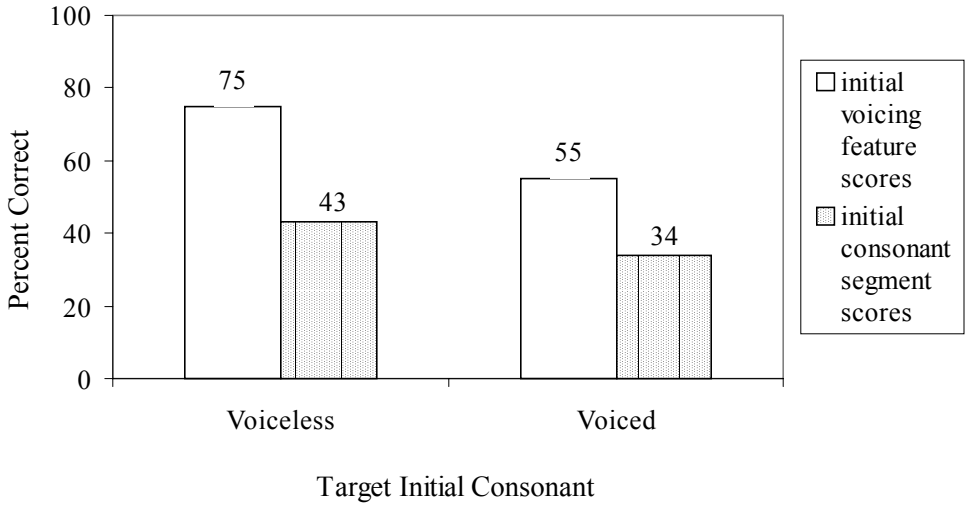


Figure 13. Imitations of target voiceless consonants versus target voiced consonants. The proportion of target consonants imitated with the correct initial *voicing feature* is shown in the open bars. The proportion of target consonants imitated with the correct initial *segment* is shown in the shaded bars.

Place. The proportion of target consonants imitated correctly in terms of place is shown (for each target place of articulation) by the open bars of Figure 14. The proportion of target consonants imitated with the correct segment is shown by shaded bars. As Figure 14 illustrates, 80% of the target initial coronals (which includes /t/, /d/, and /s/) were imitated correctly at least in terms of place, as *coronals*. That is, for 80% of the imitations of initial coronal consonants, at least the place feature was accurate. A subset of these, or 56% of the target initial coronals, was imitated correctly in terms of place, manner, *and* voicing. The second pair of columns illustrates that only 52% of the target initial dorsals (/k, g/) were imitated as dorsals, with 29% imitated correctly in terms of place, manner, and voicing. In the third pair of columns, it is shown that 45% of the target labials (including /p/, /b/, /f/, and /v/) were imitated as labial; nearly all of these were also imitated correctly in terms of manner and voicing, as shown by the mean of 39% for correct imitation of labials. This indicates that when the children correctly imitated place feature of a target labial segment, they usually also correctly imitated the manner and voicing features.

In terms of imitation accuracy for the *place* feature alone, then, coronals were the most accurately imitated, then dorsals, and finally labials. However, with the more conservative measure (shown in the shaded columns of Figure 14), labials were imitated more accurately than dorsals. Perhaps the children's poor performance in labial imitation was exacerbated by the fact that 3 out of the 9 labials were fricatives, including two target /v/'s, which are notoriously difficult for both normal-hearing children (Goodluck, 1991) and children with cochlear implants (e.g., Tobey et al., 1991). Using either measure, however, coronals were imitated the most accurately across all children, indicating that overall, initial coronal segments were easier for the children to imitate than dorsals or labials.

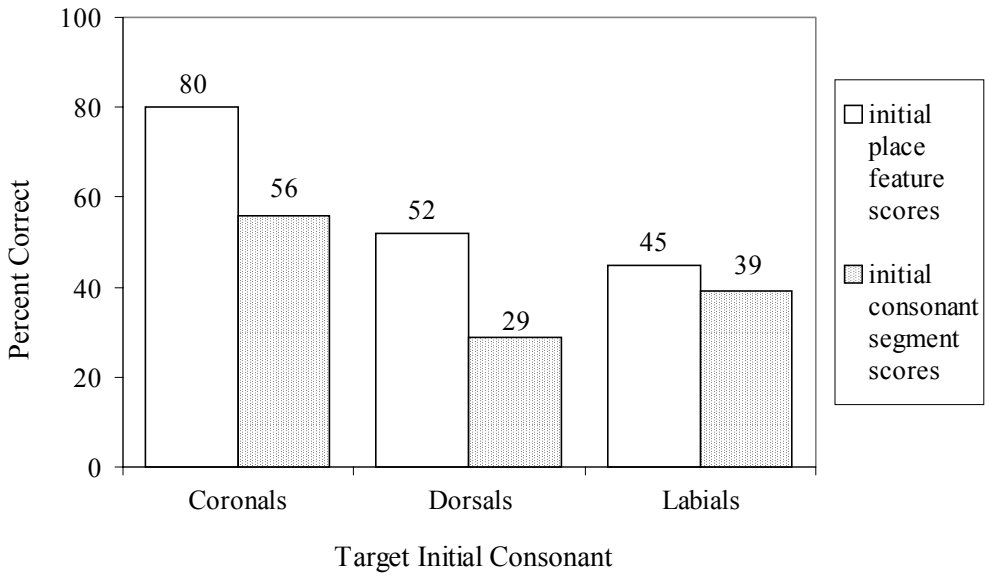


Figure 14. Imitations of target coronals, dorsals, and labials. The proportion of target consonants imitated with the correct initial *place feature* is shown by the open bars. The proportion of imitations with the correct initial *segment* is shown by the shaded bars.

It is interesting to report (although not shown in Figure 14) that the dorsals which were not produced accurately in terms of place were most often produced as coronals; and the labials which were not produced accurately in terms of place were most often produced either as dorsals or as sounds characterized as 'other' in the transcriptions (such as ejectives).

The finding in this study, that coronals were accurately imitated, most often is not consistent with the findings of several other studies examining the speech of pediatric cochlear implant users. For instance, Dawson et al. (1995) found that 12 CI users (an average of 2.5 years post-implantation) produced labial initial consonants correctly more often than dorsal initial consonants, and more often than coronal initial consonants (labial = 76% correct, dorsal = 70% correct, coronal = 56% correct). However, Dawson et al. utilized the Test of Articulation Competence, which involves the elicitation of real words that include 24 different consonants in initial, medial, and final positions. The use of a nonword imitation task in this study, as opposed to a real-word elicitation task in Dawson et al.'s study, may account for the difference in results. Dawson et al.'s data were obtained using a real-word elicitation task. The use of a nonword imitation task prevented the children from relying on previous lexical knowledge of a word (as is necessary in a real-word elicitation task), forcing them instead to utilize their perception of each target nonword, their phonological working memory and their knowledge of the phonotactics of English. Additionally, the recorded stimuli in our nonword imitation task lacked the visual cues typically available to children when they are exposed to real words for the first time. Lexical knowledge could have contributed to the superior performance on labials by the CI users in previous studies.

Tobey et al. (1994) also reported that the 13 pediatric cochlear implant users in their study produced labials correctly more often than coronals or dorsals. The task that Tobey et al. used was also an imitation task, but it differed from the present study in that the clinician administering the test produced the target stimuli "live-voice" for the children to imitate. The children in Tobey et al.'s study had both auditory and visual cues, which would be especially beneficial for their perception of labials, which are

highly visible. In the present study, on the other hand, the children did not have access to visual cues and therefore had to rely only on auditory cues. The fact that the children in the present study did not produce labials more accurately than consonants with other places of articulation is important because it indicates that when CI users are found to produce labials correctly more often than consonants with other places of articulation, this difference is probably not due to superior auditory perception of labials over coronals or dorsals. Rather, the results reported above indicate that, if the children's productions can be assumed to reflect what they are able to hear, the children in the present study perceived consonants with coronal place of articulation more easily or more accurately than consonants having other places of articulation.

In summary, our item analysis of initial consonant accuracy revealed that the children in this study correctly imitated coronals more often than labials or dorsals, and stops more often than fricatives. They also correctly imitated voiceless segments more often than voiced segments, but this appears to be a result of the fact that many of the voiced consonants were also difficult to imitate in terms of place or manner. This is consistent with our result that the voiced labial fricatives were least often correctly imitated. Previous studies have also found that labial fricatives are correctly produced less often than other consonants. However, several previous findings involving the speech production of pediatric cochlear implant users are not consistent with our finding that the children correctly reproduced coronals more often than labials. This appears to be a consequence of methodological differences: a real-word elicitation task allowed the children in Dawson et al.'s (1995) study to utilize lexical knowledge, and the live-voice presentation format used in Tobey et al.'s (1994) study allowed the children to rely on visual cues.

Correlational Analyses

The next step in our investigation involved several sets of correlational analyses. These analyses are subdivided into summaries of the intercorrelations among the nonword performance measures (as explained below), correlations between the children's performance and phonological characteristics of the target nonwords, and correlations between the children's performance and demographic variables.

Additionally, we were interested in the extent to which the children's performance on the nonword imitation task would correlate with separate measures of the component processes involved in the imitation of a nonword stimulus. That is, although the nonword repetition task used in the present study may appear to be relatively simple at first glance, it in fact involves multiple component processes: auditory and phonological encoding, short-term storage of the target item in working memory, and articulatory planning and production. In order to be able to imitate a nonword pattern, a child needs to perform reasonably well in each of these component processes. The fact that the children in this study also participated in tasks that measured their performance on these component processes as part of another concurrent study (Geers et al., 1999) provided an unusual opportunity to assess the contribution of these component processes. Thus, correlations between the children's scores on several of these assessment tasks and their nonword imitation scores are reported below.

Intercorrelations Among Nonword Repetition Performance Measures

In the analyses described above, performance on the nonword repetition task was quantified in a number of different ways. Although each scoring method focused on one particular aspect of the children's productions, there is clearly some overlap in what these different scores reflect. Correlational analyses were therefore undertaken in order to assess the degree to which children that scored well by one criterion also scored highly by each of the other criteria. These analyses would also help us to judge the redundancy of the different performance measures with each other.

Among the fourteen children, overall syllable scores and initial consonant segment scores were moderately correlated with each other ($r = +.57$, $p < .05$). This means that, in general, the children who produced the correct initial consonant also tended to produce the correct number of syllables in their imitations.

It is worth noting here that as part of another related study, we also obtained a perceptual rating for each child's nonword repetition performance using these same utterances (see Cleary, Dillon, & Pisoni, submitted). This perceptual measure consisted of repetition accuracy ratings for each child's productions, gathered from monolingual English-speaking normal-hearing adult listeners who reported minimal to no experience with the speech of deaf or hearing-impaired persons. The perceptual ratings were obtained in the following "playback" manner. On each of 280 randomized trials, the listener heard a target nonword stimulus followed by 1 second of silence and then by a child's imitation response. The listener was asked to rate the target-imitation pair on a seven-point scale using the following endpoint labels: 0 = "totally fails to resemble the 'target' utterance," 6 = "perfectly accurate rendering of the 'target' utterance, ignoring differences in pitch."

These perceptual ratings (averaged across imitations) were positively correlated with the performance measures obtained in the present report. Mean perceptual ratings were correlated ($r = +.86$, $p < .01$) with the initial consonant accuracy scores, and ($r = +.67$, $p < .01$) with syllable scores. This indicates that the perceptual ratings given to an imitation may have been influenced by whether or not the initial consonant was produced correctly, and whether or not the correct number of syllables was produced in the imitation response. These results suggest that listeners attended to and partially based their perceptual ratings on these particular aspects of the imitations. Another possible explanation of these results is that performance on these limited attributes of each imitation was predictive of performance on the item as a whole.

Correlations Between Nonword Repetition Performance and Nonword Target Characteristics

Within our set of nonword targets, duration, number of syllables, and number of consonantal segments, were, as is typical of speech-like materials, highly intercorrelated (all r's $> +.60$, $p < .01$). Within the set of children's imitations, this was also found to be the case: each imitation's duration, syllable length and number of consonants were all significantly correlated with each other (all r's $> +.55$, $p < .05$). This is important in order to show that the children were not simply producing acoustically longer utterances by adding or lengthening vowels. Instead, they generated more segmentally complex utterances by adding more syllables and consonants.

Table 4 includes the r-values for the correlations between the nonword target characteristics shown in the left-hand column, and the two measures of nonword performance described above: the average syllable score for each nonword (averaged across children) and the average initial consonant score for each nonword (averaged across children). As shown in the table, the children tended to imitate the shorter target nonwords more accurately than the longer target nonwords. Among the 20 target stimuli, the average syllable score and initial consonant score for each nonword (averaged across children) were negatively correlated with that target nonword's length in syllables ($r = -.47$, $p < .05$; $r = -.52$; $p < .05$). Although the remaining correlations shown in Table 4 did not reach statistical significance, they were all negative and indicate a trend toward better imitation of shorter target nonwords.

	Syllable Score	Initial Consonant Score
Target Duration (ms)	-.31	-.40
Target Syllable Length	-.47*	-.52*
Target # of Consonants	-.20	-.41

Table 4. Correlation *r*-values. *$p < .05$

The children's better performance in imitating shorter target nonwords was reflected in the perceptual ratings previously described. That is, among the 20 target stimuli, significant correlations were observed between the average perceptual rating for each nonword (averaged across children) and the nonword's length in milliseconds, syllables, and number of consonants. These correlations were all negative ($r = -.54$, $p < .05$; $r = -.62$, $p < .01$, and $r = -.54$, $p < .05$, respectively), indicating that the children's imitations of the shorter target nonwords generally received higher perceptual ratings.

Correlations Between Nonword Repetition Performance and Demographic Variables

Correlations were calculated between nonword repetition performance and the following demographic variables: (1) age in years at time of testing, (2) degree of exposure to an oral-only communication environment (based on Communication Mode scores), (3) age at onset of deafness, (4) duration of deafness prior to implantation, and (5) duration of CI use.

The only demographic variable that was significantly correlated with any of the nonword imitation measures was the age at onset of deafness, which correlated with the children's syllable scores ($r = +.60$, $p < .05$). This correlation must be viewed cautiously, as 11 of the 14 children in this study were congenitally deaf. Nevertheless, this moderate positive correlation indicates that children whose age at onset of deafness was later tended to produce the correct number of syllables in more of their imitations.

We were surprised to find that none of the other demographic measures correlated well with any of the other measures of nonword repetition performance. We suspect that the relative homogeneity of the demographic characteristics of the children in this study might have prevented statistically significant correlations. In looking more closely at the children's demographic characteristics, we found that Children 103, 214, and 301 had experienced the shortest durations of deafness prior to implantation and also earned the highest syllable scores and initial consonant scores. In terms of duration of CI use, we did not find any clear pattern of results (i.e., we did *not* find that children who had used their CIs for the longest period of time relative to the group performed the best). This is similar to the findings of Dawson et al. (1995), who reported that changes in speech production post-implantation did not seem to be related to the duration of CI use in the 12 CI users they studied.

Correlations Between Nonword Repetition Performance and Measures of Speech Perception

Several tests of speech perception and spoken word recognition were administered to the children as part of a larger project by CID clinicians within a few days of the nonword repetition recordings. Table 5 displays the *r*-values for the correlations between the measures of perception shown in the left-hand column, and the two measures of nonword performance described in this report: the average syllable score for each child (averaged across nonword) and the average initial consonant score for each child (averaged across nonword).

	Syllable Score	Initial Consonant Score
Speech Feature Discrimination		
VIDSPAC Manner	-.18	-.15
VIDSPAC Voicing	.03	.58*
VIDSPAC Place	.12	.29
Speech Perception/ Word Recognition		
WIPI	.48	.44
LNTE	.29	.63*
LNTH	.07	.60*
MLNT	.40	.42
BKB	.50	.05

Table 5. Correlation *r* values. *$p < .05$

The VIDSPAC assesses hearing-impaired children's perception of speech pattern contrasts and includes scores for discrimination of consonantal voicing, manner, and place (Boothroyd, 1997; Geers et al., 1999). As shown in Table 5, correlations between the VIDSPAC scores and the measures of nonword repetition generally did not reach statistical significance. Only the VIDSPAC Voicing scores were significantly correlated with the initial consonant scores.

The Word Intelligibility by Picture Identification (WIPI) test is a closed set test of spoken word identification test involving a pointing response (Ross & Lerman, 1979). We calculated correlations between WIPI scores and syllable scores, and between WIPI scores and initial consonant scores. Both correlations were positive, although neither reached statistical significance (WIPI and syllable scores, $r = +.48$, $p = .09$; WIPI and initial consonant scores, $r = .+44$, $p = .12$).

The Lexical Neighborhood Test (LNT; Kirk, Pisoni, & Osberger, 1995) is an open-set test of spoken word identification consisting of 100 monosyllabic words divided into four lists of 25 words each. Two of the lists contain words that are "lexically easy" (i.e., are phonetically similar to very few other words) and two of the lists contain words that are "lexically hard" (i.e., are phonetically confusable with many other words). A child is typically tested on one "easy" word list and one "hard" word list, with separate percent-correct scores obtained for each list. The Multisyllabic Lexical Neighborhood Test (MLNT), is analogous to the LNT, but uses multisyllabic words of 2 or 3 syllables.

Scores on LNT "easy" words and LNT "hard" words were significantly correlated with the initial consonant segment scores ($r = .63$, $p < .05$ and $r = .60$, $p < .05$), but their correlations with syllable scores were not significant ($r = .29$, $p = .31$; $r = .07$, $p = .82$). Scores on the MLNT were not significantly correlated with either the syllable scores or the initial consonant segment scores ($r = .41$, $p = .15$; $r = .41$, $p = .14$).

Lastly, we calculated correlations between measures of nonword repetition performance and performance on the Bamford-Kowal-Bench Sentence List Test (BKB), an open-set task involving spoken repetition of a target sentence (Bench, Kowal & Bamford, 1979). The correlations between syllable scores

and BKB scores, and between initial consonant scores and BKB scores, were both positive, although neither reached statistical significance (r's = .50, p's = .07).

In summary, the general pattern of correlations suggests that children who scored higher on measures of spoken word recognition tended to produce more imitations with the correct initial consonant. We speculate that with a larger sample size, more of these correlations would have reached significance.

Correlations Between Nonword Repetition Performance and a Measure of Language Comprehension

The battery of tests administered by CID also included the Test of Auditory Comprehension of Language Revised (TACL-R), a language comprehension measure that assesses children's receptive vocabulary, morphology, and syntax (Carrow-Woolfolk 1985). In the present study, the TACL-R was administered to all children using total communication (both speech and sign), and an age-equivalency score was obtained for each child. TACL-R age-equivalent scores were found to be highly correlated with the children's syllable scores (r = +.69, p < .01) and the initial consonant segment scores (r = +.65, p < .05), indicating that better performance on the nonword repetition task used in the present study also appears to correspond to higher language comprehension scores in terms of receptive vocabulary, morphology, and syntax.

Correlations Between Nonword Repetition Performance and a Measure of Working Memory

A measure of working memory was also obtained from the children using the WISC Digit Span Supplementary Verbal Sub-test of the Wechsler Intelligence Scale for Children, Third Edition (WISC-III) (Wechsler, 1991). For the purpose of the present study, we were interested in the "digits forward" subsection of this memory span task, in which a child listens to and repeats lists of digits as spoken live-voice by the experimenter at a rate of approximately one digit per second (WISC-III Manual, Wechsler, 1991). Two lists are administered at each list length, beginning with two digits. The list length is increased one digit at a time until the child fails to correctly repeat both lists administered at a given length. The child receives points for correct repetition of each list, with no partial credit. The children's WISC scores were found to be strongly correlated with their initial consonant scores (r = +.73, p < .01) and to be moderately correlated with their syllable scores (r = +.57, p < .05), indicating that a longer digit span as measured by the WISC task corresponds to higher scores on the measures of nonword repetition performance used in the present study.

Correlations Between Nonword Repetition Performance and Measures of Meaningful Speech Production

As part of the larger study at CID, a measure of speech intelligibility was also obtained from each child using the McGarr Sentence Intelligibility Test (McGarr, 1983). This test involves eliciting sentences containing either 3, 5 or 7 syllables in length. The child was provided with spoken and/or signed models of each sentence as well as the printed text of each sentence, and was prompted to speak as intelligibly as possible. The children's utterances were recorded and later played back to naive listeners who were asked to transcribe the children's speech using standard orthography. This provided an objective measure of speech intelligibility. Each child's productions were also submitted to an acoustic analysis. Included among the various acoustic measures was a simple measure of sentence duration. Pisoni and Geers (2000) reported that CI children's speaking rate on the McGarr sentences, particularly, the longer seven syllable sentences, was strongly correlated with measures of working memory as well as

with speech intelligibility. For this reason, in the present study, we examined the relationships between nonword repetition performance and McGarr Intelligibility. We also examined the relations between nonword repetition performance and sentence duration (duration being inversely related to speaking rate).

Strong correlations were found between nonword repetition performance and both speech intelligibility and sentence duration. McGarr intelligibility scores were correlated with the initial consonant segment scores ($r = +.68$, $p < .01$), indicating that the children who produced more intelligible speech on the McGarr task also tended to more often correctly reproduce initial consonants in their nonword imitations. Speaking rate was also related to nonword repetition performance. Mean sentence duration on the seven-syllable McGarr sentences was negatively correlated with the initial consonant segment scores ($r = -.64$, $p < .05$) and the syllable scores ($r = -.67$, $p < .01$). That is, the children who spoke more slowly in the McGarr task, tended to correctly imitate the initial consonants and the number of syllables in the nonwords *less often* than the children who spoke less slowly in the McGarr task. Both patterns replicate and extend the patterns of Pisoni and Geers (2000).

Summary of Correlational Analyses

In summary, the children who produced the correct initial consonants also tended to produce the correct number of syllables in their imitations. Overall, the children tended to obtain higher initial consonant scores and syllable scores for their imitations of shorter target nonwords.

The children who were not congenitally deaf tended to produce the correct number of syllables in more of their imitations. The children who scored higher on measures of spoken word recognition and the children who produced more intelligible speech on the McGarr task also tended to produce more imitations with the correct initial consonant. Better performance on the nonword repetition task, in terms of initial consonant scores and syllable scores, was associated with higher language comprehension scores, longer digit spans as measured by the WISC task, and faster speaking rates in the McGarr task.

The correlations summarized above reflect the close correspondence between the children's speech perception, speech production, and language skills. We also observed strong correlations between the children's nonword repetition performance scores and direct perceptual ratings, suggesting a correspondence between phonological skills and speech intelligibility. We speculate that with a larger sample size, nonword repetition performance might be even better predicted by demographic characteristics and the children's scores on other speech and language measures.

General Discussion

In the present study, we observed considerable variation among prelingually-deafened pediatric cochlear implant users in terms of their ability to imitate the duration, number of syllables, initial consonant segments, and features of the initial consonant of a set of 20 nonword patterns. While the children differed in terms of the number of errors they made, it is possible to make several generalizations about the types of errors that occurred. Our results show that the children correctly reproduced the number of syllables in a target more often than they correctly reproduced the initial consonant. The children correctly imitated the non-segmental characteristic that we measured (i.e., number of syllables) in 66% of the responses, but that they imitated the segmental characteristic that we measured (i.e., initial consonant accuracy) in only 39% of the responses. In other words, the children seemed to be able to reproduce the non-segmental characteristics of the target more easily than they could reproduce the more detailed segmental characteristics.

Previous studies of nonsegmental characteristics in the speech of cochlear implant users have focused on the loudness, pitch, and duration of children's imitations of words spoken by a clinician model (Tobey et al., 1991; Tobey & Hasenstab, 1991). Tobey and Hasenstab (1991) studied the speech productions of a group of 78 Nucleus multichannel CI users whose average age was 8.5 years (ranging from 2.3 to 17.7 years) and whose average age at onset of profound hearing impairment was 1.4 years. Each child's imitations were rated for accuracy. The children received 0 points for consistently incorrect productions, 1 point for inconsistently correct productions, and 2 points for consistently correct productions (in terms of loudness, pitch and duration). Tobey and Hasenstab found that overall, the CI users were significantly better at imitating nonsegmental aspects of speech (pitch, duration and loudness) than segmental aspects, and that the children's scores improved as they gained experience with their cochlear implants. Tobey and Hasenstab suggested that prosodic features of speech are more immediately salient to CI users, who show rapid post-implantation improvement in terms of *prosodic* aspects of speech production, which plateaus as the children begin to focus on improving their production at the segmental level. The results of the present study are consistent with this suggested pattern of development. The children correctly imitated the number of syllables in a target more often than they correctly imitated the initial consonant segment of a target.

Although nonsegmental aspects of production may indeed be less difficult than segmental aspects of production, previous studies have nevertheless found that "hearing-impaired and deaf speakers appear to have difficulty controlling suprasegmental features of speech, such as prosody and stress." (Tobey et al., 1991, p. 165; see also McGarr and Osberger, 1978). In our study, we did not analyze the children's productions for intonation or stress, but we found that the majority of the syllables omitted from the imitation responses were unstressed syllables. Since unstressed, or 'weak', syllables tend to have shorter durations and lower amplitudes than stressed syllables, the omission of weak syllables by the children in the present study may have resulted from greater difficulty in simply perceiving the weak syllables in the target stimuli.

Alternatively, the children's performance could be interpreted to mean that while they were able to perceive the prosodic structure of the target stimuli, they omitted and reduced syllables so that their productions conformed to rhythmic stress patterns or syllable templates, as proposed in accounts of normal-hearing children's speech (e.g., Carter, 1999/2000; Echols, 1993; Kehoe & Stoel-Gammon, 1997). Overall, the present findings regarding syllable insertion and omission, segmental errors, and durational differences between the imitations and the targets, are generally consistent with past research on the phonological development of younger normal-hearing children (e.g., Dinnsen, Barlow & Morrisette, 1997; Gerken, 1996).

Additionally, the children's performance in imitating initial consonant features corresponds with linguistic universals in several ways. For example, languages tend to have more stops than fricatives, and more voiceless than voiced obstruent phonemes (Lass, 1984). The pattern of responses observed for the children in this study is consistent with these universal characteristics of language. The children correctly imitated initial stops more often than fricatives, and voiceless targets more often than voiced targets. In terms of place, Lass (1984) reports that, across languages, phonemic coronals occur more frequently than labials or dorsals, and that if a given language contains only one place of articulation for a certain manner (stop or fricative), it is most likely to be coronal. This pattern is consistent with the children's higher scores in imitating coronals than labials or dorsals. More specifically, in the present study the children correctly imitated /s/ more often than /f/, and performed most poorly in imitating initial /v/. Cross-linguistically, for fricatives, /s/ occurs more frequently than /f/, and all other fricatives (such as /v/) are even less common (Lass, 1984).

The results of the present study indicate that when nonword stimuli are presented to pediatric cochlear implant users in an auditory-only mode, coronal consonants are imitated more accurately than labial and dorsal consonants (as shown in Figures 11 and 14). These results conflict with the earlier findings reported by Tobey et al. (1994) that, on average, pediatric cochlear implant users most often correctly produced labial consonants, followed by coronals, and lastly dorsals. An important difference between Tobey et al.'s study and the present investigation is in the presentation of the stimuli: recorded stimuli were presented in an auditory-only format in the present study, while a live-voice auditory-visual format was used in Tobey et al.'s study. When the children had access to visual cues in addition to the auditory cues, their performance in the imitation of labials surpassed that of coronals. However, when visual cues were not available, children more often correctly imitated coronals than labials (at least in terms of the initial target consonants). This difference is consistent with previous findings in the literature regarding variation in stimulus presentation format. For example, in a recent study of 27 pediatric cochlear implant users, Lachs, Pisoni, and Kirk (2001) found that audiovisual stimulus presentation elicited better overall speech perception performance than auditory-alone stimulus presentation.

The correlation we observed between the "playback" perceptual ratings of the nonword repetitions and initial consonant segmental accuracy scores lend support to previous claims that "poor segmental control may be related to poor overall speech intelligibility, [leading] many to conclude [that] the greater number of segmental errors, the poorer speech intelligibility will be in a speaker." (Tobey et al., 1991, p.165, cites Levitt & Stromberg, 1973; Parkhurst & Levitt, 1978; and Smith 1975). The present results support previous findings that speech intelligibility is related to the segmental accuracy of hearing-impaired persons' speech. Furthermore, the correlational analyses revealed a trend towards better imitation of shorter target nonwords, which also generally received higher perceptual ratings.

It is important to keep in mind that the consonant analyses reported in this study were based on the *initial* consonants only. Previous studies have found that cochlear implant users and normal-hearing children correctly produce initial consonants more often than medial or final consonants (Chin et al., 1997 and Dawson et al., 1995 for CI users; Ingram, 1989 for normal-hearing children). In addition, the findings reported in the present study must be viewed with a certain degree of caution because of the small number of children who participated and the use of a set of stimuli that were not phonologically balanced.

The analyses reported here were based on utterances obtained from an imitation study, which may naturally lead to questions about the generalizability of our results to the children's spontaneous or elicited speech production performance. Although we did not analyze non-imitative productions by these children, it should be noted that in a longitudinal study, Tobey et al. (1991) found that children's productions in imitated speech and in elicited spontaneous speech improved with increased implant use, suggesting that a common set of phonological skills is used in both imitated and spontaneous speech.

In analyzing nonword imitation responses, many variables need to be taken into consideration, including presentation mode, the phonetic similarity of the nonword to real words, and phonotactic properties of the nonword pattern. In examining segmental accuracy within an imitation, the results could be affected by the target segment's position within the syllable (Turk, 1993) and by its phonetic environment (Goodluck, 1991). While it is probably not feasible to control all, or even most, of these factors in any single study, it is crucial that all of these factors be addressed in studies of the speech of pediatric cochlear implant users, and that researchers publish as many methodological details as possible about the participants, stimulus materials, and testing conditions to permit the replication of specific findings and verification of generalizations. Further testing, with the use of new stimulus materials that are phonologically-controlled, would be useful in confirming the present findings and extending our results.

Conclusion

In summary, the present investigation analyzed the utterances of fourteen deaf children with cochlear implants who completed a nonword repetition task. The children who participated in this study demonstrated the ability to imitate unfamiliar but "word-like" targets by using their knowledge of the phonological patterns present in their ambient language. These fourteen children exhibited systematic error patterns that often resembled those found in the developing speech of normal-hearing children. Their responses also reflected linguistic universals. The children's difficulty in imitating initial labials in this study suggests that other reports of superior performance on labials are likely the result of perceptual enhancement due to the presence of visual cues or children's prior lexical knowledge. The strength of the correlations between our phonological measures and direct perceptual ratings of the imitations by naive listeners suggests that detailed linguistic analyses can help us to understand which aspects of the imitations listeners attend to when judging accuracy. Taken together, the results of this study demonstrate that the nonword repetition task can provide new insights into the speech production skills and underlying linguistic abilities of pediatric cochlear implant users. With further analytic studies of this type, we hope to better understand the relations between auditory, cognitive, and articulatory processes used in the perception and production of spoken language, and how these develop and change in deaf children following cochlear implantation.

References

Avons, S.E., Wragg, C.A., Cupples, L., & Lovegrove, W.J. (1998). Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics, 19,* 583-601.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13,* 108-112.

Block, S. & Killen, D. (1996). Speech rates of Australian English-speaking children and adults. *Australian Journal of Human Communication Disorders, 24,* 39-44.

Boothroyd, A. (1997). Vidspac 2.0: A video game for assessing speech pattern contrast perception in children. San Diego, Arthur Boothroyd.

Carlson, J.L., Cleary, M., & Pisoni, D.B. (1998). Performance of normal-hearing children on a new working memory span task. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 251-273). Bloomington, IN: Speech Research Laboratory, Indiana University.

Carrow-Woolfolk, E. (1985). *Test for Auditory Comprehension of Language-Revised (TACL-R).* Austin, TX: Pro-Ed.

Carter, A.K. (2000). *An Integrated Acoustic and Phonological Investigation of Weak Syllable Omissions*. (Doctoral dissertation, University of Arizona, 1999). *Dissertation Abstracts International, 60/09,* 3339.

Chin, S.B., & Finnegan, K.R. (1998). Minimal pairs in the perception and production of speech by pediatric cochlear implant users: A first report. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 291-303). Bloomington, IN: Speech Research Laboratory, Indiana University.

Chin, S.B., Kirk, K.I., & Svirsky, M.A. (1997). Sensory aid and word position effects on consonant feature production by children with profound hearing impairment. In *Research on Spoken Language Processing Progress Report No. 21* (pp. 455-470). Bloomington, IN: Speech Research Laboratory, Indiana University.

Chin, S.B., Pisoni, D.B., & Svec, W.R. (1994). An emerging phonetic-phonological system two years post cochlear implant: A preliminary linguistic description. In *Research on Spoken Language Processing Progress Report No. 19* (pp. 253-270). Bloomington, IN: Speech Research Laboratory, Indiana University.

Cleary, M., Dillon, C.M. & Pisoni, D.B. (submitted). Imitation of nonwords by deaf children following cochlear implantation. In *Annals of Otology, Rhinology & Laryngology, Supplement-Proceedings of the 8th Symposium on Cochlear Implants in Children*, March 2001.

Dawson, P.W. (1995). A clinical report on speech production of cochlear implant users. *Ear & Hearing, 16,* 551-561.

Dinnsen, D.A., Barlow, J.A., & Morrisette, M.L. (1997). Long-distance place assimilation with an interacting error pattern in phonological acquisition. *Clinical Linguistics & Phonetics, 11,* 319-338.

Echols, C.H. (1993). A perceptually-based model of children's earliest productions. *Cognition, 46,* 245-296.

Edwards, J. & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19,* 279-309.

Ertmer, D.J., Kirk, K.I., Sehgal, S.T., Riley, A.I., & Osberger, M.J. (1997). A comparison of vowel production by children with multichannel cochlear implants or tactile aids: Perceptual evidence. *Ear & Hearing, 18,* 307-315.

Goodluck, H. (1991). *Language Acquisition: A Linguistic Introduction.* Cambridge, MA: Blackwell.

Gathercole, S.E. (1995). Is non-word repetition a test of phonological memory or long-term knowledge? It all depends on the non-words. *Memory and Cognition, 23,* 83-94.

Gathercole, S.E. & Baddeley, A.D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81,* 439-454.

Gathercole, S.E., & Baddeley, A.D. (1996). *The Children's Test of Non-word Repetition*. London: Psychological Corporation.

Gathercole, S.E, Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Non-word Repetition: A test of phonological working memory. *Memory, 2,* 103-127.

Geers, A.E., Nicholas, J., Tye-Murray, N., Uchanski, R., Brenner, C., Crosson, J., Davidson, L.S., Spehar, B., Torretta, G., Tobey, E.A., Sedey, A., & Strube, M. (1999). Center for Childhood Deafness and Adult Aural Rehabilitation, Current research projects: Cochlear implants and education of the deaf child, second-year results. In *Central Institute for the Deaf Research Periodic Progress Report No. 35* (pp. 5-20). St. Louis, MO: Central Institute for the Deaf.

Geers, A.E., & Tobey, E. (1992). Effects of cochlear implants and tactile aids on the development of speech production skills in children with profound hearing impairment. *The Volta Review, 94,* 135-163.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language, 21,* 565-584.

Hesketh, L.J., Fryauf-Bertschy, H., & Osberger, M.J. (1991). Evaluation of a tactile aid and a cochlear implant in one child. *The American Journal of Otology, 12* (Suppl.), 183-187.

Ingram, D. (1989). *Phonological disability in children* (2nd ed.). London: Cole & Whurr.

Johnson, J.S., & Salidis, J. (1995). The shape of early words. *Proceedings of the Boston University Conference on Language Development, 20,* 386-396.

Kehoe, M., & Stoel-Gammon, C. (1997). Truncation patterns in English-speaking children's word productions. *Journal of Speech, Language, and Hearing Research, 40,* 526-541.

Kirk, K.I., Diefendorf, E., Riley, A., & Osberger, M.J. (1995). Consonant production by children with multichannel cochlear implants or hearing aids. In Uziel, A.S., & Mondain, M. (Eds.), *Advanced Otorhinolaryngology, 50,* 154-159.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16,* 470-481.

Lachs, L., Pisoni, D.B., & Kirk, K.I. (2001). Use of audio-visual information in speech perception by pre-lingually deaf children with cochlear implants: A first report. *Ear and Hearing, 22,* 236-251.

Lass, R. (1984). *Phonology*. Cambridge: Cambridge University Press.

Levitt, H., & Stromberg, H. (1983). *Speech of the Hearing-Impaired: Research, Training and Personnel Preparation*. Baltimore: University Park Press.

Lyxell, B., Andersson, J., Andersson, U., Arlinger, S., Bredberg, G., & Harder, H. (1998). Phonological representation and speech understanding with cochlear implants in deafened adults. *Scandinavian Journal of Psychology, 39,* 175-179.

McGarr, N.S. & Osberger, M.J. (1978). Pitch deviance and intelligibility of deaf speech. *Journal of Communication Disorders, 11,* 237-247.

McGarr, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26,* 451-458.

Miyamoto, R.T., Kirk, K.I., Robbins, A.M., Todd, S., & Riley, A. (1996). Speech perception and speech production skills of children with multichannel cochlear implants. *Acta Otolaryngol, 116,* 240-243.

O'Donoghue, G.M., Nikolopoulos, T.P., Archbold, S.M., & Tait, M. (1999). Cochlear implants in young children: the relationship between speech perception and speech intelligibility. *Ear & Hearing, 20,* 419-425.

Osberger, M.J. & McGarr, N. (1982). Speech production characteristics of the hearing-impaired. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice*. New York: Academic Press.

Osberger, M.J., Robbins, A.M., Berry, S.W., Todd, S.L., Hesketh, L.J., & Sedey, A. (1991). Analysis of the spontaneous speech samples of children with cochlear implants or tactile aids. *The American Journal of Otology, 12* (Suppl.)*,* 151-164.

Osberger, M.J., Maso, M., & Sam, L.K. (1993). Speech intelligibility of children with cochlear implants, tactile aids, or hearing aids. *Journal of Speech and Hearing Research, 36,* 186-203.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30,* 331-347.

Parkhurst, B., & Levitt, H. (1978). The effect of selected prosodic errors on the intelligibility of deaf speech. *Journal of Communication Disorders, 11,* 246-256.

Pisoni, D.B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear & Hearing, 21,* 70-78.

Ross, M. & Lerman, J. (1979). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13,* 44-53.

Sehgal, S.T., Kirk, K.I., Ertmer, D.J., & Osberger, M.J. (1998). Imitative consonant feature production by children with multichannel sensory aids. *Ear & Hearing, 19,* 72-184.

Serry, T.A., & Blamey, P.J. (1999). A 4-year investigation into phonetic inventory development in young cochlear implant users. *Journal of Speech, Language, and Hearing Research, 42,* 141-154.

Serry, T., Blamey, P., & Grogan, M. (1997). Phoneme acquisition in the first 4 years of implant use. *The American Journal of Otology, 18,* 122-124.

Slobin, D.I. & Welsh, C.A. (1973). Elicited imitation as a research tool in developmental psycholinguistics. In C.A. Ferguson & D.I. Slobin (Eds.), *Studies in Child Language Development*. New York: Holt, Rinehart & Winston.

Smith, C. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research, 18*, 795-811.

Stoel-Gammon, C. & Cooper, J.A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language, 11,* 247-271.

Tobey, E.A., Angelette, S., Murchison, C., Nicosia, J., Sprague, S., Staller, S.J., Brimacombe, J.A., & Beiter, A.L. (1991). Speech production performance in children with multichannel cochlear implants. *The American Journal of Otology, 12* (Suppl.)*,* 165-173.

Tobey, E., Geers, A., & Brenner, C. (1994). Speech production results: Speech feature acquisition [Monograph]. *The Volta Review, 96,* 109-129.

Tobey, E.A., & Hasenstab, M.S. (1991). Effects of a Nucleus multichannel cochlear implant upon speech production in children. *Ear and Hearing, 12,* 48-54.

Tye-Murray, N., Spencer, L., Bedia, E.G., & Woodworth, G. (1996). Differences in children's sound productions when speaking with a cochlear implant turned on and turned off. *Journal of Speech and Hearing Research, 39,* 604-610.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children - III*. San Antonio, TX: The Psychological Corporation.

Zamuner, T.S. (2001). *Input-based Phonological Acquisition*. Unpublished doctoral dissertation, University of Arizona.

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Speech Perception and Implicit Memory: Evidence for Detailed Episodic Encoding of Phonetic Events[1]

Lorin Lachs, Kipp McMichael and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH RO1 DC00111 and NIH T32 DC00012. Thanks to Luis Hernàndez and Darla Sallee for assistance with the final preparation of this paper.

[2] Also Devault Otologic Research Laboratory, Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Speech Perception and Implicit Memory: Evidence for Detailed Episodic Encoding of Phonetic Events

Abstract. Recent investigations into the nature of memory for spoken words have demonstrated that detailed, episodic information about the voice of the talker is encoded along with more abstract information about the linguistic content of an utterance. These findings present serious challenges to traditional views of speech perception, in which the process of abstraction plays a major role. We first outline the traditionalist view of speech perception, a theoretical framework largely based on the constructs of formal linguistics. Next, we elaborate the recently emerging "detailed encoding" view and summarize recent evidence from perceptual and memory experiments that demonstrate that "linguistic" and "extra-linguistic" information in spoken language are inseparable in the acoustic signal and in the representation of speech in long-term memory.

Introduction

Nearly every aspect of human speech - our accents, word choice, and even the very language we utter – is influenced by past experience. The perceptual process occurs very quickly and often appears to be carried out almost automatically. For the most part, we rarely, if ever, have any conscious awareness of our linguistic knowledge or our previous experience during speech production or perception. These general observations about speech perception suggest that implicit memory processes may play a pervasive and perhaps indissociable role in both speech perception and production. Yet despite a widespread acceptance by researchers that all behavior is ultimately grounded in prior, long-term experience, the role of implicit memory in speech production and perception has only been the focus of experimental inquiry by cognitive psychologists within the last few years.

To explain why implicit memory research in speech perception has only recently emerged, we begin this chapter with a review and discussion of the theoretical and meta-theoretical notions that underlie the traditional, abstractionist characterization of speech perception. Once we have described the traditional framework, we move on to an emerging view of speech processing and memory where both explicit and implicit effects find a unified, straightforward explanation. Finally, we will expand this emerging view to show that it is highly compatible with a seamless, undichotomized human memory system that incorporates both implicit and explicit memory components.

Speech Perception: The Abstractionist Perspective

At its inception, the field of speech science borrowed many of its constructs and conceptualizations about language from formal linguistics. Perceptual units such as phonemes, allophones, morphemes, and even words themselves were simply direct transplantations from linguistic theory. Even after extensive analysis of speech spectrograms made it clear that speech was nothing like a discrete sequence of idealized segments (Liberman, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), researchers continued to maintain that speech was, in essence, a discrete, symbolic signal (Licklider, 1952).

Under this view, then, speech is reduced to spoken text. The viewpoint was so widely accepted that Morris Halle, the noted linguist, went so far as to say:

> 'There can be no doubt that speech is a sequence of discrete entities, since in writing we perform the kind of symbolization just mentioned, while in reading aloud we execute the inverse of this operation; that is, we go from a discrete symbolization to a continuous acoustic signal.' (1956)

Not all views of speech perception have had such a literalist reading of the 'speech as spoken text' hypothesis, but accepted meta-theoretical notions about the discrete, idealized, symbolic nature of speech have been the dominant influence on research in speech perception and production for more than fifty years. Under this view, outlined quite explicitly in early work by phoneticians like Abercrombie (1967), a fundamental distinction is drawn between the *language* and the *medium* that mediates between speech production and reception. For example, the written word is a visible medium that transfers the 'language' produced by a writer to that received by the reader. Likewise, the audible signal generated by the talker's vocal tract during speech production transfers 'language' via acoustic medium to the listener. Because a physical medium has properties that are not related to the communication of language, a dichotomy arises concerning information contained in the physical signal:

> '…all that is necessary for linguistic communication is that the contrasts on which the patterns are based should not be obscured. Usually, therefore, many things about a medium which is being used as vehicle for a given language are not relevant to linguistic communication. Such 'extra-linguistic' properties of the medium, however, may fulfill other functions which may sometimes even be more important than the linguistic communication, which can never be completely ignored.' (Abercrombie, 1967, p.5)

While certainly acknowledging the utility of 'extra-linguistic' variation in the speech signal, this passage illustrates two very important aspects of the traditionalist view of speech and language. First, the passage states clearly that the primary function of any language pattern is the communication of contrast, which can be used to recover the linguistic content of a message. Second, and more importantly, the extra-linguistic properties of the signal are *defined* as the exclusive complement to linguistic properties encoded in the signal. That is, any property of a medium that was *not relevant* to signaling the linguistic content was considered to be extra-linguistic. By this view, extra-linguistic information is simply a source of undesirable 'noise' created in the physiological realization of the idealized speech signal.

Linguistic content, then, is information specifying the underlying, linguistic representation of an utterance, such as segments, phonemes, syllables or other idealized, symbolic units like words. Extra-linguistic content is everything else in the signal. Abercrombie (1967) describes the importance of extra-linguistic content, pointing out that it may contain signs or indices of other, non-linguistically important information about the talker. These *indexical* features of the speech signal –as opposed to linguistic features - might include such things as the talker's gender, dialect, or affect. However, it is precisely the dissociation between linguistic and extra-linguistic information in speech that, in our view, makes this traditional account of spoken language questionable at the present time. Over the last few years, many new findings about the contribution of extra-linguistic information to speech perception have been reported in the literature. These findings suggest that the traditional dichotomy between linguistic and extra-linguistic information in the speech signal may be somewhat misleading and possibly an incorrect characterization of the sensory information that human listeners perceive, encode and store about their language.

Reconstruction and Abstraction

The notion that speech is a noisy and highly degraded signal that fails to perfectly transmit the intended utterance of the speaker led to reconstructionist accounts of speech perception. In the words of

Neisser (1967), 'There must be a kind of filtering, of feature-detection, which precedes the active construction of detail.' (p.196). According to this view, the impoverished acoustic signal is processed extensively to uncover the underspecified linguistic message that is encoded in the speech waveform. Based on rules or schema derived from acquired linguistic knowledge, the speech signal is further processed to construct an accurate perception of the intended utterance. This view of speech was extremely compatible with the information-processing framework of early cognitive psychology (Studdert-Kennedy, 1974; Studdert-Kennedy, 1976), even as J.J. Gibson's approach to perception challenged the notions of underspecification and reconstruction in the field of perception more generally (Gibson, 1966).

The process of speech perception is, according to traditional accounts, a cleaning up or filtering mechanism that uncovers sequences of idealized units such as phonemes, or words. These abstractionist accounts of speech (Pisoni, 1997) make extra-linguistic information unavailable for encoding into memory for speech events – unless some ad hoc reintegration process is proposed before storage. Thus, the long-term memory store of spoken words and knowledge about words - the mental lexicon - necessarily becomes a formalized, idealized, abstract database of linguistic information, a large collection of symbolic representations of words in the language.

This view of speech has motivated a very specific set of research questions and encouraged the development of experimental methodologies that have been prevalent over the last 50 years. Because extra-linguistic variation was thought to obscure the 'real' objects of speech perception–the underlying, abstract, symbolic linguistic units–factors related to the talker's voice, speaking rate, dialect and affect were either eliminated from experimental designs or explicitly controlled so that effects of these 'irrelevant' factors would not obscure the 'interesting' phenomena more directly related to linguistic communication. Hundreds of experiments on speech perception have studied the perception of utterances spoken by a single talker or the perception of highly controlled 'minimal' units of language, like features or phonemes, in CV nonsense syllables using highly controlled synthetic speech signals (see Liberman et al., 1967).

As a consequence, this research paradigm has provided very little information relating to the human listener's remarkable ability to perceive speech accurately and robustly under a wide variety of conditions and circumstances. We take this ability to deal with enormous stimulus variability in the signal to be of paramount importance to the process of spoken word recognition (Klatt, 1989). Indeed, the usefulness of a linguistic system is severely, if not totally, called into question if it is highly susceptible to drastic and unrecoverable interference as a result of the seemingly limitless conditions under which spoken language is used. Ironically, the lack of research into speech variation and variability and the ways in which listeners deal with these perceptual problems is potentially quite damaging to our understanding of spoken communication. In our view, the traditional abstractionist, symbolic, or "symbol-processing" framework can no longer be accepted without serious question as to its utility. We now turn to an alternative theoretical framework in which the importance of stimulus variation is acknowledged and made explicit: the detailed encoding perspective.

Speech Perception: A Detailed Encoding Perspective

The time-varying acoustic signal that impinges upon the ears of the listener is not one that is neatly divided into linguistic and extra-linguistic information. The acoustic signal of speech simultaneously carries information about the linguistic utterance as well as information about the source of the utterance and the listener's communicative circumstances. In other words, linguistic and extra-linguistic information are mixed together and fundamentally inseparable in their initial acoustic form.

In contrast to the traditional views of speech and speech perception, then, one can consider the object of speech processing as a very rich, detailed representation of the original articulatory events that created the signal (Fowler, 1986; Goldinger, 1998). Since this representation incorporates both linguistic and extra-linguistic information, we need not puzzle over how the abstract, idealized, and formalized units of language are first separated from the extra-linguistic information in the speech signal and later recombined for subsequent semantic processing, where information such as gender, dialect or affect become relevant.

Rather than viewed as a filtering or abstracting mechanism, the nature of speech perception and processing in a detailed encoding framework is variation-preserving. Under this novel view, speech processing yields a representation of the speech signal much like the original signal itself: a very rich, interleaved collection of information about the underlying events that generated the acoustic signal, in which linguistic and extra-linguistic variation are both inextricably linked.

Detailed Encoding and Stimulus Variability

The acoustic signal that carries linguistic and extra-linguistic information provides a rich and very detailed source of information about the speaker, speaking environment, and the linguistic message. This proposal is nicely illustrated in Figure 1, a schematic diagram taken from Hirahara and Kato (1992). The figure describes some of the encoding processes that take place when an incoming acoustic signal is processed by the nervous system. Of particular interest to the present discussion is the top level of the figure, where the composite form of linguistic and extra-linguistic information is illustrated by two transformations of the incoming signal. On the left, particular frequencies in the signal are represented using a Bark scale. These frequencies correspond to the resonances of the vocal tract and can be grouped into three primary clusters, commonly referred to as formants. The location and absolute frequency of these formants provide the distinctive cues to talker identification, an 'extra-linguistic' feature of the signal. On the right, the same acoustic signal is transformed and represented on a Bark difference scale, showing the relationships between these formants. These relative differences are necessary and sufficient for vowel identification, which is based on the 'linguistic' features of the signal. Thus, different analyses of the same acoustic signal yield two distinct sources of information about its production. It is important to point out here that both of these analyses are based on a frequency analysis of the components of the acoustic signal. It is not that there are two different sources of information buried in the signal for each of these sets of attributes. Rather, the two properties of speech—the linguistic and indexical—are carried simultaneously and in parallel by the same acoustic signal.

By conceptualizing the speech signal as a rich source of information, we adopt an ecological approach to speech perception (Fowler, 1986; Gaver, 1993). According to this view, speech is neither under specified nor noisy; potentially, all of the variation in the acoustic signal can be utilized during the process of speech perception and spoken word recognition. Variation is assumed to be lawful and highly informative about the articulatory events that underlie the production of speech (Fowler, 1986; Pisoni, 1997).

The initial stages of speech perception within such a framework are then stages of information detection and recognition rather than reconstruction and rule application. Essentially, the detailed-encoding framework embraces the fact that any dichotomy between linguistic and extra-linguistic information in the speech signal is arbitrary. The distinction between linguistic and extra-linguistic information becomes merely a convenient way of discussing the different kinds of tasks that can be carried out on an acoustic speech signal by a listener. Moreover, the increased emphasis on processing of the variation in the speech signal intuitively explains the retention of this information in memory – without the need for re-integration of separate sources of linguistic and extra-linguistic information.

Because extra-linguistic information is not lost or filtered from the incoming signal, it is encoded in memory and available for use at later levels of processing.

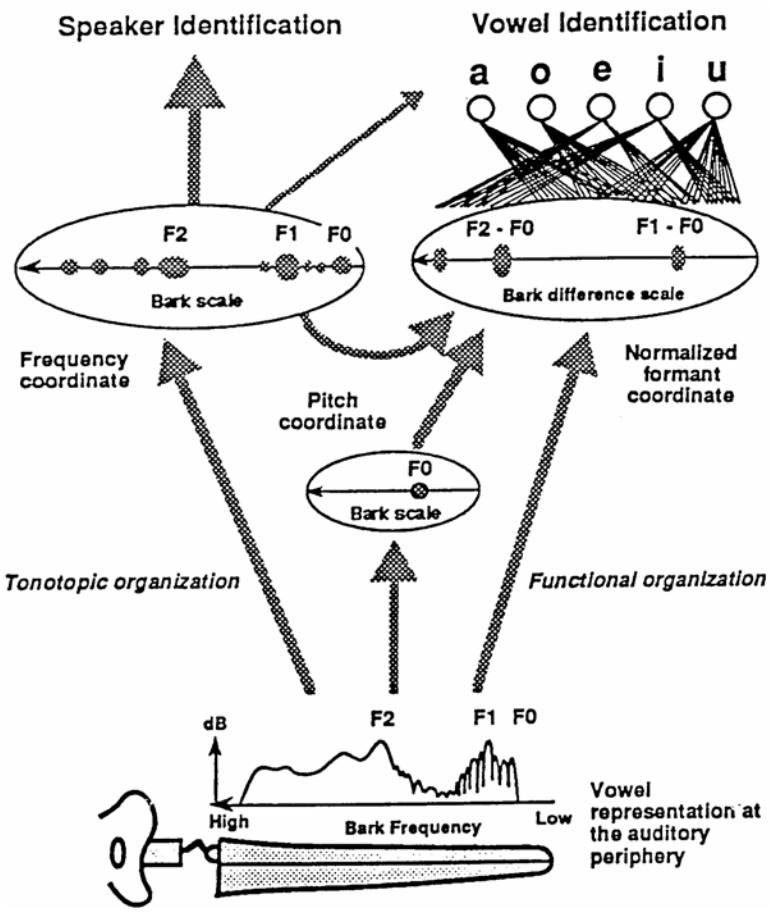


Figure 1. Information for talker and vowel identity is carried in parallel by an acoustic signal. Absolute frequencies contain information useful in talker identification, represented here along a Bark scale. Simultaneously, relative frequencies provide information useful in vowel identification (Taken from Hirahara & Kato, 1992).

This fundamental reconceptualization of the distinctive information available in the speech signal is not simply convenient or philosophically intriguing. This emerging view was necessitated by the results of a variety of novel experiments conducted over the last 10 years. In the next section, we summarize some of these findings and place them in a somewhat broader framework. We consider perceptually based phenomena in speech and describe how they affect both implicit and explicit memory processes.

Processing Dependencies: Effects of Stimulus Variation in Speech Perception

Although early work in speech perception suggested that the effects of extra-linguistic variation on the perception of spoken language were minimal (Creelman, 1957; Peters, 1955), these conclusions must be interpreted in light of the meta-theoretical notions that influenced the research agenda of the day. For example, although Creelman (1957) found that the accuracy of spoken word recognition under three

different signal-to-noise ratios decreased with an increased number of talkers, he dismissed this effect as 'relatively minor'—only a difference in performance of 7%. A difference of this size probably seemed like a small effect back in the middle 1960s. In an analogue, magnetic audiotape era, large corpora of stimuli spoken by multiple talkers were difficult to create and use in behavioral experiments with human listeners. The complex presentation schemes required to uncover effects of stimulus variation were likewise virtually intractable before the advent of computer controlled experiments. Computer control and the digital audio format have provided the tools to examine and understand the nature of stimulus variation in speech processing and encoding.

Of course, much of this discussion on the composite nature of linguistic and extra-linguistic information in memory for speech would be moot if there were not evidence that the two forms of information show demonstrable effects upon each other during processing. Early studies showed that simply changing the voice of the talker from one trial to the next affected the identification of vowels (Verbrugge, Strange, Shankweiler, & Edman, 1976), consonants (Fourcin, 1968), and words (Creelman, 1957; Mullennix, Pisoni, & Martin, 1989). In addition, changes in the talker's voice also affect speed of processing. In one study, Cole, Coltheart and Allard (1974) had participants make same-different judgments on pairs of syllables. The items in each pair were spoken either by the same talker or by different talkers. Despite the fact that the task required access to and use of what would traditionally be called 'linguistic' information, Cole et al. found that reaction times were slower when different talkers spoke the two syllables in a pair than when the same talker was used. Obviously, then, variation along an extra-linguistic dimension affects the performance in even the simplest of linguistic tasks like determining if a pair of words is the same or different.

A more detailed investigation was carried out by Mullennix and Pisoni (1990) to assess the codependencies of processing linguistic and extra-linguistic information. Using a Garner speeded classification task (Garner, 1974), they constructed sets of stimuli that varied along two dimensions. One dimension, the 'word' dimension, varied the cues to phonetic categorization of the initial segment of a word (e.g., 'b' vs. 'p'). The other dimension, the 'voice' dimension, varied the cues to the identity of the talker uttering the word (e.g., 'male' and 'female'). Mullennix and Pisoni asked subjects to make several judgments about the stimuli using one dimension at a time, while manipulating the variation along the irrelevant dimension. In the 'orthogonal' conditions, the irrelevant dimension was varied randomly from trial to trial. Subjects were asked to classify stimuli as either 'b' or 'p', while the stimuli varied in terms of the gender of the talker speaking the token. For example, in this condition, sometimes the 'b' token would be spoken by the male talker, and sometimes by the female talker. In the 'correlated' conditions, the irrelevant dimension varied consistently along with the relevant dimension. In other words, a male talker might always speak the 'b' tokens, while a female talker would always speak the 'p' tokens. Finally, in the 'control' conditions, the irrelevant dimension was always held constant, while subjects made judgments about the relevant dimension (i.e., the male or the female spoke all the tokens). Response latencies were collected so that patterns of processing speed could be assessed across these different conditions.

Mullennix and Pisoni found consistent differences in reaction time that depended on the variation in the irrelevant dimensions. Response times were fastest in the correlated conditions, slower in the control conditions and slowest in the orthogonal conditions. Correlated variation along the irrelevant dimension produced a 'redundancy gain' and facilitated classification times, while orthogonal variation along the irrelevant dimension inhibited classification and slowed down response times. The pattern of speeded classification data was consistent with the proposal of mutually dependent processing of the two stimulus dimensions. In other words, the perceptual aspects of a spoken word that are associated with phonetic information and those attributes that are associated with talker information are not analyzed independently, but rather are perceived and processed in a mutually dependent fashion. Interestingly,

Mullennix and Pisoni also manipulated the 'extent' of variation along each dimension in several additional experiments in which the number of response alternatives along each dimension was varied from 2 to 4, 8 or 16. While the general pattern of results was similar across all four conditions, they found that the amount of interference in the orthogonal condition increased as a function of stimulus variability. The results of Mullennix and Pisoni's study provide further evidence that increases in stimulus variation produce reliable effects on perceptual processing time and suggest that fine details of the stimulus patterns are not lost or discarded in a speeded classification task.

Thus, stimulus variability has an effect on speech processing. More importantly, the information about a talker's voice in an acoustic signal is processed in a dependent or contingent fashion along with the information specifying the linguistic content of the message. But precisely what kind of information about a talker's voice is available, and how does that information contribute to speech perception? In a measurement study of the acoustic correlates of talker intelligibility, Bradlow, Torretta and Pisoni (1996) found that while global characteristics of speech such as fundamental frequency and speaking rate had little effect on speech intelligibility, detailed changes in the acoustic-phonetic properties of a talker's voice, such as the amount of vowel space reduction and the degree of 'articulatory precision', were strong indicators of overall speech intelligibility. Their findings suggest that the indexical properties of a talker may be completely intermixed with the phonetic realization of an utterance and there may be no real dissociation between the two sources of information in the speech signal itself.

More direct evidence for the parallel encoding of linguistic and extra-linguistic information in the speech signal comes from other studies using sinewave replicas of speech. Sinewave speech is created by generating independent sinusoidal signals that trace the center frequencies of the three lowest formants in naturally produced utterances. The resulting pattern sounds perceptually unnatural, but the signal can be perceived by listeners as speech and the original linguistic message can be recovered (Remez, Rubin, Pisoni, & Carrell, 1981). Indeed, not only is the linguistic content of the utterance perceptible, but specific aspects of a talker's unique individual identity and speaking style are also preserved in sinewave replicas of speech.

Remez, Fellowes, and Rubin (1997) reported that listeners could explicitly identify specific familiar talkers from sinewave replicas of their utterances. Their findings on familiar talker recognition are remarkable because sinewave speech patterns preserve none of the traditional 'speech cues' that were thought to support the perception of vocal identity, such as fundamental frequency, or the average long-term spectrum. In creating sinewave speech patterns, an utterance is essentially stripped of all of the redundant acoustic information in an utterance except the time-varying properties of the vocal resonances generated by articulatory motion. While these skeletonized versions of speech have been shown to be sufficient for accurate identification of the linguistic content of a message, the new findings by Remez and colleagues demonstrates that sinewave speech patterns are also sufficient for the accurate identification of extra-linguistic information about familiar voices as well. These time-varying sinewave speech patterns preserve individual, talker-specific cues needed for voice recognition.

Thus, even in its most basic forms, linguistic and extra-linguistic sources of information appear to be inextricably bound to one another. Because sinewave speech patterns preserve little of the original signal other than the acoustic variation corresponding to the kinematics of articulatory motion, we suggest that the link between linguistic and extra-linguistic information derives from the common underlying articulatory events and movements of the speech articulators that produce speech. As we have argued, these links produce consistent effects on speech perception. But do the links between linguistic and extra-linguistic sources of information affect the memory processes that are so crucial to spoken word recognition and lexical access? We suggest they do in the next section.

Detailed Encoding Effects in Implicit and Explicit Memory

The integration of linguistic and extra-linguistic attributes in the speech signal and the mutually dependent perceptual processes that encode and process these cues has several important implications for the representation of speech in memory. According to the detailed encoding perspective, the mental representation of speech preserves the same sorts of information found in the original speech event (Goldinger, 1998). Rather than a static word-store of idealized, abstract, formalized units, Goldinger (1998) has proposed that the mental lexicon should be viewed as an extremely detailed set of instance-specific episodes. Extra-linguistic and linguistic information are preserved in the lexicon just as they are encoded in the auditory signal - in an integrated, holistic composite of linguistic and extra-linguistic properties. Evidence supporting this 'episodic' view of the lexicon comes from a series of recent memory experiments that show effects of extra-linguistic variation, even when the specific task only requires access to and use of linguistic information alone. The specific memory demands of the task—whether the task measures or assesses explicit or implicit memory—should not matter. If the basic representation of speech events in memory is highly detailed and episodic in nature, then any behavior that requires access to these memory representations should show contingent effects of these detailed composite representations.

While written words have been the primary focus in implicit memory research (Bowers, 2000), *spoken* words have received much less attention in the implicit memory literature. In the next section, we review some of the recent work that has been done on implicit effects of extra-linguistic variation in speech. We take as our starting point the operational definition summarized by Schacter (1987) that 'implicit memory is revealed when previous experience facilitates performance on a task which does not require conscious or intentional recollection of those experiences.' (p.501). The results of experiments using different memory paradigms are important in establishing the generality of these findings. Thus, we summarize experiments that examine the role of variability in both implicit and explicit memory for speech events.

Effects of Stimulus Variation on Implicit Memory

In a perceptual identification experiment conducted by Goldinger (1992), several groups of subjects were first asked to repeat words spoken to them in the quiet over headphones. The original set of stimuli was spoken by pools of 2, 6 and 10 talkers. Subjects then returned 5 minutes, 1 day or 1 week following the initial exposure and again identified spoken words in the quiet. Goldinger found that subjects were faster and more accurate in repeating words spoken by old talkers who were used at the time of the initial presentation than new talkers. Figure 2 shows the difference between test phase and study phase accuracy for words in the three talker pools across the three delay periods. Overall, Goldinger's data show evidence for a 'repetition effect'. That is, the identification of words was more accurate when those words were repeated in the same voice that spoke them at the time of study than in a novel voice. In addition, the advantage conferred by a repeated voice did not significantly decline as the delay between training and testing increased from 5 minutes to 1 day to 1 week. Goldinger's findings demonstrate that long-term memory representations of speech events not only include extra-linguistic information, but also preserve these instance-specific details for long periods of time. For talker similarity to have any effects on repetition accuracy, a record or memory trace of the extra-linguistic attributes of the talkers' speech had to persist in memory along with the more abstract, symbolic linguistic information encoded in the signal.

Goldinger also found that the differences in repetition accuracy for old and new voices were related to the perceptual similarity of the talker's voices. Words produced by talkers who had perceptually distinctive voices resulted in larger repetition effects for repeated talkers than words produced by talkers

who had less distinctive voices. These latter lists showed smaller, but still significant, effects for repeated talkers. The 'graded' effects that similarity had on the repetition effect was interpreted by Goldinger as evidence for an episodic view of memory for spoken items. Even when subjects were not instructed to attend to the voices of the stimuli, their memories for these speech events were detailed enough for the relative similarity among the talkers' voices to produce differential effects on performance in this task. Such findings are inherently compatible with exemplar models of categorization, in which similarity is represented continuously as distance in a perceptual space (Nosofsky, 1986; Nosofsky, 1987). Because points in the perceptual space represent individual tokens and not prototypical, idealized, abstract categories, the exemplar view of the lexicon provides a representational basis for making predictions sensitive to the graded similarity between voices.

In another study, Schacter and Church (1992) found consistent effects of voice information on implicit memory for words. In their experiments, subjects completed a study phase in which they made simple judgments about the enunciation or intonation of lists of words spoken by multiple talkers. Subjects then completed several implicit and explicit memory tasks in a test phase. Test stimuli were composed of tokens from the original study phase and 'new' tokens derived from study stimuli by changing the voice, intonation, or fundamental frequency of old items used in the study phase. In both an auditory identification task and a stem completion task, Schacter and Church found that study-to-test changes in all three of these stimulus attributes yielded significant reductions in subjects' accuracy.

The impairment in performance observed in the implicit memory tasks from this experiment is particularly interesting because performance on explicit recall and recognition tasks showed little, if any, effects of study-to-test changes. Similar effects had previously been observed by Church and Schacter (1994) for word identification when stimuli were presented in white noise and for stem completion tasks when stimuli were presented in the clear. As with the Goldinger experiment reported above, the findings of Schacter and Church show that extra-linguistic variation in speech is encoded and retained in memory for speech events and is an important enough component of this representation to produce reliable implicit effects on the recognition of spoken words even when such tasks do not mention these extra-linguistic dimensions at the time of initial encoding or even call attention to these attributes of the stimulus materials.

The long-term storage of extra-linguistic information about a talker's voice in implicit memory has also been demonstrated in a series of studies that examined the learning of unfamiliar voices (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1995). In one experiment, Nygaard and Pisoni (1998) trained participants to identify a set of novel talkers from their voices alone. Once the participants had learned the names of the voices using a set of training stimuli, Nygaard and Pisoni found that the knowledge of talker characteristics obtained under these conditions also generalized to new stimuli that were never used in training. More importantly, Nygaard and Pisoni found that the perceptual learning of the trained voices transferred to a novel task: words spoken by familiar voices were recognized in noise more accurately than words spoken by unfamiliar voices. Thus, performance on the transfer task was facilitated by prior experience with the voices of the talkers with whom the participants were trained. Because there was no explicit reference to previous episodes or to recognizing words during training, Nygaard and Pisoni's findings provide evidence for the implicit encoding and use of information about a talker's voice in an entirely different task - recognizing spoken words.

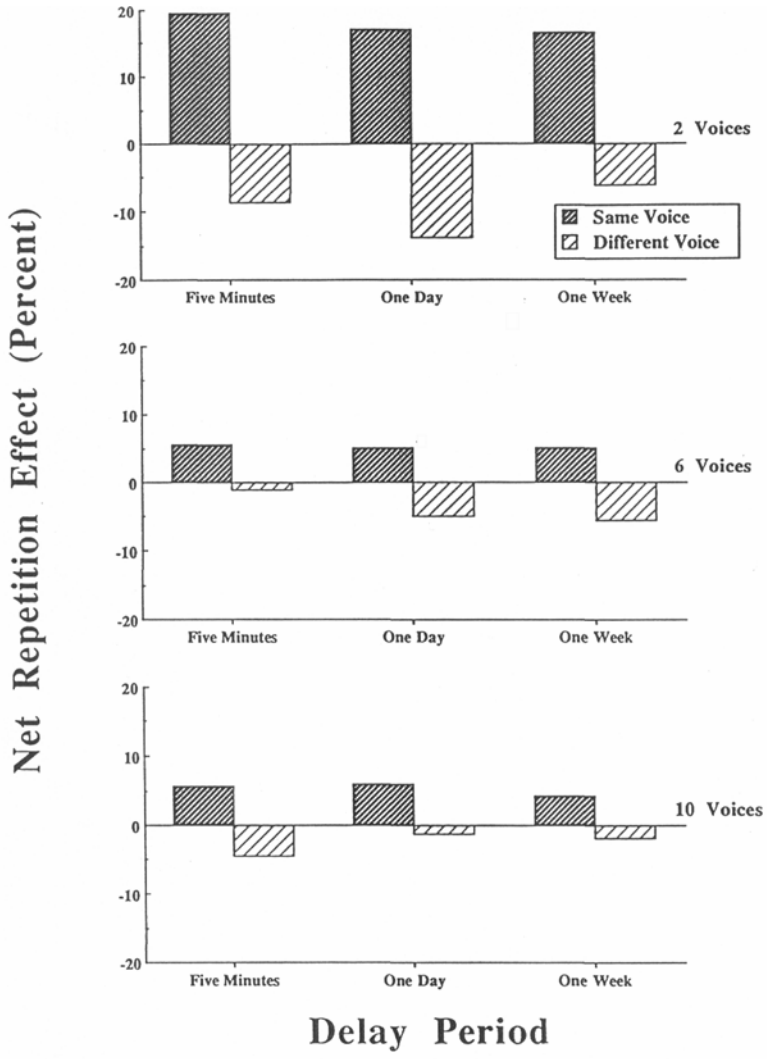


Figure 2. Net repetition effects observed in perceptual identification as a function of delay between sessions and repetition voice. The y-axis shows the difference in word identification accuracy for the original and subsequent presentation of the word. There was a benefit to word identification if the word was repeated in the same voice as it was presented with originally. This effect did not dissipate over time. Increasing the number of voices in the experiment decreased the effect somewhat, due to a decrease in the perceptual distinctiveness of the voices used (from Goldinger, 1992).

Further evidence for implicit encoding and storage of extra-linguistic variation in speech perception comes from a study on the learning of English [r] and [l] by native speakers of Japanese. In a series of perceptual learning experiments in which Japanese listeners were trained to recognize the English /l/ and /r/ distinction, Logan, Lively, and Pisoni (1991) showed implicit effects for variation in the original training sets. In a follow-up study, Lively, Pisoni, Yamada, Tohkura and Yamada (1992) found that the English /l/ and /r/ contrast was better retained by Japanese listeners when they were exposed to a large corpus of stimuli spoken by many different talkers during training. Compared to a group of listeners who had been trained using stimuli spoken by a single talker, listeners who had been trained using tokens

produced by multiple talkers were better able to distinguish /l/ and /r/ in the speech of entirely new speakers. Although not originally designed to study implicit memory effects, these perceptual learning results satisfy the standard definition of implicit memory since it is unlikely that subjects explicitly recalled their earlier training experience when required to identify novel speech samples. Moreover, the subjects were never explicitly told to attend to the different voices used in the training phases of the experiment. All they were required to do was categorize each word they heard as having an /r/ or /l/ in it.

Effects of Stimulus Variation on Explicit Memory

Although research findings on implicit memory for speech are limited, a variety of other experimental paradigms have uncovered effects that parallel the results of these implicit memory experiments. These experiments do not measure implicit memory in the standard sense laid out above. Although most of these experiments did require subjects to consciously recall their previous experiences, the results from these experiments are important because they demonstrate the same kinds of effects of encoding detail in speech memory that were revealed in implicit memory research.

In one study, Goldinger, Pisoni, and Logan (1991) examined the effects of talker variability on explicit memory for spoken words using a serial recall task. They manipulated the number of talkers used to create the stimulus lists and the rate at which items within a list were presented. They measured recall accuracy for items at the various serial positions in the list. The results showed that presentation rate interacted with the number of talkers used in the stimulus lists. At the fastest presentation rates, talker variability caused a decrease in accuracy across all serial positions in the list. Recall of single talker lists was better than recall of multiple talker lists. But, as the presentation rate decreased and the items were presented more slowly, however, the original pattern reversed. At the slowest presentation rates, subjects were more accurate at recalling lists of words spoken by multiple talkers than by a single talker, especially in the earlier portions of the list. Goldinger et al. concluded that information about voices must be incidentally encoded in memory at the time of presentation. At faster rates, this incidental encoding of voice features interferes with the perceptual encoding of items, leading to lower recall performance. At slower presentation rates, however, multiple talker lists contain additional distinctive cues that can be used to retrieve items from memory, thus yielding higher recall scores at test.

In another experiment that examined the encoding of extra-linguistic information in memory, Palmeri, Goldinger, and Pisoni (1993) used a continuous recognition memory procedure in which subjects listened to long lists of words spoken by multiple talkers. In this recognition task (see Shepard & Teghtsoonian, 1961), participants are required to judge each stimulus item in a list as ‘old’, if they have previously experienced the stimulus in the list, or ‘new’, if they have not. By varying the lag between the initial stimulus and its subsequent presentation, the effects of time and decay can be measured. In their experiment, Palmeri, et al. added a variant to the standard recognition memory paradigm by repeating old words in either the same voice or in a different voice from the initial presentation.

Palmeri et al.’s results were consistent with the findings we have reported thus far. Subjects showed the highest recognition accuracy for words that were presented in a repeated voice. Interestingly, subjects also showed the worst performance when talkers of a different gender repeated the words, indicating that highly dissimilar voices (as in cross-gender talker changes) were unable to function as reliable ‘cues’ to recognition of the words. The effects of lag between study and test in this experiment were surprising. As expected, recognition accuracy decreased overall with increasing lags between initial and subsequent presentations of the stimulus. However, the advantage for ‘same voice’ repetition did not interact with the lag between initial and subsequent presentations of an item. In other words, the encoding of extra-linguistic information facilitated the recognition of words regardless of the time between the initial encoding of the word and test. This pattern of results indicates that extra-linguistic information is

preserved in memory to the same extent that linguistic information is preserved. Although the memory trace for a word may decay over time, many of the fine details of the memory representation are not lost over time and can be used to facilitate subsequent recognition.

In another study, Lightfoot (1989) reported that subjects who had previously been trained to identify a set of talkers using common names showed better cued recall scores for lists of words when the words were spoken by multiple talkers compared to single talkers. Unlike the interaction observed by Goldinger et al. (1991), however, Lightfoot found that multiple talkers helped recall even at relatively fast presentation rates. Because the listeners in Lightfoot's experiment had been explicitly trained for several days to learn the voices of the talkers to a criterion beforehand, they were more familiar with these voices than participants in Goldinger's experiment. Both experiments provide support for the same conclusion. Detailed information about the talker's voice is encoded in memory along with the more abstract, symbolic linguistic content of the signal, and these instance-specific attributes facilitate the later recall and recognition of spoken words.

Explicit memory research using sentences has also revealed effects that suggest that detailed encoding of linguistic and extra-linguistic information is retained in memory. Geiselman and Bellezza (1977) presented one group of subjects with a set of sentences spoken by a male and a female talker. They also presented a control group with a set of sentences spoken only by the male talker or only by the female talker. Subjects were instructed either to attend only to the content of the sentences ('incidental gender encoding' condition) or to remember *both* the content and gender of the sentences ('explicit gender encoding' condition) for a subsequent memory test. Geiselman and Bellezza found that recall of the sentences was not significantly different for the experimental and control groups even though experimental subjects remembered the gender of sentences at higher than chance levels under both incidental and explicit gender encoding instructions.

Geiselman and Bellezza considered two possible explanations for how gender information could be encoded without detrimental effects on encoding of the linguistic content of the sentences. According to their 'voice-connotation' hypothesis, the meaning of a sentence may be encoded such that information about the speaker's voice is automatically encoded without increasing demands on processing resources. In contrast, their 'dual-hemisphere parallel-processing hypothesis' explained the encoding of gender information without increased processing costs by positing that both content and gender information are encoded in parallel by the left and right hemispheres of the brain. In subsequent research, Geiselman and Bellezza replicated their initial results (unpublished experiment mentioned in Geiselman & Bellezza, 1977) and found support for the voice-connotation hypothesis: the 'voice attribute is not 'attached' to the code for the item in memory. Rather, it may become an integral part of the code itself…' (Geiselman & Bellezza, 1977). These findings are important because they show that the composite encoding in memory of linguistic and extra-linguistic information is not constrained to isolated spoken words, but generalizes to larger linguistic units, like sentences.

Recently, McMichael and Pisoni (2000) obtained additional evidence for implicit encoding and retention of voice information in sentence-length stimuli. In a series of four discrete recognition memory experiments, they presented listeners with a list of 40 sentences in a 'study' phase. In the 'Intentional encoding' conditions, subjects were specifically told that their memory for sentences would be tested following the study phase. In the 'Incidental encoding' conditions, subjects received a surprise recognition memory test. During the study phase, 5 male and 5 female talkers spoke the list of sentences. In the test phase, listeners were asked to judge a list of 80 sentences as 'old' (i.e., heard at the time of study) or 'new' (i.e., not heard at the time of study). The forty 'old' sentences were spoken by either the same talker used during study ('Repeated Voice') or an entirely new talker that had not been heard during

the study phase ('Non-repeated Voice'). The forty 'new' sentences were also spoken by either a talker that had been heard during the study phase or by an entirely new talker.

McMichael and Pisoni also manipulated the encoding task during the study phase, in order to determine whether instructions focusing attention on voice attributes would affect recognition memory performance. In one task, subjects simply hit the 'enter' key on a keyboard after hearing each sentence. In the other task, subjects indicated the gender of the speaker by typing in 'm' or 'f' after each sentence was played. Both study tasks were run under either 'incidental encoding' instructions or 'intentional encoding' instructions, producing four combinations of instructions and study task.

Figure 3 shows the recognition memory results from these experiments. Each set of four columns within a panel represents the probability of a correct response for the four different types of sentences at the time of test. The pattern of results shows consistent repetition effects based on the voice of the talker. Sentences were recognized more accurately as 'old' when they were presented at test in the same voice that was used at study ('Old/Repeated') than when they were presented in a different voice ('Old/Non-repeated'). This voice repetition effect was statistically significant across all four experiments, showing that even for sentences, voice information is encoded and stored in memory along with linguistic information.

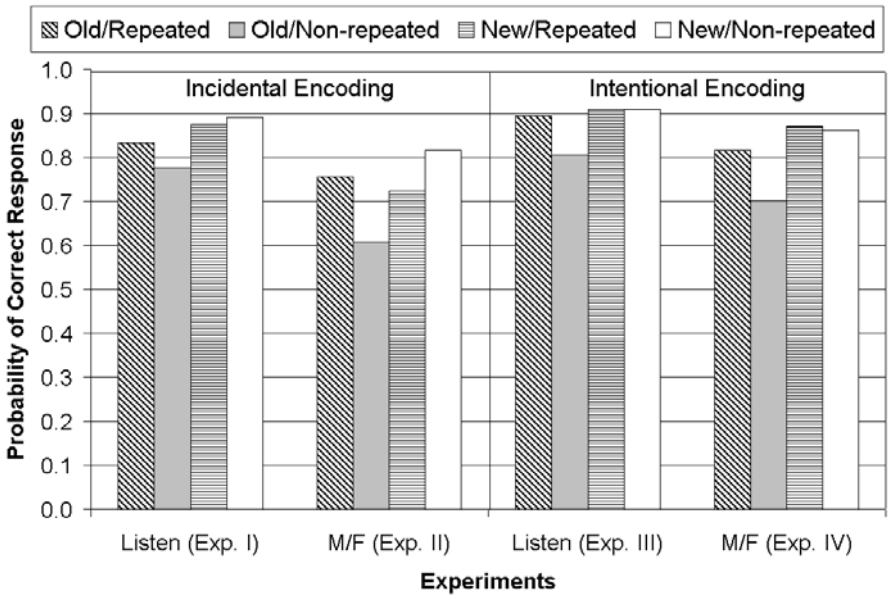


Figure 3. Probability of correct response from four recognition memory experiments using sentences. Each group of columns represents the results from a different experiment. In the 'Intentional Encoding' experiments, participants were instructed that there would be a test of recognition memory following the 'study' task. In 'Incidental Encoding' experiments, participants received a surprise recognition test after study. The four columns associated with each experiment show the various conditions under which the sentences were presented. 'Old' sentences are items that were presented during the study phase; 'New' sentences were not presented during the study phase. 'Repeated' voices were voices used during the study phase; 'Non-repeated' voices were novel voices that were not presented during study. The results show that across all four experiments, old sentences were more accurately identified as old when they were repeated in the same voice as original presentation than when they were presented in a new voice (adapted from McMichael & Pisoni, 2000).

Surprisingly, new sentences—those that were never presented during study—also showed a voice repetition effect in one of the four experiments. Under the "incidental encoding" instructions, when the study task involved identifying the gender of the speaker in Experiment II, new sentences spoken in non-repeated voices ('New/Non-repeated') were more accurately identified as new than new sentences spoken by repeated voices ('New/Repeated'). These recognition memory results suggest that listeners did encode information about the specific attributes of the talkers used at the time of study—even when the instructions used at the time of study never mentioned there would be a subsequent recognition memory test for the sentences.

These results indicate that specific details regarding the voice of the talker were encoded at the time of study and this information was later retrieved and utilized in the test phase recognition task. Additionally, the fact that these effects for voice information were observed for both sets of instructions - one that explicitly focused attention on the talker's voice at study, and one that did not - demonstrates that these recognition memory effects were not simply a result of instructions that encouraged voice information encoding. Even without the explicit intention of the listeners to encode voice information, details of the talker's voice were encoded that were sufficient to facilitate performance in the recognition memory task. Thus, the earlier results obtained with isolated words clearly generalized to sentence length materials as well.

The effects of stimulus variation uncovered in these experiments, although obtained under what would traditionally be called explicit memory paradigms, are notable because subjects were never explicitly instructed to pay attention to the talker's voice during study or test. Although subjects were required to explicitly recall or recognize the words or sentences they heard, they might not have been consciously aware of incidental variation in voice information while performing these tasks. To the extent that specific details of the original speech events affected performance in these explicit memory tasks, the results have a clear and direct relationship with the implicit memory effects summarized above.

The range of stimuli investigated in these experiments runs from isolated phonemes to words to sentence length speech samples. The fact that comparable effects of extra-linguistic variation are observed across these different stimulus materials suggests that similar perceptual and memory processes may be involved in the encoding and storage of phonemes, words and sentences. The similarity of the effects of talker variability in these experiments suggests a close link between implicit and explicit memory and raises important theoretical issues about how to explain and account for the pattern of these findings in a unified and coherent fashion.

Some Final Thoughts About Implicit Memory and Detailed Encoding

Although the history of research on speech perception has been dominated by abstractionist, information-processing approaches that consider extra-linguistic information irrelevant to the primary task of uncovering the idealized linguistic signal from beneath a wide range of noisy transformations, an emerging line of research suggests that extra-linguistic information and variation actually plays an important role in the process of speech perception. Several results of this line of research are particularly important to emphasize here:

Scale Invariance. Whether the experiment used phonemes, words or sentences, similar effects of variation in linguistic and extra-linguistic information have been obtained with units of differing lengths. The similarity of these findings suggests that all levels of speech representation may rely on a common substrate that incorporates both linguistic and extra-linguistic information. If the mental lexicon is conceived of as an abstract word-store that encodes only word 'types' and not word 'tokens', then current accounts of lexical memory will have great difficulty in explaining why units of differing length show

effects of extra-linguistic variation. Effects for stimuli that are shorter and longer than words are difficult to explain since it is unclear how an abstracted store of word information could generate episodic effects for phoneme or sentence length stimuli if the fine instance-specific details of speech events are lost or discarded from memory at the time of initial encoding via the process of normalization.

Rather, a conceptualization of the mental lexicon as an integrated, functionally-identical subsection of a larger, more general memory system, in which all experience with speech is encoded and preserved in a detail-rich, complex and multidimensional representation, seems more appropriate as a way to account for these results.

Parallel Transmission. In contrast to the traditional view of speech, in which linguistic and extra-linguistic sources of information were viewed as separate components of the speech signal (Abercrombie, 1967), the research summarized in this chapter suggests that these two sets of attributes may be inseparable. In both speech perception and memory tasks, subjects are consistently affected by variation in both sources of information even when they are not explicitly instructed to attend to one set of attributes or the other.

It is important to keep in mind that the dissociation between linguistic and extra-linguistic information in speech is arbitrary and has been handed down to speech scientists from pre-existing meta-theoretical notions inherited from the study of linguistics, where human performance had been explicitly ruled as irrelevant to the study of language by Chomsky's competence/performance distinction (Chomsky, 1965).

The finding that human listeners encode and retrieve both linguistic and extra-linguistic information is not surprising—after all, how else could we learn to recognize and identify the voices of our friends, or the slight nuances of affect that allow us to negotiate the complex rules of pragmatic discourse, unless we encode and retain very detailed extra-linguistic information in memory. It is precisely the inseparable relationship between linguistic and extra-linguistic information that is important - that is, variation in linguistic and extra-linguistic information is not simply a helpful source of information when listeners happen to have access to it. It is rather an integral part of understanding and remembering the meaning and intent of speech events. Variation in speech is so important that even when listeners are explicitly instructed to ignore differences in linguistic or extra-linguistic information, their performance in speech perception and memory tasks appears to be influenced by all aspects of the original signal, including attributes not relevant to the specific task at hand. The processing of extra-linguistic detail without conscious awareness sounds much like the obligatory or mandatory processing needed for module status under Fodor's modularity hypothesis (Fodor, 1981). However, we do not wish to imply that speech processing is undertaken by a specialized module. Rather, we think it more reasonable to claim that the conjoint processing of linguistic and extra-linguistic attributes follows naturally from their simultaneous and inextricable production by the vocal articulators.

Parallels in Explicit and Implicit Memory. In both implicit and explicit memory paradigms, similar effects of stimulus variation in the speech signal have been obtained across a variety of tasks. Utterances spoken in the same voice as earlier presentations increase accuracy in explicit memory for words and reduce response latencies in implicit tasks such as word identification.

That the same variation in the original speech signal can have parallel effects in both explicit and implicit tasks suggests that these two memory systems rely on the same types of representations for speech events. These representations are not based on abstract, idealized, contrastive linguistic units, but rather carry with them detailed episodic, instance-specific information about the circumstances of vocal articulation that produce speech. Furthermore, the similarity of these effects suggests that the traditional

separation of these two types of memory may not be a valid conceptualization of memory for spoken language.

A detailed encoding, or 'exemplar,' perspective provides an alternative view that can account for findings. With regard to scale invariance, the specific, rich detail with which speech events are retained in memory preserves information about the dialect, gender or other indexical properties of the talker at any scale, whether the units are phonemes, words, sentences, or even units like discourse segments. Just as speech is perceived and produced in a consistent manner across scales from words to extended discourse, the memory representation for speech may be similar across different sized chunks of speech. Whether these units are phonemes or sentences, the detail of these speech events in memory would allow for the observed effects of stimulus variability.

The approach advocated here is consistent with a composite form of mental representations for speech. Since speech is produced and perceived as a unitary event that carries both linguistic and extra-linguistic information, the composite representation of a detailed memory representation falls out naturally from the physics of speech motor control and behavior. The intended message of a speaker preserves a form of parity with the production of that message via vocal articulation. Rather than being a source of noise, however, the complex interaction of the speech articulators lawfully varies the speech signal in ways that are informative and distinctive to the listener. We need not posit that different 'entries' for information about linguistic content, gender, dialect and affect are stored in a complex associative memory system. Rather, the fact that this information arrives encoded and packaged in a unitary speech signal provides a *de facto* explanation of its storage together in speech memory. Since our memories for speech events are integrated, unitary composites of both linguistic and extra-linguistic information, behavioral tasks that assess these memories may also be affected by the rich, redundant information stored therein. Just as the detailed encoding perspective questions the validity of a distinction between linguistic and extra-linguistic information, so too does this perspective challenge the distinction between implicit and explicit memory.

For the purposes of speech research, implicit and explicit memory have largely been distinguished based on the kinds of speech information relevant to each memory system. Extra-linguistic information such as speaking rate or gender has been the traditional focus of implicit memory for speech experiments (Church & Schacter, 1994; Goldinger, 1992; Schacter & Church, 1992) while explicit memory research has focused on the more abstract, idealized linguistic information such as phonemes or words (Liberman, 1957; Peters, 1955). The result of this divergence of research has been the tacit assumption that implicit and explicit memory for speech events reflect the operation of functionally distinct, separate memory systems that deal with different types of speech information.

In a detailed encoding perspective, however, there is no valid distinction between linguistic and extra-linguistic information. Without this information-based distinction, the difference between implicit and explicit memory for speech events begins to blur. If the same memory representation underlies behavior in both implicit and explicit memory tasks and if behavior in these tasks shows similar effects of variation in the speech signal, then we can rightly question whether these two memory systems are, in fact, separate and distinct.

To apprehend the meaning of a given speech event and to recover the talker's intended message, it is necessary for the listener to know who is speaking, what they said, and under what conditions the articulatory events that produced an utterance occurred. The traditional perspective on speech perception, as well as the accepted distinction between implicit and explicit memory, assumes that the information in these representations is processed, stored, and accessed separately. The episodic view of speech perception, which is intimately tied to a description of the underlying events and their consequences,

takes the integration of this information as an important constraint on the way speech events are processed and stored in memory. This approach incorporates both implicit and explicit memory phenomena as reflecting aspects of a single complex memory system that retains highly detailed, instance-specific information in a perceptual record containing all of our experiences—speech and otherwise.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, IL: Aldine Publishing Company.

Bowers, J.S. (2000). In defense of abstractionist theories repetition priming and word identification. *Psychonomic Bulletin & Review, 7*, 83 - 99.

Bradlow, A.R., Torretta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255 - 273.

Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.

Church, B.A., & Schacter, D.L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 521 - 533.

Cole, R.A., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology, 26*, 1 - 7.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29*, 655.

Fodor, J.A. (1981). The mind-body problem. *Scientific American, 224*.

Fourcin, A.J. (1968). Speech-source interference. *IEEE Transactions in Audio Electroacoustics, ACC-16*, 65 - 67.

Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3 - 28.

Garner, W.R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gaver, W.W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*(1), 1 - 29.

Geiselman, R.E., & Bellezza, F.S. (1977). Incidental retention of speaker's voice. *Memory and Cognition, 5*, 658 - 665.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.

Goldinger, S.D. (1992). Words and voices: Implicit and explicit memory for spoken words, *Research on Speech Perception Technical Report No. 7* . Bloomington, IN: Indiana University.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251 - 279.

Goldinger, S.D., Pisoni, D.B., & Logan, J.S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 152-162.

Halle, M. (1956). *For Roman Jakobson: essays on the occasion of his sixtieth birthday, 11, Oct 1956*. The Hague: Mouton.

Hirahara, T., & Kato, H. (1992). The effects of F0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 89 - 112). Tokyo: Ohmsha Publishing.

Klatt, D.H. (1989). Review of selected models of speech perception. In W.D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 201 - 262). Cambridge, MA: MIT Press.

Liberman, A.M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America, 29*, 117 - 123.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431 - 461.

Licklider, J.C.R. (1952). On the process of speech perception. *Journal of the Acoustical Society of America, 24*, 590 - 594.

Lightfoot, N. (1989). Effects of talker familiarity on serial recall of spoken word lists, *Research on Speech Perception, Progress Report No. 15*. Bloomington, IN: Indiana University.

Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., & Yamada, T. (1992). Training Japanese listeners to identify English [r] and [l]: III. Long-term retention of the new phonetic categories, *Research on Speech Perception Progress Report No. 18* (pp. 185-216). Bloomington, IN: Indiana University.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify the English [r] and [l]: A first report. *Journal of the Acoustical Society of America, 89*, 874 - 886.

McMichael, K., & Pisoni, D.B. (2000). Talker-specific encoding effects on recognition memory for spoken sentences. Manuscript in preparation.

Mullennix, J.W., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*, 379-390.

Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology : General, 115*, 39 - 57.

Nosofsky, R.M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 87-108.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355 - 376.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989 - 1001.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309 - 328.

Peters, R.W. (1955). *The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise* (Joint Project Report No. 56). Pensacola, FL: U.S. Navel School of Aviation Medicine.

Pisoni, D.B. (1997). Some thoughts on "Normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego: Academic Press.

Remez, R.E., Fellowes, J.M., & Rubin, P.E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 651 - 666.

Remez, R.E., Rubin, P.E., Pisoni, D.B., & Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947 - 950.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501 - 518.

Schacter, D.L., & Church, B.A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 915-930.

Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology, 62*, 302 - 309.

Studdert-Kennedy, M. (1974). The perception of speech. In T.A. Sebeok (Ed.), *Current trends in linguistics* (Vol. XII, pp. 2349 - 2385). The Hague: Mouton.

Studdert-Kennedy, M. (1976). Speech perception. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243 - 293). New York: Academic Press.

Verbrugge, R.R., Strange, W., Shankweiler, D.P., & Edman, T.R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60*, 198-212.

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Effects of Speaking Style on the Perceptual Learning of Novel Voices: A First Report[1]

James D. Harnsberger, Richard Wright, and David B. Pisoni

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant DC00012 and NIH-NIDCD Research Grant R01-DC00111 to Indiana University. We would like to thank Kipp McMichael for his technical assistance.

Effects of Speaking Style on the Perceptual Learning of Novel Voices: A First Report

Abstract. This study examined the effects of speaking style on the perceptual learning of novel voices in the laboratory. Listeners participated in a voice learning experiment. In the training phase, listeners were asked to learn the names of either seven male or seven female talkers from samples of citation or hyperarticulated speech. In the test phase, listeners were presented with the same stimuli as in the training phase and were asked to identify the talker, with no feedback. In the sentence generalization phase, listeners were asked to identify the same voices producing new sentences in the same speaking style as that used in the previous phases. In the speaking style generalization phase, listeners were asked to identify the same voices in either the same speaking style as the previous phases or in a novel speaking style. The results showed that female voices were easier to learn in a hyperarticulated speaking style relative to a citation speaking style in the training and test phases. For the male voices, no such effect was observed. In addition, voice identification scores increased from the training to the test phase. However, voice identification scores did not improve in subsequent phases, which lacked the feedback provided during the training phase. In the style generalization phase, training with the female voice hyperarticulated tokens provided a greater advantage in identifying voices in a novel style relative to training with female voice citation tokens. No such effect was observed for listeners trained with the male voices. This gender interaction was further explored in a similarity scaling experiment, using stimuli from the first experiment. Listeners were presented with pairs of stimuli that differed in talker but matched in speaking style and gender. The citation sentences of both male and female talkers were rated as significantly more similar than the hyperarticulated sentences. However, for a subset of the stimuli, the difference in mean similarity for the female citation and hyperarticulated sentences was significantly greater than the corresponding difference in the male voices, indicating that the female talkers may have produced a more perceptually distinct hyperarticulated style than the male talkers. These differences may have contributed to the gender effect observed in learning in the first experiment. Taken together, the results of both experiments show that speaking style exerts a strong influence on the learning of novel voices, but its exact role is unclear given the interaction of speaking style and gender of the talker.

Introduction

Prior work on voice perception has shown that numerous sources of variability in the stimulus materials and in the listeners tested affect the encoding and retention of voice information in long-term memory (Kreiman, 1997). In terms of stimulus characteristics, voice perception has been shown to be influenced by the phonetic characteristics of the stimulus materials (see Bricker & Pruzansky, 1976 for a review), the filtering of glottal or vocal tract characteristics (Kubawara & Takagi, 1991; Tartter, 1991), and changes in the speaking rate or any signal distortion (Van Lancker, Kreiman, & Wickens, 1985). Several characteristics of listener groups have also been manipulated. For example, mismatches between the native dialect or language of the listeners and that of the talkers who produced the utterances have resulted in poorer voice recognition performance relative to controls that match in terms of linguistic background (Hollien, Majewski, & Doherty, 1982; Thompson, 1987; Goggin, Thompson, Strube, & Simental, 1991). Thus, linguistic information influences the voice recognition process, much like talker information has been shown to affect speech perception and spoken word recognition (Nygaard & Pisoni, 1998).

Individual listeners have been shown to vary widely in the specific acoustic cues they use to judge the relative similarity of voices (Hollien, Majewski, & Doherty, 1982; Kreiman, Gerratt, Precoda, & Berke, 1992). Listeners also vary in the strategies used to identify novel male and female voices that potentially reflect what some researchers have referred to as "cultural stereotypes" concerning male and female voices. Examples of such stereotypes include the association of "breathiness" with female voices and "hoarseness" with male voices (Singh & Murray, 1978), or associations between different distributions of vowel categories and gender that cannot be explained solely by differences in vocal tract size (Mattingly, 1966). In examining the voice perception literature overall, it is clear that the attributes of the voices used in experiments and the attributes associated with listeners' experience with talkers in their native language both play significant roles in the discrimination and recognition of novel voices.

While several sources of variability have been investigated in voice perception research, an important source of variation in speech production has been neglected in these earlier studies. Specifically, variation in speaking style has not been studied in any great depth. Listeners on a daily basis not only encounter a variety of talkers, including ones whose voices are unfamiliar to the listener, but they are also exposed to a wide variety of speaking styles. Speaking styles can vary along a continuum from casual to careful speech (Labov, 1970). Moreover, speaking styles are specific to a given situation or goal (e.g., performance speech; speech directed to authority figures; speech in noisy environments). Variation in speaking style results in significant changes in the acoustic characteristics of the speech signal that are linguistically relevant. This variation also provides information about the identity of the talker, through such cues as f0, formant frequencies, nasality, duration, and breathiness (Murray & Singh, 1980). Thus, it is possible that information about speaking style variation is encoded with voice information and, therefore, may influence the process of learning novel voices.

The goal of this study was to examine the effects of speaking style on the learning of novel voices. The speaking styles used in this study were *citation* speech, corresponding to read speech commonly elicited in the laboratory, and *hyperarticulated* speech, a style involving a high degree of articulatory precision produced by a talker who is attempting to speak clearly. Samples of citation style speech were obtained by simply having talkers read aloud sentences from a computer screen. Samples of hyperarticulated speech were elicited via the methods developed recently by Brink, Wright, and Pisoni (1998), in which talkers were asked to repeat a sentence more clearly after having produced it once before in a citation style. These two speaking styles were selected because in earlier pilot studies with utterances from five female voices, a significant increase was observed in the rate of learning of female voices from citation sentences compared to hyperarticulated sentences.

In the present study, we sought to extend this work to male voices, and to additional learning tasks. Specifically, we included a sentence generalization task (same voices, same style, novel sentences) and a speaking style generalization task (same voices, novel style, novel sentences). The sentence generalization task was added to determine whether the effects of speaking style were specific to the stimulus materials used in the training and test phases. The speaking style generalization task was administered to determine if either the citation or the hyperarticulated speaking style conveyed the most information that could be used by listeners in learning voices in novel styles.

For this study, two hypotheses were assessed concerning the effects of speaking style on novel voice learning. First, voices would be easier to learn in the hyperarticulated speaking style relative to the citation speaking style. This prediction was based on the pilot results, as well as prior research in speech perception and spoken word recognition in which hyperarticulated speech is typically found to be more intelligible and more informative in the identity of words than more reduced styles, including citation speech (Lindblom, 1990; Moon & Lindblom, 1994; Picheny, Durlach, & Braida, 1985; Picheny, Durlach,

& Braida, 1986). This hypothesis was termed the *clear speech hypothesis*. An alternative prediction was also entertained, that voices would be easier to learn in a citation rather than a hyperarticulated style. Voices in a citation speech style were hypothesized to be more distinctive because, under more casual speaking styles, idiosyncratic gestural strategies emerge that are otherwise suppressed by the use of stereotyped articulatory gestures in hyperarticulated speech. Such idiosyncrasies could be the product of the vocal anatomy of the talker, or could simply be an individual strategy to reach a common acoustic/auditory target (Johnson, Ladefoged, & Lindau, 1993). In either case, talker gestural idiosyncrasies could be available to the listener and constitute useful cues in learning to identify individual talkers. This hypothesis was termed the *idiosyncratic articulation hypothesis*.

Both hypotheses were tested in a voice learning task in which subjects were trained to identify novel male or female voices (Experiment 1). Subjects were trained on a set of sentences from each talker and tested on the same sentences. Subjects were also asked to identify the same voices producing novel sentences and, in some conditions, novel speaking styles. The role of speaking style in voice learning was further examined in a voice similarity task (Experiment 2), in which subjects were asked to rate on a seven-point scale the relative similarity of novel male and female voices producing sentences in a citation and hyperarticulated style. The voice similarity task was designed to measure the relative similarity of male and female voices in different speaking styles in an effort to account for patterns observed in the voice learning experiment.

EXPERIMENT 1

Methods

Participants

One hundred and six native speakers of American English, 52 females and 54 males ranging in age from 18 to 23 ($M = 20$), participated in this study. Participants received course credit for participating in a single one-hour test session. None of the listeners reported any history of a speech or hearing disorder at the time of testing.

Stimulus Materials

Recordings of sentences from 14 native speakers of American English, seven females and seven males ranging in age between 18 and 30, were used in this study. Participants used for creating the stimulus materials received $15 total compensation for participating in two one-hour sessions. None of the participants reported any history of a speech or hearing disorder at the time of testing. Participants were recorded reading 34 sentences chosen from the 200 sentences comprising the SPIN set (Kalikow, Stevens, & Elliot, 1977). The SPIN sentences are short sentences, five to eight words in length, ending in a high frequency monosyllabic noun (e.g., “The farmer harvested his crop”). The 34 SPIN sentences selected for this study are listed in Appendix A. The recording session took place in a sound-attenuated chamber (IAC Audiometric Testing Room, Model 402) using a head-mounted Shure (SM98) microphone positioned one inch away from the subject's chin. The recordings were digitized at 22.05 kHz (16 bit sampling) using a Tucker-Davis Technologies System II and stored on an IBM-PC 486 computer. The 34 sentences were produced using three speaking styles via a method used by Harnsberger and Pisoni (1999), a modified version of a method first developed by Brink, Wright, and Pisoni (1998). Those speaking styles were reduced (i.e., casual, hypoarticulated), citation (i.e., read speech), and hyperarticulated (i.e., careful speech). Of the 34 sentences in these three styles, 15 sentences in two styles, citation and hyperarticulated, were used in the present voice learning study. The fifteen sentences presented across all four phases appear in Appendix B.

Procedures

The participants for the voice learning experiment were randomly assigned to one of eight experimental conditions. Each condition consisted of four phases, in the following order: (1) a Training phase, (2) a Test phase, (3) a Sentence Generalization phase, and (4) a Speaking Style Generalization phase. The eight conditions differed in terms of the gender of the talkers whose stimuli were presented in the four phases (male or female), the speaking style of the stimuli presented in the first three phases, and the speaking style of the stimuli presented in the Speaking Style Generalization phase. All eight conditions are outlined in Table 1.

Gender	Style		Name of condition
	Phases 1 - 3	Phase 4	
Male	Citation	Citation	Male Cit-Cit
	Citation	Hyperarticulated	Male Cit-Hyp
	Hyperarticulated	Citation	Male Hyp-Cit
	Hyperarticulated	Hyperarticulated	Male Hyp-Hyp
Female	Citation	Citation	Female Cit-Cit
	Citation	Hyperarticulated	Female Cit-Hyp
	Hyperarticulated	Citation	Female Hyp-Cit
	Hyperarticulated	Hyperarticulated	Female Hyp-Hyp

Table 1. The eight conditions of Experiment 1.

In Table 1, *Gender* refers to the gender of the talkers who produced the stimuli in that condition. *Style* refers to the speaking style used by the talkers in that condition, in *Phases 1 - 3* of the experiment (i.e., Training, Test, and Sentence Generalization) and in *Phase 4* (Speaking Style Generalization). The *Name of condition* column lists the name assigned to each condition of the experiment. For example, listeners who learned male citation voices in all four phases of the experiment had been assigned to the "Male Cit-Cit" condition.

Training Phase. In this phase, participants were presented with a block of sentences and asked to identify the talker who produced each sentence. Two repetitions of five different sentences from each of the seven talkers in a given condition were presented in random order, for a total of 70 trials. Participants were asked to identify each voice by pressing one of seven buttons on a keyboard. Each button was labeled with a name (Ben, Greg, James, Kyle, Matt, Max, and Steve for the male voice sentences; Jenny, Kim, Lynn, Mary, Paula, Susie, and Trixie for the female voice sentences). After each response, the correct name of the talker appeared on the computer screen.

Test Phase. After participants had completed the Training phase, they were presented with one repetition of the same 35 sentences used in the Training phase. Participants were asked to identify each voice by pressing the appropriate button on the keyboard. In this phase, no feedback was provided to the participants.

Sentence Generalization Phase. Following the Test phase, participants were presented with five new sentences from the same seven talkers who produced the stimuli used in the Training and Test phases. As before, participants were asked to identify each voice by pressing the appropriate button on the keyboard. No feedback was provided.

Speaking Style Generalization Phase. Following the Sentence Generalization phase, participants were presented with five new sentences from the same seven talkers who produced the stimuli used in the previous three phases. The sentences used in this condition were produced in either the same speaking style used in the previous phases or in a novel speaking style, depending on which condition the individual participants were assigned to. Participants were asked to identify each voice by pressing the appropriate button on the keyboard. No feedback was provided.

Predictions

Two hypotheses concerning voice learning were entertained in this study: the *clear speech hypothesis* and the *idiosyncratic articulation hypothesis*. According to the *clear speech hypothesis*, voices should be easier to learn in a hyperarticulated style of speech than in a citation style because hyperarticulated sentences are produced with the more extreme articulatory gestures, resulting in longer segments, lengthy pauses between words, and an expanded vowel space. Thus, we expected significantly higher identification scores in the hyperarticulated speech conditions (Hyp-Cit, Hyp-Hyp) than in the citation speech conditions (Cit-Cit, Cit-Hyp) in the first three phases of the experiment. In the Speaking Style Generalization phase, experience with hyperarticulated speech was predicted to transfer more easily to learning voices in a novel style than experience with citation speech. Thus, the mean identification scores for participants in the Hyp-Cit conditions were predicted to be significantly higher than those of the participants in the Cit-Hyp conditions.

According to the *idiosyncratic articulation hypothesis*, less monitored and more naturally produced speaking styles should display talker-specific idiosyncratic speech production strategies that may provide more diverse cues to the identity of the talker. In this case, we expected to see significantly higher identification scores in the citation speech conditions relative to the hyperarticulated speech conditions. Moreover, in the Speaking Style Generalization phase, participants who learned to identify voices from citation speech should display an advantage in identifying voices in a novel speaking style relative to participants who learned from hyperarticulated speech. Thus, the mean identification scores for participants in the Cit-Hyp conditions should be significantly higher than those of the participants in the Hyp-Cit conditions.

Results

Training Phase

The proportion of correct responses to stimuli in the training phase of the experiment appear in Figure 1, listed by four conditions (Male Citation, Male Hyperarticulated, Female Citation, Female Hyperarticulated) rather than the full eight conditions due the lack of a significant three-way interaction between Gender, Speaking Style, and Speaking Style Generalization (see below). The four conditions were calculated by averaging together results from conditions that shared a common Gender and Speaking Style in the first three phases of the experiment (see Table 1). Thus, the results from the Female Cit-Cit and Female Cit-Hyp groups were combined in a Female Citation mean score. In a similar manner, Female Hyp-Cit and Female Hyp-Hyp; Male Cit-Cit and Male Cit-Hyp; and Male Hyp-Cit and Male Hyp-Hyp were combined in Female Hyperarticulated, Male Citation, and Male Hyperarticulated scores, respectively.

The results show an unanticipated effect of the gender of the talkers on voice learning across different speaking styles. Listeners who were assigned to the female voice hyperarticulated stimuli were more successful on average in learning new voices than the listeners assigned to the female voice citation

stimuli. In contrast, listeners assigned to the male voice citation stimuli were more successful in learning new voices than the listeners assigned to the male voice hyperarticulated stimuli.

The results of the training phase of the experiment were submitted to an ANOVA with three between-subjects factors: Gender (male, female); Speaking Style (citation, hyperarticulated); and Speaking Style Generalization, which refers to the stimulus set presented in the Speaking Style Generalization condition (citation, hyperarticulated). None of the main effects were significant. Of the two- and three-way interactions, only the Gender by Speaking Style interaction was significant (F (1, 98) = 4.97, $p < .05$). A simple effects analysis showed that the effect of Speaking Style was significant within the female voice sets but not the male voice sets (Female: F (1, 98) = 5.65, $p < .05$; Male: F (1, 98) = 0.7, $p = .4$). A simple effects analysis of Gender within Speaking Style showed no significant differences.

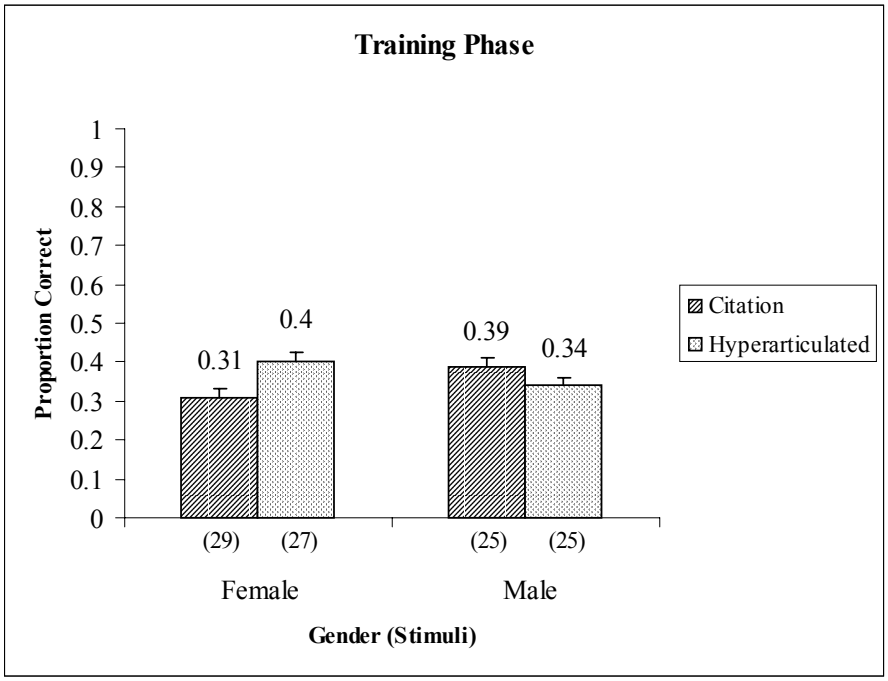


Figure 1. The mean proportion correct responses in the training phase for the four groups assigned to four different stimulus sets (Female: Citation, Hyperarticulated; Male: Citation, Hyperarticulated). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

Test Phase

Figure 2 shows the results from the test phase of the experiment. As in the training phase, the three-way interaction between Gender, Speaking Style, and Speaking Style Generalization was not significant. Once again, the results are reported in Figure 2 in four conditions, averaging over differences in Speaking Style Generalization. As in the training phase, a gender effect was observed in the identification scores for the test phase. Listeners who were assigned to the female voice hyperarticulated stimuli were more successful on average in learning new voices than the listeners assigned to the female voice citation stimuli. In contrast, listeners assigned to the male voice citation stimuli were more successful in learning new voices than the listeners assigned to the male voice hyperarticulated stimuli.

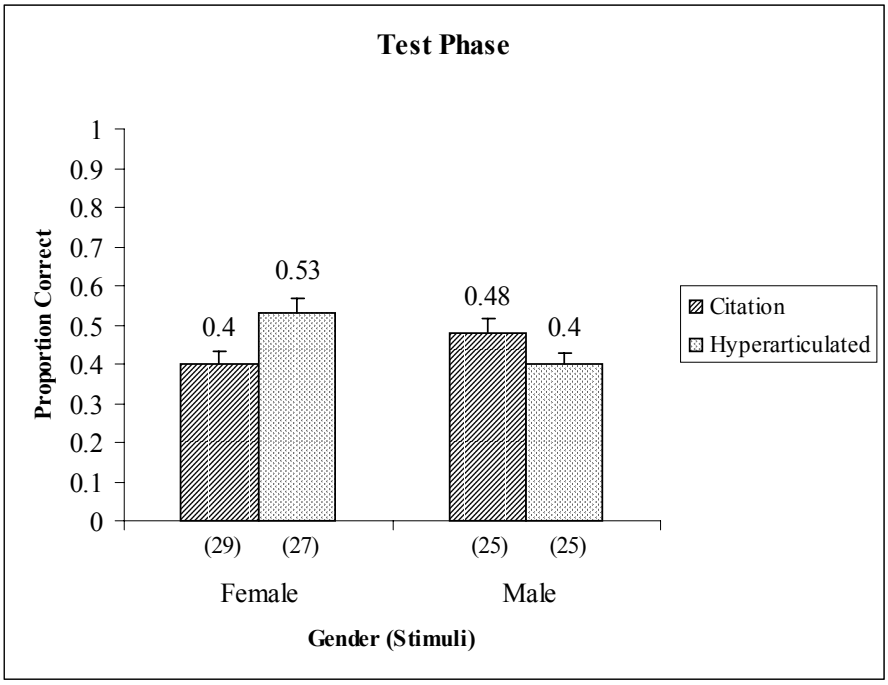


Figure 2. The mean proportion correct responses in the test phase for the four groups assigned to four different stimulus sets (Female: Citation, Hyperarticulated; Male: Citation, Hyperarticulated). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

The results of the test phase of the experiment were submitted to an ANOVA with three between-subjects factors: Gender (male, female), Speaking Style (citation, hyperarticulated), and Speaking Style Generalization (citation, hyperarticulated). None of the main effects were significant. Of the two-way interactions, only the Gender by Speaking Style interaction was significant (F (1, 98) = 8.47, $p < .01$). A simple effects analysis showed that the effect of Speaking Style was significant within the female voice sets but only marginally significant in the male voice sets (Female: F (1, 98) = 5.68, $p < .05$; Male: F (1, 98) = 3.09, $p = .08$). A simple effects analysis of Gender within Speaking Style also showed that the effect of Gender was significant within the hyperarticulated style sets but not the citation style sets (Hyperarticulated: F (1, 98) = 6.22, $p < .05$; Citation: F (1, 98) = 2.61, $p = .11$).

The results from the training and the test phases were also compared to determine if training resulted in significant improvement in the learning of novel voices. In a four-way ANOVA with Phase as a within-subject factor (training, test) and Gender, Speaking Style, and Speaking Style Generalization as between-subject factors, the main effect of Phase (F (1, 98) = 66.23, $p < .001$) and the interaction of Phase by Gender by Speaking Style (F (1, 98) = 4.29, $p < .05$) were significant. No other main effects or interactions were significant.

The significant effect of Phase showed that listeners did improve overall in the ability to identify novel voices. The mean proportion correct scores for the training and test phases were 0.36 and 0.46, respectively. The significant three-way interaction was further explored using simple effects analyses. The results of these analyses showed that all four listener groups (Female Citation, Female Hyperarticulated, Male Citation, and Male Hyperarticulated) improved significantly ($p < .01$ in all cases) from the training phase to the test phase.

Sentence Generalization Phase

Figure 3 displays the results of the sentence generalization phase of the experiment. As in the previous phases, listeners in the Female Hyperarticulated condition achieved higher identification scores than listeners in the Female Citation condition. The pattern for the male voice conditions showed the opposite effect of gender and speaking style. In a three-way ANOVA, none of the main effects proved to be significant. Of the interactions, only the Gender by Speaking Style interaction was significant (F (1, 98) = 4.01, $p < .05$). Given that the three-way interaction was not significant, the results in Figure 3 combine the groups that differ only in Speaking Style Generalization. The significant two-way interaction was assessed using simple effects analyses. Unlike the results of the previous phases, none of the tests of Speaking Style within Gender or Gender within Speaking Style were significant, although the difference between Female Hyperarticulated and Male Hyperarticulated approached significance (F (1, 98) = 3.57, p = .06).

The results from the sentence generalization and the test phases were also compared to determine if the performance of the subjects improved with additional exposure to the voices, without the feedback available in the training phase. The results of a four-way ANOVA with Phase as a within-subject factor (sentence generalization, test) and Gender, Speaking Style, and Speaking Style Generalization as between-subject factors showed that none of the main effects or interactions were significant. The results of this analysis revealed that the listener groups plateaued in their capacity to identify novel voices. Without additional feedback, simple exposure to the stimuli in the sentence generalization trials did not improve voice learning.

Speaking Style Generalization Phase

Figure 4 shows the results of the speaking style generalization phase of the experiment. The results are displayed in terms of the eight groups to which the listeners were assigned (Female: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp; Male: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp). Overall, for both the male and female stimulus sets, identification scores were highest when listeners were presented with the same speaking style in the speaking style generalization phase that they were exposed to in the preceding phases (i.e., listeners in the Male and Female Cit-Cit and Hyp-Hyp conditions). When listeners were presented with stimuli in a speaking style that differed from that used in preceding phases, identification scores were lower. Moreover, these lower scores varied by condition. Listeners who had learned voices from Female Hyperarticulated stimuli had higher identification scores in a novel speaking style (i.e., Citation) than listeners who had learned voices from Female Citation stimuli and were tested with Female Hyperarticulated stimuli in the speaking style generalization phase. An opposite pattern was observed in male voice stimuli: listeners in the Male Cit-Hyp condition had slightly higher identification scores than listeners in the Male Hyp-Cit condition. Thus, the hyperarticulated speaking style was the most informative style for learning female voices and generalizing to voices in novel speaking styles, while the citation speaking style was the most informative for male voices.

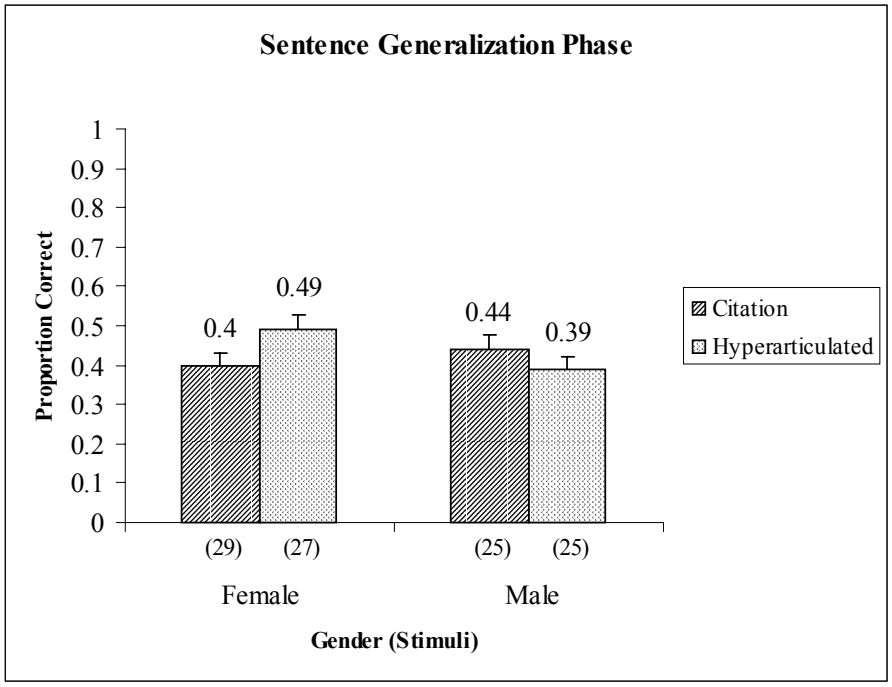


Figure 3. The mean proportion correct responses in the sentence generalization phase listed by four groups assigned to four different stimulus sets (Female: Citation, Hyperarticulated; Male: Citation, Hyperarticulated). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

A three-way ANOVA of the speaking style generalization results showed that none of the main effects of Gender, Style, or Speaking Style Generalization were significant. Of the two-way interactions, Gender by Style (F (1, 98) = 5.22, $p < .05$) and Style by Speaking Style Generalization (F (1, 98) = 10.9, $p < .001$) were significant. A simple effects analysis of the Gender by Style interaction showed that listeners who had learned female hyperarticulated voices in the previous phases were significantly better in identifying female voices than listeners who had learned female citation voices in the previous phases. Listeners who had learned female hyperarticulated voices in previous phases averaged 0.48 proportion correct on female voices in this phase, as compared with 0.40 proportion correct for listeners who learned female voices in a citation style in previous phases.

A simple effects analysis of the Speaking Style by Speaking Style Generalization interaction showed that, in several cases, performance in voice learning decreased when listeners were asked to identify voices in a novel speaking style. Listeners who learned citation voices in previous phases were significantly better in identifying citation voices in the speaking style generalization phase (i.e., Cit-Cit listeners) than listeners who learned hyperarticulated voices in the previous phases (Cit-Hyp listeners) (F (1, 98) = 4.9, $p < .05$). Cit-Cit and Cit-Hyp listeners averaged 0.49 and 0.35 proportion correct, respectively. Cit-Cit listeners also outperformed listeners who learned hyperarticulated voices and were presented with citation voices in this phase (Hyp-Cit listeners, who averaged 0.38 proportion correct) (F (1, 98) = 9.29, $p < .01$). Finally, Hyp-Hyp listeners identified voices significantly better than Hyp-Cit listeners, averaging 0.46 and 0.38 proportion correct, respectively (F (1, 98) = 6.05, $p < .05$).

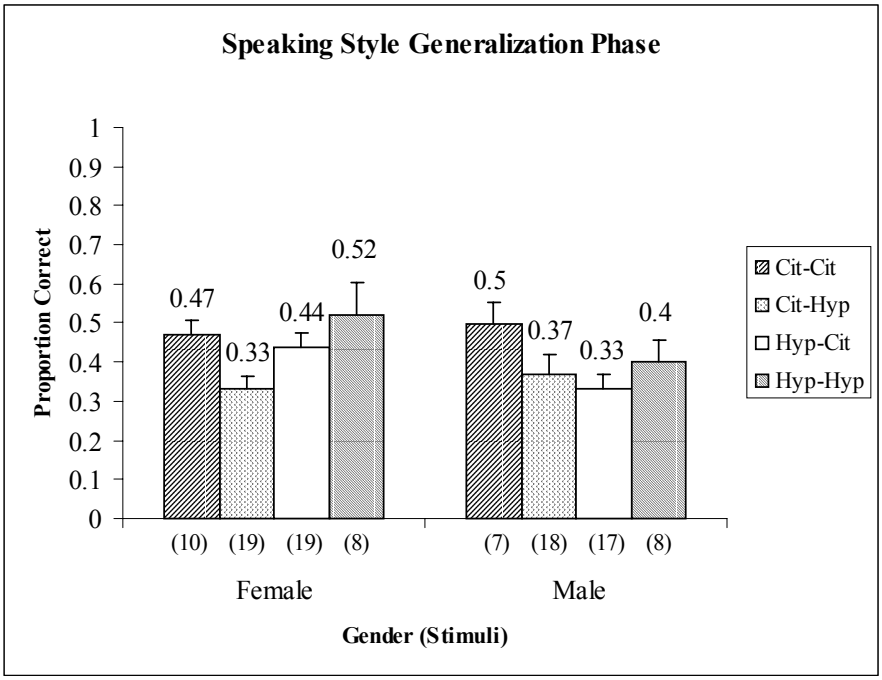


Figure 4. The mean proportion correct responses in the speaking style generalization phase listed by eight groups assigned to eight different stimulus sets (Female: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp; Male: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

The results from the speaking style and sentence generalization phases were also compared to determine if the performance of the subjects improved with additional exposure to the voices. A four-way ANOVA with Phase as a within-subject factor (speaking style generalization, sentence generalization) and Gender, Style, and Speaking Style Generalization as between-subject factors showed only a significant effect of Phase (F (1, 98) = 5.44, $p < .05$). Performance on the speaking style generalization phase was significantly lower than performance in the sentence generalization phase, reflecting the poorer scores of listeners who identified voices in a novel speaking style.

Discussion

In the first experiment, we tested two hypotheses concerning the effects of speaking style on learning novel voices. One hypothesis predicted that participants would be more successful in learning voices from citation speech than hyperarticulated speech because of the absence of talker-specific gestural strategies that are suppressed when producing hyperarticulated speech. The second hypothesis made the opposite prediction. The results showed an unanticipated gender effect that makes evaluating hypotheses about speaking style difficult because the effects are influenced by gender.

With the seven female voices, the results in all training and test phases of the experiment supported the *clear speech hypothesis*. We found that it was easier to learn female voices in a hyperarticulated speaking style than female voices in a citation speaking style in the training and test phases. Learning from a female voices in a hyperarticulated speaking style also generalized more readily to identifying the same voices in a novel speaking style than learning from female voices in a citation speaking style. No significant effects of speaking style were found for the male voices, although there

was a trend for male voices in a citation speaking style to be easier to learn than male voices in a hyperarticulated speaking style. In addition, for the listeners who learned female voices, exposure to hyperarticulated voices in the training, test, and sentence generalization phases improved performance in identifying voices in a novel speech style relative to exposure to citation voices. Finally, across the four phases of the experiment, learning in general was enhanced by the feedback available in the training phase, but did not increase in subsequent phases.

In interpreting the results of the first experiment, it should be noted that some of the marginally significant or nonsignificant results with both the male and female voice sets might be a function of relatively low power. In the Female Cit-Cit, Female Hyp-Hyp, Male Cit-Cit, and Male Hyp-Hyp conditions, only seven to ten listeners were tested. In the original design of the experiment, we did not anticipate that gender would play a role in novel voice learning and we planned to combine data from both gender conditions. Thus, the results from the matching Female and Male conditions (e.g., Female Cit-Cit and Male Cit-Cit) were to be combined into analyses of style-based conditions (e.g. Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp). Using this approach, an adequate number of listeners were tested in the Female Cit-Cit, Female Hyp-Hyp, Male Cit-Cit, and Male Hyp-Hyp conditions. However, given the gender effect that was observed, a larger number of participants should be recruited in future studies in all conditions, to ensure that the lack of results is not due to inadequate power.

The results of the experiment demonstrate an effect of speaking style on learning new voices, although the effect of gender makes interpreting the results more complex. In formulating the two hypotheses, we assumed that in either the citation or the hyperarticulated styles, greater talker-specific detail would be available for learners to encode. According to the *idiosyncratic articulation hypothesis*, talkers in a less careful, less monitored speaking style would produce more idiosyncratic gestural strategies and show greater overall articulatory/acoustic variability, all of which would constitute important and useful cues in distinguishing different voices. According to the *clear speech hypothesis*, hyperarticulated speech constituted a clear, information-rich signal that delineates the entire gestural space used by the individual talker.

Both hypotheses assumed that the attributes of one of the two styles would provide more information for voice learning. With the gender effect observed here, if the attributes of the speaking style account for the ease of learning female voices in a hyperarticulated speaking style, then the male and female participants may have adopted quite different strategies when prompted to produce the two speaking styles. If the two gender groups adopted somewhat different strategies in producing the two speaking styles, then presumably we would see a gender effect on the perceptual judgments of phonetically trained and naïve listeners with these two speaking styles (see Harnsberger & Pisoni, 1999) and on the acoustic analysis of sentences produced in these speaking styles (see Harnsberger & Goshert, 2000). In fact, no gender effect was observed in either of these studies. In particular, Harnsberger and Goshert (2000) found consistent differences across all talkers regardless of gender between male and female talkers' citation and hyperarticulated speech. Citation speech, on average, had shorter keyword durations and shorter sentence durations than Hyperarticulated speech. Moreover, vowels in keywords produced in a Citation style were more centralized, resulting in a smaller vowel space than vowels in keywords produced in a Hyperarticulated style, for all talkers regardless of gender. The male and female talkers may differ in their production of the two styles in dimensions not measured by Harnsberger and Goshert, although the link between such dimensions and the rate of learning novel voices is unclear.

While an explanation for the gender effect focusing on the attributes of the stimuli would be the most satisfactory one, there are other possible explanations that focus on the listeners' prior experience with male and female voices, and how such experience might influence novel voice learning. The gender effect revealed here is reminiscent of a frequently reported dichotomy in male and female speaking styles.

Specifically, the characteristics associated with hyperarticulated speech (expanded vowel space, maintenance of consonant clusters) appear to occur more frequently in female speech, while male speech typically shows more instances of "reduction" phenomena, such as a compressed vowel space and consonant cluster reduction/deletion (Fischer, 1958; Trudgill, 1974; Byrd, 1994; Bradlow, Toretta, & Pisoni, 1996). This gender-based dichotomy has been observed in several acoustic-phonetic studies of "laboratory speech," sociolinguistic studies (using impressionistic coding), and in transcriptions of spontaneous speech corpora. For instance, Byrd (1994) examined gender-based and regional dialect patterns in the TIMIT database, a large sentence database incorporating 630 talkers who read a total of 2,342 sentences. She found gender differences in speaking rate; the release of sentence-final stops; and the frequency of occurrence of schwa, glottal stops, syllabic [n], voiceless schwa, and [h]. In all of these gender differences, male speech displayed the more reduced forms (e.g., faster speaking rate, less frequent release of final stops, more likely to use schwa).[2] Bradlow, Toretta, and Pisoni (1996) measured several acoustic-phonetic characteristics in their study of the correlates of intelligible speech in a multitalker database. They observed that female speech was significantly more intelligible than male speech, and that female speech and male speech differed in some, but not all, of the attributes associated with careful, clear speech, such as fundamental frequency range or the timing relationship between adjacent segments. Several studies have shown that women use phonological forms associated with more "standard" speech more frequently than men (Fischer, 1958; Trudgill, 1974).

Overall, along a continuum of casual to careful speech, the typical speaking style employed by men may be found closer to the casual end of the continuum than the typical speaking style employed by women. This trend may be related to the gender effect observed in this study. Our prior experience with male and female voices in different speaking styles may influence the encoding of novel male and female voices in long-term memory which, in turn, influences the learning of novel voices. In other words, listeners possess a greater familiarity with male reduced speech and female hyperarticulated speech than their corresponding counterparts, male hyperarticulated speech and female reduced speech. This familiarity implies that listeners may have a bias in extracting cues to identify individual male and female talkers in different styles. To account for the results of this study, this bias would have to be rather robust, and would presumably operate under a variety of noisy conditions as well as the clear condition in which the stimuli were presented in this study. To date, such a robust effect of prior experience with male and female voices has not been documented in other studies. However, this effect might be an example of a "cultural stereotype" that listeners have acquired for male and female speech (Singh & Murray, 1978).

Given the unexpected gender effect observed in Experiment 1, the role of speaking style and gender in the learning of novel voices was explored in a second study. The purpose of the second experiment was to test two possible explanations for the gender effect. First, the gender effect may reflect acoustic properties of the stimuli not measured by Harnsberger and Goshert (2000). The female and male talkers who produced the stimuli for Experiment 1 may have adopted somewhat different strategies in differentiating the citation from the hyperarticulated style, resulting in two stimulus sets (male and female voices) that differed in their relative similarity, and thus learnability. Alternatively, the gender effect may reflect the frequency with which listeners have encountered relatively "careful" female speech versus relatively "reduced" male speech. These two hypotheses can be distinguished by comparing male and female citation and hyperarticulated stimuli using perceptual similarity tests. In these tests, naïve listeners are asked to rate pairs of sentences in terms of their similarity. It is possible that male voices in a citation style are judged as equally similar to one another as male voices in a hyperarticulated style. In contrast, it is possible that female voices in a hyperarticulated style are judged to be less similar to one another than female voices in a citation style. Findings such as these would indicate that specific acoustic-phonetic

[2] Male and female talkers did not significantly differ in one phonetic process associated with reduced speech, specifically the palatalization of alveolar obstruents.

properties of the sentences not measured by Harnsberger and Goshert (2000) may be responsible for the gender effect observed in the voice learning experiment. Alternatively, if the results of such a similarity test fail to show any gender effects (as predicted by the acoustic analysis), then the patterns observed in the voice learning experiment may reflect prior experience with male and female speaking styles. To assess these hypotheses, we administered a scaling experiment to obtain measures of perceptual similarity for four sets of stimulus materials: female citation, female hyperarticulated, male citation, and male hyperarticulated speech

EXPERIMENT 2

Methods

Participants

Thirty-two native speakers of American English, 24 females and eight males ranging in age from 18 to 29 (M = 20), participated in this study. The listeners received either course credit or five dollars for participating in a single one-hour test session. None of the listeners reported any history of a speech or hearing disorder at the time of testing.

Stimulus Materials and Procedures

A subset of the stimulus materials used in Experiment 1 was presented to participants in this experiment. Specifically, the Citation and Hyperarticulated readings of two sentences, "The beer drinkers raised their mugs" and "I made the phone call from a booth," were selected. The seven male and seven female talkers' Citation and Hyperarticulated readings of these sentences were used, for a total of 28 different stimuli per sentence. Participants were divided randomly into two groups of 16 listeners each. The two groups were presented with tokens of either the first or the second sentence. The 28 stimuli represented four sets of stimulus materials: Male Citation, Female Citation, Male Hyperarticulated, and Female Hyperarticulated. On an individual trial, listeners heard a pair of sentences differing only in talker, and were asked to rate on a 1 – 7 scale the perceptual similarity of the two sentences. The sentence pairs always included talkers of the same gender producing the same sentence in the same style (e.g., two different male talkers producing citation speech or two different female talkers producing hyperarticulated speech). All possible talker combinations (within a gender), in both orders (i.e., A-B, B-A) were used, for a total of 168 trials. The interstimulus interval was 1 s. All trials were randomized for each participant and presented auditorily. Participants rated the sentence pairs by clicking on one of seven labeled buttons on a computer screen. No feedback was provided.

Results

The results of the voice similarity test, listed by stimulus set (Gender-Style) and individual sentence (Sentence 1 and Sentence 2), are given in Figure 5. For both male and female sets and for both sentences, voices in the citation style were judged as more similar to one another than voices in the hyperarticulated style. Of the four stimulus sets, the Female Citation voices were judged as more similar to one another than the Female Hyperarticulated voices. The difference in similarity ratings between the Male Citation and Male Hyperarticulated voices was smaller for Sentence 1 than Sentence 2. However, overall, Male Citation voices were still judged as more similar to one another than Male Hyperarticulated voices.

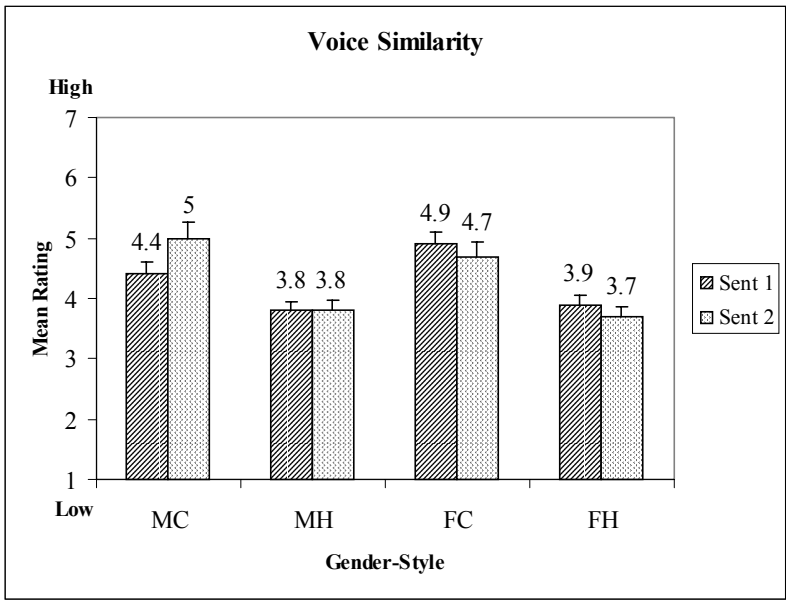


Figure 5. The mean similarity scores for the four stimulus sets, Male Citation (MC), Male Hyperarticulated (MH), Female Citation (FC), and Female Hyperarticulated (FH). Higher values denote greater similarity. "Sent" = Sentence. Error bars denote the standard error of the mean.

The mean perceptual ratings of the individual participants were submitted to a repeated-measures ANOVA, with two within-subjects factors (Gender, Speaking Style) and one between-subjects factor (Sentence). Of the main effects, only Speaking Style was significant (F (1, 30) = 64.9, $p < .0001$). Voices in a citation speaking style elicited significantly higher ratings than voices in a hyperarticulated speaking style. The Gender by Sentence (F (1, 30) = 16.2, $p < .0001$) and Gender by Speaking Style by Sentence (F (1, 30) = 11.3, $p < .01$) interactions were also significant.

The three-way interaction was further analyzed by running two two-way ANOVAs on the individual sentence data. For the participants who rated tokens of Sentence 1, both main effects and their interaction were significant (Gender: F (1, 15) = 13.4, $p < .01$; Speaking Style: F (1, 15) = 31.6, $p < .0001$; Gender by Speaking Style: F (1, 15) = 11.2, $p < .01$). In post hoc simple effects analyses, the effect of Speaking Style for both the male and female stimulus sets was significant ($p < 0.01$), while the effect of Gender within Speaking Style was only significant for the citation style stimuli ($p < 0.001$). For the participants who rated tokens of Sentence 2, Speaking Style was the only significant main effect. (F (1, 15) = 34, $p < .0001$), although Gender was marginally significant (F (1, 15) = 4.4, $p = .053$). The Gender by Speaking Style interaction was not significant.

In addition to analyzing the mean ratings, difference scores for the male and female voice sets were calculated to compare the effect of gender on the similarity of the citation and hyperarticulated speaking styles for each sentence. The mean rating for male and female hyperarticulated sentences of each subject was subtracted from the corresponding mean ratings for citation sentences (e.g., Listener 1's mean rating for Male Citation Sentence 1 minus that listener's mean rating for Male Hyperarticulated Sentence 1). The mean difference scores are shown in Figure 6. For Sentence 1, the mean difference score for the male stimulus set was lower than that for the female stimulus set, indicating that the effects of speaking style were larger for the female voices than the male voices in that sentence. In contrast, for Sentence 2, the mean difference score for the male stimulus set was higher than that for the female stimulus set.

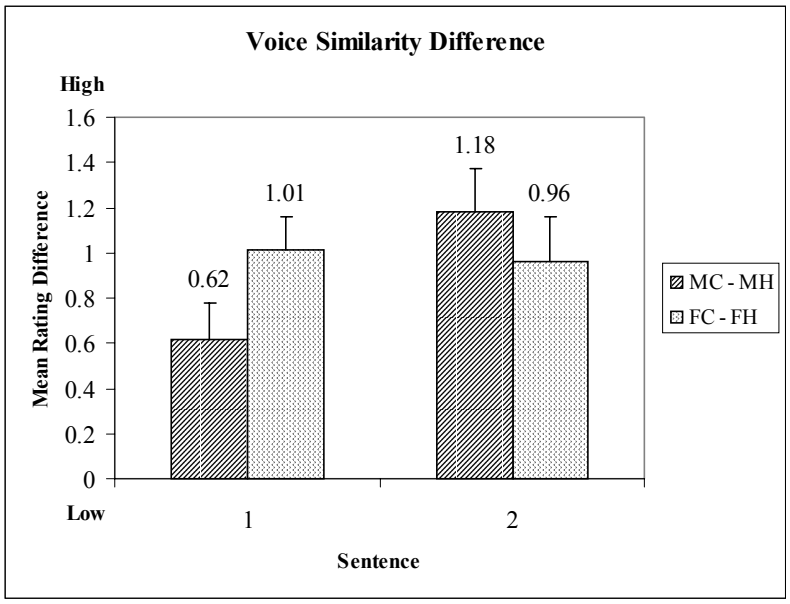


Figure 6. The mean difference scores for the two stimulus sets, Male Citation (MC) minus Male Hyperarticulated (MH) and Female Citation (FC) minus Female Hyperarticulated (FH). Higher values denote greater similarity. Error bars denote the standard error of the mean.

The difference scores of the individual participants were submitted to a repeated-measures ANOVA, with one within-subjects factor (Gender) and one between-subjects factor (Sentence). Neither of the main effects was significant. However, the Gender by Sentence interaction was significant (F (1, 30) = 11.3, $p < .01$). In post hoc simple effects analyses, the effect of Gender within both the male and female stimulus sets differed by sentence. For Sentence 1, the mean difference score for the male stimulus set was significantly lower than that for the female stimulus set ($p < 0.01$). For Sentence 2, the mean difference score for the male stimulus set was not significantly higher than that for the female stimulus set ($p = 0.09$).

General Discussion

The results of the perceptual similarity experiment using the scaling procedures were largely consistent with the acoustic properties of these speaking styles. Harnsberger and Goshert (2000) observed that sentences in citation and hyperarticulated speaking styles differed in overall duration, keyword duration, and vowel dispersion in the same manner for both male and female talkers. Likewise, listeners in the perceptual similarity experiment judged talkers producing citation speech as more similar to one another than talkers producing hyperarticulated speech, regardless of the gender of the talker. The greater similarity of the citation sentences suggests that voices in a citation speaking style should be harder to learn than the same voices in a hyperarticulated speaking style because they are perceptually less distinctive. We observed a significant effect of gender for Sentence 1 due to the relatively high mean similarity score elicited from female citation speech. However, this gender effect was not the same effect found in the first experiment (i.e., female hyperarticulated voices were easier to learn than female citation voices), because male citation and hyperarticulated speech also differed significantly in perceived similarity.

The gender effects observed in learning voices could be accounted for by the similarity difference scores shown in Figure 6. For one of the sentences, the difference between the female hyperarticulated and citation speaking styles was greater than the equivalent difference in the male speaking styles. It is possible that the female talkers produced a distinctive enough hyperarticulated speaking style that voice learning was influenced, while the male talkers made a more modest, though significant, distinction that was less informative. Given that the gender effect in the similarity difference scores varied by sentence, a scaling procedure for the entire stimulus set of Experiment 1 is required to determine whether the perceptual similarity of voices in Sentence 1 or 2 (or neither) were representative of the stimulus set as a whole. To further clarify the source of the gender effect, a replication of the first experiment is also needed with a new set of seven male and seven female talkers. In addition, other acoustic attributes of the stimuli should be examined, to ensure that male and female talkers manipulated the same gestural properties when shifting from citation to hyperarticulated speech. Such studies could take the form of a more extensive acoustic analysis, patterned after Brink et al. (1998) or Bradlow et al. (1996). Finally, a new voice learning experiment involving training over a longer period of time should be conducted, to determine if the gender effect on voice learning only emerges very early during stages of learning unfamiliar voices.

Regardless of the findings of subsequent studies, it is clear from the present results that speaking style influences the learning of novel voices. The results of the present study show that the relationship between speaking style and voice learning is complex, and may involve both attributes of the signal as well as listeners' prior experience with different kinds of voices. Further studies would serve to clarify our understanding of this process and, in turn, enrich our understanding of the nature of the perceptual learning process.

Summary and Conclusions

This study examined the effects of two speaking styles on the learning of novel voices. Listeners participated in a voice learning experiment consisting of four phases, (1) training, (2) test, (3) sentence generalization, and (4) speaking style generalization. In all four phases, listeners were presented with sentences produced by either seven male or seven female talkers in either a citation or hyperarticulated speaking style. The experiment was designed to assess two hypotheses concerning the effects of speaking style on novel voice learning: the *clear speech hypothesis* and the *idiosyncratic articulation hypothesis*. According to the *clear speech hypothesis*, hyperarticulated voices should be easier to learn than citation voices because hyperarticulated speech has been shown in prior work to be highly informative of other aspects of the speech signal, particularly, its linguistic content. According to the *idiosyncratic articulation hypothesis*, citation voices should be easier to learn than hyperarticulated voices because idiosyncratic gestural styles emerge that are normally suppressed in hyperarticulated speech. The results supported neither of these hypotheses directly because of an unanticipated effect of gender on the voice identification results. As it happened, female voices were easier to learn in a hyperarticulated style relative to a citation style in the training and test phases. In the style generalization phase with female voices, training with the hyperarticulated tokens provided a greater advantage in identifying voices in a novel style relative to training with citation tokens. In contrast, no differences were observed in the learning of male voices in different styles.

Several accounts of the gender effect were offered, including the possible role of prior experience in listening to male and female voices, which have been reported to differ in terms of the frequency of occurrence of casual and hyperarticulated speech forms. The gender interaction in voice learning was examined further in a similarity scaling experiment in which listeners rated the similarity of male and female voices in citation and hyperarticulated speaking styles. Both male and female citation sentences were rated as significantly more similar than the corresponding hyperarticulated sentences. However, for

one sentence, the difference in mean similarity for the female citation and hyperarticulated voices was significantly greater than the corresponding difference in the male voices, indicating that the female talkers may have produced a perceptually more distinct hyperarticulated style than the male talkers. Such a distinct style may have contributed to the gender effect observed in the voice learning experiment. Overall, the results show that speaking style affects voice learning, although the interpretation of the results is complicated by differences in learning male and female novel voices in the laboratory using this experimental paradigm.

References

Bradlow, A.R., Torreta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic characteristics. *Speech Communication*, *20*, 255-272.

Bricker, P.D. & Pruzansky, S. (1976). Speaker recognition. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295-326). New York: Academic Press.

Brink, J., Wright, R., & Pisoni, D.B. (1998). Eliciting speech reduction in the laboratory: Assessment of a new experimental method. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 396-420). Bloomington, IN: Speech Research Laboratory, Indiana University.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39-54.

Fischer, J.L. (1958). Social influences on the choice of a linguistic variant. *Word*, *14*, 47-56.

Goggin, J.P, Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19*, 448-458.

Harnsberger, J.D. & Goshert, L. (2000). Reduced, citation, and hyperarticulated speech in the laboratory: An acoustic analyses. In *Research on Spoken Language Processing Progress Report No. 24* (pp. 357-368). Bloomington, IN: Speech Research Laboratory, Indiana University.

Harnsberger, J.D. & Pisoni, D.B. (1999). Eliciting speech reduction in the laboratory II: Calibrating cognitive loads for individual talkers. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 339-349). Bloomington, IN: Speech Research Laboratory, Indiana University.

Hollien, H., Majewski, W., & Doherty, E.T. (1982). Perceptual identification of voices under normal, stress, and disguise conditions. *Journal of Phonetics*, *10*, 139-148.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *Journal of the Acoustical Society of America, 94,* 701-714.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Kreiman, J. (1997). Listening to voices: Theory and practice in voice perception research. In K. Johnson and J.W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 85-108). San Diego: Academic Press.

Kreiman, J., Gerratt, B.R., Precoda, K. & Berke, G.S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research*, *35*, 512-520.

Kuwabara, H. & Takagi, T. (1991). Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method. *Speech Communication*, *10*, 491-495.

Labov, W. (1970). The study of language in its social context. *Studium Generale*, *23*, 30-87.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W.J. Hardcastle and A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Mattingly, I. (1966). Speaker variation and vocal-tract size. *Journal of the Acoustical Society of America, 39,* 1219.

Moon, S.-J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96,* 40-55.

Murray, T. & Singh, S. (1980). Multidimensional analysis of male and female voices. *Journal of the Acoustical Society of America, 68,* 1294-1300.

Nygaard, L.C. & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*, 355-376.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*, 96-103.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434-446.

Singh, S. & Murray, T. (1978). Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America, 64,* 81-87.

Tartter, V.C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perception and Psychophysics*, *49*, 365-372.

Trudgill, P. (1974). *The social differentiation of English in Norwich. Cambridge*: Cambridge University Press.

Thompson, C.P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*, 121-131.

Van Lancker, D., Kreiman, J., & Wickens, T.D. (1985). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics*, *13*, 39-52.

Appendix A:

The SPIN sentences recorded for this study

The farmer harvested his crop.
His boss made him work like a slave.
He caught the fish in his net.
Close the window to stop the draft.
The beer drinkers raised their mugs.
I made the phone call from a booth.
The cut on his knee formed a scab.
The railroad train ran off the track.
They drank a whole bottle of gin.
The airplane dropped a bomb.
I gave her a kiss and a hug.
The soup was served in a bowl.
The cookies were kept in a jar.
How did your car get that dent?
The baby slept in his crib.
The cop wore a bulletproof vest.
No one was injured in the crash.
The hockey player scored a goal.
How long can you hold your breath?
At breakfast he drank some juice.
The king wore a golden crown.
He got drunk in the local bar.
The doctor prescribed the drug.
The landlord raised the rent.
Playing checkers can be fun.
Throw out all this useless junk.
Her entry should win first prize.
The stale bread was covered with mold.
I ate a piece of chocolate fudge.
The story had a clever plot.
He's employed by a large firm.
The mouse was caught in the trap.
I've got a cold and a sore throat.
The judge is sitting on the bench.

Appendix B:

The SPIN sentences presented in the first experiment

Sentences used in the Training and Test Phases

- I made the phone call from a booth.
- The railroad train ran off the track.
- No one was injured in the crash.
- The landlord raised the rent.
- The beer drinkers raised their mugs.

Sentences used in the Sentence Generalization Phase

- Playing checkers can be fun.
- Her entry should win first prize.
- The stale bread was covered with mold.
- He's employed by a large firm.
- The judge is sitting on the bench.

Sentences used in the Speaking Style Generalization Phase

- His boss made him work like a slave.
- They drank a whole bottle of gin.
- The hockey player scored a goal.
- How long can you hold your breath?
- The doctor prescribed the drug.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Some Effects of Phonotactic Probabilities on the Processing of Spoken Words and Nonwords by Post-Lingually Deafened Adults with Cochlear Implants[1]

Michael S. Vitevitch, David B. Pisoni,[2] Karen Iler Kirk,[2]
Marcia Hay-McCutcheon,[2] and Stacy Yount[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported in part by grants from the American Speech Language and Hearing Foundation and the Acoustical Society of America to the first author, and from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (R01 DC 00111, T32 DC 00012, K08 DC 00126-3).
[2] DeVault Otologic Research Laboratory, Department of Otolaryngology, Indiana University School of Medicine, Indiana University Purdue University Indianapolis

Some Effects of Phonotactic Probabilities on the Processing of Spoken Words and Nonwords by Post-Lingually Deafened Adults with Cochlear Implants

Abstract. Probabilistic phonotactics refers to the frequency with which segments and sequences of segments occur in syllables and words. Knowledge of phonotactics has been shown to be an important source of information in segmenting and recognizing speech in normal hearing listeners. A post-perceptual task (nonword rating) and two on-line tasks (an auditory same-different and an auditory lexical decision task) were used in the present set of experiments to examine the use of phonotactic information by post-lingually deafened adults who have received a cochlear implant. The results of all three experiments showed that both normal-hearing and hearing-impaired listeners are sensitive to differences in phonotactic information to varying degrees. Furthermore, cochlear implant patients with better word recognition abilities (as measured by the NU-6) tended to be more sensitive to phonotactic information than cochlear implant patients with poorer word recognition abilities. The implications of these results for outcome assessments and clinical interventions are discussed.

Phonotactic information refers to the sequential arrangement of phonetic segments in morphemes, syllables, and words (Crystal, 1980). Sounds and sequences of sound that are found in a given language are said to be *legal* within that language, whereas sounds and sequences of sound that are not found in a given language are said to be *illegal* within that language. Awareness of the sounds that are legal in one's native language occurs very early in life. For example, Jusczyk, Frederici, Wessels, Svenkerud, and Jusczyk (1993) showed that Dutch and American children as young as nine months of age listen longer to lists of words with patterns of segments and sequences allowed in their native language than to lists of words with patterns from the other language. These results show that listeners are sensitive early in life to the sounds and sequences of sound that are legal in their native language.

Although phonotactic information is often described as a set of rules—or a "phonological syntax" (Malmkaer, 1991)—specifying the sequences of segments that are legal or illegal in a language, recent work has explored the *probabilistic* nature of phonotactic constraints (Kessler & Treiman, 1997; Treiman, Kessler, Knewasser, Tincoff, & Bowman, 1996; Vitevitch & Luce, 1998, 1999; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). That is, rather than using stimuli that contained either legal or illegal sequences, these researchers created stimuli that were completely legal in a given language, but that varied in how common the segments and sequences were in that language. Jusczyk, Luce, and Charles-Luce (1994) demonstrated that nine-month old infants are also sensitive to the *probabilities* of sound patterns within their native language. Using the same procedure as Jusczyk et al. (1993), Jusczyk, Luce, and Charles-Luce (1994) found that American infants listened longer to lists of nonwords that contained high probability segments and sequences of segments than to lists of nonwords that contained low probability segments and sequences in English. These results suggest that sensitivity to *probabilistic* phonotactic information may also develop early in life and may be important for the processing of spoken language later in life.

For example, sensitivity to the phonotactic probabilities of the ambient language may assist children in acquiring and building a lexicon. Computational (e.g., Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997) and experimental investigations (e.g., Mattys, Jusczyk, Luce & Morgan, 1999; Saffran, Newport, & Aslin, 1996) suggest that phonotactic information may play a role in the segmentation of words from continuous speech. Sensitivity to the patterns of segments that occur only within words (such as /tl/), or only at the edges of words (e.g., /ŋ/ does not occur in the initial portion of

English words) may allow a child to identify the beginning and endings of words. With the beginning and ending of a word identified, the child can isolate an individual word from the continuous stream of speech and begin acquiring a lexicon. Other research suggests that older children may also use phonotactic information to add new words to the lexicon (e.g., Gathercole, Willis, Emslie, & Baddeley, 1991; Gupta & MacWhinney, 1997; Storkel & Rogers, 2000). Thus, phonotactic probabilities in language are valuable sources of information early in life for processing spoken language.

Phonotactic information is not only used in the acquisition of language. Adults are also sensitive to phonotactic information and may use it to process spoken language. For example, Vitevitch, Luce, Charles-Luce, and Kemmerer (1997; see also Messer, 1967) created bisyllabic nonword stimuli containing segments and sequences of segments that were completely legal in English, but varied in how common they were in English. Stimuli comprised of segments and sequences of segments that occur frequently in English, such as /kikrig/, are said to have high phonotactic probability. Stimuli comprised of segments and sequences of segments that occur less frequently in English, such as /j^ʃð^tʃ/, are said to have low phonotactic probability. The researchers asked participants to rate how "good" each item would be if it were a real word in English. Their results showed that participants' subjective ratings of the spoken nonwords followed the objective measure of phonotactic probability: Nonwords with high probability patterns were rated as being more word-like than low probability patterns. These results suggest that adults are sensitive to fine-grained probabilistic phonotactic information within their native language, and can access and use this information in tasks requiring explicit judgment about nonword patterns.

Vitevitch et al. (1997) also asked another group of participants to repeat the same nonwords presented auditorily. An analysis of the response latencies showed that nonwords with high probability patterns were repeated more quickly than nonwords with low probability patterns, suggesting that probabilistic phonotactic information may play a role in spoken word recognition in normal hearing listeners (see also Vitevitch & Luce, 1998, 1999; Vitevitch, Luce, Pisoni, & Auer, 1999). That is, phonotactic information may be one of several sources of information, such as word frequency (e.g., Savin, 1963; Solomon & Postman, 1952) or the stress pattern of a word (e.g., Cutler & Norris, 1988) that normal hearing listeners use to understand spoken language.

In the present study, we were interested in determining whether a group of post-lingually deafened adults who have subsequently received a cochlear implant also make use of phonotactic probabilities to understand spoken words. Doyle et al. (1995), for example, reported that cochlear implant users have difficulty distinguishing among segments varying in manner of articulation, voicing, and place of articulation. Given the difficulty in discriminating fine phonetic details in speech, cochlear implant users may no longer consistently rely on information or representations related to segments or sequences of segments to process spoken words. Post-lingually deafened adults who used a cochlear implant for at least one year participated in the present set of experiments. Our goal was to determine if these patients are able to make use of information about phonotactic probabilities and whether these cognitive processing strategies help cochlear implant users recognize isolated spoken words.

The post-lingually deafened adults who participated in this set of experiments were all patients who had acquired language with normal hearing. Later in life these individuals became profoundly deafened through trauma or disease and had subsequently received and used a cochlear implant for at least a year. A cochlear implant is a sensory aid--a surgically implanted prosthetic device that bypasses the damaged inner hair cells and transduces an auditory signal into an electrical signal that stimulates the auditory nerve (Wilson, 2000). A cochlear implant provides patients who have profound hearing loss with useable forms of auditory stimulation. A typical multi-channel cochlear implant consists of a microphone that receives auditory input, a speech processor that uses one of several possible preset algorithms to

process incoming auditory signals, and an array of electrodes that are surgically implanted into the cochlea to electrically stimulate the auditory nerve. Electrical stimulation of the auditory nerve by the implant results in the perception of spectral information via the tonotopic arrangement of the electrodes in the cochlea. The stimulation also provides durational and intensity information about the auditory signal (Wilson, 2000). The outcome measures of the effectiveness of cochlear implants in adults (across the several types of systems and several processing strategies) ranges from being able to follow a conversation on the telephone to being able to merely detect the presence or absence of sound (e.g., Blamey et al., 1987; Cohen, Waltzman, & Shapiro, 1989; Dowell, Mecklenburg, & Clark, 1986; Gantz et al., 1988; Skinner et al., 1991; Geier, Fisher, Barker, & Opie, 1999; Hollow et al., 1995; Holden, Skinner, & Holden, 1997; Staller et al., 1997).

To examine whether cochlear implant users are still able to make use of phonotactic information to recognize spoken words, we used the nonwords of Vitevitch et al. (1997) with a slightly modified methodology. In the present experiment, participants were presented with bisyllabic nonwords varying in phonotactic probabilities and were asked to repeat the nonword as accurately as possible. After the repetition response, they heard the stimulus again but were asked to rate the goodness of each item as if it were a real word in English. Participants used a scale of 1 (“Bad sounding English word”) to 5 (“Good sounding English word”).

If cochlear implant users are able to access phonotactic information, we would expect to find a difference in the ratings of the nonwords that is similar to that observed by Vitevitch et al. (1997). Specifically, nonwords with high-probability phonotactics should be rated as better sounding English words than nonwords with low-probability phonotactics by the cochlear implant users. Moreover, patients with better word recognition skills (as assessed by scores on the NU-6) may be able to more finely discriminate sound patterns and sequences varying in phonotactic probability and therefore would be more likely to use this more detailed information than those with poorer word recognition abilities (i.e., lower NU-6 scores). We further predicted that the ratings would reflect this difference in word recognition ability. Specifically, patients with poorer word recognition abilities should not be able to make fine-grained discriminations among segments and sequences of segments making it difficult to distinguish a real word from a nonword. This pattern would be expected in cochlear implant patients with poorer word recognition abilities. They would rate all of the nonwords as being "better words" than cochlear implant users with better word recognition ability or normal hearing listeners.

For repetition accuracy, we predicted that if cochlear implant users were able to use phonotactic information, the accuracy with which the nonwords were repeated would also vary as a function of phonotactic probabilities. Specifically, nonwords with high phonotactic probability should be repeated more accurately than nonwords with low phonotactic probability, as in Vitevitch et al. (1997). Finally, we predicted that the cochlear implant users with better word recognition ability would repeat the nonwords more accurately than the cochlear implant users with poorer word recognition ability.

Methods

Participants

Eight adult patients with cochlear implants and four normal-hearing adults participated in this experiment. Based on the preliminary analysis of the repetition data from the present experiment and feedback from the cochlear implant patients, no more than eight cochlear implant patients were tested in this difficult task. Four normal-hearing adults were recruited in order to have equal numbers of participants in each group based on perceptual ability. The normal-hearing listeners were recruited from introductory Psychology classes at Indiana University-Bloomington, reported no history of speech or

hearing disorders, and received partial credit toward the fulfillment of a course requirement. All participants were native English speakers. The mean age of the normal-hearing participants was 19.75 years old.

The eight adult cochlear implant users were outpatients at Riley Hospital, Indianapolis, Indiana who were paid for their participation in the study. All the patients were post-lingually deafened adults. The mean age of the participants who used cochlear implants was 44.8 years old. The mean age of onset of deafness was 29.4 years old. The mean age at which implantation of a cochlear implant device took place was 41.5 years. The 12.1 year difference between the age of onset of deafness and the age at which implantation took place does not mean the participants were without auditory stimulation for an average of 12.1 years; all of the post-lingually deafened participants used hearing aids for some period of time before being implanted with the cochlear implant. Five participants used the Nucleus device, two used the Clarion device, and one used the MedEl device. See Table I for individual participant information.

Table I. Individual Characteristics of Cochlear Implant Users in Experiment 1

	Age	Age at Onset of Deaf	Etiology	Age at Implant-ation	Type of Implant and processing strategy	Years of CI Use	NU-6 Word	NU-6 Phon	NU-6 Cond.
1	50	24	u.k	39	Nucleus-22, SPEAK	11	68	85	HIGH
2	37	35	trauma	37	Nucleus-22, SPEAK	1	34	61	HIGH
3	35	24	u.k	31	Nucleus-22, SPEAK	4	58	77	HIGH
4	52	49	u.k	51	Nucleus-22, SPEAK	1	50	69	HIGH
5	63	39	trauma men.	57	Nucleus-22, SPEAK	6	8	29	LOW
6	42	19	otscl.	41	Combi40, CIS	1	28	53	LOW
7	44	42	ototox	43	Clarion, CIS	2	34	56	LOW
8*	37	3	u.k	33	Clarion, CIS	4	34	59	LOW
	45.0	29.4		41.5		3.75			

Note: Two listeners also participated in Experiments 2 and 3; they are indicated by * next to the participant number. u.k. = unknown

All of the participants were divided into three groups based on word recognition ability: the normal-hearing adults, the cochlear implant patients who had above average speech perception as measured by the NU-6, a standard test of word recognition abilities (High-NU-6 scores), and the cochlear implant patients who had average speech perception (Low-NU-6 scores). A median split on the NU-6 scored by percent words correct and by percent phonemes correct for each patient served as the criterion to divide the cochlear implant patients into the two groups of four participants each. Patients in the High-NU-6 group had a mean NU-6 scored by percent words correct of 52.5%, and a mean NU-6 scored by percent phonemes correct of 73.0%. Patients in the Low-NU-6 group had a mean NU-6 scored by percent words correct of 26.0%, and a mean NU-6 scored by percent phonemes correct of 49.2%. The differences between the NU-6 scored by percent word correct (F (1,6) = 7.84, $p < .05$) and by percent phoneme correct (F (1,6) = 7.65, $p < .05$) between the groups were significantly different.

Although the two groups of cochlear implant patients differed in their word recognition abilities, the two groups did not differ in their hearing thresholds as measured by pure-tone averages (F (1,6) < 1). A pure-tone average is the mean sound level for detecting a pure-tone at 500, 1000, and 2000 Hz. Patients in the High-NU-6 group had a pure-tone average of 28.33 dB SPL, and patients in the Low-NU-6 group had a pure-tone average of 29.33 dB SPL suggesting that the two groups had comparable abilities in detecting sound.

Materials

Two-hundred-forty bisyllabic nonwords with the stress on the first syllable were selected from the stimuli constructed by Vitevitch et al. (1997). These nonword stimuli were divided into two lists of 120 stimuli each. One list had nonwords with syllables in the order A-B, whereas the other list had the same syllables forming nonwords, but the order of the syllables in the nonwords was B-A. No syllable was used more than once in a list. Examples of the stimuli are listed in Table II.

Table II. Examples of bisyllabic nonword stimuli varying in phonotactic probability

Condition	List 1	List 2
High-High	`f^ltʃ^n	`tʃ^nf^l
High-Low	`l^nð^z	`ð^zl^n
Low-High	`gaɪbsaɪk	`saɪkgaɪb
Low-Low	`ðaɪbdʒaiz	`dʒaizðaɪb

Phonotactic probability was defined as in Jusczyk et al. (1994) and Vitevitch et al. (1997). The phonotactic probability of a nonword CVC syllable was based on the following statistics: (1) positional segment frequency (i.e., how often a phonetic segment occurs in a particular position in a word), and (2) biphone probability (i.e., the segment-to-segment co-occurrence probability). Log-frequency weighted values were used to compute positional segment frequency and biphone probability from a computer-readable version of Webster's Pocket Dictionary, which contains approximately 20,000 words (see Auer, 1993). Because frequency-weighted values were used in our computations, the segment and biphone statistics can be viewed as being based on word *token* counts, not word *type* counts.

High probability nonword patterns consisted of segments with high segment positional probabilities and frequent biphone probabilities. For example, in the high probability pattern /kik/ ("keek"), the consonant /k/ is relatively frequent in the initial position, the vowel /i/ is relatively frequent in the medial position, and the consonant /k/ is relatively frequent in the final position. The probabilities of the initial consonant-vowel (/ki_/) and the vowel-final consonant (/_ik/) co-occurring were also relatively high.

Conversely, low probability nonword patterns consisted of segments with low segment positional probabilities and less common biphone probabilities. Despite being relatively rare, none of the patterns formed were phonotactically illegal in English. Each of the five vowels used in the CVCs, /^, aɪ, i, e, ɜ/ occurred in equal proportions in each of the syllable types. The same vowel appeared in the first and second syllable of each nonword.

The average segment probability was .1926 for the high-probability pattern list and .0543 for the low probability pattern list. The average biphone probability was .0143 for the high-probability list and .0006 for the low-probability list. The difference in the magnitudes of the segment and biphone

probabilities reflects the fact that there are more biphones than segments. This results in biphones having a lower probability of occurrence overall than segments because the same total probability (i.e., 1.00) is divided among many more possible outcomes for the biphones than for the segments.

The same stimulus tokens used in Vitevitch et al. (1997) were also used in this experiment. A trained phonetician originally recorded all the stimuli, which were spoken in isolation. The stimuli were then low-pass filtered at 4.8 kHz and digitized at a sampling rate of 10 kHz using a 12-bit analog-to-digital converter. All nonwords were edited into individual sound files and stored on computer disk using a digital waveform editor. A trained speech scientist measured the amplitude of the vowel of each syllable with a digital waveform editor to confirm correct stress placement by the speaker.

Procedure

Participants were tested individually. Each participant was seated in front of a 200MHz Gateway 2000 Pentium computer that controlled stimulus presentation and response collection. All stimuli were presented in random order one at a time. Cochlear implant users were tested in an IAC sound booth in the DeVault Otologic Laboratory at the IU School of Medicine and heard the stimuli at 70 dB SPL over an Advent AV570 speaker. The normal-hearing participants were tested in a sound attenuated booth using an identical computer system in the Speech Research Laboratory in Bloomington. Normal-hearing participants heard the stimuli over a pair of Beyerdynamic DT-100 headphones. Because of the mechanics of the cochlear implant, headphones could not be used with the eight cochlear implant patients.

Each participant received one of the two lists of 120 randomly ordered stimuli. A scale from 1, labeled “Bad English Word” to 5, labeled “Good English Word” was attached to the first five buttons of a seven-button response box. The sixth button was deactivated for response, and the seventh button was labeled “Play Again.”

A trial proceeded as follows: A prompt appeared on the computer screen, and one of the test signals was presented at 70 dB SPL over the headphones or speaker. The participant was asked to repeat the nonword as accurately as possible into a Shure 5755 microphone connected to a Marantz tape recorder. Because of technical considerations, response latencies were not recorded from these patients as they were in Vitevitch et al. (1997). Specifically, cochlear implant users cannot be presented with auditory stimuli over headphones. Rather, the stimuli must be presented free-field. Unfortunately, such presentation would trigger a voice-key interfaced with a microphone normally used to record reaction times in similar experiments. Thus, only the accuracy of the response was examined in the present study. The participant pressed the labeled button on the response box to hear the stimulus again. After the second presentation of the stimulus, the participant rated the item as quickly as possible by pressing one of the five numbered buttons on the response box. After recording the response, the computer began the next trial.

Results

To examine sensitivity to phonotactic information as a function of word recognition ability, a mixed design ANOVA with the Greenhouse-Geisser correction was performed on the mean ratings with phonotactic probability as a within-participants factor and word recognition ability as a between participants factor for each scoring criterion. The mean ratings for the four phonotactic conditions as a function of the three groups of listeners are shown in Figure 1. Ratings on a scale of 1 (“BAD”) to 5 (“GOOD”) are plotted on the *y* axis. The three groups of listeners are represented on the *x* axis. “High-High” refers to nonwords with high phonotactic probability initial and final syllables and is represented by dotted bars. “High-Low” refers to nonwords with high probability initial and low probability final

syllables and is represented by the gray bars. "Low-High" refers to nonwords with low probability initial and high probability final syllables and is represented by the clear bars. Finally, "Low-Low" refers to nonsense words with low probability initial and final syllables and is represented by the striped bars. Figure 1 shows the mean rating for *all* stimuli regardless of whether they were correctly repeated or not.

Ratings to All the Nonwords

Examination of the ratings to all the stimuli revealed a main effect of phonotactic probability (F (3,27) = 7.43, $p < .001$). Stimuli in the High-High condition (mean = 3.02) were rated higher than stimuli in the Low-Low condition (mean = 2.56, F (3,27) = 21.06, $p < .001$). Stimuli in the High-High condition were also rated higher than stimuli in the High-Low condition (mean = 2.76, F (3,27) = 6.57, $p < .05$). Finally, stimuli in the Low-High condition (mean = 2.87) were rated higher than stimuli in the Low-Low condition (F (3,27) = 9.46, $p < .01$). No other comparisons or interactions were significant (all $F < 1$). These results confirm our initial prediction and suggest that cochlear implant patients are able to access phonotactic information. These results also replicate the findings of Vitevitch et al. (1997) who examined sensitivity to phonotactic information in normal-hearing listeners.

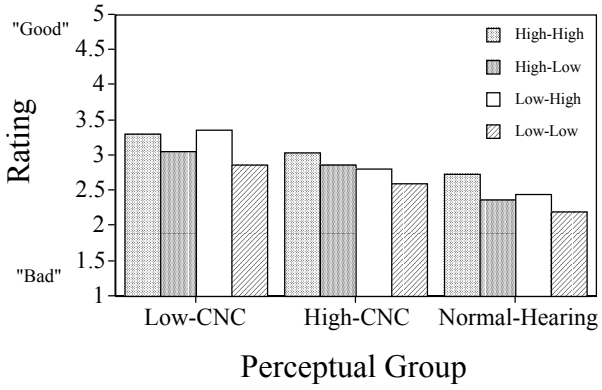


Figure 1. The mean ratings of nonwords on a scale from 1 (BAD English word) to 5 (GOOD English word) as a function of perceptual group for all the stimuli.

A main effect of word recognition ability was also found (F (2,27) = 5.56, $p < .05$). The Low-NU-6 group (mean = 3.15) had higher nonword ratings than the normal-hearing group (mean = 2.43; F (2,27) = 11.07, $p < .01$). The High-NU-6 group (mean = 2.83) also had higher nonword ratings than the normal-hearing group, but this effect was only marginally significant (F (2,27) = 3.41, $p = .09$). The nonword ratings for the Low-NU-6 group were not significantly different from the High-NU-6 group ($F <$ 1). Although there was no statistically significant difference in nonword ratings between the two groups of cochlear implant patients, the two groups of cochlear implant patients did have higher nonword ratings than the normal hearing group. That is, normal hearing listeners rated the nonword stimuli as being less like English words than the cochlear implant patients. These results partially support our initial prediction regarding the ability of listeners varying in word recognition skill to make fine-grained discriminations among segments and sequences of segments. Both groups of cochlear implant patients were not as good as the normal hearing listeners at making fine-grained discriminations among segments and sequences of segments. The poorer ability of the cochlear implant patients to make fine-grained discriminations made it

difficult for them to distinguish possible real words from nonwords, resulting in the nonwords being rated as "better words" than normal hearing listeners.

Ratings to Nonwords using an Accuracy Criterion

Repetition of the nonwords proved to be a very difficult task for the cochlear implant users. When the repetitions were scored with a strict criterion (all phonemes repeated correctly), a mean value of 6% of the nonwords was correctly repeated across participants and conditions. When a less strict criterion was used, in which a majority of the stimulus (4 out of the 6 phonemes in the stimulus) was repeated correctly, the mean value across participants and conditions of correct repetitions rose to 45%. Figure 2 shows the mean rating from stimuli that were correctly repeated using this criterion. The accuracy rates with which four of the six phonemes in the stimuli were repeated are also shown in Figure 2.

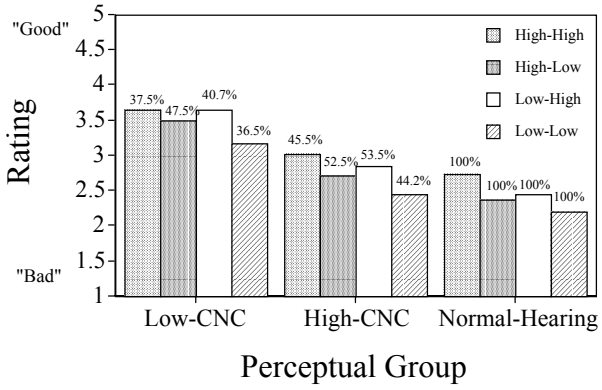


Figure 2. The mean ratings of nonwords on a scale from 1 (BAD English word) to 5 (GOOD English word) as a function of perceptual group for only the stimuli in which four out of the six phonemes were correctly repeated.

Analyses of the ratings for the stimuli in which four of the six phonemes were correctly repeated revealed a similar pattern of results as the analyses of the ratings to all the stimuli. A main effect of phonotactic probability was found for the correctly repeated stimuli (F (3,27) = 6.44, $p < .01$). Stimuli in the High-High condition (mean = 3.13) were rated higher than stimuli in the Low-Low condition (mean = 2.61, F (3,27) = 18.07, $p < .001$). Stimuli in the High-High condition were also rated higher than stimuli in the High-Low condition (mean = 2.86, F (3,27) = 4.88, $p < .05$). Stimuli in the Low-High condition (mean = 2.98) were rated higher than stimuli in the Low-Low condition (F (3,27) = 9.06, $p < .01$). Finally, stimuli in the High-Low condition were rated higher than stimuli in the Low-Low condition (F (3,27) = 4.16, $p < .05$). No other comparisons or interactions were significant in the analysis of correctly repeated nonwords (all $F < 1$). These results further suggest that cochlear implant patients are able to access phonotactic information.

A main effect of word recognition ability was also found (F (2,27) = 5.21, $p < .05$). The Low-NU-6 group (mean = 3.49) had higher ratings than the High-NU-6 group (mean = 2.75) and the normal-hearing group (mean = 2.44). Pairwise comparisons show that the difference between the Low-NU-6 group and the High-NU-6 group was statistically significant (F (2,27) = 4.78, $p < .05$), as was the difference between the Low-NU-6 group and the normal-hearing group (F (2,27) = 9.92, $p < .01$).

However, the difference between the High-NU-6 group and the normal-hearing group was not statistically significant ($F < 1$). These results also provide partial support for our initial prediction regarding the ability of listeners varying in word recognition skill to make fine-grained discriminations among segments and sequences of segments. Normal hearing listeners and cochlear implant patients with High-NU-6 scores were better than cochlear implant patients with Low-NU-6 scores at making fine-grained discriminations among segments and sequences of segments. The poorer ability of the cochlear implant patients with Low-NU-6 scores to make fine-grained discriminations made it difficult for them to distinguish possible real words from nonwords, resulting in the nonwords being rated as "better words" than cochlear implant patients with High-NU-6 scores and normal hearing listeners.

Accuracy Analysis of Repeated Nonwords

Analysis of the accuracy rates for the stimuli in which four of the six phonemes were correctly repeated showed a main effect of word recognition ability (F (2,27) = 16.44, $p < .01$). The normal-hearing group correctly repeated more nonwords (mean = 100%) than the Low-NU-6 group (mean = 40.6%) and the High-NU-6 group (mean = 48.9%). Pairwise comparisons show that the difference between the normal-hearing group and the Low-NU-6 group was statistically significant (F (2,27) = 28.05, $p < .001$), as was the difference between the normal-hearing group and the High-NU-6 group (F (2,27) = 20.70, $p < .01$). The difference in repetition accuracy between the Low-NU-6 group and the High-NU-6 group was not statistically significant ($F < 1$). These results suggest that listeners with a better ability to make fine-grained discriminations among segments and sequences of segments (i.e., normal hearing listeners) are more accurate in their repetition of those segments and sequences of segments.

The main effect of phonotactic probability was not significant, nor was the interaction between perceptual group and phonotactic probability (Fs < 1). The lack of a difference between nonwords varying in phonotactic probability suggests that the four types of nonwords were equally perceptible for each of the three groups of listeners.

The Phonotactic Sensitivity Index

To further assess the relationship between the use of phonotactic information and spoken word recognition performance, we developed a global index of *phonotactic sensitivity* and correlated it with a measure of spoken word recognition ability. Each cochlear implant patient's NU-6 scored by percent words correct was used as the measure of spoken word recognition. Phonotactic sensitivity was calculated by computing a difference score between the nonword ratings each participant gave to stimuli in the High-High condition and to stimuli in the Low-Low condition. We hypothesized that individuals who were more sensitive to phonotactic information in these patterns should display a larger difference between the ratings of High-High stimuli and Low-Low stimuli; the High-High stimuli would be rated much higher (i.e., as better possible words in English) than the Low-Low stimuli. Conversely, we predicted that individuals who were less sensitive to phonotactic information in these patterns would display a smaller difference between the ratings of High-High stimuli and Low-Low stimuli; the High-High and Low-Low stimuli would not be well discriminated and would be rated similarly, thereby producing small differences in the ratings.

The measures of phonotactic sensitivity and word recognition performance were only weakly related ($r = + .32$), and the correlation was not significant. (An analysis using the NU-6 scored by percent phonemes correct showed a similar pattern with a somewhat weaker correlation.) A scatterplot of this relationship is displayed in Figure 3. Examination of these individual data points shows that no participant rated the Low-Low stimuli higher than the High-High stimuli; this would have resulted in a negative difference score. When the data for the two participants who obtained a phonotactic sensitivity

measure of zero (one had a difference of .03, the other had a difference of 0.0) were excluded from the analysis, a stronger correlation was observed (r = + .66), although this did not reach statistical significance most likely because the sample size was too small. (An analysis using the NU-6 scored by percent phonemes correct shows a similar increase in the correlation coefficient when the difference scores near zero are removed.) Although suggestive, this trend indicates that success in using phonotactic information in this nonword rating task may be related to performance in recognizing isolated spoken words.

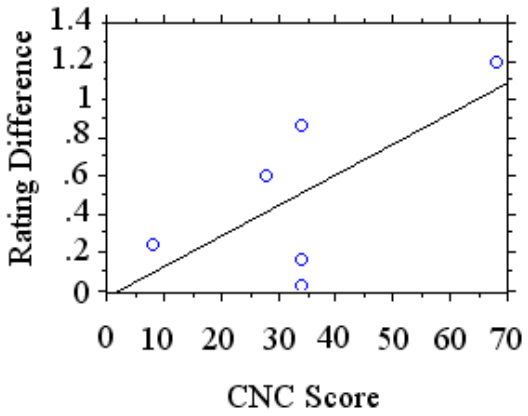


Figure 3. Scatterplot for a measure of phonotactic sensitivity (the difference in ratings to High-High and Low-Low nonword stimuli) and a measure of spoken word recognition ability (NU-6) for 8 cochlear implant users.

Discussion

The results of this experiment confirmed several predictions we made regarding the sensitivity of cochlear implant patients to phonotactic information in isolated nonword patterns. First, all three groups of listeners demonstrated via their "goodness" ratings sensitivity to differences in phonotactic probabilities among the nonword stimuli that corresponded with the objective measures of segment and sequence frequency. That is, nonword patterns comprised of segments and sequences of segments that are common in English (high phonotactic probability) were rated as being more word-like than nonwords comprised of segments and sequences of segments that are less common in English (low phonotactic probability). This result, which was observed for all three groups of listeners, replicates the finding of Vitevitch et al. (1997) in which normal hearing listeners rated the same nonwords in a rating task with a slightly modified methodology. In the present study, differences in ratings among the nonwords varying in phonotactic probability were observed when the ratings to all of the stimuli were analyzed as well as when only those stimuli that had four out of six phonemes correctly repeated were analyzed. These results suggest that cochlear implant users, like normal-hearing listeners (Vitevitch et al., 1997), still have access to and can use phonotactic information to make judgments about the sound patterns of spoken stimuli.

The results of this study also demonstrate that ratings of nonwords varying in phonotactic probabilities differ as a function of word recognition ability. Normal hearing listeners consistently rated the stimuli as being less word-like than the cochlear implant users with low word recognition ability regardless of whether the ratings from all stimuli were included in the analysis or just the ratings from those stimuli that were correctly repeated to an accuracy criterion. Although normal-hearing listeners tended to have lower ratings overall (i.e., less word-like) than the cochlear implant users with high word

recognition ability, this difference approached statistical significance only when the ratings from all, rather than just the accurately repeated, nonwords were analyzed. Similarly, the cochlear implant users with high word recognition ability tended to have somewhat lower ratings (i.e., less word-like) than the cochlear implant users with low word recognition ability. However, this difference was significant when only the correctly repeated nonwords were analyzed. These results suggest that access to and optimal use of phonotactic information in nonword pattern may be related to performance in recognizing isolated spoken words.

The sensory information that cochlear implant patients rely on in this rating task may be different from the information that normal-hearing listeners have access to. In the present experiment, cochlear implant users generally rated the nonwords as better "English words" than the normal hearing listeners rated them. This trend in the rating data may reflect the fact that cochlear implant users may have more broadly or coarsely defined representations of acoustic-phonetic input compared to normal-hearing listeners, thus, many more nonword patterns sound like a possible word in English and therefore are rated as more word-like. Normal-hearing listeners, however, are able to make much finer-grained phonetic distinctions in their encoding of the initial acoustic-phonetic input in these nonword patterns. Consequently, normal-hearing listeners may have different equivalence classes than cochlear implant users; these different perceptual categories may contribute to the overall difference in ratings as a function of word recognition ability.

The analysis of the nonword repetitions showed that the cochlear implant users were less accurate than the normal hearing listeners in the repetition of the nonword stimuli, regardless of the word recognition ability of the cochlear implant users. There were no differences in repetition accuracy between the two groups of cochlear implant users based on word recognition ability. We also found no differences in repetition accuracy as a function of phonotactic probability, in contrast to the significant difference observed among normal hearing listeners in Vitevitch et al. (1997). It should be noted, however, that the significant effect of repetition accuracy as a function of phonotactic probability observed in Vitevitch et al. (1997) was due to the extremely poor performance of participants repeating stimuli containing low phonotactic probability segments and sequences in both syllables (LOW-LOW condition). On average, the stimuli in the remaining three conditions (HIGH-HIGH, HIGH-LOW, and LOW-HIGH) were repeated approximately 10% more accurately than stimuli in the LOW-LOW condition in Experiment 2 of Vitevitch et al. (1997). That is, repetition performance was approximately equivalent across conditions, except when attempting to repeat stimuli that contained segments and sequences of segments in both syllables that are not common in English. The poorer performance in the LOW-LOW condition may also have been a function of the speeded nature of the task used by Vitevitch et al. (1997).

Furthermore, a strict accuracy criterion (*all six phonemes* had to be repeated correctly) was used in Vitevitch et al. (1997), further contributing to the difference in the accuracy results between that study and the present experiment. When a less stringent criterion is used to score stimulus repetitions among normal hearing listeners—such as four out of the six phonemes in the stimulus being repeated correctly—the performance of normal hearing listeners (as seen in the present experiment) reaches ceiling across all four of the phonotactic conditions. When the normal hearing listeners were removed from the analysis and the repetition accuracy of the High- and Low-NU-6 groups of patients are compared, the differences as a function of phonotactic probability still fail to reach significance ($F < 1$), although the differences are in the predicted direction. The equivalent repetition performance across the four conditions of phonotactic probability for the normal hearing listeners and for both groups of cochlear implant users suggests that the stimuli in each condition are equally perceptible. That is, LOW-LOW nonword patterns were not more difficult to perceive than HIGH-HIGH nonword patterns. This finding contrasts with our initial prediction, perhaps because the present task was not a speeded task as in Vitevitch et al. (1997). Finally,

the significant difference in repetition accuracy between the normal hearing listeners and the cochlear implant users underscores the difficulty that these patients had in performing this task.

In summary, the results of the first experiment suggest that cochlear implant users still have access to and may use phonotactic information--knowledge of the sounds and sequences of sounds in a word or syllable--to process spoken language. Furthermore, the results of Experiment 1 suggest that the extent to which phonotactic information is used by these patients may vary as a function of spoken word recognition ability, as measured by scores on the NU-6. To further investigate how cochlear implant users access and use phonotactic information to process spoken words, we presented stimuli varying in phonotactic probability in two additional tasks that measure on-line processing by using reaction time in addition to accuracy rates as dependent measures. Tasks that measure online processing may be more sensitive to certain aspects of linguistic representations and processes than offline or post-perceptual tasks, such as the rating task used in Experiment 1. Furthermore, these online measures may lead to new methods and assessments that can be used to develop clinical outcome measures or for treatment purposes (Tompkins, 1998).

Experiment 2

An example of a task that can be used to measure on-line processing is the AX or same-different task. In this task, a participant hears two stimuli separated by a short interval (e.g., 200msec). After hearing both signals, the listener must decide as quickly and as accurately as possible if the two stimuli they heard were the same or were different, and indicate their decision by pressing a button on a response box. The time required for the participant to respond (measured from the beginning of the second stimulus to the press of the button) and the accuracy with which the participant responds constitute the dependant measures. In Vitevitch and Luce (1999) and in the present study, the label for the "SAME" response was placed under the dominant hand on the response box. Due to the wider variability of non-dominant hand responses compared to dominant-hand responses, only reaction times from dominant hand responses were analyzed.

Using the same-different task, Vitevitch and Luce (1999) were able to better describe the influence of phonotactic information on the processing of spoken words than by using the rating task--a task that can be influenced by post-perceptual processes. For example, in the post-perceptual rating task used by Vitevitch et al. (1997) and in Experiment 1 of the present study, responses to the nonword patterns could have been based on activation from sublexical representations, lexical representations, or both types of representations. Specifically, high probability segments and sequences might have been activated to a greater degree than low probability segments and sequences. Responses based on the level of activation solely at the sublexical level may have resulted in the significant effects of phonotactic probability on the nonword ratings. Alternatively, the segments and sequences in the nonwords may have partially activated whole words in the mental lexicon (i.e., lexical representations). Nonwords with high probability segments and sequences, which are common to many words (Vitevitch, Luce, Pisoni, & Auer, 1999), would have partially activated more lexical representations than nonwords with low probability segments and sequences. Responses based on the level of activation at the lexical level may also have produced the observed results. Finally, given that there was no time pressure to make a response, listeners may have developed a cognitive strategy that combined the activation among lexical and sublexical representations, also producing the observed results. Thus, it is possible that the pattern of results observed in Experiment 1 may not have been due to the direct access of phonotactic information, but to information about multiple words indirectly activated in the lexicon.

With a task that measured on-line processing activities, Vitevitch and Luce (1999; see also Pitt & Samuel, 1995) found evidence for the hypothesis that two levels of representation are involved in the

process of spoken word recognition. In one study, Vitevitch and Luce (1999) presented normal-hearing listeners with monosyllabic words or monosyllabic nonwords varying in phonotactic probability. For *nonwords* varying in phonotactic probability, they found stimuli with high probability patterns were responded to ("SAME") more quickly than stimuli with low probability patterns. In contrast, for *real words,* stimuli with low probability patterns were responded to ("SAME") more quickly than real word stimuli with high probability patterns.

Based on these results, Vitevitch and Luce (1999) concluded that normal hearing listeners use *two* types of representations to process spoken language: lexical and sublexical. Lexical representations consist of phonological word forms, whereas sublexical representations consist of units smaller than a whole word, such as segments or sequences of segments. When lexical representations are used to process spoken stimuli, competition among similar sounding word forms results in stimuli with common sequences to be responded to more slowly than spoken stimuli with less common sequences. Note that there is a correlation between the frequency of a segment or a sequence of segments and the number of words that are activated and compete among each other for recognition. Common patterns of segments and sequences of segments are found in many words, whereas rare patterns of segments and sequences of segments are found in few words (see Luce & Pisoni, 1998; Vitevitch, Luce, Pisoni & Auer, 1999). On the other hand, when sublexical representations are used to process spoken patterns of segments, stimuli with common segments and sequences of segments are processed more quickly than stimuli with less common segments and sequences of segments. In the same-different task, Vitevitch and Luce (1999) found that participants used lexical representations to process the spoken words they heard and sublexical representations to process the nonwords they heard.

Just as Vitevitch and Luce (1999) found that a task that measured on-line processing was more sensitive to certain aspects of linguistic representations and processes in normal hearing listeners than an offline or post-perceptual task (such as the offline rating task used in Vitevitch et al., 1997), we predicted that the use of similar on-line tasks might reveal additional information about the processes and representations that cochlear implant users rely on in processing spoken language. In the present experiment, we used a subset of the monosyllabic words and nonwords varying in phonotactic probability used in Vitevitch and Luce (1999), and presented them to cochlear implant patients who differed in their spoken word recognition abilities (as measured by the NU-6) in a same-different task.

If listeners with cochlear implants use representations and processes that are similar to the representations and processes used by normal hearing listeners to process spoken words (Vitevitch & Luce, 1999), we would expect that the pattern of results for the two groups of listeners should be similar. Specifically, if cochlear implant users rely on sublexical representations to process nonwords as normal hearing listeners do in the same-different task (Vitevitch & Luce, 1999), we would expect that patients with a cochlear implant should respond more quickly to nonwords with high phonotactic probability than to nonwords with low phonotactic probability. Similarly, if cochlear implant patients rely on lexical representations to process real words as normal hearing listeners do in the same-different task (Vitevitch & Luce, 1999), then we would expect them to respond more quickly to words with low phonotactic probability than to words with high phonotactic probability.

As in the previous study, we predicted that the representations and processes used by patients with cochlear implants may vary as a function of spoken word recognition ability as measured by scores on the NU-6 test of spoken word recognition. Specifically, patients with cochlear implants who have good word recognition abilities should have more detailed lexical and sublexical representations and may use both types of representation in an optimal way, producing a pattern of results that is fundamentally similar to the normal hearing listeners of Vitevitch and Luce (1999).

In contrast, cochlear implant patients with poor word recognition abilities may not be able to construct detailed lexical and sublexical representations, or may not use both types of representation in an optimal manner. Some listeners may try to process words using sublexical representations. Others may try to process nonwords with lexical representations, or they may switch back and forth on a trial-by-trial basis between lexical and sublexical representations regardless of lexical status. An attenuation of the effect of phonotactic probability on processing for the real words in Experiment 2 of Vitevitch and Luce (1999) demonstrates such non-optimal processing in normal-hearing listeners when words and nonwords are mixed together rather than blocked in the same-different task. In the present experiment, if cochlear implant users with poor word recognition skills are unable to make optimal use of both types of processes and representation we would expect a similar attenuation of the effects of phonotactic probability for listeners in this group.

Methods

Participants

Eighteen adult users of a cochlear implant, all outpatients at Riley Hospital, Indianapolis, Indiana, were paid for their participation this experiment. Two of the participants in the present experiment had also participated in Experiment 1, which was conducted at least six months prior to participation in the present experiment. All participants were right-handed, native English speakers. Data from two other participants were not included in the final analysis because one participant was pre-lingually deafened, and the other participant experienced technical problems during testing because the battery in the processor ran out. The remaining participants were post-lingually deafened adults who had used their cochlear implant for at least a year prior to testing.

The mean age of the participants was 55.9 years old. The mean age of onset of deafness was 34.0 years old. The mean age at which implantation of a cochlear implant device took place was 53.3 years. Nine participants used the Nucleus device, 5 used the Clarion device, and 2 used the MedEl device. See Table III for individual participant information.

The cochlear implant patients were divided into two groups based on word recognition ability. A median split on the NU-6 scored by percent words correct for each user served as the criterion to divide the cochlear implant users into the two groups of eight participants each. Patients who had above average speech perception as measured by the NU-6 were in the High-NU-6 group and had a mean NU-6 scored by percent words correct of 46.75%. Patients who had average speech perception as measured by the NU-6 were in the Low-NU-6 group and had a mean NU-6 scored by percent words correct of 12.50 %. The difference in the NU-6 scores between the groups was significantly different (F (1,14) = 22.64, $p < .001$).

As in Experiment 1, the two groups of cochlear implant users did not differ in their hearing thresholds as measured by pure-tone averages, even though their speech perception abilities did differ (F (1,14) < 1). Users in the High-NU-6 group had a pure-tone average of 32.15 dB SPL, and users in the Low-NU-6 group had a pure-tone average of 31.05 dB SPL suggesting that the two groups had comparable abilities in detecting sound.

Table III. Individual Characteristics of Cochlear Implant Users in Experiments 2 and 3

	Age	Age at Onset of Deafness	Etiology	Age at Implant	Type of Implant and processing strategy	Years of CI use	NU-6 Word	NU-6 Phon	NU-6 Cond.
1	60	18	unknown	56	Clarion, CIS	4	0	13	LOW
2	44	27	otosclerosis	42	MedEl, SPEAK	2	28	53	HIGH
3*	53	49	unknown	51	Nucleus-24, SPEAK	2	50	69	HIGH
4	69	21	unknown	67	Nucleus-24, SPEAK	2	12	33	LOW
5	71	10	unknown	63	Nucleus-22, SPEAK	8	14	30	LOW
6	37	30	hereditary	34	Clarion, CIS	3	76	90	HIGH
7	68	43	cryoglubli.	62	Clarion, CIS	6	68	83	HIGH
8	69	40	miniere's	68	Clarion, CIS	1	38	61	HIGH
9	48	44	unknown	47	Nucleus-22, SPEAK	1	24	41	LOW
10*	39	3	unknown	33	Clarion, CIS	5	34	59	HIGH
11	71	63	hereditary	68	MedEl, SPEAK	3	2	32	LOW
12	40	38	unknown	39	Nucleus-24, ACE	1	34	57	HIGH
13	77	45	neuroma	75	Nucleus-24, SPEAK	2	46	70	HIGH
14	60	30	unknown	59	Nucleus-24, CIS	1	0	17	LOW
15	49	45	trauma	48	Nucleus-22, SPEAK	1	22	48	LOW
16	40	38	infection	39	Nucleus-24, ACE	1	26	59	LOW
	55.9	34.0		53.3		2.68			

Note: Two listeners also participated in Experiment 1; they are indicated by * next to the participant number. Also note that participant #10 in the present experiment was classified in the "Low-NU-6" group in Experiment 1.

Materials

Fifty of the words and 50 of the nonwords used in Vitevitch and Luce (1999) were used in this experiment. Phonotactic probability was calculated with the same two measures--positional segment frequency and biphone frequency--and with the same computerized dictionary used in Experiment 1. Words and nonwords that were classified as high-probability patterns consisted of segments with high segment positional probabilities. Words and nonwords that were classified as low-probability patterns consisted of segments with low segment positional probabilities and low biphone probabilities. For the words, the average segment and biphone probabilities were .1740 and .0070 for the high probability lists and .0960 and .0030 for the low probability lists in the present experiment. For the nonwords, the average segment and biphone probabilities were .1550 and .0050, respectively, for the high probability lists and .0670 and .0010 for the low probability lists in the present experiment.

Similarity Neighborhoods. Frequency-weighted similarity neighborhoods were computed for each stimulus by comparing a given phonemic transcription (constituting the stimulus pattern) to all other transcriptions in the lexicon (see Luce & Pisoni, 1998). A neighbor was defined as any transcription that could be converted to the transcription of the stimulus word by a one phoneme substitution, deletion, or addition in any position. The log frequencies of the neighbors were then summed for each word and nonword, rendering a frequency-weighted neighborhood density measure. The mean log-frequency-weighted neighborhood density values for the high and low probability nonwords were 41 and 13 respectively. The same values for the high and low probability words were 45 and 30 respectively.

Word Frequency. Frequency of occurrence (Kucera & Francis, 1967) was matched for the two probability conditions for the words. Average log word frequency was 2.004 for the low probability words and 2.005 for the high probability words ($F < 1$).

Durations. The average durations of the stimuli in the two phonotactic conditions were equivalent. For the words, the high probability items had a mean duration of 650 ms and the low probability items had a mean duration of 657 ms (F (1,48) <1). For the nonwords, the high probability items had a mean duration of 699 ms and the low probability items had a mean duration of 697 ms (F (1,48) < 1).

The words and nonwords were spoken one at a time in a list by the same trained phonetician who made the recordings used in Experiment 1. All the stimuli were treated in the same way as the stimuli in Experiment 1.

Procedure

Participants were tested individually. Each participant was seated in front of a Macintosh Performa 6200CD computer equipped with a PsyScope response box (with three response buttons) and an Advent AV570 speaker. The computer program PsyScope 1.2.2 (see Cohen, MacWhinney, Flatt, & Provost, 1993) controlled stimulus presentation and response collection. The response box had the label "DIFFERENT" on the left button and the label "SAME" on the right button (the middle response button was deactivated).

An experimental trial proceeded as follows: The word "READY" appeared in the center of the computer screen for 500ms to indicate the beginning of a trial. Participants were then presented with two of the spoken stimuli at 70dB SPL. The inter-stimulus interval was 150 ms. Reaction times were measured from the onset of the second stimulus in the pair to the button press response. If the maximum reaction time (3 s) expired, the computer automatically recorded an incorrect response and presented the next trial. Participants were instructed to respond as quickly and as accurately as possible on each trial. SAME responses were made with the dominant hand.

The words and nonwords were presented blocked in separate lists. Order of list presentation was counterbalanced across participants. Half of the trials consisted of two identical stimuli (constituting SAME trials) and half of the trials consisted of different stimuli. Half of the SAME pairs had high phonotactic probabilities and half had low probabilities. Non-matching stimuli were created by pairing stimulus items from the same phonotactic category. For the DIFFERENT stimulus pairs, items with the same initial phoneme and (when possible) the same vowel were paired.

Prior to the experimental trials, each participant received ten practice trials. These trials were used to familiarize the participants with the task and were not included in the final data analysis.

Results

To examine the on-line processing of phonotactic information as a function of word recognition ability, a mixed design ANOVA was performed on the mean reaction times with phonotactic probability and lexicality as within-participant factors and word recognition ability as a between participants factor. The mean reaction times for each phonotactic condition as a function of lexicality and word recognition ability are shown in Figure 4. The top panel shows data plotted from the normal hearing listeners that participated in Experiment 1 in Vitevitch and Luce (1999) for comparison. These data were not included in the statistical analyses below. The middle panel shows the reaction times from the High-NU-6 group. The bottom panel shows the reaction times from the Low-NU-6 group. Lexicality is represented on the x axis. Reaction time in milliseconds is represented on the y axis. Words and nonwords with high phonotactic probability are represented by the clear bars. Words and nonwords with low phonotactic probability are represented by the stripped bars. Accuracy rates for responding SAME in each condition

are also presented in the figure. There were no significant differences in the accuracy rates (all *F*s < 1) indicating that participants did not sacrifice speed for accuracy in making their responses.

For the reaction times, the results showed no significant main effects (all *F*s < 1) for Lexicality, Word Recognition Ability (comparing only the High- and Low-NU-6 groups), or Phonotactic Probability (F (1,14) = 3.08, p = .10). However, the non-significant main effects should be considered in the context of significant interactions between Lexicality and Phonotactic Probability (F (1,14) = 8.34, p < .05) and between Lexicality, Phonotactic Probability, and Word Recognition Ability (F (1,14) = 6.30, p < .05).

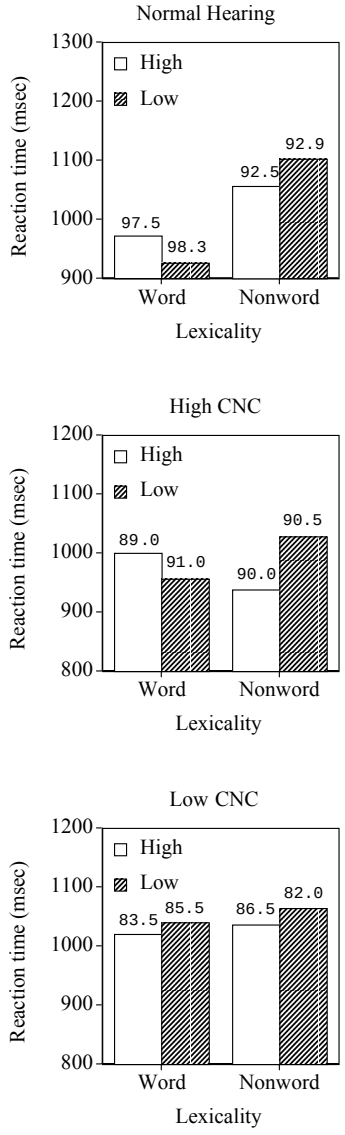


Figure 4. Mean reaction times and accuracy rates to "same" responses for normal-hearing listeners from Vitevitch and Luce (1999; top panel), better than average cochlear implant users (High-NU-6; middle panel) and average cochlear implant users (Low-NU-6; bottom panel) in the SAME-DIFFERENT task.

Subsequent analyses of the Lexicality X Phonotactic Probability interaction revealed that for real words, low probability stimuli tended to be responded to more quickly (997ms) than high probability stimuli (1009ms); however, this difference did not reach statistical significance (F (1,14) < 1). However, for the nonwords the opposite pattern was observed. High probability nonwords were responded to significantly more quickly (987ms) than low probability nonwords (1046ms; F (1,14) = 11.53, p < .01). Although not statistically significant, this pattern is fundamentally similar to the pattern of data for normal hearing listeners found in Vitevitch and Luce (1999), which is displayed in the top panel of Figure 4.

Consider now the Lexicality X Phonotactic Probability X Word Recognition Ability interaction. For the Low-NU-6 group, there were no significant differences in the response times for words and nonwords, or between high and low probability stimuli (all Fs < 1). However, for the High-NU-6 group, a different pattern of results was observed. Listeners in the High-NU-6 group tended to respond more quickly to real words with low phonotactic probability (955ms) than to real words with high phonotactic probability (999ms; F (1,14) = 4.62, p = .07). On the other hand, for nonwords, listeners in the High-NU-6 group responded significantly more quickly to high probability nonwords (937ms) than low probability nonwords (1028ms; F (1,14) = 20.06, p < .01). The pattern of data observed for the High-NU-6 group, but not the Low-NU-6 group, was similar to the pattern of data found in Vitevitch and Luce (1999) and displayed in the top panel of Figure 4.

We also computed an index of *phonotactic sensitivity* that was similar to the measure of phonotactic sensitivity developed in Experiment 1. We subtracted the mean reaction time to low phonotactic stimuli from the mean reaction time to high phonotactic stimuli for each listener for words and nonwords separately. We predicted that listeners who are more sensitive to phonotactic information would show a larger difference between the means, whereas listeners who are less sensitive to phonotactic information during on-line processing would show a smaller difference between the means. For words, this difference should be negative: listeners should respond more quickly to low probability real words than high probability real words because of competition among lexical representations. In contrast, for nonwords, this difference should be positive: listeners should respond to high probability nonwords more quickly than low probability nonwords because of facilitation among sublexical representations. Furthermore, measures of spoken word recognition ability, such as the NU-6, should be related to this index of phonotactic sensitivity for real-words if phonotactic information in the lexicon is used in the processing of spoken words. The NU-6 should not be correlated with this index of phonotactic sensitivity for nonwords if different representations are used to process nonwords, as predicted based on earlier research (Vitevitch & Luce, 1999).

A correlational analysis of phonotactic sensitivity (i.e., the difference in reaction time to stimuli with high and low phonotactic probability) and NU-6 scores (a measure of spoken word recognition) was performed to examine these predictions for the sixteen participants. For the nonwords, the index of phonotactic sensitivity and NU-6 scores were not significantly correlated (r = -.10, Z < -1, p = .69). However, for real words, the index of phonotactic sensitivity and NU-6 scores were significantly correlated (r = +.55, Z = 2.20, p < .05). Patients with higher scores on the NU-6 were better able to take advantage of the differences in phonotactic probability among the words, and therefore, had a greater difference in reaction time between the words with high and low phonotactic probability. Patients with lower scores on the NU-6 were not able to take full advantage of the differences in phonotactic probability among the words, and therefore, had a smaller difference in reaction time between the words with high and low phonotactic probability. This relationship is displayed in Figure 5.

It should be noted that the measure of on-line sensitivity to phonotactic information is not equivalent to the measure of phonotactic sensitivity developed in Experiment 1. In Experiment 1, the

correlation between phonotactic sensitivity for *nonword* stimuli and word recognition performance as scored by the NU-6 approached significance, suggesting that spoken word recognition ability might be related to sensitivity to the sequences of sound patterns found in *nonwords*. Recall, however, that the task in Experiment 1 was a word-likeness rating task. The word-likeness rating task did not have time demands (i.e., it was a post-perceptual task) and listeners were required to rate nonwords in relation to real words. Thus, participants may have had time for lexical representations to become partially activated and may have relied on the activation of those partially activated lexical representations to perform the task, even though the sound patterns they were presented with were nonwords (see also Vitevitch et al., 1997). For example, nonwords that had common segments and sequences of segments may have activated more lexical representations than nonwords that had less common segments and sequences of segments, resulting in the difference in ratings as a function of phonotactic probability observed in Experiment 1.

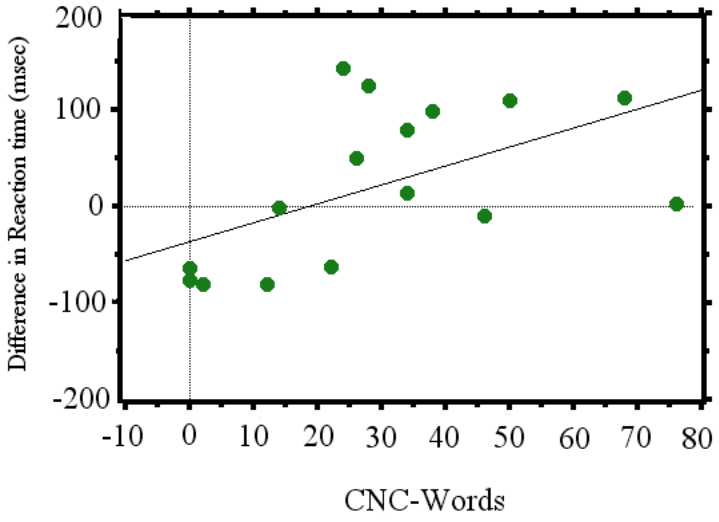


Figure 5. Scatterplot for a measure of on-line phonotactic sensitivity (the difference in reaction times to high and low probability nonword stimuli in the AX task) and a measure of spoken word recognition ability (NU-6) for the cochlear implant users in Experiment 2.

In contrast, the listeners in the present experiment were under time-pressure to respond to the nonwords quickly. This time pressure made decisions to nonwords based on partial activation of lexical representations more difficult. Thus, responses to nonwords in the present experiment were based on more completely activated sublexical representations. Real words, however, activate lexical representations more completely, allowing for decisions regarding real-words to be based on activation among lexical representations. The activation among lexical representations for real-words, and the absence of this activation for the nonwords in the present same-different task accounts for the relationship between the measure of on-line sensitivity to phonotactic information for words and the NU-6 scores, and for the lack of a relationship between the measure of on-line sensitivity to phonotactic information for nonwords and the NU-6 scores.

Discussion

The results from the same-different task used in Experiment 2 show that cochlear implant listeners with average word recognition ability (Low-NU-6 group) did not differentially respond to

stimuli varying in lexicality (word or nonword) and phonotactic probability. In contrast, cochlear implant listeners with better than average word recognition ability (High-NU-6 group) tended to respond more quickly to words with low rather than high phonotactic probability. In the case of nonwords, the group of listeners with High-NU-6 scores responded more quickly to nonwords with high rather than low phonotactic probability.

The pattern of results for the cochlear implant users with High-NU-6 scores replicates the pattern of results obtained by Vitevitch and Luce (1998) with normal-hearing listeners. Vitevitch and Luce (1998) suggested that normal-hearing listeners were making optimal use of two types of information--lexical and sublexical--to process words and nonwords. Cochlear implant patients with better than average word recognition ability (High-NU-6 scores) are also able to make optimal use of detailed lexical and sublexical representations to process spoken words and nonwords. In contrast, cochlear implant patients with average word recognition ability (Low-NU-6 scores) may not be able to construct such detailed representations to optimally process the spoken stimuli. The significant correlation between word recognition score (NU-6) and the measure of on-line sensitivity to phonotactic information for words further suggests that optimal use of detailed lexical and sublexical representations in the processing of spoken words may also be required for the accurate recognition of isolated words, especially under speeded conditions.

Less than optimal use of lexical and sublexical representations may be due to one of several factors. One possibility is that some listeners may switch back and forth between lexical and sublexical representations to process the input. In Experiment 2 of Vitevitch and Luce (1999), the researchers again used a same-different task, but mixed word pairs and nonword pairs together instead of separating them into distinct blocks as they did in Experiment 1. They found that the difference in reaction time to real words as a function of phonotactic probability was greatly attenuated. They hypothesized that the normal-hearing participants might have been switching back and forth between lexical and sublexical representations either on a trial-by-trial basis or at some unspecified point in the experiment, resulting in the observed attenuation of effects. The cochlear implant patients with poor word recognition skills in the present experiment may also have been switching between lexical and sublexical representations attempting to process the input, resulting in the attenuation of effects observed in the present study for cochlear implant users with poor word recognition abilities.

An alternative, but not necessarily independent, account may be that some cochlear implant patients may construct representations that are more coarsely coded. That is, some patients may not be able to distinguish between phonological segments that differ in voicing or place of articulation (e.g., Doyle et al., 1995). The inability to discriminate among speech sounds varying on a particular dimension may decrease the utility of phonotactic information in processing. For example, a sequence containing an initial stop, the vowel /^/, and a final stop may represent a word with high or low phonotactic probability, such as the word *cup* and the word *tug* respectively. Although there is still sequential information about the sounds contained in the words in the coarsely coded representations, the fine-grained phonetic details that allows one to discriminate between them is absent. Given the coarse coding of segmental information, listeners may be forced to rely on other types of representations, perhaps using only *lexical* information, to process spoken input. To further examine the efficiency with which lexical representations are used by patients with cochlear implant to process words and nonwords, we presented a different set of nonwords varying in phonotactic probability to the same sample of cochlear implant listeners in an auditory lexical decision task in Experiment 3.

Experiment 3

In an auditory lexical decision task, a listener hears a stimulus--either a real word in English or a nonword pattern--and must press an appropriately labeled button on a response box as quickly and as accurately as possible to indicate whether they heard a real word or a nonword. Typically the speed (i.e., reaction time) and accuracy with which listeners respond to the words are the dependent variables. However, in this case, we measured these variables in response to the specially created *nonwords* varying in phonotactic probability. Note that reaction times from only the dominant-hand are typically used in lexical decision experiments. Reaction times from the non-dominant hand are often slower and have much greater variability than responses made by the dominant hand. Thus, the label for nonwords was under the dominant hand in this task.

Our reason for focusing on the processing of nonwords in a lexical decision task comes from Experiment 3 of Vitevitch and Luce (1999; see also Vitevitch, Luce, Pisoni & Auer, 1999). In that experiment, Vitevitch and Luce hypothesized that the demands of the task--discriminating a nonword that does not have a lexical representation from a real word that does have a lexical representation--would require that only lexical representations be used for processing. If the stimulus item activates a lexical representation, listeners will respond that it was a real word. If the stimulus item fails to activate a lexical representation, listeners will respond that it was a nonword. Although sublexical representations by themselves may be useful in assessing whether two stimuli are the same or different as in Experiment 2, sublexical representations alone cannot be used to assess whether a string of phonemes is a real word or a nonword. Rather, a representation in lexical memory must be activated above some threshold for a sequence to be recognized as a real word.

If lexical representations are used to assess the specially constructed set of nonwords varying in phonotactic probability in the lexical decision task, then we might expect to see a reversal in the pattern of reaction times for the nonwords observed in the same-different task. Recall that in the same-different task, sublexical representations were used to process nonwords. In that task, listeners responded to high probability nonwords more quickly than low probability nonwords. In the present experiment, we predict that the nonwords should now be responded to as if they were real words. That is, listeners should now respond to low probability nonwords more quickly than high probability nonwords due to differences in lexical competition.

Furthermore, we predicted that listeners with above average word recognition skills should demonstrate a greater difference in reaction time between the nonwords varying in phonotactic probability than listeners with average word recognition ability. Recall that listeners with above average word recognition ability (the High-NU-6 group) are hypothesized to have more robust and detailed lexical and sublexical representations. It is further hypothesized that listeners with average word recognition skills (the Low-NU-6 group) have lexical and sublexical representations that are not as fine-grained or detailed as the representations of listeners with above average word recognition ability. Listeners who rely on only one type of representation, or on less robust, or more incomplete and underspecified representations may not be able to determine whether a sequence of sounds is a real word in English or a nonsense word to the same extent as listeners with more distinct lexical and sublexical representations. Listeners with more robust, well-specified representations may be more efficient at identifying and recognizing spoken words because these two types of representation interact during processing to further discriminate among possible candidates activated in memory. To further investigate the on-line use of phonotactic information by patients with cochlear implants, listeners were asked to listen to sound patterns and determine as quickly and as accurately as possible whether the sequence was a real word in English or a nonsense word.

Method

Participants

The same listeners who took part in Experiment 2 also participated in the present experiment. Data from the two listeners that were excluded from Experiment 2 were also not analyzed in the present experiment.

Materials

A different set of 50 real words and 50 nonwords used in Vitevitch and Luce (1999) were used in this experiment. The stimuli used in the present experiment were not presented in Experiment 2. Words and nonwords that were classified as low-probability patterns consisted of segments with low segment positional probabilities and low biphone probabilities. For the words, the average segment and biphone probabilities were .2170 and .0110 for the high probability lists and .1440 and .0050 for the low probability lists in the present experiment. For the nonwords, the average segment and biphone probabilities were .1730 and .0070, respectively, for the high probability lists and .0570 and .0010 for the low probability lists in the present experiment.

Similarity Neighborhoods. Frequency-weighted similarity neighborhoods were computed for each stimulus in the same manner as in Experiment 2. The mean log-frequency-weighted neighborhood density values for the high and low probability words were 52 and 39 respectively. The same values for the high and low probability nonwords were 44 and 12 respectively.

Word Frequency. Frequency of occurrence (Kucera & Francis, 1967) was matched for the two probability conditions for the words. Average log word frequency was 2.33 for the low probability words and 2.30 for the high probability words ($F < 1$).

Durations. The durations of the stimuli in the two phonotactic conditions were equivalent. For the words, the high probability items had a mean duration of 665 ms and the low probability items had a mean duration of 671 ms (F (1,48) <1). For the nonwords, the high probability items had a mean duration of 691 ms and the low probability items had a mean duration of 689 ms (F (1,48) < 1).

The words and nonwords were spoken one at a time in a list by the same trained phonetician who made the recordings in Experiment 1. All the stimuli were treated in the same way as the stimuli in Experiment 1.

Procedure

Participants were tested individually with the same equipment used in Experiment 2. In the present experiment, the response box had the label “WORD” on the left button and the label “NONWORD” on the right button. Note that the responses to words and nonwords in Vitevitch and Luce (1999) were made by different groups of participants. One group of participants had the WORD label under the dominant hand and the other group had the NONWORD label under the dominant hand. The WORD and NONWORD responses in the present investigation were made by the same group of cochlear implant users with the WORD label under the non-dominant hand and the NONWORD label under the dominant hand. Thus, one must exercise caution in interpreting the WORD responses in the present experiment.

A trial proceeded as follows: The word "READY" appeared in the center of the computer screen for 500ms to indicate the beginning of a trial. Participants were then presented with one of the randomly selected spoken stimuli at 70dB SPL. Reaction times were measured from the onset of the stimulus to the button press response. If the maximum reaction time (3 s) expired, the computer automatically recorded an incorrect response and presented the next trial. Participants were instructed to respond as quickly and as accurately as possible. NONWORD responses were made with the dominant hand.

Half of the trials consisted of real words in English, half of the trials consisted of the nonwords. Also, an equal number of words and nonwords had high and low phonotactic probabilities. Prior to the experimental trials, each participant received ten practice trials. These trials were used to familiarize the participants with the task and were not included in the final data analysis.

Results

To examine the on-line processing of phonotactic information as a function of word recognition ability, a mixed design ANOVA was performed on the mean reaction times with phonotactic probability as a within-participants factor and word recognition skill as a between participants factor. The word recognition skill condition consisted of the same two groups of cochlear implant users as in Experiment 2. Recall that listeners in the High-NU-6 group had significantly higher scores on the NU-6 than listeners in the Low-NU-6 group as determined by a median split of the NU-6 scores. Also recall that the two groups of listeners did not differ in their hearing thresholds as measured by pure-tone averages.

The mean reaction times for each phonotactic condition as a function of lexicality and word recognition ability are shown in Figure 6. The top panel shows data plotted from the normal hearing listeners that participated in Experiment 3 in Vitevitch and Luce (1999). These data are presented for comparison only and were not included in the following analyses. The middle panel shows the reaction times from the High-NU-6 group. The bottom panel shows the reaction times from the Low-NU-6 group. Lexicality is represented on the *x* axis. "High" refers to words and nonsense words with high phonotactic probability. "Low" refers to words and nonsense words with low phonotactic probability. Accuracy rates for responding NONWORD to the nonwords and WORD to the words in each condition are also presented in the figure.

For the reaction times among the patients with cochlear implants, the main effect of word recognition skill was not significant. That is, there was no significant difference in overall reaction time (F (1,14) = 1.11, $p > .30$) between the two groups of patients (High- and Low-NU-6).

There was a significant main effect of lexicality (F (1,14) = 5.80, $p < .05$), such that words (1349 msec) were responded to more quickly than nonwords (1446 msec) by the cochlear implant patients, even though WORD responses were made with the non-dominant hand that are typically slower than dominant hand responses. More interestingly, the interaction of lexicality and word recognition ability was significant (F (1,14) = 5.04, $p < .05$). Words were responded to more quickly than nonwords for the High-NU-6 group, but not for the Low-NU-6 group. Additional analyses (F (1,7) = 6.47, $p < .05$) confirmed that listeners in the High-NU-6 group responded to words (1251 msec) more quickly than to nonwords (1438 msec), whereas listeners in the Low-NU-6 group did not differentially respond (F (1,7) < 1) to words (1447 msec) and nonwords (1454 msec). These results suggest that patients in the High-NU-6 group were able to discriminate between words and nonwords at some level of processing. Listeners in the Low-NU-6 group were unable to discriminate any differences between words and nonwords. None of the other main effects or interactions were significant for the reaction times (all $p > .10$). Finally, there were no significant differences among the accuracy rates (all $p > .10$).

As in Experiment 2, we calculated a measure of *on-line phonotactic sensitivity* by subtracting the mean reaction time to low phonotactic stimuli from the mean reaction time to high phonotactic stimuli for each listener for the nonwords separately. Because only the High-NU-6 group responded differentially to real words and nonwords in terms of reaction time, measures of on-line phonotactic sensitivity to the nonword stimuli in the lexical decision task were calculated only for the High-NU-6 group. Also, a measure of on-line sensitivity to the real words was not calculated because real word responses were made with the non-dominant hand.

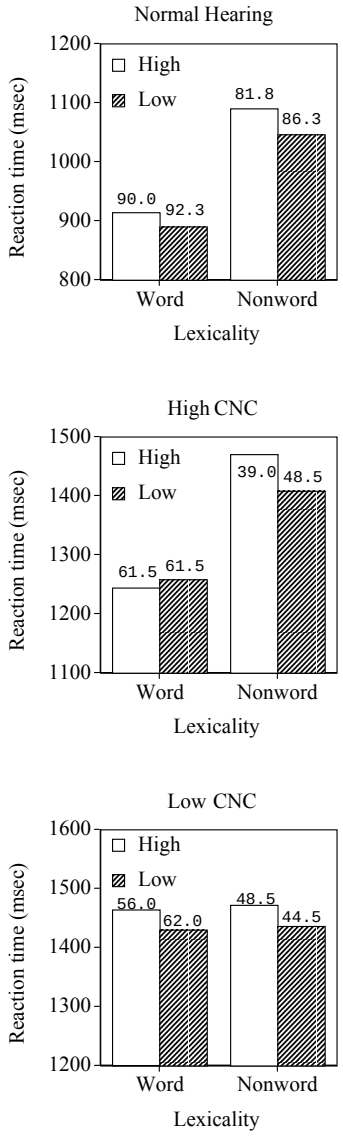


Figure 6. Mean reaction times and accuracy rates to "nonword" responses for normal-hearing listeners from Vitevitch and Luce (1999; top panel), better than average cochlear implant users (High-NU-6; middle panel) and average cochlear implant users (Low-NU-6; bottom panel) in the lexical decision task. "Word" responses for the cochlear implant users were made with the non-dominant hand.

We predicted that the patients in the High-NU-6 group who were more sensitive to the phonotactic information in the nonwords would show a greater difference between the reaction time means. In contrast, the patients who were less sensitive to the phonotactic information in the nonwords during on-line processing would show a smaller difference between the means. If these patients rely primarily on lexical representations rather than sublexical representations to process the nonword stimuli, the difference in the response times between high and low probability nonwords should be negative. That is, listeners should respond more quickly to low probability nonwords than to high probability nonwords because of competition among lexical representations that have been activated by the nonwords. Measures of spoken word recognition skill, such as the score on the NU-6, should be related to the on-line measure of phonotactic sensitivity for the nonwords in the lexical decision task if phonotactic information among lexical representations is used in the processing of spoken words. Listeners with better word recognition ability, even within the High-NU-6 group, should then show greater sensitivity to phonotactic information, whereas listeners with poorer word recognition ability should show less sensitivity to phonotactic information.

An examination of the on-line phonotactic sensitivity for the nonwords in the lexical decision task and the NU-6 scores for the High-NU-6 group revealed that the correlation between these two measures for the nonwords approached significance (r = -.67, Z = -1.84, p = .06). This relationship is displayed in Figure 7. Although the correlation was not statistically significant at the traditional p-value of .05, the consistency of this result with the pattern of results obtained across the other experiments suggests that this marginal effect is more than just Type-I error. Overall, the results suggest that the sensitivity to phonotactic information in nonwords and the skills used to recognize isolated words are closely related and draw on the same types of information.

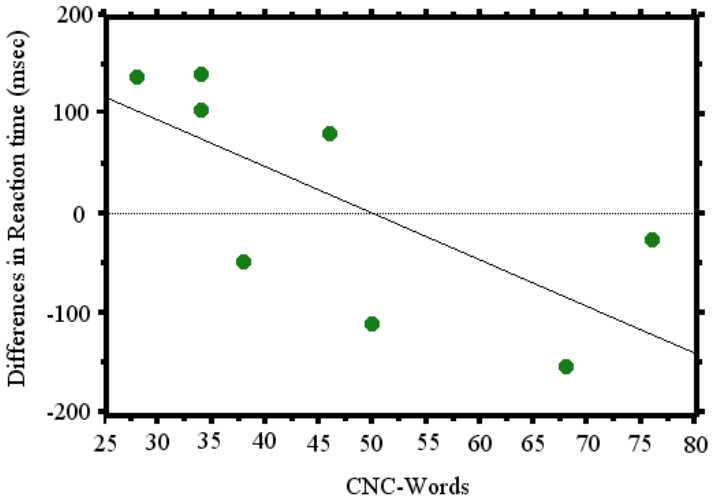


Figure 7. Scatterplot for a measure of on-line phonotactic sensitivity (the difference in reaction times to high and low probability nonword stimuli in the lexical decision task) and a measure of spoken word recognition ability (NU-6) for above average cochlear implant users in Experiment 3.

Discussion

The results from the lexical decision task used in Experiment 3 show that patients with cochlear implants who have average word recognition skills (Low-NU-6 group) were unable to respond differentially to stimuli varying in lexicality (word or nonword) or phonotactic probability. In contrast, cochlear implant patients with better than average word recognition skill (High-NU-6 group) responded more quickly to words than to nonwords, replicating a pattern commonly found in normal-hearing listeners (e.g., Chambers & Forster, 1975; Forster & Chambers, 1973), despite making the response to words with their non-dominant hand. Although listeners in the High-NU-6 group tended to respond to low probability nonwords more quickly (1408 msec.) than high probability nonwords (1469 msec.), this difference was not statistically significant at the $p < .05$ level. When an on-line measure of phonotactic sensitivity was calculated for the nonword responses from the High-NU-6 listeners, a negative correlation ($r = -.67$) that approached significance ($p = .06$) was found between this measure and listeners' scores on the NU-6. This pattern of results suggest that patients with cochlear implants who have better than average word recognition skills are able to make optimal use of detailed lexical and sublexical representations to process the spoken stimuli. That is, information about the sounds and sequences of sounds in a word (i.e., phonotactic information) may still be included in the cognitive repertoire of some cochlear implant users.

In contrast, cochlear implant patients with average word recognition skill may not have such detailed representations to optimally process spoken input and may rely on alternative cognitive strategies. The inability of average users of a cochlear implant (the Low-NU-6 group) to even differentially respond to words and nonwords further suggests that these listeners are not relying on optimal processes and representations of sound-based information. At present, the exact nature of the processes and representations used by average cochlear implant listeners is unclear. Such listeners may be relying on either lexical or sub-lexical representations that are more coarsely coded than analogous representations in normal-hearing listeners or better than average cochlear implant users. Alternatively, listeners may switch back and forth between lexical and sublexical representations to process the input, or may rely solely on alternative representations to process spoken input. As stated earlier, both accounts may ultimately interact and influence each other.

General Discussion

The results of Experiments 1-3 demonstrate the importance of using behavioral tasks that measure on-line processing along with tasks that measure post-perceptual processing. The results of Experiment 1 suggested that patients with cochlear implants who have average and above average word recognition skills were able to access information related to the sequences of sounds (i.e., phonotactic information) to make judgments of spoken nonwords. In contrast, the results of Experiments 2 and 3 indicate that only those cochlear implant patients with above average word recognition skill use this information for the on-line processing of spoken input. Although cochlear implant patients with only average word recognition skills can access phonotactic information to make explicit judgments of spoken nonwords, they may not rely on this information consistently or optimally, and they may not use this information to process spoken input in real time under speeded conditions. We hypothesized that one possible reason cochlear implant users with average word recognition ability may not rely on phonotactic information may be related to the ability to discriminate among the finer details of lexical or sublexical representations. The ability to discriminate among the finer details of lexical or sublexical representations should not be confused with the ability of the patients to detect sounds. Recall that in all three experiments, patients in the High- and Low-NU-6 groups had equivalent hearing thresholds as measured by pure-tone averages, suggesting that the locus of these effects are not in peripheral or sensory systems.

Rather, this research examined the ability of cochlear implant patients to encode and represent the fine phonetic details of speech.

Further experimentation examining the time course of processing phonotactic information is required. The present experiments, however, are important in demonstrating that some patients with cochlear implants can access and process information about the probability of segments and sequences of segments in words and nonwords (i.e., phonotactic information), much like normal-hearing listeners. Furthermore, this source of information is correlated with performance in recognizing isolated spoken words. The relationship observed here between phonotactic sensitivity and spoken word recognition suggests that interventions that explicitly focus attention on phonotactic relationships among sound patterns in words and nonwords may help less successful users develop improved spoken word recognition abilities, and therefore receive greater benefit from their cochlear implant. Additional work will be required to examine the efficacy of such interventions and identify the locus of any effects of these methods in changing the word recognition and comprehension skills of patients with cochlear implants. The three tasks that were used and the index of phonotactic sensitivity developed in the experiments reported could also offer a new method for measuring and assessing outcome of word recognition and comprehension skills for patients who receive cochlear implants.

References

Auer, E.T. (1993). *Dynamic processing in spoken word recognition: The influence of paradigmatic and syntagmatic states*. Unpublished doctoral dissertation, University at Buffalo, Buffalo, NY.

Blamey, P.J., Dowell, R.C., Brown, A.M., Clark, G.M., & Seligman, P.M. (1987). Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. *Journal of the Acoustical Society of America, 82*, 48-57.

Brent, M.R. & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93-125.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111-153.

Chambers, S. M. & Forster, K. I. (1975). Evidence for lexical access in a simultaneous matching task. *Memory and Cognition, 3*, 549-559.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, and Computers, 25*, 257-271.

Cohen, N.L., Waltzman, S. & Shapiro, W.H. (1989). Telephone speech comprehension with use of the Nucleus cochlear implant. *Annuls of Otology, Rhinology and Laryngology, 98, Supplement 142*, 8-11.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113-121.

Crystal, D. (ed.) (1980). *A First Dictionary of Linguistics and Phonetics*. London: Andre Deutsch.

Dowell, R.C., Mecklenburg, D.J. & Clark, G.M. (1986). Speech recognition for 40 patients receiving multichannel cochlear implants. *Acta Otolaryngologica, 12*, 1054-1059.

Doyle, K.J., Mills, D., Larky, J., Kessler, D., Luxford, W.M. & Schindler, R.A. (1995). Consonant perception by users of Nucleus and Clarion multichannel cochlear implants. *The American Journal of Otology, 16*, 676-681.

Forster, K.I. & Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Gantz, B.J., Tyler, R.S., Knutson, J.F., Woodworth, G.C., Abbas, P., McCabe, B.F., Hinrichs, J., Tye-Murray, N., Lansing, C., Kuk, F., & Brown, C. (1988). Evaluation of five different cochlear implant designs: Audiologic assessment and predictors of performance. *Laryngoscope, 98*, 1100-1106.

Geier, L., Fisher, L., Barker, M., & Opie, J. (1999). The effect of long-term deafness on speech recognition in postlingually deafened adult Clarion cochlear implant users. *Annals of Otology, Rhinology & Laryngology, 108*, 80-83.

Gathercole, S.E., Willis, C., Emslie, H., & Baddeley, A.D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349-367.

Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267-333.

Holden, L.K., Skinner, M.W., & Holden, T.A. (1997). Speech recognition with the MPEAK and SPEAK speech-coding strategies of the Nucleus Cochlear Implant. *Otolaryngology-Head and Neck Surgery, 116*, 163-167.

Hollow, R.D., Dowell, R.C., Cowan, R.S.C., Skok, M.C., Pyman, B.C., & Clark, G.M. (1995). Continuing improvements in speech processing for adult cochlear implant patients. *Annals of Otology, Rhinology, & Laryngology, 166*, 292-294.

Jusczyk, P.W., Frederici, A.D., Wessels, J.M.I., Svenkerud, V.Y. & Jusczyk, A. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory & Language*, 32, 402-420.

Jusczyk, P.W., P.A. Luce, & J. Charles-Luce (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language, 33*, 630-645.

Kessler, B. & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory & Language, 37*, 295-311.

Kucera, H. & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Malmkaer, K. (ed.) (1991). *The Linguistics Encyclopedia*. Routledge: London.

Mattys, S.L., Jusczyk, P.W., Luce, P.A., & Morgan, J.L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

Messer, S (1967). Implicit phonology in children. *Journal of Verbal Learning and Verbal Behavior, 6*, 609-613.

Pitt, M.A. & Samuel, A.G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29*, 149-188.

Saffran, J.R., Newport, E., & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory & Language, 35*, 606-621.

Savin, H.B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America, 35*, 200-206.

Skinner, M., Holden, L., Holden, T., Dowell, R., Seligman, P., Brimacombe, J., & Beiter, A. (1991). Performance of postlingually deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the Nucleus multi-channel cochlear implant. *Ear & Hearing, 12*, 3-22.

Solomon, R.L. & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology, 43*, 195-201.

Staller, S., Menapace, C., Domico, E., Mills, D., Dowell, R.C., Geers, A., Pijl, S., Hasenstab, S., Justus, M., Bruelli, T., Borton, A.A., & Lemay, M. (1997). Speech perception abilities of adults and pediatric Nucleus implant recipients using Spectral Peak (SPEAK) coding strategy. *Otolaryngology-Head and Neck Surgery, 117*, 236-242.

Storkel, H.L. & Rogers, M.A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics and Phonetics, 13*, in press.

Tompkins, C.A. (1998). Special forum on online measures of comprehension: Implications for Speech-Language pathologists. *American Journal of Speech-Language Pathology, 7,* 48.

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (1996). English speakers' sensitivity to phonotactic patterns. *Paper for volume on Fifth Conference on Laboratory Phonology*.

Vitevitch, M.S. & Luce, P.A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9*, 325-329.

Vitevitch, M.S. & Luce, P.A. (1999). Probabilistic phonotactics and spoken word recognition. *Journal of Memory & Language, 40*, 374-408.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47-62.

Vitevitch, M.S., Luce, P.A., Pisoni, D.B., & Auer, E.T. (1999). Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language, 68*, 306-311.

Wilson, B.S. (2000). Cochlear implant technology. In J. Niparko et al. (eds.) *Cochlear Implants: Principles and Practices*. Philadelphia: Lippincott, Williams & Wilkins. (pp. 109-127).

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Effects of Talker Variability and Lexical Competition on Audiovisual Word Recognition by Adult Users of Cochlear Implants [1]

Adam R. Kaiser,[2] Karen Iler Kirk,[2] Lorin Lachs and David B. Pisoni[3]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1]This work was supported by NIH-NIDCD Grants K23 DC00126, R01 DC00111, and T32 DC 00012. Support also was provided by Psi Iota Xi National Sorority. We thank Marcia Hay-McCutcheon and Stacey Yount for their assistance in data collection and management. We also are grateful to Luis Hernandez and Marcelo Areal for their development of the software used for stimulus presentation and data collection.

[2] DeVault Otologic Research Laboratory, Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN

[3] Also DeVault Otologic Research Laboratory, Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN

Effects of Talker Variability and Lexical Competition on Audiovisual Word Recognition by Adult Users of Cochlear Implants

Abstract. The present study examined how postlingually deafened adults with cochlear implants (CIs) combine visual information from lip-reading with auditory cues in an open set word recognition task. Adults with normal hearing served as a comparison group. Word recognition performance was assessed using lexically controlled word lists presented under audio-alone (A), visual-alone (V), and combined audiovisual (AV) presentation formats. Effects of talker variability were studied by manipulating the number of talkers producing the stimulus tokens. Lexical competition was investigated using sets of lexically easy and lexically hard test words. To assess the degree of audiovisual integration above and beyond simple additive cues, an index, I, was developed. A measure of visual enhancement, R, was also used to assess the gain in performance provided in the AV condition relative to the maximum possible performance obtainable in the audio-alone format. Results showed that word recognition performance was highest for AV presentation followed by A and then V. Performance was better for single-talker lists than for multiple-talker lists, particularly under the AV presentation format. Word recognition performance was better for the lexically easy than for the lexically hard words regardless of presentation format. Visual enhancement scores were higher for single talker conditions compared to multiple talker conditions and tended to be somewhat better for lexically easy words than for lexically hard words. The pattern of results suggests that information from the auditory and visual modalities is used to access common, multimodal lexical representations in memory. The findings are discussed in terms of the complementary nature of auditory and visual sources of information that specify the same underlying gestures and articulatory events in speech.

Cochlear implants (CIs) are electronic auditory prostheses for individuals with severe to profound hearing impairment that enable many of implant users to perceive and understand spoken language. However, the benefit to an individual user varies greatly. Audio-alone performance measures have demonstrated that some users of cochlear implants are able to communicate successfully over a telephone even when lip-reading cues are unavailable (Dorman, Dankowski, McCandless, Parkin, & Smith, 1991). Other users display little benefit in open-set speech perception tests under audio-alone listening conditions, but find that the cochlear implant helps them understand speech when visual information also is available. One source of these individual differences is undoubtedly the way in which the surviving neural elements in the cochlea are stimulated with electrical currents provided by the speech processor (Fryauf-Bertschy, Tyler, Kelsay, Gantz, & Woodworth, 1997). Other sources of variability, however, may result from the way in which these initial sensory inputs are coded and processed by higher centers in the auditory system. For example, listeners with detailed knowledge of the underlying phonotactic rules of English may be able to use limited or degraded sources of sensory information in conjunction with linguistic knowledge to achieve better overall performance.

Fortunately, in everyday experience, speech communication is not limited to input from only one sensory modality. Optical information about speech obtained from lip reading improves speech understanding in listeners with normal hearing (Sumby & Pollack, 1954) as well as persons with CIs (Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997b). While lip reading cues enhance speech perception, the sensory, perceptual, and cognitive processes underlying this gain in performance are not well understood at this time. In one of the first studies to investigate audiovisual integration, Sumby and Pollack (1954) demonstrated that lip-reading cues greatly enhance the speech perception performance of

normal hearing listeners, especially when the acoustic signal is masked by noise. They found that performance on closed-set word recognition tasks increased substantially under audiovisual presentation compared to audio-alone presentation. This increase in performance was comparable to the gain observed when the auditory signal was increased by 15 dB SPL under audio-alone perception conditions (Summerfield, 1987).

Numerous studies have demonstrated that visual information from lip reading improves speech perception performance over audio-alone conditions for adults with normal-hearing (Massaro & Cohen, 1995) and for adults with mild to moderate hearing impairment (Grant, Walden, & Seitz, 1998; Massaro & Cohen, 1999). The cognitive processes by which individuals combine and integrate auditory and visual speech information with lexical and syntactic knowledge has become an important area of research in the field of speech perception. Audiovisual speech perception appears to be more than the simple addition of auditory and visual information (Bernstein, Demorest, & Tucker, 2000; Massaro & Cohen, 1999). A well-known example of the robustness of audiovisual speech perception is the "McGurk effect" (McGurk & MacDonald, 1976). When presented with an auditory /ba/ stimulus and a visual /ga/ stimulus many listeners report hearing an entirely new stimulus: a perceptual /da/. The McGurk and MacDonald study demonstrates that information from separate sensory modalities can be combined to produce percepts that differ predictably from either the auditory or the visual percept alone. However, these findings are not universal across all individuals (see Massaro & Cohen, 2000). Grant and Seitz (1998) suggested that listeners who are more susceptible to the McGurk effect also are better at integrating auditory and visual speech cues. Grant and his colleagues proposed that some listeners could improve consonant perception skills by as much as 26% by sharpening their integration abilities (Grant et al., 1998). Their findings on audiovisual speech perception may have important clinical implications for deaf and hearing-impaired listeners because consonant perception accounted for approximately half of the variance of word and sentence recognition.

Stimulus Variability and Spoken Word Recognition

In addition to differences in audiovisual integration, listeners with cochlear implants may also differ in their ability to perceive speech from a variety of different talkers and to deal with the resulting variability in the acoustic-phonetic properties of speech. Listeners with normal hearing reliably extract invariant phonological and semantic information from speech, even when the utterances are produced by different talkers using different speaking rates or dialects, different styles, or under adverse listening environments (Pisoni, 1993; Pisoni, 1996). The processes by which listeners recognize words and extract meaning from widely divergent acoustic signals is often referred to as perceptual constancy or perceptual normalization. The ability of listeners with normal hearing to normalize speech from different talkers under audio-alone presentation has been demonstrated (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow & Pisoni, 1999; Nygaard & Pisoni, 1995; Nygaard & Pisoni, 1998). However, little is known about the effects of talker variability on speech perception under audiovisual presentation conditions by normal-hearing listeners (Demorest & Bernstein, 1992; Lachs, 1996; Lachs, 1999), or by hearing-impaired listeners with cochlear implants.

One of the first studies to examine the effects of talker variability on spoken word recognition was performed by Creelman (1957). He presented lists of words consisting of tokens produced by 1 to 8 talkers to normal-hearing listeners in noise. He found poorer speech intelligibility for lists containing tokens produced by two or more talkers than lists produced by only one talker. Subsequent studies have demonstrated similar findings for normal hearing listeners (Mullennix, Pisoni, & Martin, 1989; Sommers, Nygaard, & Pisoni, 1994) and for listeners with hearing loss (Kirk, Pisoni, & Miyamoto, 1997; Sommers, Kirk, & Pisoni, 1997).

It is not clear, however, if these findings can be generalized to adult users of cochlear implants. One reason is that some talker specific attributes, such as a talker's fundamental frequency, may not be well represented in the electrical stimulation pattern provided by the current generation of multichannel cochlear implants. If the percepts elicited by changes in fundamental frequency play a role in mediating talker effects, then one might expect the effects of talker variability to be different for listeners with normal hearing than for hearing-impaired listeners with cochlear implants. Cochlear implant users may be unable to discriminate the subtle differences between similar talkers.

One explanation of the effects of talker variability on spoken word recognition is that perceptual normalization increases processing demands and may divert limited cognitive resources that are normally used for speech perception (Mullennix et al., 1989; Sommers et al., 1994). This hypothesis can account for the decrease in speech perception performance for word lists produced by multiple talkers. An alternative explanation of these findings is that listeners learn to focus on the acoustic cues present in a particular individual's voice and then use these "talker-specific" cues to help them in perceiving single-talker word lists (Nygaard & Pisoni, 1998). The first proposal suggests that multi-talker lists produce a decrement in speech perception performance because cognitive resources are diverted from the normal processes of speech perception to operations needed for perceptual normalization (Mullennix & Pisoni, 1990). The second account suggests that the perceptual advantage for single-talker lists over multiple-talker lists is due to processes related to perceptual learning, attunement, and talker-specific adaptation or adjustment to an individual talker's voice (Nygaard & Pisoni, 1998).

Lexical Effects on Spoken Word Recognition

Audiovisual speech perception is a complex process in which information from separate auditory and visual sensory modalities is combined with prior linguistic knowledge stored in long-term memory. Several researchers have argued that the process of speech perception is fundamentally the same regardless of the conditions under which it is performed, implying that the nervous system processes optical and auditory cues in a similar manner using the same perceptual and linguistic mechanisms (Stein & Meredith, 1993). Recently, Grant et al. (1998) have proposed a conceptual framework that can be used to assess this proposal. Their approach combines top-down cognitive processes and bottom-up sensory processes to account for performance on audiovisual speech perception tasks. Grant et al. argue that audiovisual integration takes place prior to the influence of higher-level lexical factors. While their approach acknowledges that lexical factors may influence the perception of auditory and visual speech cues, they claim that the largest increases in audiovisual speech perception occur when the information present in the auditory and visual signals are complementary and specify the same underlying phonetic events expressed in the talker's articulation.

The Neighborhood Activation Model of Spoken Word Recognition

One way to investigate the perception of audio and visual speech cues and assess the effects of the lexicon on word recognition is to measure speech perception and audiovisual integration abilities using words that have different lexical properties. The Neighborhood Activation Model (NAM) provides a theoretical framework for understanding how spoken words are recognized and identified from sensory inputs (Luce & Pisoni, 1998). More specifically, NAM provides a theoretical basis for explaining why some words are easy to identify and other words are hard to identify. The NAM assumes that a stimulus input activates a set of similar acoustic-phonetic patterns in memory, known as a lexical neighborhood. The activation level of each word pattern is proportional to the degree of similarity between the acoustic-phonetic input of the target word and the acoustic-phonetic patterns stored in memory in a multidimensional acoustic-phonetic space. Lexical properties also strengthen or attenuate these levels of activation for particular sound patterns. In the NAM, a word's level of activation is proportional to its

word frequency, the frequency with which it occurs in the language. The probability of matching a given sensory input to a particular stored lexical pattern is based on the activation level of the individual pattern and the sum of the activation levels of all of the sound patterns selected (Luce & Pisoni, 1998).

The NAM uses information about a word's lexical neighborhood, its acoustic-phonetic similarity space, to predict whether it will be relatively easy or relatively hard to perceive. In one version of the NAM, words are considered to be lexical neighbors, (i.e., part of the same activation set), if they differ from a target word by the addition, deletion, or substitution of a single phoneme. For example, *scat, at,* and *cap* are neighbors of the target word *cat.* For a given target word, the number of lexical neighbors is called the neighborhood density of the word. Words from "dense" lexical neighborhoods have many similar sounding words, whereas words from "sparse" neighborhoods have fewer similar sounding words. Neighborhood frequency is the average frequency of all the words in the neighborhood of a target word. Using these lexical characteristics and word frequency, it is possible to construct two sets of words that differ in lexical discriminability. Lexically easy words are high frequency words from low-density lexical neighborhoods with low neighborhood frequency whereas lexically hard words are low frequency words from high-density lexical neighborhoods with high neighborhood frequency. Luce and Pisoni (1998) have shown that lexically easy words are identified faster and more accurately than lexically hard words under auditory only presentation.

Lexical and Talker Effects on Audiovisual Speech Integration

Although there has been a great deal of research on audiovisual integration and multimodal speech perception in both normal-hearing and hearing-impaired listeners in the last few years, the contribution of the lexicon and knowledge of the sound patterns of words in the language has not been studied before and may provide important new insights into the large individual differences in outcome in patients with cochlear implants.

New knowledge about the process of audiovisual integration in deaf patients with CIs can be obtained by comparing the intelligibility of lexically easy and lexically hard words under different presentation formats. If the differences in intelligibility between lexically easy and lexically hard words are similar regardless of presentation format, this would suggest that the processes of audiovisual integration take place prior to lexical selection and contact with stored knowledge about words in long-term memory. If the differences in intelligibility between lexically easy and lexically hard words are influenced by presentation format, this would suggest that extensive interactions between sensory processing, multimodal integration, and lexical selection take place at a very early stage in the word recognition process.

It is well known that audiovisual speech perception often provides large benefits to individuals with hearing impairment, including cochlear implant recipients (Erber, 1972; Erber, 1975; Tyler et al., 1997a). In every day activities, listeners with cochlear implants perceive speech in a wide variety of contexts including television, face-to-face conversation, and over the telephone. Success in recognizing words and understanding the talker's intended message may differ quite substantially under these diverse listening conditions. The primary goal of this study was to examine the ability of cochlear implant users to integrate the limited auditory information they receive from their implant with visual speech cues during spoken word recognition. To achieve this goal, we examined the effects of lexical and talker variability on word recognition under three presentation formats: audio-alone (A), visual-alone (V) and auditory-plus-visual (AV).

Methods

Participants

Forty-one adults served as listeners in this study and were paid for their participation. Twenty were postlingually deafened adult users of cochlear implants who were recruited from the clinical population at Indiana University (Table 1). All of these listeners had profound bilateral sensorineural hearing losses and had used their cochlear implant for at least six months. Their mean age at time of testing was 50 years. The comparison group consisted of 21 listeners with self-reported normal hearing. They were recruited from within Indiana University and the associated campuses through newspaper and e-mail advertisements and announcements. These participants averaged 42 years of age. All of the listeners in the comparison group had pure tone thresholds below 25 dB HL at 250, 500, 1000, 2000, 3000, and 4000 Hz and below 30 dB HL at 6000 Hz. Each participant was reimbursed for travel to and from testing sessions and was paid $10.00 per hour of testing.

Participant	Age at Test (years)	Onset of Deafness (years)	CI Use (Months)	Implant Type	Processor	Strategy
CI1	69	65	30	N22	Spectra	SPEAK
CI2	43	27	24	MedEl	Combi40	n-of-m
CI3	51	50	8	N24	Sprint	SPEAK
CI4	71	36	6	MedEl	Combi40+	CIS
CI5	35	30	48	N22	Spectra	SPEAK
CI6	42	39	6	N24	Sprint	SPEAK
CI7	59	38	108	N22	Spectra	SPEAK
CI8	62	56	60	N22	MSP	MPEAK
CI9	36	34	6	N24	Sprint	SPEAK
CI10	59	50	107	N22	Spectra	SPEAK
CI11	45	40	60	N22	Spectra	MPEAK
CI12	19	5	96	N22	Spectra	SPEAK
CI13	49	41	6	N24	Sprint	ACE
CI14	34	30	36	Clarion	Clarion	CIS
CI15	66	57	63	Clarion	Clarion	CIS
CI16	74	65	34	N22	Spectra	SPEAK
CI17	68	58	12	Clarion	Clarion	CIS
CI18	40	9	18	Clarion	Clarion	CIS
CI19	37	3	57	Clarion	Clarion	CIS
CI20	44	43	12	Clarion	Clarion	CIS

Table 1. Demographics of patients with cochlear implants.

Stimulus Materials

The stimulus materials used in the present investigation were drawn from a large database of digitally recorded audiovisual speech tokens (Sheffert, Lachs, & Hernández, 1996). This database contains 300 monosyllabic English words produced by five male and five female talkers. For the present study, we created six equivalent word lists that would allow us to examine the effect of presentation format, talker variability, and lexical competition on spoken word recognition. Each test list contained 36

words. On each list, half of the words were lexically easy, and half were lexically hard. Lexical density was calculated for each word by counting the number of lexical neighbors using the Hoosier Mental Lexicon database (Nusbaum, Pisoni, & Davis, 1984). The word frequency values represented the number of times each target word occurred per one million words of text (Kucera & Francis, 1967). Two versions of each of the six original word lists were produced: one version contained tokens produced by a single talker. The second version contained tokens produced by six different talkers. This arrangement enabled us to administer a single-talker or multiple-talker version of each test list.

Balanced Word List Generation. The specific audiovisual stimulus tokens used in the 12 word lists were selected from the digital database using intelligibility data obtained from undergraduate psychology students at Indiana University in two earlier investigations (Lachs & Hernández, 1998; Sheffert et al., 1996). In these intelligibility studies, different groups of students listened to the words produced by each of the talkers in the database and typed the word they perceived into a computer. Separate groups of listeners were used for each talker under each presentation mode (A, V, and AV). The average intelligibility of each word produced by each talker was computed separately under each of the three presentation formats. In creating the final word lists, 216 words were selected from seven of the talkers using a customized computer program. This program generated equivalent word lists within a given presentation format regardless of lexical discriminability. Thus, the average intelligibility of the lexically easy words and the lexically hard words was equivalent across the six lists used under the three presentation formats. Paired t-tests revealed no significant differences in the speech intelligibility scores between any of the lists under a given presentation format.

Because one goal of this study was to investigate the effects of talker variability on word recognition, the lists were also balanced for talker effects. To accomplish this, the talker with the average visual-alone speech intelligibility score was chosen as the talker for the single-talker lists. Visual-alone intelligibility scores were used to select the single talker because the audio-alone and audiovisual intelligibility scores, which were obtained from normal-hearing listeners, were near ceiling. Once the speaker for the single-talker lists was chosen, the intelligibility scores for the tokens produced by the single talker and the remaining six talkers were used to evaluate intelligibility of the single- and multiple-talker lists respectively. This selection process was based on the audio-alone, visual-alone, and audiovisual intelligibility data for each token. Following this procedure, all six word lists were equally intelligible under a given presentation format regardless of talker condition.

Procedure

Testing was conducted in a single-walled sound treated IAC booth (Model #102249). The digitized audiovisual stimuli were presented to participants using a PowerWave 604 (Macintosh compatible) computer equipped with a Targa 2000 video board. All listeners were tested individually. The experimental procedures were self-paced. Video signals were presented with a JVC 13U color monitor. Speech tokens were presented via a loudspeaker at 70 dB SPL (C weighted) for participants using CIs. Each participant was administered three single talker and three multiple talker lists. Within each talker condition, one list was presented using an audio-alone format, one using a visual-alone format, and one using an auditory plus visual format. Visual-alone conditions were achieved by attenuating the loudspeaker and audio-alone conditions were achieved by turning off the video display monitor.

For the conditions where auditory stimulation was present in the stimulus, normal hearing participants were tested using a –5 dB signal to noise ratio (SNR) in speech spectrum noise at 70dB SPL relative to the 65 dB SPL speech tokens. This SNR was chosen during preliminary testing to prevent most of the participants with normal hearing from attaining ceiling performance on the task. All of the

participants were asked to verbally repeat the word that was presented aloud. The experimenter subsequently recorded their responses into computer files on-line. No feedback was provided.

Results and Discussion

Analysis of Raw Scores

The data from all 41 subjects were submitted to a 4-way repeated-measures ANOVA, with the factors of Presentation Mode (Visual-alone, Audio-alone, vs. Audiovisual), Talker Variability (Single vs. Multiple), and Lexical Competition (Easy vs. Hard) treated as within-subjects variables, and Group (normal hearing vs. cochlear implant) as a between-subjects variable.

Table 2 presents a summary of the raw scores obtained by the two participant groups as a function of presentation format, lexical competition and talker variability. It should be noted that, with the exception of the visual only presentation format, direct comparisons of the raw scores between the cochlear implant and normal hearing control groups are not valid. Recall that in formats where auditory speech information was presented, the cochlear implant group was tested in the quiet while the normal-hearing comparison group was tested in white noise to reduce performance below ceiling levels. It is the pattern of performance within each listener group that can be compared, not absolute scores between groups. It is interesting to note, therefore, that several commonalities between the two comparison groups emerged with respect to the manipulated factors.

Talker Condition	Lexical Discrim.	CI (N=20) Presentation Format			NH (N=21) Presentation Format		
		V	A	AV	V	A	AV
Single-Talker	Easy	23.9	34.4	75.8	18.0	54.2	75.4
	Hard	8.9	29.4	64.2	4.8	45.2	70.9
Multiple-Talker	Easy	21.7	38.6	70.0	15.6	48.9	74.3
	Hard	9.4	23.9	52.7	8.2	39.7	62.2

Table 2. Mean percent correct word recognition performance by condition.

Effects of Presentation Format

Figure 1 displays the overall performance of the two participant groups under the three presentation formats, averaged across talker condition and lexical competition. A significant main effect of Presentation Mode was observed, $F(1,39) = 352.6$, $p < 0.001$. Regardless of group membership, performance in the visual-alone condition (M = 13.81) was worse than in the audio-alone condition (M = 39.18), which was even worse than in the audiovisual condition (M = 68.20). Presentation format also interacted with the Group variable. Simple effects analyses revealed that this interaction was supported by differences in performance between the normal-hearing and cochlear implant groups in the visual-alone, $F(1,39) = 5.40$, $p = 0.03$, and the audio-alone, $F(1,39) = 9.73$, $p = 0.003$, presentation conditions. As shown in Figure 1, cochlear implant users obtained higher scores in the visual-alone condition than their normal-hearing counterparts. In contrast, normal hearing subjects tested in noise obtained higher scores in the audio-alone condition than cochlear implant users tested in the quiet.

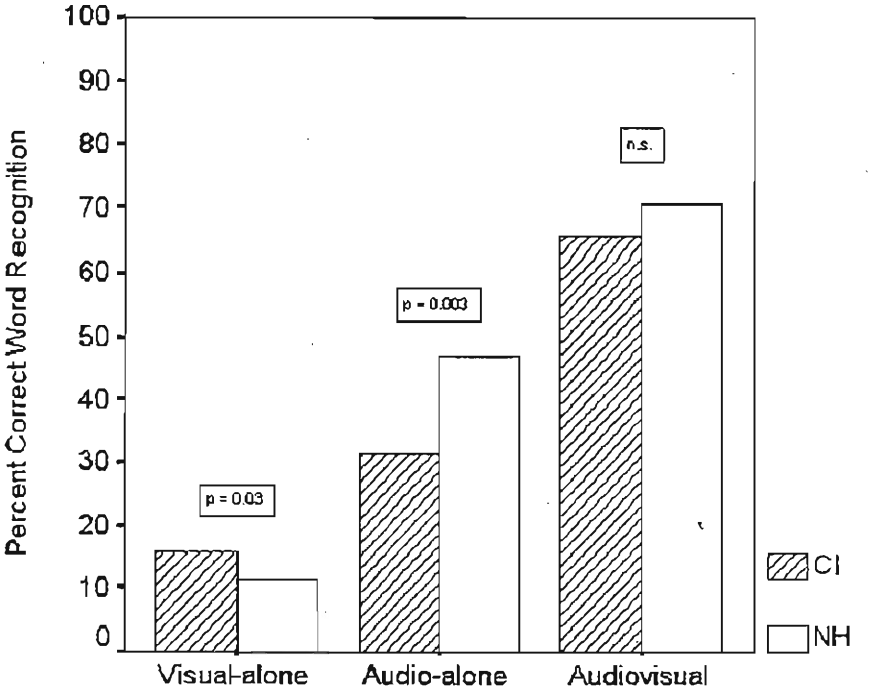


Figure 1. Percent correct word recognition performance of the CI users and the normal hearing participants under the three presentation formats averaged over talker and lexical variables. Listeners with CIs were tested in quiet and normal-hearing comparison participants were tested in noise at −5 dB SPL.

Correlations. One way to evaluate the relationship and to assess the underlying similarities among visual-alone, audio-alone, and audiovisual speech perception skills is to correlate performance for each presentation format. Previous studies have demonstrated significant correlations among visual-alone, audio-alone, and audiovisual performance on consonant perception measures (Grant & Seitz, 1998) and between visual-alone and audio-alone performance on word perception measures (Watson, Qiu, Chamberlain, & Li, 1996). Grant et al. found that for sentence materials only audio-alone and audiovisual performance were correlated. However, correlations with visual-alone scores failed to reach significance.

To examine the relations among the presentation formats, speech intelligibility scores obtained under each presentation format were correlated separately for each group of listeners. Figure 2 shows each subject's performance in the audio-alone (filled circles) and visual-alone (open circles) presentation conditions plotted as a function of audiovisual speech perception performance. The top panel shows this data for the CI group while the bottom panel shows the data for the NH group. The data are collapsed over talker and lexical variables. Significant correlations were observed between audio-alone performance and audiovisual performance for both groups of listeners, $r(20) = +0.81$, $p<.001$, for CI listeners and $r(21) = +0.67$, $p<0.001$, for NH listeners. However, the correlations between visual-alone performance and audiovisual performance were not significant for either group. Additional correlations were computed between the audio-alone and visual-alone performance for each group of listeners. None of these correlations was significant either. Inspection of Figure 2 shows that the individual scores for listeners using cochlear implants varied over a somewhat greater range than the individual performance for listeners with normal hearing tested in noise. Audiovisual speech perception scores for listeners with

normal hearing varied from 60% to 92% and performance for listeners using cochlear implants ranged from 53% to 88%, when one poorly performing listener was excluded from each group. Audio-alone performance varied between 22% and 74% for the normal hearing group, and between 10% and 61% for the cochlear implant group. Ranges in the visual-alone condition were more restricted, with performance between 3% and 22% for the normal hearing group and between 6% and 31% for the cochlear implant group. It is very likely that the lack of correlations involving visual-alone performance is due to floor effects and the absence of variability for performance in the visual-alone presentation condition. Performance in this condition ranged from 0% to 44.4%.

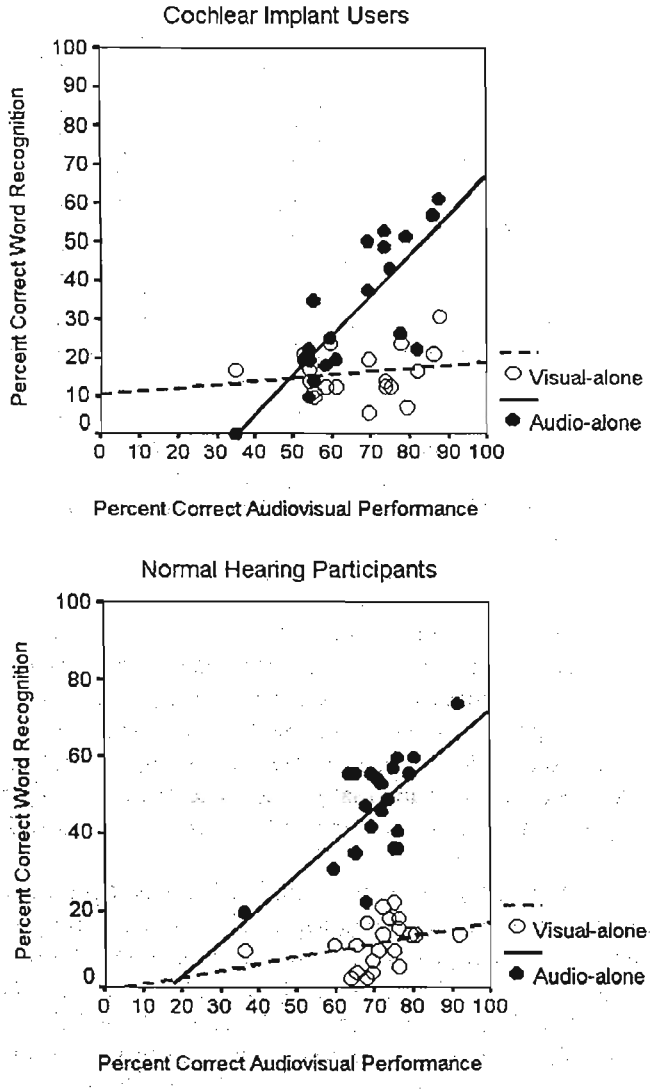


Figure 2. Scatterplot of individual participants' performance in a unimodal presentation format (audio-alone or visual-alone) and their performance in the audio-visual presentation format. Performance is shown separately for CI listeners or listeners with normal hearing. Data were averaged across talker and lexical variables. Regression lines are shown separately for correlations between AV vs. A as well as for AV vs. V performance.

Effects of Lexical Competition

The omnibus 4-way ANOVA also revealed a significant main effect of Lexical Competition, $F(1,39) = 158.89$, $p < 0.001$. Easy words (M = 45.83) were recognized better than Hard words (M = 34.96). Interestingly, this factor also interacted with the group variable, as shown in Figure 3. A simple effects analysis of this interaction showed that it was due to a difference between groups in accuracy of the Hard words $F(1,39) = 5.5$, $p = 0.024$, but not the Easy words $F(1,39) = 1.38$, n.s. As shown in Figure 3, normal-hearing subjects identified Hard words better than cochlear implant users.

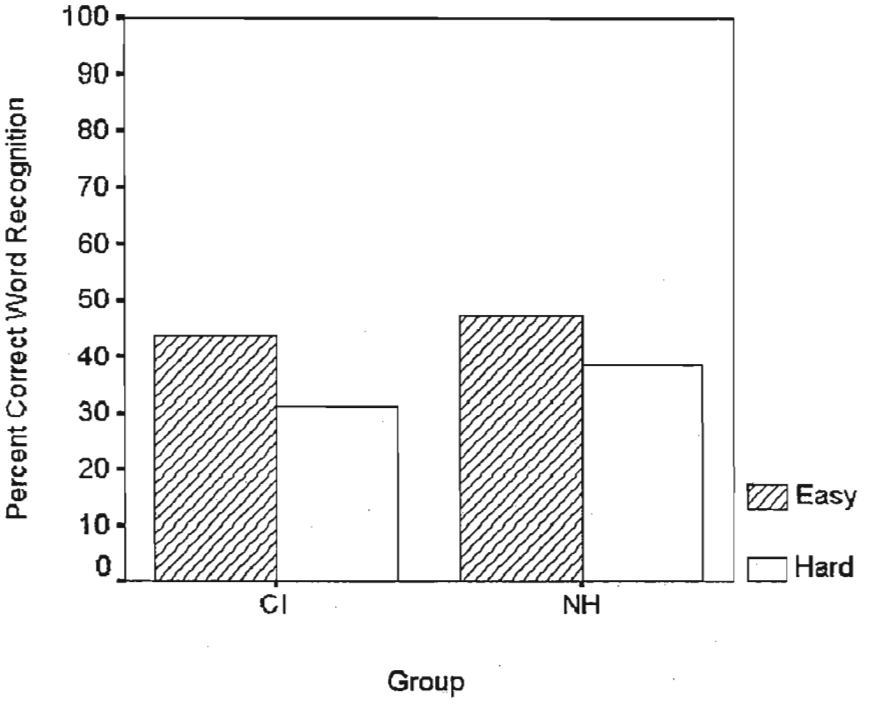


Figure 3. The percentage of words correctly identified by listeners with CIs and the control participants with normal hearing as a function of lexical competition of the stimulus words

Effects of Talker Variability

The main effect of Talker Variability was also significant, $F(1,39) = 13.69$, $p = 0.001$. Overall, single talker lists (M = 42.01) were identified better than multiple talker lists (M = 38.78). Talker Variability also interacted with Presentation Mode, $F(1,39) = 4.73$, $p = 0.01$. Figure 4 illustrates this interaction. Simple effects analyses revealed that the source of this two-way interaction was the difference between single and multiple talker lists in the audiovisual presentation condition only, $F(1,39) = 16.76$, $p < 0.001$; single talker lists were identified better than multiple talker lists.

Finally, Figure 4 shows the significant three-way interaction between Presentation Format, Talker, and Lexical Competition, $F(1,39) = 4.33$, $p = 0.011$. For ease of comparison, this figure has been split so that each hearing group is represented separately. The top two panels show the interaction for the

cochlear implant group, and the bottom two panels show the interaction for the normal-hearing group. Tests of simple effects showed a complex pattern of results. For easy words (the left panels for both groups), the difference in performance between single and multiple-talker lists did not differ in any presentation condition, although there was a marginal effect of talker in the audiovisual presentation condition, $F(1,39) = 3.4$, $p = 0.073$. For hard words (the right panels for both groups), however, the simple effects of talker were significant in both the audiovisual presentation condition, $F(1,39) = 18.06$, $p < 0.001$, and the audio-alone condition, $F(1,39) = 5.96$, $p = 0.019$. In addition, the analysis revealed a marginal effect of talker in the visual-alone presentation condition, $F(1, 39) = 3.4$, $p = 0.073$. As shown in the figure, whenever there was a significant effect of talker, performance on single talker lists was better than performance on multiple talker lists, except in the marginally significant case of visual-alone performance for lexically hard words.

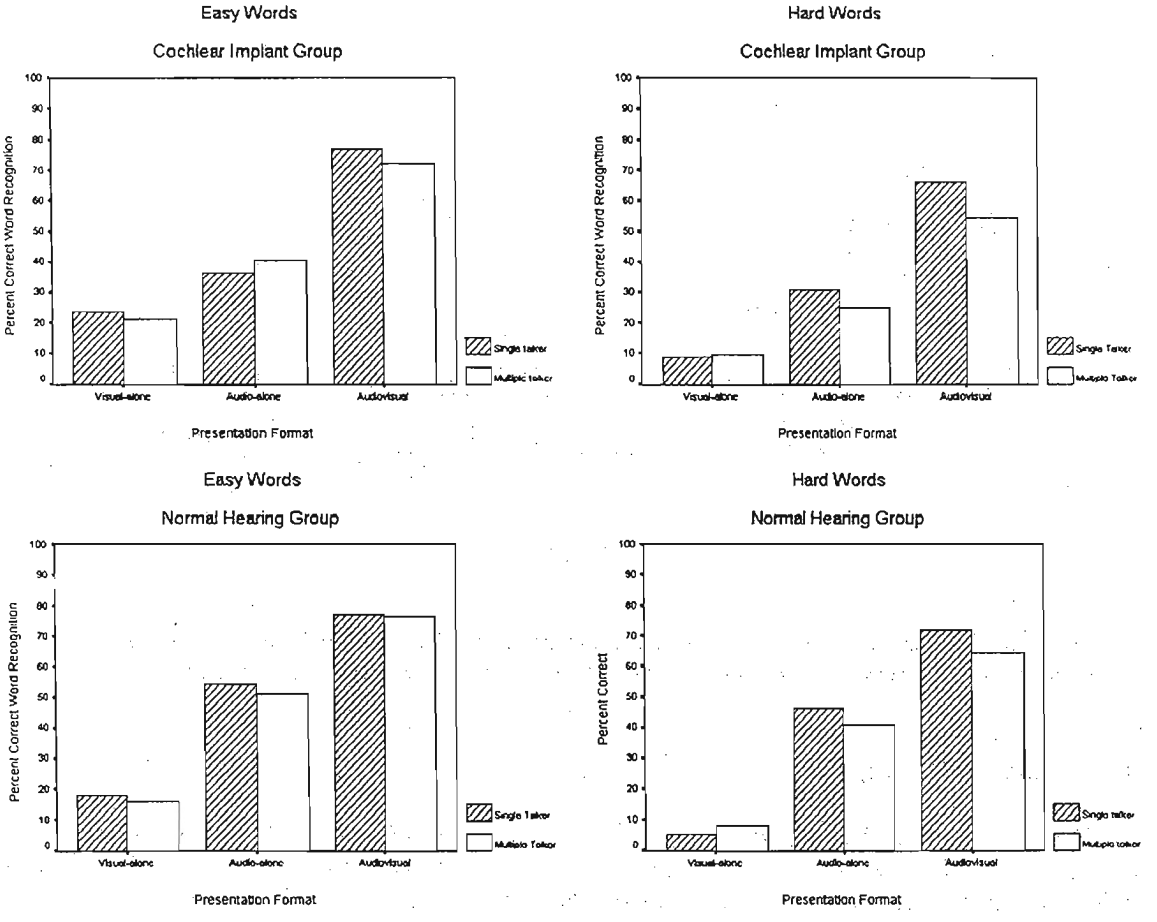


Figure 4. The percent of words correctly identified in each presentation format as a function of talker variability for lexically easy words and lexically hard words. The top two panels show data for cochlear implant users and the bottom two panels show data for the normal-hearing group.

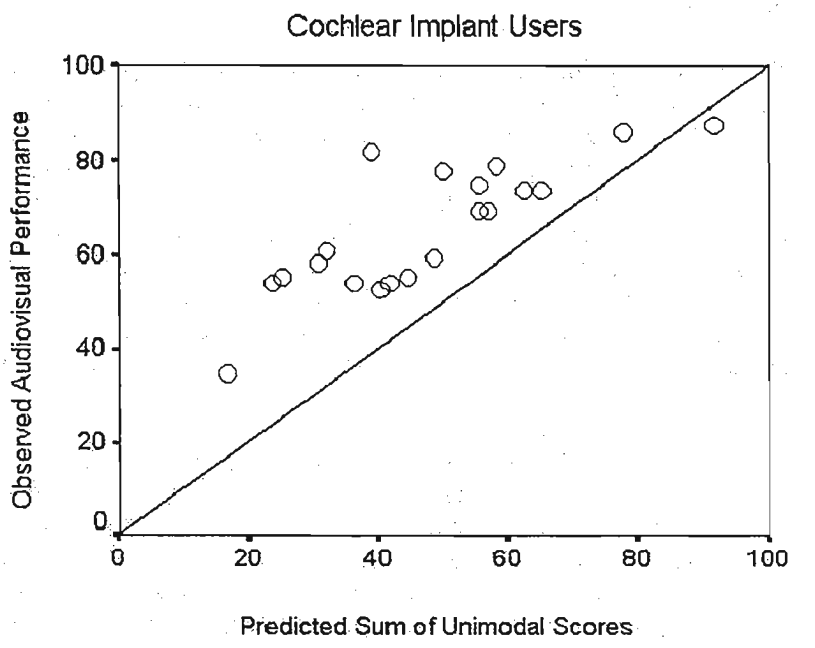
Audiovisual Integration (I)

Conceptually, audiovisual integration means that listeners can do more than simply add the perceptual cues obtained from the auditory and visual modalities to increase their speech perception scores. One way to quantify the extent of this integration process is to compare the observed audiovisual performance to an estimate of the performance that a listener could achieve assuming no integration at all. As an example, consider an individual who receives a short open-set word recognition task consisting of the four words: *lace*, *wife*, *short*, and *long*. Under audio-alone presentation, this hypothetical listener correctly perceives *lace* and *short* (50% correct). Under visual-alone presentation, this individual correctly perceives *wife* and *short* (50% correct). The best the listener could do then without any integration would be to correctly identify *lace*, *wife*, and *short* in the audiovisual modality (75% correct). Any additional gain above this level of performance would provide evidence for some form of integration.

Unlike the hypothetical case, however, the listeners in the present experiment did not receive the same tokens under each of the three presentation formats, so the observed responses could not be compared in this way. Only the raw percentage scores could be used for this purpose. We assumed that the upper bound estimate used in this experiment was simply the sum of visual-alone and audio-alone performance. The example above can be used to demonstrate that this is a more conservative estimate. Comparing the auditory and visual responses above, one would expect audiovisual performance of 75% correct (*lace, wife, short*). On the other hand, if only percentage scores were used, one would expect audiovisual performance to be 100% correct (50%+50%). Limiting the process of AV integration to simple addition would limit perception to this upper bound score.

Individual Data

Figure 5 shows individual subjects' performance as a function of the sum of audio-alone and visual-alone performance. The top panel shows the data for listeners in the cochlear implant group while the panel shows the data for listeners in the normal-hearing group. In both figures, the diagonal represents the audiovisual score that would be obtained by each listener if audiovisual performance were simply equal to the sum of the audio-alone and video-alone scores. The observed scores in Figure 5 demonstrate that almost all of the listeners were able to perform above the prediction expected from simple additive integration. Audiovisual performance in excess of this (i.e., performance above the diagonal) provides support for non-additivity of audiovisual integration.



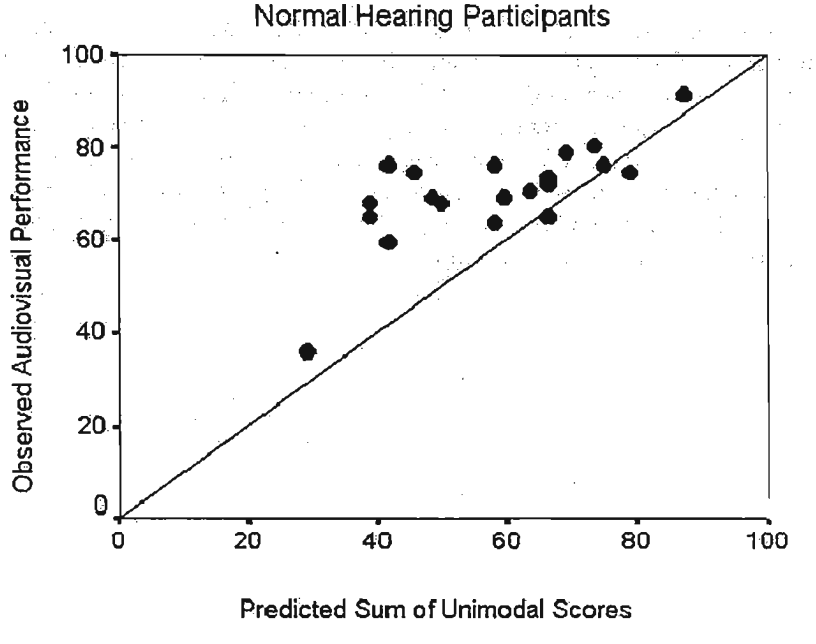


Figure 5. Scatterplot of individual participants' performance in the AV presentation mode plotted against the sum their auditory-only and visual-only scores. Scores are averaged over lexical competition and talker variability. Scores above the diagonal represent AV scores greater than the simple sum of performance in the two unimodal conditions.

Group Analysis

An index of audiovisual integration (I), can be calculated as the difference between observed audiovisual performance and the predicted performance estimated by the simple addition of visual-alone and audio-alone scores (equations 1 and 2). Again, this is simply the amount by which the observed performance exceeded that predicted by an additive model of audiovisual speech perception; it is also the distance from the diagonal to the observed score in Figure 5. These values can be used to estimate the portion of audiovisual speech intelligibility that is attributable to the processes of integration. Due to ceiling effects, this analysis can only be used when the sum of audio-alone and visual-alone performance is less than 100%. As an upper bound performance estimate, the index "I" was calculated for each participant using equation 1. Only participants whose audio-alone + visual-alone performance was less than 90% in every condition were used in this analysis. This criterion was chosen to minimize ceiling effects in the calculation of "I". The final analysis included 17 participants with cochlear implants and 17 participants with normal hearing.

$$(1) \qquad I = AV_{measured} - AV_{estimate}$$

where

$$(2) \qquad AV_{estimate} = A_{measured} + V_{measured}$$

The "I" scores from these 34 subjects were used as the dependent variable and submitted to a three-way repeated-measures ANOVA using Talker, Lexical Competition and Group as independent variables. The analysis revealed that the index of integration, "I", differed between the two groups, $F(1,34) = 6.49$, $p = 0.016$. Overall, the integration score "I" was higher for cochlear implant users (M = 20.43) than for normal hearing listeners (M = 11.84). In addition, the integration score "I" was larger for lists of lexically hard words (M = 21.03) than for lists of lexically easy words (M = 11.24), $F(1,34)=8.68$, $p=0.006$. A marginally significant difference in the integration score was also found for the Talker factor. Performance on multiple talker lists (M = 18.57) was significantly higher than on single talker lists (M = 13.70), $F(1,34) = 3.65$, $p = 0.07$. None of the other interactions reached significance.

The overall pattern of results suggests that audiovisual integration benefit (I) is greatest in the conditions where the auditory or visual portion of the stimulus alone is ambiguous or underspecified in some way, causing performance in the single modality presentation conditions to suffer. As a result, the combined information increases performance above and beyond the level that would be expected from simple addition of information from the two separate input modalities. It is interesting to note, however, that this principle applies whether unimodal stimulus ambiguity arises due to properties of the perceiver (i.e., the effect of group), properties of the stimulus item itself (i.e., the effect of lexical competition), or properties of the environment in which those stimuli are presented (i.e., the marginal effect of talker).

Visual Enhancement (R_a)

In their pioneering study of audiovisual speech perception, Sumby and Pollack (1954) developed a quantitative metric to evaluate the gains in speech intelligibility performance due to the addition of visual information from seeing a talker's face. Because speech perception scores have a theoretical maximum (i.e., perfect performance), the measure was developed to show the extent to which additional visual information about speech improved performance *relative to the amount by which audio-alone performance could possibly improve.* Their metric, R_a, can be used to assess the extent of visual enhancement for an individual perceiver in our study. To assess visual enhancement, R_a was calculated

for all 41 participants based on the recognition scores obtained in the audiovisual and audio-alone conditions using Equation 3, from Sumby and Pollack (1954). In the equation, "AV" is performance in the audiovisual presentation condition, and A is performance in the audio-alone condition. R_a was calculated separately for lexically easy and lexically hard words in each of the two talker conditions. The R_a's resulting from this analysis are reported in Table 3 and displayed graphically in Figure 6.

$$(3)\, R_a = \frac{AV - A}{1.0 - A}$$

	CI			NH		
Talker Condition	Total R	Easy R	Hard R	Total R	Easy R	Hard R
Single-Talker	.56	.64	.49	.45	.40	.46
Multiple-Talker	.43	.50	.35	.43	.49	.37

Table 3. Mean visual enhancement (R) by listener group (CI or NH), list type and presentation format.

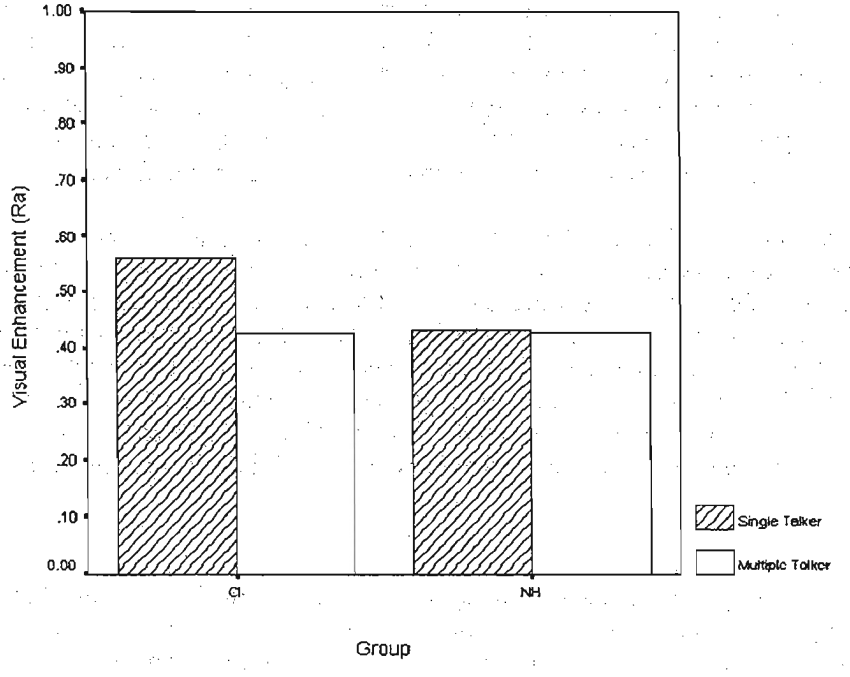


Figure 6. Visual enhancement (R_a) for the CI users and the comparison normal-hearing participants calculated separately in each of the two talker conditions. R_a represents the gain in word recognition performance due to the addition of visual speech information.

An ANOVA was used to analyze the visual enhancement scores (R_a) in each condition. Listener group was a between subject factor while talker variability and lexical competition were within subject factors. Overall, R_a was lager for single talker (M = 0.50) conditions than for multiple talker conditions (M = 0.43), $F(1,39) = 4.78$, $p = 0.04$. This result suggests that listeners realized more of the potential benefit from the addition of visual information to auditory information under single talker presentation conditions than under multiple talker conditions.

The interaction between Talker and Group was also significant, $F(1,39) = 4.05$, $p = 0.05$. Figure 6 shows R_a scores for the relevant cells of this interaction. Simple effects analysis revealed that the interaction was due to a difference in visual enhancement for single vs. multiple talker lists for cochlear implant users, $F(1,39) = 8.58$, $p = 0.006$, but not for normal-hearing participants, $F(1,39) < 1$, n.s.

There was also a marginal main effect of Lexical Status, $F(1,39) = 3.82$, $p = 0.06$. R_a scores for lexically easy words (M = 0.51) were higher than the scores for lexically hard words (M = 0.42). This result indicates that listeners obtained somewhat greater visual benefit from words that have less competition than from words that have more competition. All other main effects and interactions were not significant.

General Discussion

The results from this study of audiovisual word recognition suggest that similar factors affect the process of spoken word recognition in normal hearing listeners and postlingually deafened adults with cochlear implants. Overall, we observed only a marginal difference in the mean performance levels of both groups ($p = 0.07$), indicating that our goal of equating performance across both listener groups was met with the signal-to-noise ratio we picked for use with the normal-hearing participants. It is important to note here that using broad-band noise to degrade speech for the normal-hearing listeners is not the same type of auditory degradation experienced by cochlear implant users. However, overall similarities between the groups in terms of the effects of our manipulated factors provide some new insights into the perceptual and linguistic processes at work in both groups of listeners during spoken word recognition.

Analysis of Raw Scores

Presentation Format. Presentation Format affected both groups of listeners in similar ways: visual-alone performance was consistently below audio-alone performance. In addition, performance was always best when both auditory and visual sources of information were available for speech perception. The significant cross-over interaction between Group and Presentation Format revealed that normal-hearing listeners performed better than cochlear implant users in the audio-alone condition but that CI users performed better than normal hearing listeners in the visual-alone condition. This finding is consistent with a recent report by Bernstein, Auer, and Tucker (2001) who found reliable differences in the performance of normal-hearing and hearing-impaired speechreaders on a visual-alone speech perception task. The pattern of results observed in the present study may be due to the way lip-reading skills were acquired in these patients. The CI users in our sample were all progressively deafened post-lingually. It is possible that over long periods of time, a gradual reliance on lip-reading eventually leads to greater use of the visual correlates of speech when the auditory information in the speech signal is no longer sufficient to support word recognition. Further work on the time-course of learning speechreading skills in post-lingually deafened adults prior to implantation is needed before any definitive conclusions can be drawn.

We also found that the two groups of listeners achieved roughly the same level of performance in the audiovisual condition, even though they differed in the extent to which they were able to perceive speech from either sensory modality alone. This result illustrates the complementary nature of auditory and visual information about speech (Summerfield, 1987): when the information available in one sensory modality (e.g., audition) is noisy, degraded, or impoverished, information available in the other modality (e.g., vision) can "make up the difference" by providing complementary cues that combine to enhance overall word recognition performance in a particular task.

According to the event-based theory of speech perception (Fowler, 1986), the complementarity of the two sources of sensory stimulation about speech arises because auditory and visual sources of information in speech are structured by a unitary, underlying articulatory event. That is, when a person speaks, their articulatory patterns and gestures simultaneously shape both auditory and optic patterns of energy in very specific, lawful ways. The relations between the two modalities, then, is specified by the information relating each pattern to the common, underlying, dynamic vocal tract gestures of the talker that produced them. It is precisely this time-varying articulatory behavior of the vocal tract that has been shown to be of primary importance in the perception of speech (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Remez, Fellowes, Pisoni, Goh, & Rubin, 1998; Remez, Rubin, Pisoni, & Carrell, 1981; Remez, Rubin, Berns, Pardo, & Lang, 1994).

With this conceptualization in mind, the sensory and perceptual information relevant for speech perception is *modality-neutral* or *amodal*, since it can be carried by more than one sensory modality (Fowler, 1986; Gaver, 1993; Remez et al., 1998; Rosenblum & Saldaña, 1996). The amodal nature of phonetic information is demonstrated convincingly in studies showing that perceptual information obtained via the *tactile* modality, in the form of Tadoma, can be used and integrated across sensory modalities in speech perception (Fowler & Dekle, 1991), albeit with limited utility.

Although the *information* necessary for speech perception may be modality-neutral, the internal *representation* of speech appears to be based on an individual's experience with perceptual events and actions in the physical world (Lachs, Pisoni, & Kirk, 2001). Thus, awareness of the intermodal relations between auditory and visual information is contingent upon experience with more than one sensory modality. Because our CI participants were all post-lingually deafened adults, they had all had prior experience with the auditory properties of speech and had acquired knowledge of the lawful correspondences between auditory and visual correlates of speech. As the present findings demonstrate, the CI participants were able to make use of this experience under audiovisual presentation and were able to recognize isolated words at levels comparable to our normal hearing participants.

Lexical Competition Effects. We also found robust effects of Lexical Competition in this study. For both groups of listeners, lexically easy words were recognized better than lexically hard words, indicating that normal-hearing and cochlear implant listeners organize and access words from memory in fundamentally similar ways. Thus, phonetically similar words in the mental lexicons of CI users compete for selection during word recognition. This process is also affected by word frequency such that higher frequency words are more apt to win out among phonetically similar competitors. The finding that lexical competition affected our CI group is not surprising because the participants in this group were all post-lingually deafened and had no evidence of any central nervous system involvement prior to or after the onset of deafness. Presumably, they developed extensive lexical representations when they had normal hearing and retained some form of this information over time after their hearing loss.

Despite these overall similarities, several differences in the effects of lexical competition were found between the groups. CI users performed more poorly on lexically hard words than did normal hearing listeners. However, performance for both groups on lexically easy words was statistically

equivalent. This interaction suggests that the CI users were less able to make the fine acoustic/phonetic distinctions among words that are needed to distinguish lexically hard words from their phonetically similar neighbors. Although the cochlear implant appears to provide enough auditory information to recognize words when only gross acoustic cues are sufficient, in many patients the implant may not be sufficient to provide the more fine-grained phonetic information necessary to discriminate between very similar lexical candidates.

Talker Variability Effects. Across both groups of listeners, we observed an interesting three-way interaction between talker variability, lexical competition and presentation format. For lexically easy word lists, talker variability did not play a role in word recognition performance, although a marginal difference in performance was observed between single-talker and multiple-talker lists under audiovisual presentation. However, the effects of talker variability were robust for lists of lexically hard words. Performance on single talker lists was better than performance on multiple talker lists in both the audiovisual and audio-alone presentation conditions. The difference was only marginally significant in the visual-alone condition. The results on the effects of talker variability are consistent with the proposal that repeated exposure to a single talker allows the listener to encode voice-specific attributes of the speech signal. Once internalized, voice-specific information can improve word recognition performance (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1995).

The "single talker advantage" appears to be most helpful when there is a great deal of lexical competition among words and fine phonetic discrimination is required, as there is with lexically hard words. Talker-specific information appears to be used in conditions where a detailed perceptual representation of the acoustic/phonetic input can serve to more clearly disambiguate multiple word candidates from within the lexicon. The reduced magnitude of the talker effect in visual-alone conditions may be because the perceptual input provided by the optical display of speech itself is insufficient to specify a set of lexical candidates small enough that fine-grained phonetic information can improve word recognition. Also, it is very likely that detailed talker-specific information would be difficult to obtain for visual-alone presentation of short words presented in isolation (see Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). Nygaard et al found that it was much more difficult for participants to learn novel voices from isolated words than for short sentences. Apparently, participants are able to make use of several different acoustic correlates of talker-specific information from sentences that are not present in isolated words.

Analysis of Perceptual Errors

Up to this point, only the accuracy scores for correct word recognition have been presented. However, it is also possible to examine the errors made by participants in order to determine if their responses were randomly distributed, or if they followed a pattern that made use of partial stimulus information and some form of sophisticated guessing strategies. By determining whether confusions and error responses were lexical neighbors of the intended target words, we can assess the extent to which Presentation Format and Talker Variability affected the information available to the perceiver under the various conditions. Our earlier discussion suggested that audiovisual presentation influences lexical competition by providing more robust perceptual information as input to the word recognition process. If this prediction is correct then we would expect that even in cases when the ultimate output of the word recognition system was incorrect, the incorrect response would be perceptually more similar to the target word in those cases where the input to the system was more clearly specified.

For the purposes of this error analysis, we considered a response to be a lexical neighbor of the original target word if it differed from the target word by the insertion, deletion or substitution of a single phoneme (Luce & Pisoni, 1998). For example, if the target word was "cat", the response "sat" would be

considered a lexical neighbor by substitution, "at" would be considered a neighbor by deletion, and "scat" would be considered a lexical neighbor by insertion.

To carry out the error analysis, we converted the target words and the incorrect responses into phonetic representations. The target words and responses were transformed into a computer-readable phonetic notation using DECtalk (DCT03) Text-to-Speech System. Each target word-response pair that was incorrect was then analyzed in greater detail to determine the nature of the phonetic and lexical confusions. The proportion of errors that were in the lexical neighborhood of the target word ("neighborhood errors") was then computed. This measure may be thought of as an index of how similar an incorrect response was to the original target word. Because of the significant effects of talker variability found earlier, the data were analyzed separately for single talker and multiple talker lists. Results from both sets of analyses are reported separately below.

Single-Talker Confusions. We first present the results of the error analysis using data from the single-talker conditions. The proportion of incorrect responses that were neighbors of the target word was computed and then submitted to a 2 (Presentation: audiovisual or audio-alone) x 2 (Lexical Competition: Easy or Hard) x 2 (Group: normal hearing or cochlear implant) ANOVA. Responses in the visual-alone condition were not included in the error analysis because the criterion for neighborhood membership in visual-alone conditions is still an unresolved issue awaiting further investigation (however, see Auer & Bernstein, 1997 for preliminary work on this topic).

The ANOVA revealed a significant main effect of Presentation Format, $F(1,39) = 4.713$, $p = 0.036$. The top panel of Figure 7 shows the proportion of incorrect responses that were lexical neighbors of the target word for both groups of participants under the audiovisual and audio-alone conditions. For both groups of listeners, it can be seen that a higher proportion of errors were lexical neighbors of the target word under audiovisual presentation than under audio-alone presentation. However, this factor did not interact with listener group, indicating that a participant's hearing status did not affect this pattern. In general, the addition of visual information to auditory information helps to narrow down or constrain the set of possible lexical candidates competing for selection during word recognition.

The bottom panel of Figure 7 shows the proportion of errors that were neighbors of the target word for both groups of participants separately by lexical competition. The figure shows that incorrect responses were more often selected from the neighborhoods of hard words than easy words. This pattern was confirmed by the presence of a significant main effect of lexical competition, $F(1, 39) = 30.0$; $p < 0.001$. The interaction between lexical competition and listener group was not significant, indicating again that the pattern was similar across both groups of listeners. None of the remaining interactions was significant. Once again, the results indicate that the combination of auditory and visual information about speech produces a more detailed perceptual representation that places additional constraints on the set of possible response alternatives in the mental lexicon.

For the single talker conditions, it is apparent that both normal-hearing listeners and listeners with cochlear implants display very similar error patterns and lexical confusions. Relative to audio-alone presentation, audiovisual presentation increased the likelihood that an incorrect response would be selected from within the lexical neighborhood of the target word. This pattern of errors is consistent with the proposal that the additional visual information contained in an audiovisual signal serves to refine and constrain the sensory and perceptual information needed for lexical discrimination and selection. The additional visual information under audiovisual presentation influenced and controlled the participants' responses, even when they failed to identify the target word correctly.

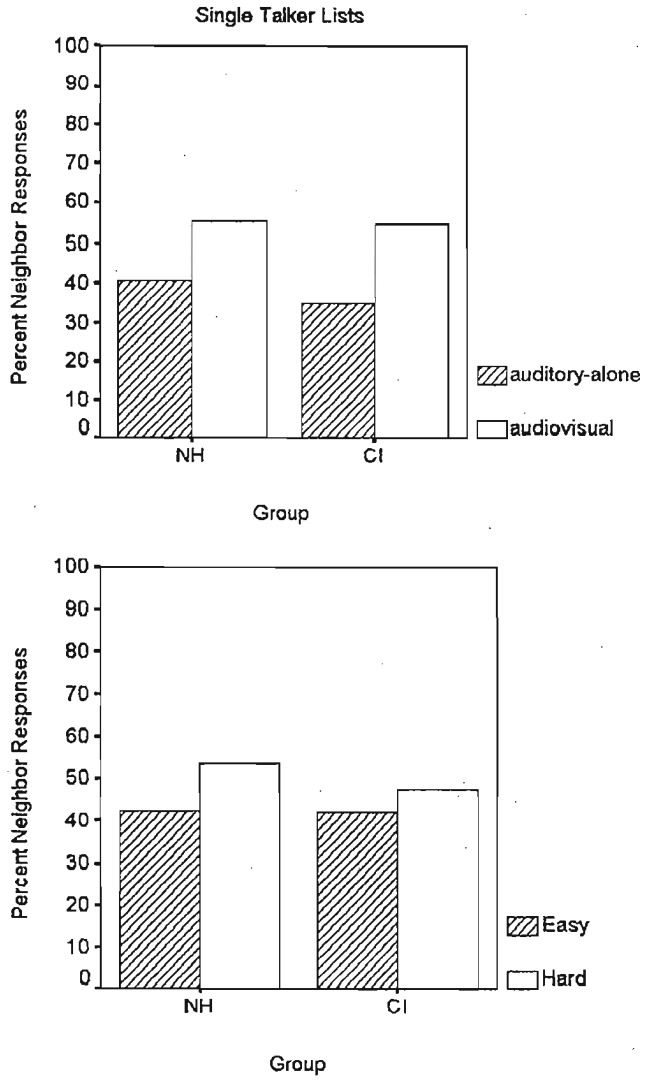


Figure 7. Percentage of error responses that were within the lexical neighborhood of the target word for the single-talker conditions. Error responses are shown separately for listeners with CIs and the normal-hearing participants as a function of presentation format and lexical competition of the target word.

The lexical properties of the target word also had an effect on the nature and number of errors that came from the same similarity neighborhood. Specifically, more errors were observed in the neighborhoods of lexically hard target words than lexically easy target words. This is not a surprising result, because lexically hard words, by definition, generate more competition during recognition and have more neighbors with which they might be confused than do lexically easy words. Because participants in this study were required to respond with only English words in an open-set response format, the response confusions that they generated were simply more likely to produce an error in the neighborhood for hard words than they were for easy words.

Multiple-Talker Confusions. The pattern of incorrect responses for the multiple-talker lists was similar to the results observed for the single talker lists. As in the previous analysis, the proportion of incorrect responses that were in the neighborhood of the target word was used as the dependent variable in a 2 (Presentation Format) x 2 (Lexical Competition) x 2 (Group) ANOVA. The results showed a main effect of Presentation Format, $F(1, 38) = 49.50$, $p < 0.001$. Audiovisual presentation was better than audio-alone presentation. Figure 8a shows this main effect, with the two listener groups displayed separately for ease of comparison. The interaction between Presentation Format and listener Group was only marginally significant, $F(1, 38) = 3.093$, $p = 0.087$. Inspection of the top panel of Figure 8 shows that participants with cochlear implants received greater benefit from audiovisual presentation than the normal hearing comparison group; once again audiovisual presentation increased the number of response confusions within a lexical neighborhood.

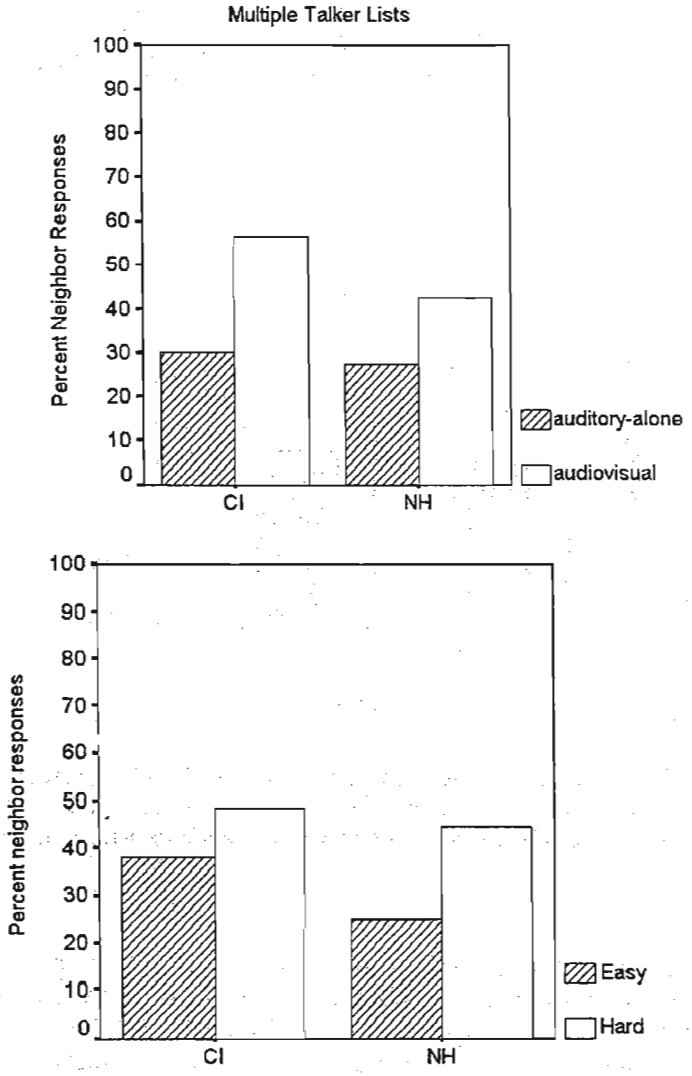


Figure 8. Percentage of error responses that were within the lexical neighborhood of the target word for the multiple-talker conditions. Error responses are shown separately for listeners with CIs and the normal hearing participants as a function of presentation format and lexical competition of the target word.

As in the single talker analysis, we also found a main effect of Lexical Competition, $F(1, 38) = 28.446$, $p < 0.001$. The proportion of confusions within a lexical neighborhood was much higher for lexically hard words than for lexically easy words. The bottom panel of Figure 8 shows the proportion of errors that were neighbors of the Easy and Hard words, separately for each listener group. Again, the interaction between Lexical Competition and listener Group was marginally significant in the multiple talker analysis, $F(1, 38) = 3.054$, $p = 0.089$. Examination of the figure shows that this result was due to an increase in the number of incorrect responses from the neighborhoods of hard words relative to easy words in the normal-hearing group. The increase in neighborhood responses for hard words in the cochlear implant group was smaller.

The analysis of the incorrect responses from the multiple-talker lists replicates the patterns observed in the single talker lists, although the differences were somewhat smaller and only marginally significant. Audiovisual presentation increased the number of lexical confusions that came from within the same lexical neighborhood as the target word. In addition, the added task load of recognizing words from multiple talkers seems to have produced smaller differences in the extent to which these factors affect participants with either normal hearing or cochlear implants. However, the overall pattern of the responses is similar across talker conditions and listener groups.

Audiovisual Integration ("I")

We also assessed the extent to which our participants combined auditory and visual information. Our measure of integration, "I", reflected the amount by which the observed audiovisual performance differed from the scores predicted by a simple, additive model of audiovisual integration. Overall, the two hearing groups differed on this measure; the cochlear implant users benefited more from combined audiovisual inputs than the normal hearing group. Although the CI group were better speechreaders than the NH group, the magnitude of the difference in performance in the visual-alone condition was not as great as the advantage the NH group had in the audio-alone condition. Thus, on the whole, the CI group performed worse in the unimodal conditions than the NH group. However, no difference in performance was observed in the audiovisual condition between the two groups. The results show that CI users were better able to combine the redundant, multimodal information than the NH group.

We also observed effects of lexical competition on the integration scores. Specifically, the measure of audiovisual integration was better for lexically hard words than for lexically easy words. Hard words require the perceiver to make fine phonetic distinctions for accurate identification. If these distinctions cannot be made based on auditory information in the acoustic signal, then the addition of visual speech information can serve to disambiguate the competing lexical entries. In contrast, easy words are much more distinct and discriminable and thus acoustic information alone may be adequate to identify these words. The addition of visual cues to highly discriminable audible patterns contributes little to the recognition of lexically easy words.

Visual Enhancement (R_a)

The second measure of audiovisual integration that we examined was R_a. This is the gain in speech intelligibility due to the presentation of combined audiovisual information relative to audio-alone presentation (Sumby & Pollack, 1954). This measure of integration was used because gains above audio-alone performance due to audiovisual presentation are necessarily limited by the theoretical maximum: 100% performance. We found that talker variability only affected visual enhancement scores for the CI users in the single talker condition. There was no effect of talker variability on visual enhancement for normal-hearing listeners. This finding does *not* mean that NH listeners were unaffected by talker

variability. However, talker variability did not affect the degree to which normal-hearing listeners could *combine* audiovisual information. The present findings suggest that CI users are better able to extract idiosyncratic talker information from audiovisual displays than NH listeners are, because they rely more on visual speech information to perceive speech in every day situations. With repeated exposure to audiovisual stimuli spoken by the same talker, the CI users exhibited a gain above and beyond that observed in normal hearing listeners. Cochlear implant users may be able to acquire more detailed knowledge of the cross-modal relations between audition and vision for a particular talker. Because NH listeners can successfully process spoken language by relying entirely on auditory cues, they may not have learned to utilize visual cues as successfully (Bernstein et al., 2001). For NH listeners, combined audiovisual information from a single talker may not provide any additional information about that talker than the cues provided by audio-alone presentation. This is especially true in a short-term laboratory experiment like the present one. The normal-hearing listeners have little prior exposure to visual-alone stimuli. If normal-hearing listeners came back repeatedly to the lab and were forced to listen to degraded speech in noise, they might develop a greater awareness of the inherent relationships between auditory and visual speech and show these effects after a short period of time.

Limitations of Present Findings

Although this is the first detailed study of audiovisual word recognition in postlingually deafened adult patients following cochlear implantation, there are a number of limitations in the experimental design that are worth mentioning here. First, we want to emphasize that direct comparisons of the raw scores between the two groups of listeners used in this study need to be made with some degree of caution. The data obtained from the normal-hearing listeners were collected under masking conditions using white noise while the data obtained from the CI patients were collected in the quiet. Masking noise was used with the normal-hearing listeners simply to reduce scores from ceiling levels of performance. The nature of the degradation resulting from the presentation of speech in noise to normal-hearing listeners is not equivalent to the transformation of speech that is processed by a cochlear implant and then presented as an electrical signal to a hearing-impaired listeners' auditory system. The two forms of signal degradation are not commensurate despite the fact that there was no overall statistical difference in the audiovisual condition between the two groups of listeners in the global analysis of variance of the main effects. Equivalent levels of performance on the word recognition task do not imply that the signals were encoded and processed in the same manner by both groups of listeners.

While there were similarities in performance across the two groups as a function of the variables under study, a number of the comparisons revealed small and consistent differences within groups. It is these differences that are informative and provide some new insights into how patients with CIs recognize spoken words and make optimal use of the available auditory information provided by their implant under different listening conditions. Both groups of listeners combine and integrate auditory and visual information about spoken words, but the CI patient appears to make somewhat better use of the visual information in more difficult listening conditions when there is ambiguity about the talker, or when they are forced to make fine phonetic discrimination among acoustically confusable words. The deaf listeners combine visual information with the available auditory information conveyed by their cochlear implant to support open-set word recognition, but they do so in somewhat different ways than the normal-hearing listeners.

It is not surprising that the same basic underlying lexical selection processes are used by both groups of listeners in this study. After all, the patients with cochlear implants were all postlingually deafened and had acquired language and lexical knowledge normally prior to the onset of their hearing loss. There is little reason to expect any large differences in the effects of lexical competition in these patients (Luce & Pisoni, 1998). Both groups showed sensitivity to the lexical manipulations used here.

The pattern of their word recognition scores demonstrates that they recognize spoken words "relationally," in the context of other words they know and have in their mental lexicons. Both groups are sensitive to frequency and phonetic similarity. The differences between the two groups of listeners occur earlier during perceptual analysis, when the initial sensory information in encoded prior to lexical selection.

In future studies using normal-hearing listeners, it may be better to use other methods of signal degradation, such as noise-band speech, to simulate the nature of the hearing loss and signal transformation produced by a cochlear implant (Dorman, Loizou, Fitzke, & Tu, 1998; Dorman, Loizou, Kemp, & Kirk, 2000; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). The use of isolated words as stimulus materials in this study also presents several limitations in terms of providing reliable and useful information about variability and the dynamics of a talker's vocal tract. Earlier studies by Nygaard et al. (1995) and Nygaard and Pisoni (1998) have shown that it is much more difficult to learn to identify novel voices from isolated words than longer sentences. Although greater in length, sentences provide additional information about prosody and timing. This may help to increase familiarity with a novel voice in a shorter period of time with less exposure.

Although we did not observe any consistent individual differences in performance across any of the experimental conditions, it is well known that listeners with cochlear implants display a wide range of individual differences in performance on various outcome measures. The patients selected for this study were all good users who were able to derive large benefits from their cochlear implants. All of them were able to respond appropriately in an open set word recognition task given the limited auditory input provided by their implant. Examination of the other implant patients who do more poorly under these conditions may reveal a wider range of scores, a different pattern of audiovisual integration skills and different levels of reliance on visual information about speech.

Clinical Implications

With recent advances in cochlear implant technology, many postlingually deafened adults are now able to achieve very high levels of spoken word recognition through listening alone (Kirk, 2000). Other patients may derive substantial benefit from a cochlear implant only when the auditory cues they receive are combined with visual information from a talker's face. Like listeners with normal hearing, most cochlear implant recipients benefit from the presence of visual speech cues under difficult listening situations, such as when the talker is speaking rapidly or has an unfamiliar dialect, or when listening in the presence of background noise.

Structured aural rehabilitation activities with a sensory aid (either a cochlear implant or a hearing aid) often rely on highly constrained and organized listening activities intended to enhance the ability to discriminate or recognize various acoustic cues in speech. Words or sentences are usually presented in the audio-alone format by a single clinician. There has been little systematic application of the findings from recent studies on variation and variability in speech perception and multi-modal perception to therapy and rehabilitation with clinical populations. The findings from the present study suggest that it may be fruitful to apply some of the knowledge gained recently about audiovisual speech perception to clinical problems associated with intervention and aural rehabilitation after a patient receives a sensory aid. Exposure to multiple talkers and a wide range of speaking styles in both audio-alone and auditory-visual modalities may provide patients with a greater range of stimulus variability during the first few months of use after receiving an implant; this in turn may help patients develop more robust perceptual strategies for dealing with speech in real world listening conditions that exist outside the clinic and research laboratory. Specially-designed word lists can be developed and used for training materials under different presentation formats to emphasize difficult phonetic contrasts that may be hard to recognize under audio-

alone conditions but easy to identify under audiovisual conditions. Similarly, activities such as connected discourse tracking using audiovisual stimuli may promote the development of robust multimodal speech representations and enhance spoken language processing. Auditory training activities using multimodal stimuli may enhance both the perception of audio-alone and visual-alone speech cues. As noted above, not all cochlear implant recipients can recognize speech through listening alone. For many of these patients, the cochlear implant serves as a sensory aid to improve lip reading skills they already have and use routinely in processing spoken language.

References

Auer, E.T., & Bernstein, L.E. (1997). Speechreading and the structure of the lexicon: computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *Journal of the Acoustical Society of America, 102*, 3704-3710.

Bernstein, L.E., Demorest, M.E., & Tucker, P.E. (2000). Speech perception without hearing. *Perception & Psychophysics, 62*, 233 - 252.

Bernstein, L.E., Auer, E.T., Jr., & Tucker, P.E. (2001). Enhanced speechreading in deaf adults: Can short-term training/practice close the gap for hearing adults? *Journal of Speech, Language, and Hearing Research, 44*, 5 - 18.

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning perception and production. *Perception and Psychophysics, 61*, 977-985.

Bradlow, A.R., & Pisoni, D.B. (1999). Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America, 106*, 2074-2085.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29*, 655.

Demorest, M. E., & Bernstein, L. E. (1992). Sources of variability on speechreading sentences: a generalizability analysis. *Journal of Speech and Hearing Research, 35*, 876-891.

Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., & Smith, L. B. (1991). Vowel and consonant recognition with the aid of a multichannel cochlear implant. *Quarterly Journal of Experimental Psychology, 43*, 585-601.

Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *Journal of the Acoustical Society of America, 104*, 3583-3596.

Dorman, M. F., Loizou, P. C., Kemp, L. L., & Kirk, K. I. (2000). Word recognition by children listening to speech processed into a small number of channels: Data from normal-hearing children and children with cochlear implants. *Ear and Hearing, 21*, 590-596.

Erber, N.P. (1972). Auditory, visual and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech and Hearing Research, 15*, 413-422.

Erber, N.P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*, 481-492.

Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3-28.

Fowler, C.A., & Dekle, D.J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 17*, 816-828.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, Gantz, B., & Woodworth, G. (1997). Cochlear implant use by prelingually deafened children: The influences of age at implant and length of device use. *Journal of Speech, Language, and Hearing Research, 40*, 183-199.

Gaver, W. W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology, 5*, 1 - 29.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America, 104*, 2438-2450.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103*, 2677-2690.

Kirk, K. I. (2000). Challenges in the clinical investigation of cochlear implant outcomes. In J. K. Niparko, K. I. Kirk, N. K. Mellon, A. M. Robbins, D. L. Tucci, & B. S. Wilson (Eds.), *Cochlear implants: Principles and practices* (pp. 225-259). Philadelphia: Lippincott Williams & Wilkins.

Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech and Hearing Research, 40*, 1395-1405.

Kucera, F., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Lachs, L. (1996). Static vs. dynamic faces as retrieval cues in recognition of spoken words, *Research on Spoken Language Processing Progress Report 22* (pp. 141 - 177). Bloomington, IN: Speech Research Laboratory.

Lachs, L. (1999). A voice is a face is a voice, *Research on Spoken Language Processing No. 23*. Bloomington, IN: Speech Research Laboratory, Indiana University Bloomington.

Lachs, L., & Hernández, L. R. (1998). Update: The Hoosier Audiovisual Multitalker Database, *Research on Spoken Language Processing Progress Report 22* (pp. 377 -388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear & Hearing, 22,236-251*.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.

Massaro, D. W., & Cohen, M. M. (1995). Perceiving talking faces. *Current Directions in Psychological Science, 4*, 104-109.

Massaro, D. W., & Cohen, M. M. (1999). Speech perception in perceivers with hearing loss: synergy of multiple modalities. *Journal of Speech, Language, and Hearing Research, 42*, 21-41.

Massaro, D. W., & Cohen, M. M. (2000). Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception. *Journal of the Acoustical Society of America, 108*, 784-789.

McGurk, H., & MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics, 47*, 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words, *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory.

Nygaard, L. C., & Pisoni, D. B. (1995). Talker- and task-specific perceptual learning in speech perception. *ICPhS 95, 1*(Section 9.5), 194-197.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355 - 376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics, 57*, 989 - 1001.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Communication, 13*, 109-125.

Pisoni, D. B. (1996). Some thoughts on "normalization" in speech perception. In K.A. Johnson & J.W. Mullinnex, *Talker variability in speech processing* (pp. 9-32): Academic Press.

Remez, R. E., Fellowes, J. M., Pisoni, D. B., Goh, W. D., & Rubin, P. (1998). Multimodal perceptual organization of speech: evidence from tone analogs of spoken utterances. *Speech Communication, 26*, 65-73.

Remez, R. E., Rubin, P., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-950.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101*, 129-156.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 318 - 331.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.

Sheffert, S. M., Lachs, L., & Hernández, L. R. (1996). The Hoosier Audiovisual Multitalker Database, *Research on Spoken Language Processing No. 21* (pp. 578 - 583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise masked normal hearing and cochlear implant listeners. I: The effects of response format. *Ear and Hearing, 18*, 89-99.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America, 96*, 1314-1324.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: The MIT Press.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of Acoustical Society of America, 26*, 212-215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* . Hillsdale.

Tyler, R. S., Fryauf-Bertschy, H., Kelsay, D. M., Gantz, B., Woodworth, G., & Parkinson, A. (1997a). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology Head and Neck Surgery, 117*, 180-187.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., Lowder, M. W., & Gantz, B. J. (1997b). Performance over time of adult patients using the Ineraid or nucleus cochlear implant. *Journal of the Acoustical Society of America, 102*, 508-522.

Watson, C. S., Qiu, W. W., Chamberlain, M. M., & Li, X. (1996). Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. *Journal of the Acoustical Society of America, 100*, 1153-1162.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

PET Imaging of Differential Cortical Activation by Monaural Speech and Nonspeech Stimuli[1]

Donald Wong,[2,3] David B. Pisoni,[3] Jennifer Learn,[4] Jack T. Gandour,[5] Richard T. Miyamoto[3] and Gary D. Hutchins[6]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This study was supported by the Departments of Otolaryngology, Radiology, NIH Grants DC 00064 (RTM) and DC 000111 (DBP). We thank Rich Fain and PET-facility staff for assistance and radionuclide production and developers of Michigan PET software for its use.

[2] Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN 46202.
[3] Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN 46202.
[4] Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202.
[5] Department of Audiology and Speech Sciences, Purdue University, West Lafayette, IN 47907.
[6] Department of Radiology, Indiana University School of Medicine, Indianapolis, IN 46202.

PET Imaging of Differential Cortical Activation by Monaural Speech and Nonspeech Stimuli

Abstract. PET imaging was used to investigate the brain activation patterns of listeners presented monaurally (right-ear) with speech and nonspeech stimuli. The major objectives were to identify regions involved with speech and nonspeech processing, and to develop a stimulus paradigm suitable for studies of cochlear-implant subjects. Scans were acquired under a silent condition and stimulus conditions that required listeners to press a response button to repeated words, sentences, time-reversed (TR) words, or TR sentences. Group-averaged data showed activated foci in the posterior superior temporal gyrus (STG) bilaterally and in or near the anterior insula/frontal operculum across all stimulus conditions compared to silence. The anterior STG was activated bilaterally for speech signals, but only on the right side for TR sentences. Only nonspeech conditions showed frontal-lobe activation in both the *left* inferior frontal gyrus [Brodmann's area (BA) 47] and ventromedial prefrontal areas (BA 10/11). An STG focus near the superior temporal sulcus was observed for sentences compared to words. The present findings show that both speech and nonspeech engaged a distributed network in temporal cortex for early acoustic and prelexical phonological analysis. Yet backward speech, though lacking semantic content, is perceived as speechlike by engaging prefrontal regions implicated in lexico-semantic processing.

Introduction

Functional neuroimaging techniques have provided a noninvasive tool for elucidating the neural circuits engaged in speech perception and spoken-language processing (Petersen et al., 1988, 1989). These studies of language processing have identified sites in the left inferior frontal cortex and posterior temporal cortex, regions classically implicated as speech/language centers from postmortem studies with aphasic patients (Geschwind, 1979). Furthermore, by comparing brain activation patterns under task conditions requiring different levels of signal processing and analysis, more extensive regions beyond the classical regions have also been identified for speech processing (Peterson & Fiez, 1993; Binder et al., 1997). Based on several recent imaging studies in speech processing, a widely distributed neural network has been hypothesized that links frontal and temporo-parietal language regions (for review, see Brown et al., 1999).

The motivation of the present investigation was twofold. First, we wanted to understand the cortical mechanisms underlying speech perception and spoken language processing. Both speech and nonspeech signals were used in order to identify cortical sites associated with complex-signal processing from sensory to lexico-semantic stages. Previous imaging studies have used time-reversed (TR) speech as a control condition to compare with speech signals (Howard et al., 1992; Price et al., 1996; Hirano et al., 1997; Binder et al., 2000). TR speech preserves some of the acoustical properties of speech sounds, but these nonspeech signals are devoid of semantic content. Second, we hoped to develop methods with normal-hearing listeners that could be used in future PET-imaging studies of patients with cochlear implants (CIs). Monaural stimulation was employed with these normal-hearing subjects to simulate the clinical conditions in which CI subjects hear on the side fitted with the prosthetic device (Naito et al., 1995, 1997; Wong et al., 1999). Another important consideration was selection of sets of stimuli that have served as standard test batteries for assessing speech perception in hearing-impaired subjects. Lists of words and sentences were presented as test signals under different scanning conditions. Listeners were required to detect a consecutive repetition in a sequence of stimuli by pressing a response button. TR words and TR sentences served as the nonspeech control conditions. Monaural stimulation was used in all

acoustic conditions in order to compare the findings with recently emerging imaging data from speech/language studies using binaural presentation.

Materials and Methods

Subjects

Five right-handed subjects (3 males, 2 females) with normal-hearing sensitivity (pure-tone air-conduction threshold $\leq$ 20 dB H.L. at 0.5, 1.0, 2.0, 4.0 kHz) and with a mean age of 28.3 $\pm$ 10.1 years (mean + SD) participated in this study. Payment was given for their participation. All subjects provided written consent to the experimental protocols for this study, which was approved by The Institutional Review Board of IUPUI and Clarian and was in accordance with the guidelines of the Declaration of Helsinki.

Stimuli and Tasks

An audiotape cassette reproduced the test signals at approximately 75 dB SPL for a total of 3-½ min during scanning under the acoustic conditions. The acoustic signals were originally recorded onto a tape and played back as free-field stimuli delivered by a high-quality loudspeaker approximately 18 inches from the right ear. The left ear was occluded using an E-A-R foam insert to attenuate sound transmission in this ear by at least 25-30 dB SPL. This monaural stimulation mimics the clinical condition found with monaural listening by CI subjects hearing with a prosthesis implanted into only one ear.

Table 1.
Stimulus Paradigm

Condition	Auditory Stimulus (Right Ear)	Examples	Motor Response (Right Thumb)
1. Baseline	Silence		No response required
2. Speech	Words	boost, fume, thread, calf, day, bind	Press button after repeated stimuli
3. Speech	Sentence	What joy there is in living. See the cat glaring at the scared mouse.	Press button after repeated stimuli
4. Non-Speech	TR Word		Press button after repeated stimuli
5. Non-Speech	TR Sentence		Press button after repeated stimuli

Prior to the scanning session, subjects were told that they would hear either speech or nonspeech signals in each of the acoustic conditions. However, they were not informed about the specific type of acoustic signal to be presented prior to each scan. Scanning conditions were either active or passive. In

the passive, silent baseline condition, the subject was instructed to relax. There was no stimulus presentation or response required during this scanning condition. In the active task conditions, subjects grasped a response device with their right hand, and were instructed to press a button with their right thumb immediately after an acoustic stimulus (e.g., word or sentence) was consecutively repeated. This detection task was designed to direct the subjects' attention to the sound pattern and to monitor the stimulus sequence for repetitions. Thus, the speech and nonspeech tasks were conceptualized as simple auditory discrimination tasks (see paradigm in Table 1). Each button-press response activated a red light, which the experimenter observed and used to score the number of correct responses during each active task condition. Subjects were debriefed after the imaging session to discuss relative task difficulty and their subjective impressions of the perceived stimuli.

Speech stimuli consisted of lists of isolated English words and meaningful sentences used for speech-intelligibility testing (Egan, 1948). The Word condition used a list of 54 phonetically-balanced, monosyllables (e.g., "*boost", "fume", "each", "key", "year", "rope"*). The Sentence condition used a list of 44 Harvard sentences (e.g., "*What joy there is in living*."; "*Those words were the cue for the actor to leave*."; "*The wide road shimmered in the hot sun*.") selected from the lists developed by Egan (1948). Six to ten words comprised each sentence. TR versions of the same words and sentences were used for the two nonspeech conditions. These stimuli consisted of the 54 words played backwards in the TR Word condition, and 44 sentences played backwards in the TR Sentence condition. The percentage of consecutively repeated stimuli averaged 21% across all acoustic conditions. TR speech was considered to be devoid of semantic content, and was therefore considered an appropriate nonspeech control. The lists were presented at a rate of 1 stimulus per 3 sec in the Word and TR Word conditions and at 1 stimulus per 4 sec in the Sentence and TR Sentence conditions. The duration of each list was approximately 3 ½ min long.

PET Image Acquisition and Processing

PET scans were obtained using a Siemens 951/31R imaging system, which produced 31 brain image slices at an intrinsic spatial resolution of approximately 6.5 mm full-width-at-half-maximum (FWHM) in plane and 5.5 mm FWHM in the axial direction. During the entire imaging session, the subject lay supine with his/her eyes blindfolded. Head movement was restricted by placing the subject's head on a custom-fit, firm pillow, and by strapping his/her forehead to the imaging table, allowing pixel-by-pixel within-subject comparisons of cerebral blood flow (CBF) across task conditions. A peripheral venipuncture and an intravenous infusion line were placed in the subject's left arm. For each condition, about 50 mCi of $H_2{}^{15}O$ was injected intravenously as a bolus; upon bolus injection, the scanner was switched on for 5 min to acquire a tomographic image. During the active acoustic conditions, sounds were played over a 3 ½ min period followed by 1 ½ min of silence. A rapid sequence of scans was performed to enable the selection of a 90-s time window beginning 35-40 s after the bolus arrived in the brain. For each condition in the experimental design, instructions were given immediately prior to scanning. Repeated scans were acquired from subjects in the following stimulus conditions: *(1) Silent Baseline, (2) Word, (3) Sentence, (4) TR Word, (5) TR Sentence.*

Seven paired-image subtractions were then performed on group-averaged data to reveal statistically significant results in the difference images: (1) Word – Silence, (2) Sentence – Silence, (3) TR Word – Silence, (4) TR Sentence – Silence, (5) Sentence – Word, (6) Sentence – TR Sentence, and (7) Word – TR Word. The Sentence – Word subtraction was designed to dissociate processing of suprasegmental or prosodic cues at the sentence-level from those at the level of isolated words. In the Sentence – TR Sentence condition, nonspeech is subtracted from speech. Regions of significant brain activation were identified by performing an analysis process (Michigan software package**,** Minoshima et al., 1993) that included image registration, global normalization of the image volume data, identification of the

intercommissural (anterior commissure – posterior commissure) line on an intrasubject-averaged PET image set for stereotactic transformation and alignment, averaging of subtraction images across subjects, and statistical testing of brain regions demonstrating significant regional CBF changes. Changes in regional CBF were then mapped onto a standardized coordinate system of Talairach and Tournoux (1988). Foci of significant CBF changes were tested by the Hammersmith method (Friston et al., 1990, 1991) and values of $p \leq 0.05$ (one-tailed, corrected) were identified as statistically significant. The statistical map of blood flow changes was then overlaid onto a high-resolution T1-weighted, structural MRI of a single subject for display purposes to facilitate identification of activated and deactivated regions with respect to major gyral and sulcal landmarks under each of the subtractions.

The multiple foci of significant peak activation in the superior temporal gyrus were distinguished by arbitrarily grouping these foci into anterior ($y \geq$ -5 mm) middle (y from –5 to –23 mm), and posterior (y from –24 to –35 mm) (Wong et al., 1999). The extent of activation was determined only in the superior temporal gyrus (STG) of each hemisphere by drawing regions of interest (ROIs) around the activation foci at the Hammersmith threshold. A single ROI was drawn on each side to include the extent of activation from all peak foci of STG activation.

Results

Behavioral Performance

The subjects scored 100% on the detection tasks for the Word and Sentence conditions. On the nonspeech tasks, a total of one error was scored in the TR Word condition, and two errors in the TR Sentence condition for all subjects.

Foci of Significant Blood Flow Increases

Compared to the silent baseline condition, in both the Sentence and Word conditions, extensive CBF increases were observed bilaterally in the STG (Table 2; **Fig. 1,** upper two panels). The STG activation pattern was generally more robust and larger on the *left* side for all baseline subtractions in this study using right-ear stimulation. The activated region was elongated in an anterior-to-posterior direction with multiple peak foci distinguishable in the anterior, middle, and posterior parts of the STG (Table 2: foci # 4-9, 14-20; **Fig. 1**). The activations in the posterior half of the STG were often in the superior temporal plane within the Sylvian fissure, presumably encompassing the primary and secondary association auditory cortex [Brodmann's area (BA) 41/42]. This activation pattern also extended ventrally onto the lateral (exposed) STG surface as far as the superior temporal sulcus (STS) and middle temporal gyrus (MTG), especially on the left side. This robust activation presumably included a part of BA 22 on the lateral surface and a part of BA 21 in the banks of the STS or on the MTG. The anterior STG activation was typically observed on the lateral surface, near the STS, and toward BA 38, a region containing the temporal pole. CBF increases were consistently found at the junction between the anterior insula and the frontal operculum on the left side (Table 2: foci #10-11, 21-22; **Fig. 1**) (bilateral for Sentence). No CBF increases were found in the inferior frontal gyrus (IFG) of the *left* frontal cortex.

Compared to the silent baseline condition, in both the TR Sentence and TR Word conditions, CBF increases were observed in the temporal lobe bilaterally (Table 3; **Fig. 1**). Compared to baseline, the TR Sentence showed a robust bilateral STG activation (Table 3: foci #13-16), a pattern similar to that observed for the Sentence minus baseline condition. The strong left posterior STG activation also extended ventrally as far as the STS/MTG (Table 3: focus #13; **Fig. 1**), a spread of activity similar to that found for the speech conditions compared to baseline. Noteworthy is the pattern of STG activation on the right side, which contains multiple anterior and posterior foci (Table 3: foci #14-16); the foci extended

along the lateral STG surface, but did not spread to the STS. Compared to baseline, the TR Word condition showed a noticeably weaker STG activation (Table 3: foci #4-6) than that observed for the TR Sentence condition (**Fig. 1**, lower two panels). The temporal-lobe activation was mainly on the left side in the posterior STG and MTG. Only a single focus was observed in the posterior STG on the right side. The activated focus observed for TR Sentence minus baseline condition was an elongated swath of activity on the left lateral STG surface along the anterior-to-posterior direction similar to that found in the Sentence minus baseline condition. In contrast, the left STG focus for TR Word-baseline was more focally confined to the posterior STG, extending ventrally rather than anteriorly (**Fig. 1**). Examination of the activation patterns of all four baseline comparisons revealed that both the Sentence and TR Sentence conditions evoked larger activations than the Word and TR Word conditions. These larger STG activations occurred extensively along the anterior-to-posterior direction, whereas the smaller activations were confined only to the posterior STG.

Compared to the baseline condition, the two nonspeech conditions showed CBF increases in foci of the frontal lobes that were not observed in the speech conditions (Table 3). For example, activation foci were found in the *left* inferior frontal gyrus (*pars orbitalis*, BA 47) (Table 3: foci #1-2, 8-9; **Fig. 2**), a region often referred to as ventral or inferior prefrontal cortex in imaging studies on language processing (see Fiez, 1997). A second pattern of frontal-lobe activation was also found bilaterally in a part of the frontopolar region; this activation was largely confined to the ventromedial frontal cortex in BA 10/11 (**Fig. 2**).

CBF increases were also isolated when the speech and nonspeech conditions were compared (Table 4). In the Sentence minus Word condition, a focus of CBF increase was found in the middle part of the left STG (BA 22) near the STS (Table 4: focus #1; **Fig. 3**). The TR Sentence minus TR Word condition was the only other comparison between two active tasks to show CBF increases in the STG (Table 4: foci #8-10). No CBF increases were found in any speech region of the left frontal lobe for comparisons between speech conditions (Sentence – Word) or between speech and nonspeech conditions (Sentence-TR Sentence). No significant CBF increase was found for Word – TR Word.

Table 6 summarizes the major similarities and differences in the patterns of CBF increases for speech and nonspeech conditions relative to silent baseline. In brief, activation patterns in posterior STG bilaterally and anterior insula/frontal operculum were found consistently across all speech and nonspeech conditions. Bilateral activation in anterior STG was observed for the speech conditions only. Activation in both the ventromedial prefrontal cortex and left inferior frontal gyrus was found only in the nonspeech conditions. When different levels of complex-sound processing were compared, a focus in the left STG/MTG in the STS was observed for the sentence condition compared to the word condition.

Foci of Significant Blood Flow Decreases

Compared to the baseline condition, all speech and nonspeech conditions showed CBF decreases typically in the medial parietal and occipital lobe in such regions as the precuneus (BA 7) and posterior cingulate (BA 31, 30, 23) (Table 7: foci #1, 6-8, 10-12). The majority of these foci were found for the TR Sentence minus baseline condition (**Fig. 4**).

Table 2.
Speech task compared to silent baseline.
Regions of significant blood flow increases*

Regions	Brodmann's Area	Coordinates (mm) x	y	z	Z score
Word - Baseline					
Frontal lobe					
1. R inferior frontal gyrus	47	48	30	-7	4.8
2. L orbital gyrus/gyrus rectus	11	-10	12	-18	4.9
3. L middle frontal gyrus	11	-19	19	-16	4.8
Temporal lobe					
4. L anterior superior temporal gyrus/STS	22/21	-46	-4	-7	4.8
5. L posterior superior temporal gyrus/STS	22	-55	-28	4	6.6
6. R anterior superior/middle temporal gyrus/STS	22/21/38	48	5	-11	4.3
7. R anterior superior temporal gyrus/STS	22/21	57	-4	-2	4.9
8. R mid superior temporal gyrus	22	53	-15	2	4.5
9. R post superior temporal gyrus/STS	22	57	-28	4	5.9
Other					
10. L anterior insula/frontal operculum	-	-35	19	0	5.2
11. L anterior insula	-	-33	5	-11	4.5
12. L posterior insula	-	-33	-31	9	5.4
13. R thalamus (dorsomedial nucleus)		1	-17	2	4.6
Sentence - Baseline					
Temporal Lobe					
14. L anterior superior temporal gyrus/STS	22/21	-48	3	-9	6
15. L transverse gyrus of Heschl	42/41	-33	-31	14	5.5
16. L posterior superior temporal gyrus/STS	22	-53	-28	4	8.4
17. R anterior superior temporal gyrus	38	46	10	-9	5
18. R mid superior temporal gyrus	22/42	53	-15	2	6.1
19. R mid superior temporal gyrus	22	55	-19	4	6.1
20. R posterior superior temporal gyrus/STS	22/21	55	-28	2	6.1
Other					
21. L anterior insula/frontal operculum	-	-35	21	2	5.4
22. R anterior insula/frontal operculum	-	39	14	0	5.1

*Significant activation foci that exceed the Hammersmith statistical criterion of significance (adjusted p threshold = .05) in normalized CBF for all subtractions. Stereotaxic coordinates, in millimeters, are derived from the human brain atlas of Talairach and Tournoux (1988). The x-coordinate refers to medial-lateral position relative to midline (negative = left); y-coordinate refers to anterior-posterior position relative to the anterior commissure (positive = anterior); z-coordinate refers to superior-inferior position relative to the CA-CP (anterior commissure-posterior commissure) line (positive = superior). Designation of Brodmann's areas is also based on this atlas. L = left; R = right.

Table 3.
Nonspeech task compared to silent baseline.
Regions of significant blood flow increases*

Regions	Brodmann's Area	Coordinates (mm) x	y	z	Z score
Time-Reversed Word - Baseline					
Frontal lobe					
1. L Inferior frontal gyrus, *pars orbitalis*	47	-35	28	-2	4.4
2. L Inferior frontal gyrus, *pars orbitalis*	47	-35	39	-7	4.5
3. R orbital gyrus	11	10	46	-18	4.4
Temporal lobe					
4. L posterior superior temporal gyrus	22/42	-51	-31	7	6.2
5. L posterior middle temporal gyrus	21/20	-51	-40	-9	4.2
6. R posterior superior temporal gyrus	42	57	-26	7	4.2
Other					
7. L insula	-	-37	1	-16	5
Time-Reversed Sentence - Baseline					
Frontal Lobe					
8. L inferior frontal gyrus, *pars orbitalis*	47	-30	41	-9	5.2
9. L inferior frontal gyrus (*pars orbitalis*)/middle frontal gyrus	47/11/10	-39	41	2	4.8
10. L orbital gyrus	11	-12	39	-18	5.1
11. R middle frontal gyrus	11	26	48	-11	4.5
12. R middle frontal gyrus	11/10	37	48	-4	4.5
Temporal lobe					
13. L posterior superior temporal gyrus	22/42	-53	-22	2	10
14. R anterior superior temporal gyrus	38	48	8	-9	5.3
15. R anterior superior temporal gyrus	22	57	-6	-2	6.8
16. R posterior superior temporal gyrus	22/42	57	-26	4	7.1
Other					
17. L anterior insula/frontal operculum		-35	23	0	5.1
18. R anterior insula/frontal operculum	-	39	19	2	5.7
19. R inferior parietal lobule	40	44	-49	45	4.3

*Significant activation foci that exceed the Hammersmith statistical criterion of significance (adjusted *p* threshold = .05) in normalized CBF for all subtractions. Stereotaxic coordinates, in millimeters, are derived from the human brain atlas of Talairach and Tournoux (1988). The x-coordinate refers to medial-lateral position relative to midline (negative = left); y-coordinate refers to anterior-posterior position relative to the anterior commissure (positive = anterior); z-coordinate refers to superior-inferior position relative to the CA-CP (anterior commissure-posterior commissure) line (positive = superior). Designation of Brodmann's areas is also based on this atlas. L = left; R = right.

Table 4.
Comparison of speech and nonspeech tasks.
Regions of significant blood flow increases*

Regions	Brodmann's Area	Coordinates (mm) x	y	z	Z score
Sentence - Word					
Temporal Lobe					
1. L mid superior temporal gyrus (in STS)	22	-53	-13	-2	4.3
2. L post thalamus (pulvinar)		-12	-35	7	4.5
Other					
3. R frontal operculum/anterior insula	-	35	-1	16	4.2
Sentence - TR Sentence					
Parietal/Occipital Lobe					
4. L post cingulate	23	-3	-55	18	4.4
5. L post cingulate	23//31	-3	-46	27	4.4
6. L precuneus	31	-6	-62	20	4.3
TR Sentence -TR Word					
Frontal Lobe					
7. L orbital gyrus	11	-15	41	-18	4.1
Temporal Lobe					
8. L mid superior temporal gyrus	22/42	-55	-17	2	6.5
9. L transverse gyrus of Heschl	41	-37	-31	7	4.3
10. R anterior superior temporal gyrus	22/38	53	-1	-7	4.1

*Significant activation foci that exceed the Hammersmith statistical criterion of significance (adjusted *p* threshold = .05) in normalized CBF for all subtractions. Stereotaxic coordinates, in millimeters, are derived from the human brain atlas of Talairach and Tournoux (1988). The x-coordinate refers to medial-lateral position relative to midline (negative = left); y-coordinate refers to anterior-posterior position relative to the anterior commissure (positive = anterior); z-coordinate refers to superior-inferior position relative to the CA-CP (anterior commissure-posterior commissure) line (positive = superior). Designation of Brodmann's areas is also based on this atlas. L = left; R = right.

Table 5.
ROI analysis of superior temporal gyrus

Subtraction	Volume (ml)		L/R
	Left	Right	Ratio
Word - Baseline	10.82	4.88	2.20
Sentence - Baseline	18.81	8.04	2.30
Time-Reversed Word - Baseline	3.90	0.02	195.00
Time-Reversed Sentence - Baseline	24.66	9.58	2.60

Table 6.
Summary of key regions engaged in speech and nonspeech tasks compared to silence.

Region	Brodmann's	SPEECH		NONSPEECH	
	Area	Sentence	Word	TR Sentence	TR Word
Anterior superior temporal gyrus	22/21/38	L/R	L/R	R	--
Posterior superior temporal gyrus	22	L/R	L/R	L/R	L/R
Anterior insula/frontal operculum	--	L/R	L	L/R	L
Inferior frontal gyrus					
(*pars orbitalis*)	47	--	R	L	L
Frontopolar region	10/11	--	--	L/R	L

Table 7.
Regions of significant blood flow decreases*

Regions	Brodmann's Area	Coordinates (mm) x	y	z	Z Score
Word - Baseline					
1. R precuneus	31/30	8	-64	11	-5.4
2. R superior parietal lobule	7	21	-49	54	-4.4
3. R superior parietal lobule	7	15	-60	47	-4.8
4. R inferior temporal gyrus/fusiform gyrus	19	37	-67	-2	-4.7
5. R medial occipital gyrus	19	28	-76	14	-4.7
Sentence - Baseline					
6. R precuneus	7	3	-67	9	-4.4
Time-Reversed Word - Baseline A					
7. R precuneus	7	8	-55	45	-4.2
Time-Reversed Sentence - Baseline B					
8. L precuneus	7	-6	-69	20	-5.3
9. L superior parietal lobule	7	-17	-49	52	-4.5
10. R post cingulate	23	1	-46	25	-5.6
11. R post cingulate	23/30/31	6	-55	14	-6.1
12. R precuneus	7	1	-64	40	-5.1
13. R superior parietal lobule	7	17	-49	54	-5.6
14. R cuneus	17/18	3	-69	9	-5.7
Sentence - Word					
15. R anterior cingulate	32	6	39	14	-4.5
Sentence - TR Sentence					
16. L transverse gyrus of Heschl	42/41	-57	-17	7	-4.1
TR Sentence -TR Word					
17. R fusiform gyrus	18	30	-85	-20	-4.2
Word - TR Word					
18. L fusiform gyrus	18/19	-24	-76	-16	-4.6

* Significant de-activation foci that exceed the Hammersmith statistical criterion of significance (adjusted *p* threshold = .05) in normalized CBF for all subtractions. See Table 2 footnote regarding Talairach coordinates and Brodmann's areas.

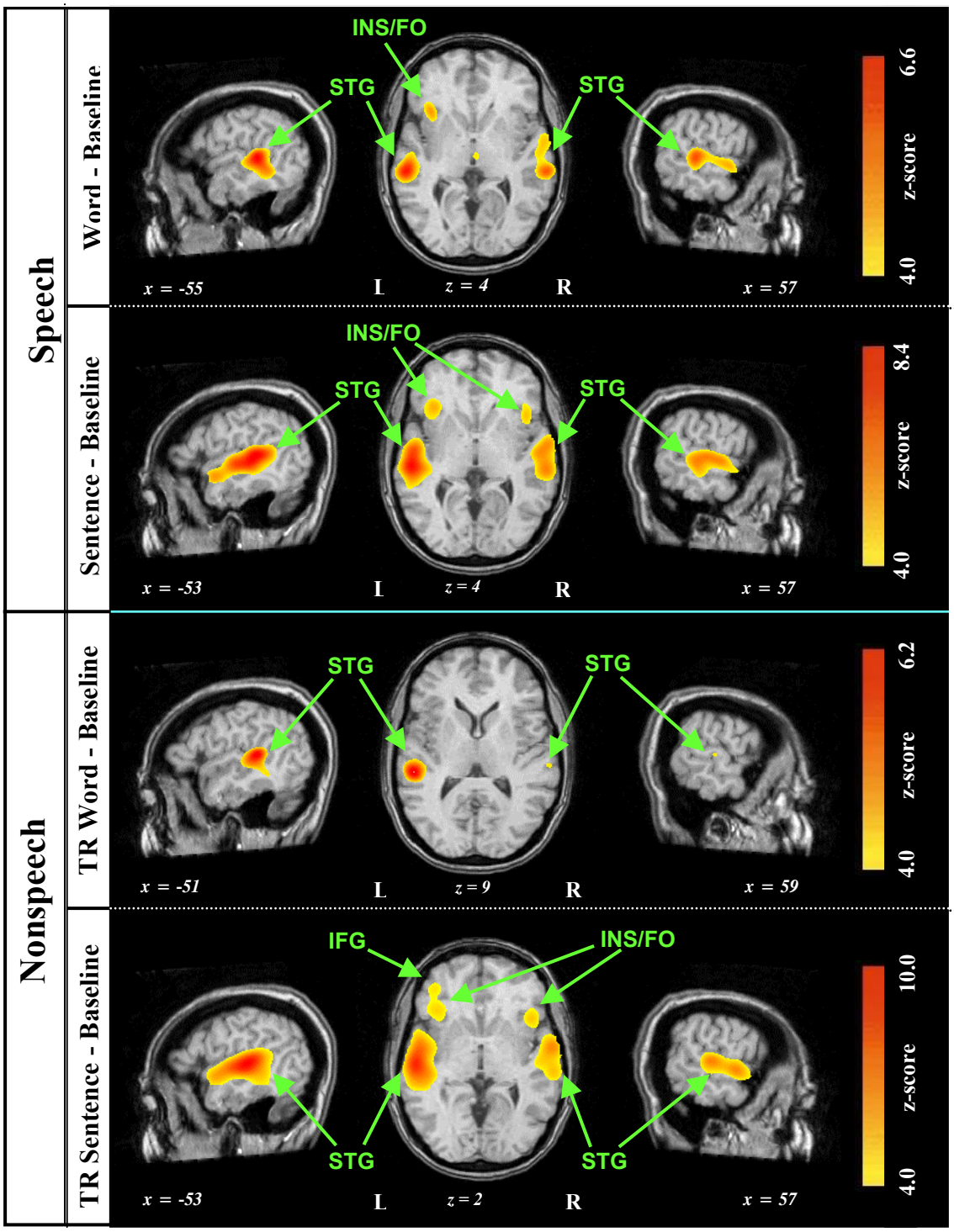


Figure. 1. Baseline subtraction of the speech and nonspeech conditions activated the superior temporal gyrus (STG) bilaterally with a more extensive activation observed on the left side. The STG activation also was more extensive for the Sentence and TR Sentence conditions than the Word and TR Word conditions. Activation also was observed at the junction of the anterior insula (INS) and the frontal operculum (FO) for all baseline subtractions. The TR Sentence minus Baseline condition shows an activation in the left inferior frontal gyrus (IFG, bottom panel; see Fig. 3). All sound stimuli were monaurally presented into the *right* ear.

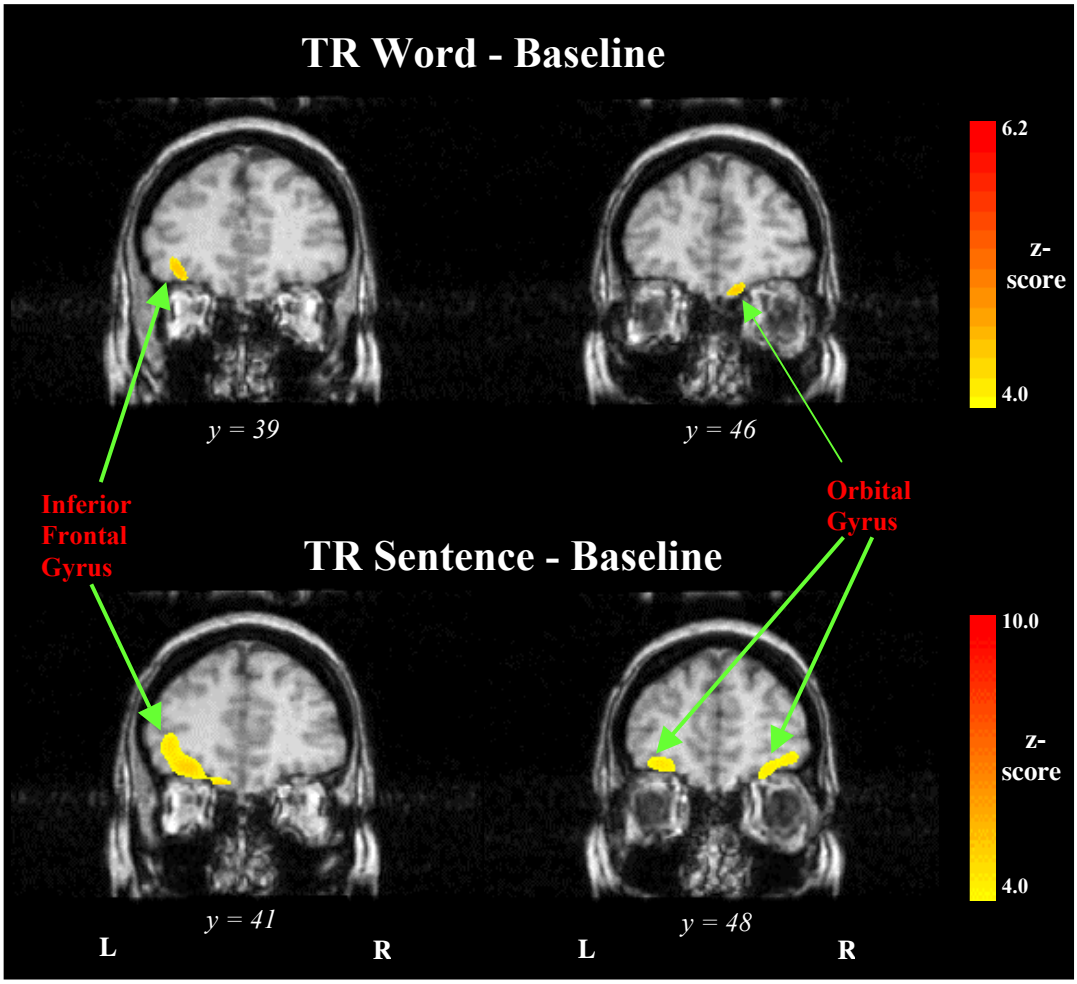


Figure. 2. Both nonspeech minus baseline conditions activated the inferior frontal gyrus, *pars orbitalis*, on the *left* side only. Other frontal activations included the orbital gyrus on both sides.

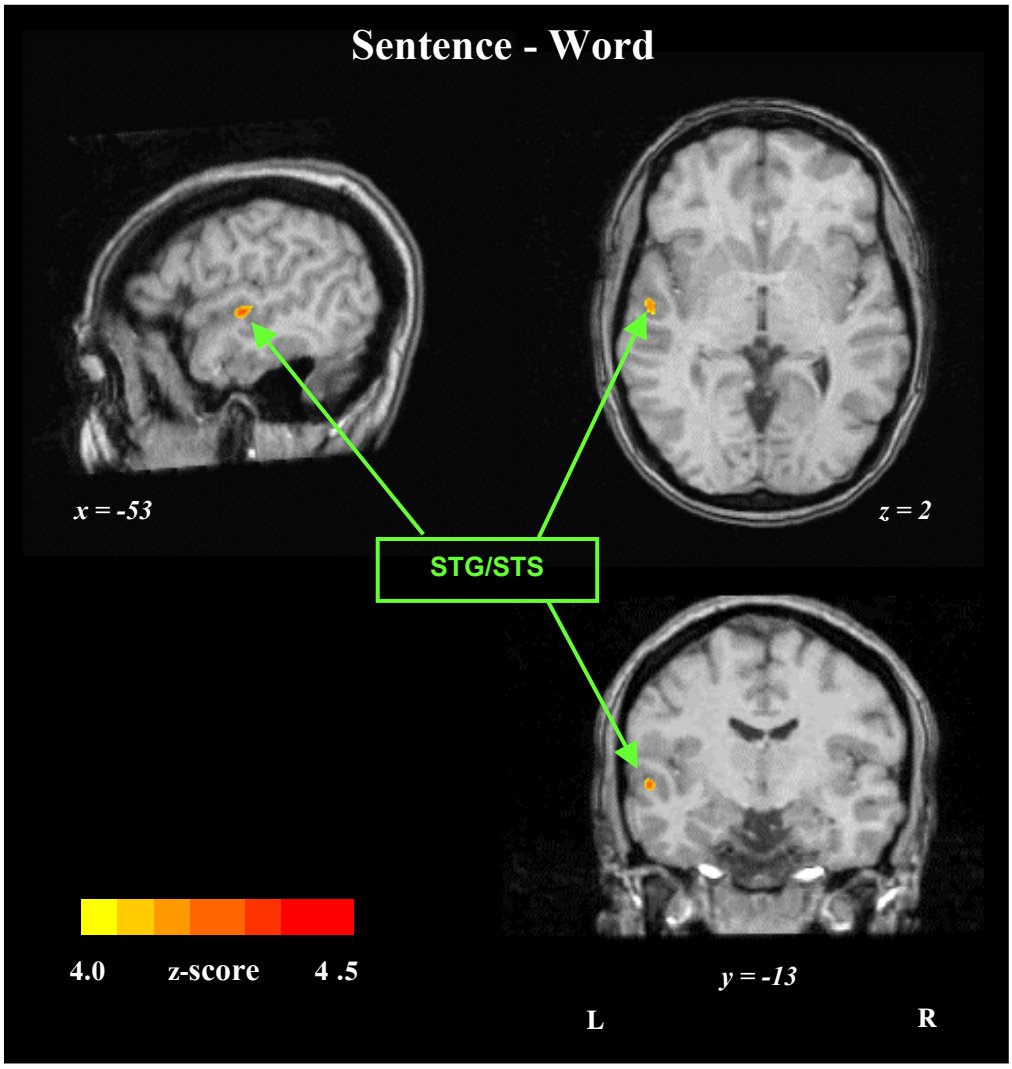


Figure. 3. The Sentence minus Word condition activated the temporal lobe on the left side only in the midportion of the superior temporal gyrus (STG) near or in the superior temporal sulcus (STS).

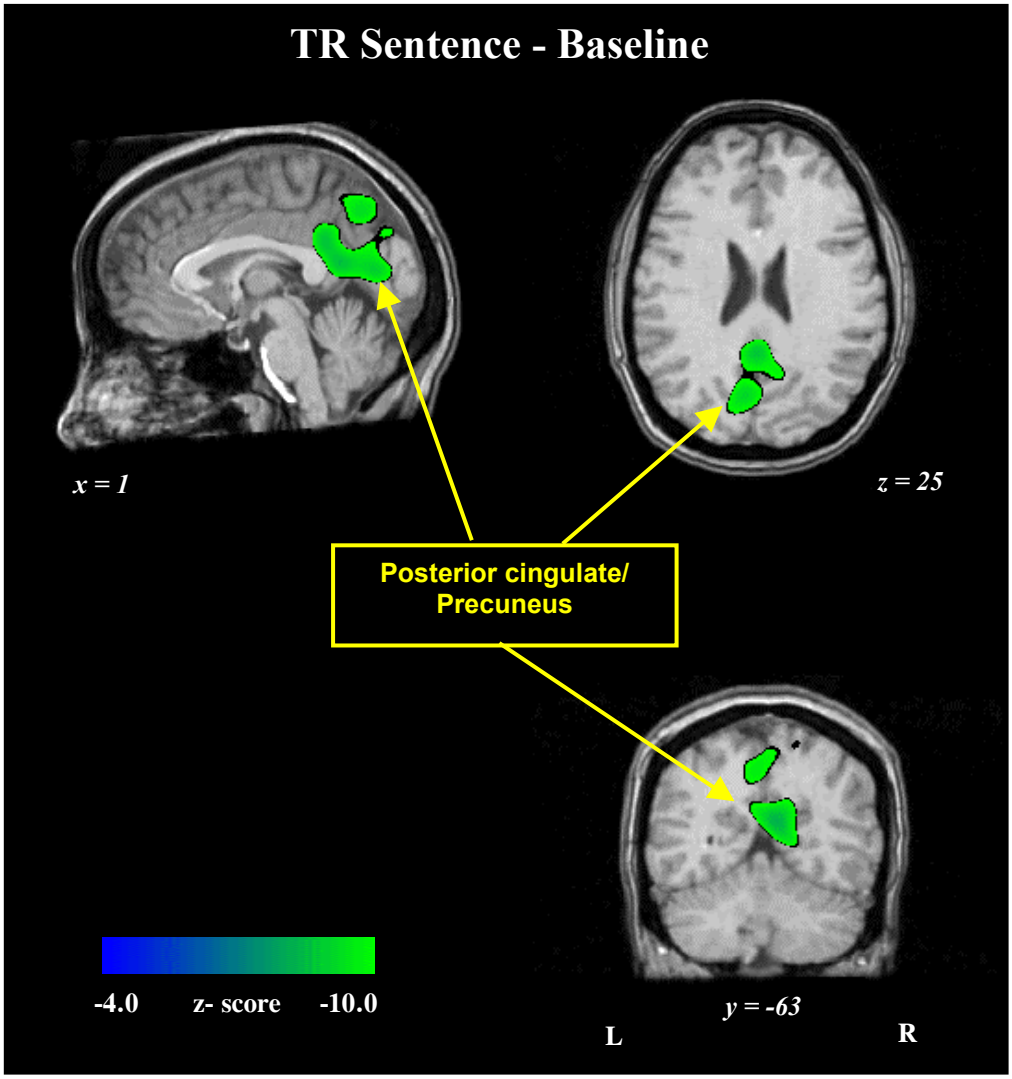


Figure. 4. The time-reversed Sentence minus baseline condition showed the most extensive deactivation mainly in the precuneus and posterior cingulate in the medial parietal lobe.

Discussion

Frontal-Lobe Activation

Two separate activations foci in the frontal lobe were found almost exclusively for the nonspeech conditions compared to the silent baseline condition: in the *left* inferior (BA 47) and ventromedial (BA 11/10) prefrontal cortex. The activation observed in the left inferior prefrontal cortex was located in the inferior frontal gyrus (IFG), *pars orbitalis* (BA 47; Table 3: foci #1-2, 8-9). This perisylvian focus is over 20 mm rostral to the activation in the frontal operculum/anterior insula (Table 3: foci #7, 17-18), and is at least 18 mm from the second prefrontal focus.

Neuroimaging studies over the last decade have demonstrated that the left inferior prefrontal cortex (BA 44, 45, 46, 47) shows the strongest activation to semantic task demands (for review see Price et al., 1999). The finding that the left IFG was activated only under the *nonspeech* conditions compared to silent baseline was somewhat unexpected. Thus, one would have expected the speech conditions to activate the part of the left IFG that contains the classically defined Broca's area (*pars opercularis, pars triangularis*: BA 44/45) and is associated mainly with phonological processing. Petersen and colleagues (1988, 1989) were the first to demonstrate that the *pars orbitalis* (BA 47) of the left IFG showed the strongest activation pattern to semantic processing in speech production tasks. Imaging data obtained across word generation and semantic decision tasks provide further support for the association of semantic processing and the left inferior frontal cortex (e.g., Peterson et al., 1989; Wise et al., 1991; Shaywitz et al., 1995; Binder et al., 1997; Poldrack et al., 1999). This left prefrontal activation has also been found to extend into the temporal-parietal cortex, from the anterior temporal pole to as far posterior as the angular gyrus (Vanderberghe et al., 1996). Thus, based on both neuropsychological data and neuroimaging studies, the semantic processing system for spoken language may be mediated by a distributed neural network linking frontal regions with temporo-parietal language centers on the left hemisphere: BA 47, a part of the inferior frontal gyrus ventral to the classical-defined Broca's area (BA 44/45), has an executive role for maintaining and controlling the effortful retrieval of semantic information from the posterior temporal areas (Fiez, 1997), and the posterior temporal regions for stored semantic knowledge (Price et al., 1999).

The lack of a left prefrontal activation to speech compared to baseline may be explained by the fact that no explicit semantic association was required for the task in the present study. The subjects treated the task as a simple detection task. Despite the available lexico-semantic information contained in the familiar speech sounds, the speech tasks were apparently performed at the earlier acoustic and prelexical levels of analysis in the brain with minimal effort. This interpretation would account for the observed activation pattern in the superior temporal region that did not extend into the left inferior prefrontal cortex. In fact, similarity in the STG activation patterns for speech and backward speech is consistent with the hypothesis that regions of the temporal lobe are largely used for acoustic-phonetic processing (see Binder et al., 2000). The additional activation of the left prefrontal regions for the nonspeech signals may be related to greater task demands for backward speech that involved an effortful, although unsuccessful attempt at semantic retrieval and interpretation of these sound patterns. The presence of errors found only in the backward speech tasks and greater effort subjectively reported during post-scan debriefing are consistent with greater task demands in these types of auditory discrimination tasks. It is also possible that the activation in BA 47 for nonspeech tasks compared to silent baseline may be due to a relatively greater task demands placed on auditory working memory to maintain representations of these nonlinguistic stimuli. Models of working memory (Baddeley, 1996; Smith & Jonides, 1999) have suggested that executive processes, which provide the attentional control of working memory, may depend on the operation of the ventral part of Broca's area in Brodmann's area 45/47. The present study cannot dissociate the multiple cortical subsystems, whether task- or speech-specific, that are engaged when listening to backward speech.

The other major prefrontal activation that we observed was on the basal surface, typically in the ventromedial frontal cortex (BA 11/10) in the gyrus rectus and orbital gyrus (basal part). The further recruitment of this prefrontal focus, in addition to the left inferior prefrontal cortex, may be related to the generally greater task demands required in performing these tasks when listening to unfamiliar auditory signals. This activation was observed in or near the part of the frontopolar region that is engaged in verbal memory tasks involving monitoring of auditory inputs (Petrides et al., 1993). The frontopolar region has extensive interconnections with auditory regions of the superior temporal gyrus (Petrides & Pandya, 1984; Barbas & Mesulum, 1985). Other researchers have suggested that processes related to retrieval effort and search in memory, whether successful or not, may also engage regions of anterior prefrontal cortex near BA 10 (e.g., see Buckner et al., 1996; Schacter et al., 1996). In the present study, the backward sentence task, presumably the more demanding of the two nonspeech tasks, evoked multiple activated foci extending bilaterally in these areas.

Activation in Temporal Lobe

In neuroimaging studies with normal-hearing listeners, both binaural (for review see e.g., Peterson et al., 1988) and monaural studies (e.g., Lauter et al., 1985; Hirano et al., 1997; Scheffler et al., 1998) have demonstrated bilateral activation of the STG. In the monaural studies, a more extensive activation was observed on the hemisphere contralateral to the side of stimulus presentation. The larger temporal-cortex activation contralateral to monaural stimulation reflects the known fact that auditory inputs transmitted along the central auditory pathway are sent predominantly to the contralateral auditory cortex (primary and secondary association areas) (e.g., see Jones and Peters, 1985). This general pattern of bilateral activation in the STG was also found in the present monaural study. When speech (words, sentences) and nonspeech (TR words, TR sentences) conditions were compared to the silent baseline condition, the posterior STG**,** which includes the auditory cortex, was activated bilaterally, but displayed a stronger and more extensive focus in the left hemisphere, contralateral to the stimuli presented to the right ear.

The sound patterns for both the sentences and backward sentences evoked the largest overall level of activation, the isolated words evoked a smaller level of activation, and the TR words evoked the lowest level of activation within the STG (Table 5). This differing extent of activation may reflect the greater complexity of the sound patterns (Sentence and TR Sentence) versus the isolated stimuli (Word and TR Word). This interpretation is consistent with studies using binaural stimuli, in which the activated auditory regions in the temporal lobe became more extensive bilaterally as the task demands increased with the complexity of the speech stimuli (e.g. sentences differing in structural complexity) (Just et al., 1996).

Another possible explanation for the larger STG activation in the Sentence condition is related to differences in presentation rate between the Sentence and Word conditions (Price et al., 1992; Binder et al., 1994; Dhankhar et al., 1997). The rate of word presentation was slower in the Word than Sentence condition (isolated words versus sequence of words of a sentence during a unit of time). In the nonspeech conditions, a presentation rate effect may also have induced the larger activation observed for the TR Sentence than the TR Word. Furthermore, both the Word and TR Word conditions (compared to baseline) showed a more confined activation in the posterior STG that was as robust on the left as the more extensive STG activation in the Sentence and TR Sentence conditions. The fact that we find common activation in the posterior STG for both Sentence and Word conditions cannot be explained by a presentation-rate effect. Instead, this posterior STG focus appears to reflect processing demands associated with speech stimuli (Price et al., 1992).

In all stimulus conditions compared to silent baseline, the left posterior STG was consistently activated. This focus extended from within the Sylvian fissure to the lateral cerebral surface as far ventrally as the STS and MTG. Hirano et al. (1997) also observed activated foci in both the STG and the MTG when (Japanese) sentences were compared to backward sentences under monaural stimulation. Furthermore, Wise et al. (1991) noted that the bilateral activation of the STG showed a similar pattern when speech and backward speech were each compared to silent baseline under binaural presentation. This pattern of activation suggests that the posterior temporal gyrus of both sides participates in the initial cortical stages of sensory analysis of complex sounds, whether isolated words or sentences perceived as speech or nonspeech. Furthermore, the left posterior STG, especially in the vicinity of the STS and the posterior MTG, has been hypothesized to be involved with prelexical phonological processing (Mazoyer et al., 1993; Price et al., 1999). Our findings suggest that both speech and nonspeech were processed cortically beyond the early sensory level to at least the prelexical phonological stages in left-lateralized speech-specific sites.

The anterior STG (BA 22) was activated bilaterally only when the speech conditions were compared to the silent baseline condition. This pattern of activity was typically located ventrally near the STS/MTG (BA 22/21) and as far anteriorly as BA 38 in a region of the temporal pole (Rademacher et al., 1992). The view that the anterior STG/temporal pole of both sides involves speech-specific processing was previously proposed because bilateral activation of the anterior STG was consistently found when subjects listened to speech sounds (Petersen et al., 1988; 1989; Wise et al., 1991; Zatorre et al., 1992; Mazoyer et al., 1993; Binder et al., 1994). In the present study using monaural stimulation, the anterior STG was activated on the right side only when the nonspeech condition (backward sentence) was compared to silence. This finding, in conjunction with the bilateral activation observed in the anterior STG for the speech conditions, suggests that the anterior STG is engaged in speech-specific processing on the left side only. The ipsilateral activation of the right anterior STG found in the backward sentence condition strongly suggests that this acoustic stream engaged right-hemispheric mechanisms specialized for the encoding and storage of prosodic and intonation cues (Zatorre et al., 1992; Griffiths et al., 1999). The absence of right-sided hemispheric activity for the backward word condition compared to baseline is unclear, although this may be due simply to the relatively lower overall level of STG activation observed for this task. The bilateral activation under the speech conditions is also consistent with the interpretation that prosodic processing of the speech stimuli is also lateralized to this right homologous region (Zatorre et al., 1992; Mazoyer et al., 1993).

Mazoyer et al. (1993) have associated bilateral activation of the temporal poles with binaural listening to continuous speech (e.g., stories). A similar bilateral activation of the anterior STG was also evoked in the present monaural study, although continuous speech stimuli were not essential. Both the Word and Sentence conditions compared to silent baseline gave rise to an anterior STG focus. Moreover, our findings support the specialized role of this region in the processing of spoken language at the lexico-semantic level. When compared to silent baseline, the relatively stronger focus observed for sentences than for words also supports their hypothesis that the greater the extent of activity in the left temporal-pole, the more levels of linguistic processing are engaged and/or the more memory demands are placed on the linguistic content of stimuli. Since the word condition is sufficient to activate the left anterior temporal region, our findings indicate that the left STG/temporal pole is a component of a distributed network involved with lexical processing. In contrast, the posterior STG of both sides appears to be part of a network by which both sensory and sublexical phonological stages of cortical processing are shared by both speech and nonspeech signals. The fact that both speech and nonspeech stimuli similarly activated this region supports the view by Binder et al. (2000) that listeners can still perceive speech features from nonspeech signals as unintelligible as backward speech.

Activation in Anterior Insula

The anterior insula/frontal operculum was activated on the left side for both speech and nonspeech conditions compared to silent baseline. Bilateral activation was found only for the conditions that required listening to stimulus patterns (Sentence or TR Sentence). Although the exact role of the insula in language processing remains controversial (see Flynn et al 1999 for review), its connections with the auditory cortex and inferior frontal gyrus strategically places this relay station as part of a network engaged in verbal communication. Previous neuroimaging studies have demonstrated bilateral insular activation at or near the junction with the frontal operculum in tasks that involve speech articulation and coordination (Wise et al., 1999), short-term verbal memory (Paulesu et al., 1993), control of vocal pitch including subvocal rehearsal (Zatorre et al., 1994), and phonological encoding (Paulesu et al., 1996). In the present study, the demands of the auditory task in detecting signal repetition require maintenance of the stimulus pattern in short-term working memory. Subvocal rehearsal of meaningful speech signals or pitch patterns in nonspeech would be consistent with the role implicated for the anterior insula/frontal operculum.

Activation Dissociated from Speech and Nonspeech Comparison

Previous neuroimaging studies have attempted to dissociate sites implicated in prelexical and lexico-semantic stages of cortical processing by directly comparing speech and nonspeech conditions. However, when using backward speech as a nonspeech control for forward speech or pseudowords as a control for real words (Wise et al., 1991; Hirano et al., 1997; Price et al., 1996; Binder et al., 2000), the brain activation patterns among these subtractions showed little if any differences, especially in left-hemispheric regions associated with semantic processing (e.g., prefrontal cortex, angular gyrus and ventral temporal lobe). These negative results suggest that these "speechlike" stimuli, even though they are devoid of semantic content, unavoidably accessed stages of processing up to possibly the lexical level, but produced less activation in this network overall than real words (Norris & Wise, 2000). Consequently, commonly activated foci would be subtracted out in speech versus nonspeech contrasts. The present study also did not isolate auditory/speech centers of significant activation when speech was compared to backward speech [Word minus TR Word; Sentence minus TR Sentence (Table 4)]. These findings are consistent with the proposal that backward speech, which is even less speechlike than pseudowords, is a complex signal that will attempt to engage the distributed network for spoken language as much as possible. In fact, it is noteworthy that subjects reported that these backward speech stimuli appeared to be language-like, and even resemble "bits of a foreign language". Yet this anecdotal finding is not inconsistent with earlier behavioral studies. For example, Kimura and Folb (1968) have demonstrated similar right-ear advantages for the perception of both forward and backward speech. Cutting (1974) noted that backward speech, as well as CV stimuli, contains transitions often unsuitable for perceiving speech segments, but yet are heard and processed as speech stimuli. In the present neuroimaging study, the similarities and differences found between the brain activation patterns for the speech and backward speech compared to silent baseline provide further insights into how brain circuits for speech may be exploited for processing complex nonspeech signals. Backward speech engaged not only most of the temporal-lobe network that mediates auditory and prelexical phonological stages of analysis of spoken language, but also additional stages of lexico-semantic processing associated with the left frontal lobe.

The TR Sentence compared to TR Word condition revealed activated foci in the left STG, right anterior STG, and basal prefrontal cortex (BA 11) (Table 4: foci #7, 8-10). These activated foci may be simply related to the greater complexity and higher presentation rate, and hence greater potency in activation, of a sound pattern associated with a stream (sentences or TR sentences) than with isolated stimuli (words or TR words). Whereas the activation on the left side may merely reflect a greater activation contralateral to the monaural stimulus, the activation on the right side (anterior STG) probably reflects the relatively greater pitch processing associated with the stimulus stream.

The present investigation was able to dissociate a cortical site related to processing at the sentence level. When the Sentence was compared to Word condition, a discrete site was isolated in the left STS at the junction between the midportion of the STG and the MTG (Table 4: focus #1**; Fig. 3**). Mazoyer et al. (1993) implicated a similar region on the left that included the STG and MTG for sentence-level processing. In their study with binaural stimuli, the STG activation became significantly more asymmetric (left-sided) to meaningful stories than to word lists, and the MTG on the left side was activated by stories but not by word lists. Our observations also provide further support for the hypothesis that the cortical stages of processing at the single-word level and higher involve more extensive areas in the temporal lobe outside the classically defined Wernicke's area in the temporo-parietal regions (Peterson et al., 1989; Binder et al., 1997).

Deactivation of Cortical Regions

For all of the silent baseline subtractions, decreases in cerebral blood flow were commonly found in the medial parietal/occipital lobe (precuneus/post cingulate gyrus in BA 7/23/31), cortical regions known to show deactivation in auditory and non-auditory tasks (Shulman et al., 1997; Binder et al., 1999). Shulman et al. (1997) suggested that the information-processing demands required in the active conditions were sufficient to result in suspension of ongoing processes (e.g., self-monitoring of external environment or unconstrained verbal thought processes), which are normally found in the silent baseline condition. Compared to the silent baseline condition, the TR Sentence condition noticeably produced multiple deactivated foci in this region. This finding is also consistent with the hypothesis that greater effort and increased attentional demands are required in performing these tasks in the nonspeech conditions.

Implications for Neuroimaging of CI Patients

Neuroimaging studies of speech and language processing in normal-hearing subjects have recognized that task performance can involve not only the intended auditory processing from early sensory analysis to linguistic processing, but other nonspecific cognitive task-demands that are automatically engaged, such as selective attention and working memory. Yet, no imaging study with CI subjects has considered these more general cognitive demands as they relate to outcomes in speech-perception tasks. Thus, future imaging studies of CI users that attempt to relate their speech-perception levels to the distributed neural network activated in task performance should consider the attentional and working-memory networks that are engaged along with those for speech processing. In a recent PET study of a new CI user (Miyamoto et al., 2000), speech stimuli evoked activated prefrontal foci (BA 11/47) near some of those activated by backward speech in the present study. CI users presumably encounter greater demands on attention and working memory when listening to speech as compared to normal listeners. Thus, the effortful attempt of CI users to make sense of *speech* may be modeled in part by observing normal-listeners' efforts to make sense of *backward speech*. These cognitive demands may initially be quite substantial as CI users attempt to recognize degraded signals fed through the device as speech. After about two years of device use, the prefrontal activation induced by speech extended into the right prefrontal regions where pitch processing of complex sounds has been implicated (Zatorre et al., 1992, 1994). It remains to be determined whether these frontal circuits will further develop and influence the speech-perception strategies and outcomes of CI users.

References

Baddeley, A., 1996. The fractionation of working memory. *Proc. Natl. Acad. Sci. USA 93*, 13468-13472.

Barbas, H., Mesulum, M.-M., 1985. Cortical afferent input to the principalis region of the rhesus monkey. *Neuroscience 15*, 619-37.

Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., Prieto, T., 1997. Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci. 17,* 353-362.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Rao, S.M., Cox, R.W., 1999. Conceptual processing during the conscious rest state: A functional MRI study. *J. Cogn. Neurosci. 11*, 80-93.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, T., 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cort. Mon. 10*, 12-528.

Binder, J., Rao, S., Hammeke, T., Frost, J., Bandettini, P., Hyde, J., 1994. Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Cogn. Brain Res. 2*, 31-38.

Brown, C.M. and Hagoort, P., (editors) 1999. *The Neurocognition of Language*. Oxford University Press, New York.

Buckner, R.L., Raichle, M.E., Miezin, F.M., Petersen, S.E., 1996. Functional anatomic studies of memory retrieval for auditory words and visual pictures. *J. Neurosci. 16,* 6219-6235.

Cutting, J.E., 1974. Two left-hemisphere mechanisms in speech perception. *Perception and Psychophysics, 16*, 601-612.

Dhankhar, A., Wexler, B.E., Fulbright, R.K., Halwes, T., Blamire, A.M., Shulman, R.G. 1997. Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rate. *J. Neurophysiol. 77*, 476-483.

Egan, J.P., 1948. Articulation testing methods. *Laryngoscope 58,* 955-991.

Fiez, J.A., 1997. Phonology, semantics, and the role of the left inferior prefrontal cortex. *Hum. Brain Mapping 5,* 79-83.

Flynn, F., Benson, F., Ardila, A., 1999. Anatomy of the insula – functional and clinical correlates. *Aphasiology 13,* 55-78.

Friston, K., Frith, C., Liddle, P., Dolan, R., Lamerstma, A., Frackowiak R., 1990. The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab. 10*, 458-466.

Friston, K., Frith, C., Liddle, P., Frackowiak, R., 1991. Comparing functional PET images: The assessment of significant changes. *J. Cereb. Blood Flow Metab. 11*, 81-95.

Geschwind, N., 1979. Specializations of the human brain. *Scientific American 241,* 158-168.

Griffiths, T.D., Johnsrude, I., Dean, J.L., Green, G.G.R., 1999. A common neural substrate for the analysis of pitch and duration patterns in segmented sounds. *NeuroReport 10*, 3815-3820.

Hirano, S, Naito, Y, Okazawa, H, Kojima, H, Honjo, I., Ishizu, K, Yenokura, Y, Nagahama, Y, Fukuyama, H., Konishi, J., 1997. Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Exp. Brain Res. 113*, 75-80.

Howard, D., Patterson, K., Wise, R. Brown, W.D., Friston, K., Weiller, C., Frackowiak, R., 1992. The cortical localization of the lexicon. Brain 115, 1769-1782.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1996. Brain activation modulated by sentence comprehension. *Science 274*, 114-116.

Jones, E.G., Peters, A., 1985. Cerebral cortex. *Association and auditory cortices. Vol., 4*. Plenum Press, New York.

Kimura, D., Folb, S., 1968. Neural processing of backwards-speech sounds. *Science 161*, 395-396.

Lauter, J.L., Herscovitch, P., Formby, C., Raichle, M.E., 1985. Tonotopic organization in human auditory cortex revealed by positron emission tomography. *Hearing Res. 20,* 199-205.

Mazoyer, B., Dehaene, S., Tzourio, N., Frak, V., Cohen, L., Murayama, N., Levrier, O., Salamon, G., Mehler, L., 1993. *J. Cogn. Neurosci. 5,* 467-479.

Minoshima, S., Koeppe, R., Mintum, M., Berger, K.L., Taylor, S.F., Frey, K.A., Kuhl, D.E., 1993. Automated detection of the intercommissural line for stereotaxic localization of functional brain imaging. *J. Nucl. Med. 34,* 322-329.

Miyamoto, R.T., Wong, D., Pisoni, D.B. Changes induced in brain activation in a prelingually-deaf, adult cochlear implant user: A PET study. Presented at ASHA, November, 2000.

Naito, Y., Okazawa, H., Honjo, I., Hirano, S., Takahashi, H., Shiomi, Y., Hoji, W., Kawano, M., Ishizu, K., Yonekura, Y., 1995. Cortical activation with sound stimulation in cochlear implant users demonstrated by positron emission tomography. *Cogn. Brain Res. 2,* 207-214.

Naito, Y., Okazawa, H., Hirano, S., Takahashi, H., Kawano, M., Ishizu, K., Yonekura, Y., Konishi, J., Honjo, I., 1997. Sound induced activation of auditory cortices in cochlear implant users with post- and prelingual deafness demonstrated by positron emission tomography. *Acta Oto-Laryngologica 117,* 490-496.

Norris, D., Wise, R., 2000. The study of prelexical and lexical processes in comprehension: Psycholinguistics and functional neuroimaging. In: *The new cognitive neurosciences*. M Gazzaniga (editor), 2nd edition, chapter 60, pp. 867-880. The MIT Press: Cambridge, MA.

Paulesu, E., Frith, C., Frackowiak, R., 1993. The neural correlates of the verbal component of working memory. *Nature 362*, 342-345.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R., Frith, C., 1996. Is developmental dyslexia a disconnection syndrome? *Brain 119,* 143-157.

Petersen, S.E., Fiez, J.A., 1993. The processing of single words studied with positron emission tomography. *Annu. Rev. Neurosci. 13*, 25-42.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintum, M., Raichle, M.E., 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature 331*, 585-589.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintum, M., Raichle, M.E., 1989. Positron emission tomographic studies of the processing of single words. *J. Cogn. Neurosci. 1*, 153-170.

Petrides, M., Pandya, D.N., 1984. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol. 273*, 52-66.

Petrides, M., Alivasatos, B., Meyer, E., Evans, A.C., 1993. Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proc. Natl. Acad. Sci. USA 90*, 878-882.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrielli, J.D.E., 1999 Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage 10*, 15-35.

Price, C., Indefrey, P., Turrennoul, M., 1999. The neural architecture underlying the processing of written and spoken word forms. In: *The Neurocognition of Language* (Colin M Brown and Peter Hagoort eds). Oxford University Press, New York.

Price, C., Wise, R., Ramsay, S., Friston, K., Howard, D. Patterson, K., & Frackowiak, R., 1992. Regional response differences within the human auditory cortex when listening to words. *Neurosci. Lett. 146*, 179-182.

Price, C.J., Wise, R.J.S., Warburton, E.A., Moore, C.J., Howard, D., Patterson, K., Frackowiak, R.S.J., Friston, K.J., 1996. Hearing and saying. The functional neuroanatomy of auditory word processing. *Brain 119*, 919-931.

Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness, V.S., 1992. Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. *J. Cogn. Neurosci. 4,* 352-374.

Schacter, D.L., Alpert, N.M., Savage, C.R., Rauch, S.L., Albert, M.S., 1996. Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. *Proc. Natl. Acad. Sci. USA 93,* 321-325.

Scheffler, K., Bilecen, D., Schmid, N., Tschopp, K., Seelig, J., 1998. Auditory cortical responses in hearing subjects and unilateral deaf patients as detected by functional magnetic resonance imaging. *Cereb. Cortex 8,* 156-163.

Shaywitz, B.A., Pugh, K.R., Constable, R.T., Shaywitz, S.E., Bronen, R.A., Fulbright, R.K., Shankweiler, D.P., Katz, L., Fletcher, J.M.S.E., Skudlarski, P., Gore, J.C., 1995. Localization of semantic processing using functional magnetic resonance imaging. *Hum. Brain Mapping 2,* 149-158.

Shulman, G.L., Fiez, J.A., Corbetta, M., Buckner, R.L., Miezin, F.M., Raichle, M.E., Petersen, S.E., 1997. Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cogn. Neurosci. 9*, 648-663.

Smith, E., Jonides, J., 1999. Storage and executive processes in the frontal lobes. *Science* 283, 1657-1661.

Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain*. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. Thieme Medical Publisher, New York, NY.

Vanderberghe, R., Price, C.J., Wise R, Josephs, O., Frackowiak, R.S.J., 1996. Functional anatomy of a common semantic system for words and pictures. *Nature 383*, 254-256.

Wise, R., Chollet, F., Hadar, U., Friston, K.J., Hoffner, E., Frackowiak, R.S.J., 1991. Distribution of cortical networks involved in word comprehension and word retrieval. *Brain 114*, 1803-1817.

Wise, R.J., Greene, J., Buchel, C., Scott, S.K., 1999. Brain regions involved in articulation. *Lancet, 353,* 1057-1061.

Wong, D., Miyamoto, R.T., Pisoni, D.B., Sehgal, M., Hutchins, G.D., 1999. PET imaging of cochlear-implant and normal-hearing subjects listening to speech and nonspeech. *Hearing Res. 132*, 34-42.

Zatorre, R.J., Evan, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science 256*, 846-849.

Zatorre, R.J., Evans, A.C., Meyer E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci. 14,* 1908-1919.

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Use of Partial Stimulus Information by Cochlear Implant Patients and Normal-Hearing Listeners in Identifying Spoken Words: Some Preliminary Analyses[1]

Lorin Lachs, Jonathan W. Weiss and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by NIH T32 Training Grant DC00012 to Indiana University.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Use of Partial Stimulus Information by Cochlear Implant Patients and Normal-Hearing Listeners in Identifying Spoken Words: Some Preliminary Analyses

Abstract. An error analysis of the word recognition responses of cochlear implant patients and normal-hearing listeners was conducted to determine the types of partial information used by these two populations when they identify spoken words under auditory-alone and audiovisual conditions. The results revealed that different types of partial information are used by the two groups in identifying spoken words under audio-alone or audiovisual presentation. Different types of partial information are also used in identifying words with different lexical properties. However, there were no significant interactions with hearing status, indicating that cochlear implant patients and normal-hearing listeners identify spoken words in a similar manner. The information available to patients with cochlear implants preserves much of the partial information necessary for accurate spoken word recognition.

Introduction

Cochlear implants are surgically inserted prosthetic devices that directly interface with the cochlea, providing electrical stimulation to the auditory nerve and thereby restoring hearing to those who had lost it. Although many users of the implants do well on standard outcome measures of word recognition, others receive less benefit from their devices (Pisoni, 1999; Pisoni, 2000). In order to improve cochlear implant (CI) design and use, it is necessary to determine the factors that give rise to this extensive variation. One possible source of variation lies in the sensitivity of cochlear implant users to the information necessary for accurate speech perception. In addition, there may be extensive individual differences in the process by which CI users utilize this sensory information during the process of spoken word recognition (Kirk, 2000; Pisoni, 2000; Pisoni, Cleary, Lachs, & Kirk, 2000). By comparing the performance of normal-hearing listeners and cochlear implant patients, we can examine the similarities and differences in the information each group perceives and the ways in which they use that information for spoken word recognition.

Several factors play well-established and important roles in speech perception and spoken word recognition. For example, although speech perception seems to be an inherently auditory process, a growing body of research has shown that *visual* information about speech can also be informative. In their pioneering study, Sumby and Pollack (1954) reported that the intelligibility of spoken words is enhanced when listeners are presented with both auditory and visual information, compared to auditory-only conditions. The addition of visual information can result in performance gains that are equivalent to an increase in signal-to-noise ratio of +15dB (Erber, 1969; Middleweerd & Plomp, 1987; Rosenblum & Saldaña, 1996; Sumby & Pollack, 1954). Under certain conditions, visual input may be a very important source of information about the speech signal, especially when acoustic information is degraded or unavailable. Several studies have shown that hearing impaired listeners (Erber, 1975; Massaro & Cohen, 1999; Tyler, Tye-Murray, & Lansing, 1988) and CI users (Lachs, Pisoni, & Kirk, in press; Tyler et al., 1997; Tyler, Opie, Fryauf-Bertschy, & Gantz, 1992) are also sensitive to the relationship between visual and auditory information, and make use of both sources during speech perception.

Another factor known to influence speech perception in normal-hearing listeners is talker variability. In clinical settings, cochlear implant patients frequently report better understanding of familiar voices, such as those of their spouses and other family members, than unfamiliar voices. Indeed, a listener's familiarity with the specific details of a talker's voice has been shown to improve speech

intelligibility scores (Nygaard & Pisoni, 1998). Other studies have found processing costs associated with perceiving speech when it is produced by multiple talkers, as compared to speech produced by a single talker (Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989). These findings suggest that dealing with talker variability is a resource demanding process, and as such, plays an important role in speech perception and spoken word recognition.

Finally, numerous studies have demonstrated that the lexical properties of words affect speech perception under auditory-only presentation conditions (Luce & Pisoni, 1998). Frequency of occurrence in the language, neighborhood density, and average neighborhood frequency all affect recognition performance (Luce & Pisoni, 1998). Word frequency is a measure of how often a particular word is used in language. Neighborhood density refers to the number of words that sound similar to a target word. One way this is estimated is by counting the number of words that differ from a target word by one phoneme. This measure can be used as an index of phonological similarity in local regions of the lexicon. Neighborhood frequency is the average word frequency of all the words in a given phonological neighborhood. In addition to auditory-only speech perception, there is some recent evidence showing that these lexical factors also affect visual-alone speech perception (Auer & Bernstein, 1997; Lachs, 1999) and audiovisual speech perception (Kirk, Pisoni, & Osberger, 1995; Lachs et al., in press). In order to examine the effects of these lexical factors on word recognition, "easy" and "hard" words can be selected using the theoretical framework developed in the Neighborhood Activation Model (Luce, 1986; Luce & Pisoni, 1998). Easy words have a high frequency of occurrence, low neighborhood density, and low neighborhood frequency. The combination of these variables makes them relatively "easy" to perceive quickly and accurately. In contrast, hard words have the opposite characteristics: low frequency of occurrence, high neighborhood density, and high neighborhood frequency, resulting in these words being perceived more slowly and less accurately. The easy/hard lexical distinction - the most extreme conditions formed by the orthogonal combination of word frequency, neighborhood density, and neighborhood frequency - is a useful tool for examining speech perception performance under different presentation conditions.

In a recent paper, Kaiser, Kirk, Pisoni, and Lachs (2000) examined the spoken word recognition skills of cochlear implant patients and a group of normal-hearing listeners as a function of lexical discriminability, presentation mode, and number of speakers. All stimuli were isolated English monosyllabic words, presented under three conditions: audiovisual (AV), auditory-alone (A), and visual-alone (V). In addition, the words were grouped according to their lexical discriminability into two classes, Easy and Hard. In order to equate performance levels across the hearing groups, normal-hearing listeners heard auditory-alone and audiovisual stimuli in speech spectrum noise at a signal-to-noise ratio of –5 dB SPL. During the course of the experiment, listeners were presented with several lists of stimuli and asked to repeat the word being spoken. These lists varied according to the third factor included in the design, with some lists spoken by a single talker and other lists spoken by multiple talkers.

Several intriguing discoveries were made. First, Kaiser et al. confirmed that both normal-hearing listeners and cochlear implant patients correctly perceived spoken words with the greatest accuracy under audiovisual presentation. As expected, accuracy was lower under audio-alone presentation and even worse under visual-alone presentation conditions. In addition, Kaiser et al. found decreased word recognition accuracy for multiple-talker lists relative to single-talker lists. Confirming that, like normal-hearing listeners, CI patients also incur processing costs during the presentation of multiple-talker lists. Finally, Kaiser et al. found that both normal-hearing listeners and CI patients identified lexically easy words more accurately than lexically hard words under auditory-only, visual-only and audiovisual presentation conditions.

The results reported by Kaiser et al. demonstrate that CI patients are affected by many of the same factors that affect normal-hearing listeners during speech perception. In order to investigate potential differences more closely, Kaiser et al. also performed an analysis on the errors made by normal-hearing listeners and CI patients. Every error response was analyzed in order to determine if it was a phonological neighbor of the target word. The errors for both groups of listeners showed that under audiovisual conditions, a higher proportion of errors came from the target word's phonological neighborhood than under audio-alone conditions. This pattern suggests that responses to audiovisual stimuli, even when incorrect, were more accurate than responses to audio-alone stimuli. Because there were no overall differences in the error patterns between the two groups, the error analysis indicates that CI users are making use of partial information for word identification, resulting in responses that are phonetically similar to the target word, just as normal-hearing listeners do.

In order to examine the detailed characteristics of the partial information that cochlear implant patients used in the Kaiser et al. study, we performed several additional error analyses using broad phonetic categories. Broad coding eliminates distinctions between phonemes along particular perceptual dimensions and preserves distinctions along others (Huttenlocher & Zue, 1984; Miller & Nicely, 1955; Shipman & Zue, 1982). For example, the phonemes /p/, /b/, /m/, /β/, /ɸ/, /t/, /d/, /s/, /z/ and /n/ can be grouped together into two larger categories, /p b ɸ β m/ and /t d s z n/ by eliminating distinctions between phonemes based on their manner of articulation and by preserving distinctions based on place of articulation. In this case, one broad category consists of segments with bilabial place of articulation and the other consists of segments with alveolar place of articulation, but both contain segments that are stops, fricatives or nasals. The idea here is that broad coding allows the investigator to determine which featural distinctions are perceived and which ones are missed based on partial information. To continue with the example, if response accuracy in a set of trials increased after transcription with the broad categories outlined above, then one could conclude that responses were made based on perceived place of articulation, but not on perceived manner of articulation. This technique could be extremely useful in determining the type and quality of information transmitted by the cochlear implant under a variety of conditions.

Summerfield (1987) described the "Visual: Place, Auditory: Manner" (VPAM) model of audiovisual speech perception, which is a rough approximation of the types of information that can be obtained from the auditory and visual sensory modalities. VPAM is based on the assumption that during audiovisual speech perception, the place of articulation of an utterance is obtained through visual information and the manner of articulation is obtained through auditory information. Of course, this is an extremely simplified account of the actual process of audiovisual speech perception. However, in general, such a model is consistent with experimental evidence about the perceptual confusability of phonemes under audio- and visual-alone conditions (Summerfield, 1987; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). In general, phonemes that are highly confusable under auditory-alone conditions tend to be highly distinct under visual-alone conditions, and vice-versa. Confusions made under audio-alone conditions tend to be along the place of articulation dimension. For example, /f / and /θ/ are the most confusable phonemes in auditory noise (Miller & Nicely, 1955; Summerfield, 1987). Note that both are unvoiced fricatives, but differ in their place of articulation (/f/ is labiodental, /θ/ is dental). Visually, however, these two phonemes are highly distinct. In contrast, the phonemes /b/ and /m/ are virtually indistinct under visual-alone conditions, but are highly distinct under audio-alone conditions. To a rough approximation, this pattern is observed for the relationships among all phonemes: place distinctions are more easily made with visual information, while manner distinctions are more easily made with auditory information (Summerfield, 1987).

The present broad coding analysis was carried out to examine the patterns of errors made under various presentation conditions by cochlear implant patients and normal-hearing listeners to reveal both similarities and differences in spoken language processing for the two groups of listeners. We used two different broad categorization methods to represent the use of place and manner information during speech perception. In order to examine place of articulation, we broad coded the target and response words using eight categories based on the International Phonetic Association's places of articulation, collapsing across manner distinctions. The data were also scored using the broad categories described by Shipman and Zue (1982). In this classification method, six broad categories roughly approximate the different manners of articulation in English. Each target-response pair was examined using both of these broad coding methods to examine how partial phonetic information is used in an open-set word recognition task.

Method

Participants

Details of the methodology used for data collection in Kaiser et al. are provided here for convenience. Twenty postlingually deafened adult users of cochlear implants and nineteen normal-hearing adults participated in the original study. The hearing impaired adults had profound bilateral sensorineural hearing loss and a mean age of 50 years. All patients had more than six months of experience using their cochlear implant device and were recruited from the clinical population at Indiana University School of Medicine. The normal-hearing adults had a mean age of 40 years and were recruited from staff and students at Indiana University and the associated campuses.

Materials and Equipment

Six pairs of lexically-balanced word lists were formed from the Hoosier Audiovisual Multitalker Database (HAVMD), a digital database of audiovisual recordings of eight talkers speaking isolated monosyllabic English words (Lachs & Hernández, 1998; Sheffert, Lachs, & Hernández, 1996). Stimuli in the HAVMD were digitized at 640 x 480 resolution at 30 frames per second (fps). Audio tracks in the stimuli were digitized at 22 kHz with 16-bit resolution. In order to examine the effects of lexical discriminability, each list contained thirty-six words of which half were lexically hard and half were lexically easy. Lexical discriminability was determined by examination of the lexical characteristics of the 20,000 words in *Webster's Pocket Dictionary* (Nusbaum, Pisoni, & Davis, 1984). To examine the effects of talker variability, each pair of lists contained the same words but had a different number of talkers producing the words. One list in a pair was created using only a single talker; the other list was created using six different talkers each producing six words for the list. Using data obtained from normal-hearing adult listeners under visual-only presentation (Lachs & Hernández, 1998), the lists were then equated so that visual intelligibility was balanced across the various experimental conditions.

Each subject was tested in an IAC single-walled sound-treated booth. A PowerWave 604 (Macintosh compatible) computer with a Targa 2000 video board was used to present the digitized audiovisual stimuli. Each listener was presented with six different lists. Because the present analyses focused on responses to audio-alone and audiovisual stimuli, only four of these lists were used in the present analysis. A detailed analysis of the visual-only responses can be found in Kaiser et al. Each list contained eighteen lexically easy words and eighteen lexically hard words. A single talker produced two of the lists and groups of multiple talkers produced the other two. Within each talker condition, one list was presented in the audiovisual mode and the other list was presented in the audio-only mode. For the cochlear implant patients, each speech token was presented at 70 dB SPL (C weighted). Normal-hearing listeners were tested in speech spectrum noise at a –5 dB signal-to-noise ratio. All subjects were asked to

repeat the word that was presented on each trial. The response to each speech token was recorded on-line by the experimenter, who typed the response into a file containing all the responses for a particular subject.

For the current analysis, all responses to all stimuli were compiled into one of four text files according to the Presentation format (audiovisual or audio alone) and Talker (single or multiple) conditions. The resulting four text files contained the target words and all the responses to those words. These text files were fed into a DECtalk DTC03 Text-to-Speech System, configured such that it could output an ASCII-based phonemic transcription of each target word and response (Bernstein, Demorest, & Eberhardt, 1994).

Place			"Zue"		
Name	Real	Random	Name	Real	Random
Vowels	ɯ ju ʊ ə˞ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ w	ɯ ju ʊ ə˞ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ	Vowels and syllabic consonants	ɯ ju ʊ ə˞ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ w ju l̩ m̩	ɯ ju ʊ ə˞ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ w ju l̩ m̩
Bilabial	b p m m̩	p d ʒ ŋ n	Stops	p t k b d g	p θ w l ɹ b g v
Labiodental	f v	t ʃ ɹ m̩ z	Nasals	m n ŋ	t ʃ t͡s z m n
Dental	ð θ	k θ j l b	Strong fricatives	f θ h v ð	k j ð
Velar	k h ŋ g	f s ð m	Weak fricatives	s ʃ t͡s ʒ d͡ʒ z	f s h ʒ n
Alveolar	s l l̩ ɹ t d z n n	t͡s w d͡ʒ	Glides and semi-vowels	w j l ɹ	d d͡ʒ ŋ
Postalveolar and affricates	ʃ t͡s ʒ d͡ʒ	h l̩			
Palatal	j	g v n			

Table 1. Broad categories used in the present analysis. The "Real" columns outline categories patterned in a principled manner, as described in the text. The "Random" columns outline categories to which phonemes were randomly assigned.

The phonemic transcriptions of the four text files were then recoded using two "real" ("place" and "Zue") and two "random" ("random place", and "random Zue") broad coding methods. Table 1 shows the assignment of each phoneme to a broad category for each of the four coding methods. The method used to broad code by "place" was patterned after the International Phonetic Alphabet. All speech sounds in the target words and responses were classified according to their place of articulation. Of the speech sounds used in the target words and responses, only seven places of articulation were represented: 1) bilabial, 2) labiodental, 3) dental, 4) alveolar, 5) post-alveolar, 6) palatal, and 7) glottal. All vowels were broad coded into an eighth group.

The method used to broad code by "Zue" was constructed by using the 6-way classification of phonemes proposed in Shipman and Zue (1982). This classification scheme was chosen for two reasons. First, there is empirical evidence that such a broad coding scheme can preserve much of the information in the lexicon by maintaining a large proportion of unique words. Second, the classification system

corresponds roughly to a classification system based on the manner of articulation of speech sound. The six resulting broad categories were defined as follows: 1) vowels and syllabic consonants, 2) stops, 3) nasals, 4) strong fricatives, 5) weak fricatives, and 6) glides and semi-vowels.

	Phonetic	Place		Phonetic	Zue	
		real	random		real	random
Target	b æ t	/bilabial/ V /alveolar/	/4/ V /3/	k æ t	/stop/ V /stop/	/4/ V /3/
Response	m æ d	/bilabial/ V /alveolar/	/5/ V /2/	p æ k	/stop/ V /stop/	/2/ V /4/
Correct?	NO	YES	NO	NO	YES	NO

Table 2. Scoring methods used for target-response pairs. The "Target" row contains the various transcriptions of a target word. The "Reponse" row contains the various transcriptions of a response to the target word above it. The "Correct?" row shows whether the target-response pair was graded correct or incorrect under the relevant coding method. The columns denote the different coding methods. Numbers in the "random" columns denote the set to which the phoneme was randomly assigned, as outlined in Table 1.

We expected that by loosening the criterion for a correct response, overall accuracy would increase. However, there are two possible sources for this improvement in accuracy. First, accuracy might increase because the broad categories used might more accurately reflect the information perceived by the listener. For example, it is well known that the distinction between /b/, /p/, and /m/ is very hard to make when speechreading (Summerfield, 1987). It is common in analyses of speechreading data, therefore, to group these phonemes into a larger equivalence class, or "viseme," and count as correct any responses that substitute one member of the class for another (Bernstein et al., 1994; Walden et al., 1977). Accuracy scores improve because it is relatively easy to distinguish bilabially articulated phonemes from other phonemes, while it is more difficult to make distinctions based on voicing and nasality during speechreading. The drawback to this approach, however, is that the odds of randomly picking a correct segment increase. If there are only six response categories, the chance of randomly choosing the correct feature is much greater than if there are more than 40 response categories (as there are with phonemes).

In order to control for improvements in accuracy due to chance, two "random" broad coding methods were constructed. The "random place" and "random Zue" broad coding methods contained the same number of broad categories as their "real" counterparts. However, for each random coding method, each phoneme in the target set was randomly assigned to a broad category. Using the random coding as a benchmark, we can then determine how much improvement is due to the use of partial information in the stimulus and how much is due to just a decrease in the number of response categories.

The phonemic transcriptions of each of the target-response pairs, along with the broad-coded versions of the transcription, were analyzed using a custom-designed scoring program. Each target-response pair under each broad coding method was scored as correct if the target and response were identical. Table 2 shows two representative target-response pairs under each transcription and whether the target-response pair would be considered correct.

As discussed earlier, target words were presented under eight experimental conditions, based on three factors (Lexical Disciminability, Number of Talkers, and Presentation Modality) with two levels each. An index of broad coding enhancement (Y) was determined for each subject in each of the eight conditions. Y is therefore a measure of any possible improvement due to broad coding, normalized by the amount of gain that could have occurred. Y can also be conceptualized as the proportion of error

responses that are scored as correct due to the broad transcription process. Y was calculated by the following formula:

$$Y = (b - p) / (100 - p)$$

where "b" is the percent correct performance for a given broad coding method and "p" is the percent correct performance under phonetic transcription.

Results and Discussion

The enhancement score, Y, for each subject under each experimental condition was calculated and submitted as the dependent variable to a six-way (Number of Talkers, Presentation Mode, Lexical Discriminability, Hearing Group, Broad Categorization and Coding) ANOVA. The first four factors were from the original design. Talker had two levels: single talker (ST) and multiple talker (MT). The two levels of Presentation Mode were auditory-only (A) and audiovisual (AV). Lexical Discrimination also had two levels: Easy (E) and Hard (H). Finally, Hearing Group had two levels: normal hearing (NH) and cochlear implant (CI) patients. The last two factors were relevant to the present error analysis. The first was Broad Categorization, which had two levels: "place" and "Zue". The second was Coding and had two levels: "real" and "random".

There were no interactions between the Number of Talkers factor and the other factors in the analysis. The findings below are collapsed across the two levels of this factor.

Effects of Presentation Mode

Figure 1 shows the enhancement scores (Y) for responses to words presented under audiovisual or audio-alone conditions for "real" and "random" coding methods, separated by broad categorization and collapsed across Talker, Lexical Discriminability, and Hearing Group. The left panel shows the data from "place" transcription and the right panel shows the data from "Zue" transcription. Within each panel, the left set of bars shows audiovisual presentation data and the right set of bars shows audio-alone presentation data. Within each set of bars, the dark shaded bars show data from "real" coding and the open bars show data from "random" coding. The ANOVA revealed a significant three way interaction of Coding by Broad Categorization by Presentation, $F(1,38) = 26.503$, $MSE = 0.312$, $p < 0.001$, $\eta^2 = 0.411$. This interaction indicates that the relative effectiveness of each coding scheme, as compared with it's random counterpart, was different based on the Presentation Modality. Furthermore, these three factors did not interact with Hearing Group; thus, the patterns below reflect the behavior of both normal-hearing listeners and cochlear implant patients. The three-way interaction was probed in depth by splitting the data along the Broad Categorization (Place vs. Zue) factor.

For target-response pairs transcribed by preserving place of articulation (i.e., for the data in the left-hand panel of Figure 1), we found a main effect of Coding, $F(1, 39) = 307.56$, $MSE = 0.546$, $p < 0.001$, $\eta^2 = 0.887$. Simple effects analysis revealed that relative difference scores in the "real" coding condition were greater than those in the "random" coding condition for both audiovisual, $F(1,39) = 208.133$, $p < 0.001$, $\eta^2 = 0.842$, and audio-alone, $F(1,39) = 25.335$, $p < 0.001$, $\eta^2 = 0.394$, presentations. In other words, regardless of the presentation modality, responses were correct more often under principled coding methods than they were under randomly constructed ones. This is not surprising given the fact that the principled coding methods are based on the selective collapsing and/or preservation of perceptually discriminable dimensions, whereas the random coding method follows no such constraints. It is based solely on the qualification that there be a set number of response categories.

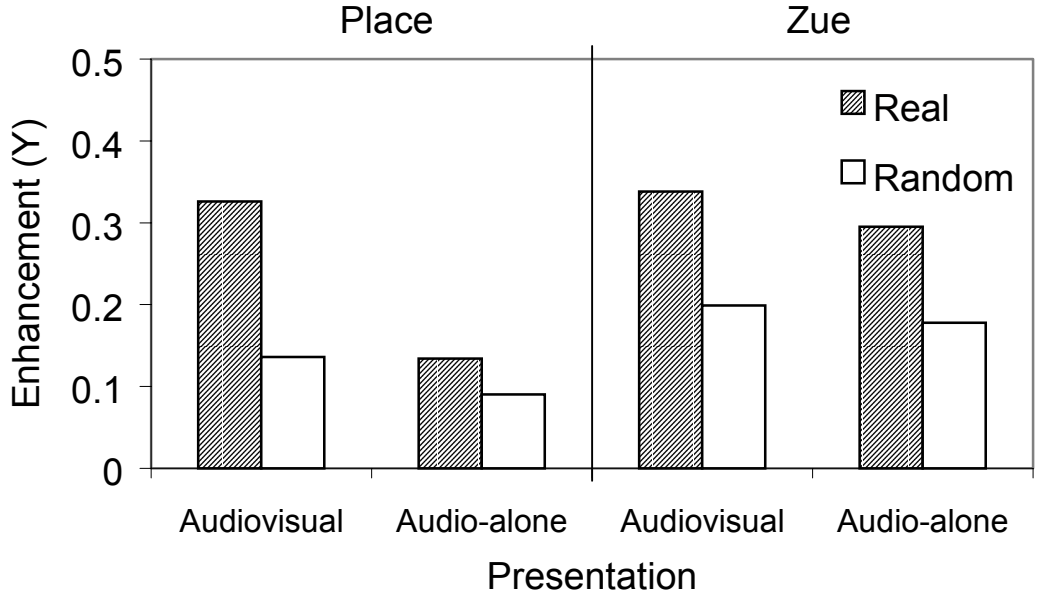


Figure 1. The relative difference scores for gains in accuracy due to broad transcription (Y) for responses to words presented under audiovisual or audio-alone conditions for "real" and "random" coding schemes separated by Broad Categorization. The data is collapsed across Talker, Lexical Discriminability, and Hearing Group.

More importantly, we also found a significant interaction between Presentation and Coding, $F(1,39) = 67.125$, $MSE = 0.215$, $p < 0.001$, $\eta^2 = 0.633$. It is clear from Figure 1 that, relative to the "random"-coding baselines, audiovisual responses benefited more from place-transcription than did audio-alone responses. Because Y represents the proportion of incorrect responses that became correct under transcription, this result indicates that in audiovisual conditions responses were closer to the target *even when they were phonetically incorrect* than they were under audio-alone conditions. In other words, audiovisual presentation elicited more accurate responses than did audio-alone presentation. Because this transcription method preserved distinctions by place of articulation, we can conclude that these "more accurate" responses to audiovisual stimuli were more accurate because they preserved the place of articulation of the segments in the target word. Even when subjects responded inaccurately in the audiovisual condition, their error responses were based closely on place of articulation: a perceptually salient dimension of visual speech. In other words, these subjects perceived partial information about the stimulus.

A slightly different picture emerges for target-response pairs transcribed with the "Zue" coding method (the right-hand panel of Figure 1). As with the "place" categorization, we observed a main effect of Coding, $F(1,39) = 116.604$, $MSE = 0.655$, $p < 0.001$, $\eta^2 = 0.749$. Simple effect analysis revealed that relative difference scores in the "real" coding condition were greater than those in the "random" coding condition for both audiovisual, $F(1,39) = 63.638$, $p < 0.001$, $\eta^2 = 0.62$, and audio-alone, $F(1,39) = 77.64$, $p < 0.001$, $\eta^2 = 0.666$, presentation. Again, this is not a surprising result, but is necessary to establish the validity of gains in accuracy due to the broad coding procedure.

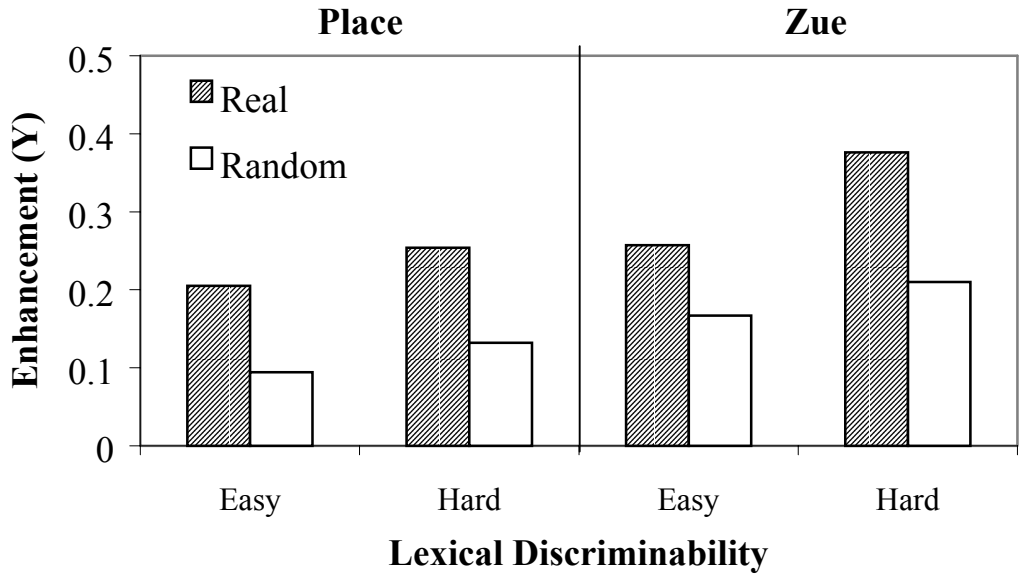


Figure 2. The relative difference scores of percent correct (Y) for responses to easy or hard words for "real" and "random" coding schemes separated by Broad Categorization. The data is collapsed across Talker, Presentation, and Hearing Group.

In contrast to the pattern observed with place transcription, the interaction of Coding by Presentation for responses transcribed with the "Zue" method was not significant, $F(1,39) = 1.171$, $MSE = 0.0047$, n.s. In general, there was no difference between the improvement in audiovisual scores and the improvement in audio-alone scores when they were classified according to the "Zue" method. This is interesting because it implies that, to a rough approximation, the responses elicited under either audiovisual or audio-alone presentation equally preserved the manner of articulation of the segments in the target word. Because the auditory stimulus was identical in the two Presentation conditions, this fact is not surprising; to a rough approximation, the manner of articulation is most perceptually salient via the acoustic modality.

Thus, we can conclude that the Place coding method accurately represented the kinds of partial information available under audiovisual presentation, as opposed to auditory-alone presentation, for both normal-hearing listeners and cochlear implant patients. The fact that these factors did not interact with Hearing Group indicates that both groups were equally sensitive to the additional place information available under audiovisual presentation.

Effects of Lexical Discriminability (Easy vs. Hard)

Figure 2 shows the enhancement scores (Y) for responses to easy or hard words for "real" and "random" coding methods separated by Broad Categorization. The data in Figure 2 is collapsed across Talker, Presentation, and Hearing Group. The left panel shows the data from "place" transcription and the right panel shows the data from "Zue" transcription. Within each panel, the left set of bars shows data from lexically easy target words and the right set of bars shows data from lexically hard target words. Within each set, the dark shaded bar shows data from "real" coding and the open bar shows data from "random" coding. The ANOVA showed a significant three way interaction of Coding by Broad

Categorization by Lexical Discriminability, $F(1,38) = 8.945$, $MSE = 0.0319$, $p < 0.01$, $\eta^2 = 0.191$. The interaction was probed by splitting the data along Broad Categorization.

For target-response pairs transcribed by place, we found a main effect of Coding, $F(1,39) = 307.562$, $MSE = 0.546$, $p < 0.001$, $\eta^2 = 0.887$. As above, the main effect of coding demonstrates that "real" scores were greater than the "random" scores, confirming that scores did not improve simply due to a reduction in the number of response categories.

We also found an effect of Lexical Discriminability, $F(1,39) = 8.472$, $MSE = 0.07426$, $p < 0.01$, $\eta^2 = 0.178$, indicating that incorrect responses (phonetically) to Hard words were more accurate than were incorrect responses to Easy words. Although this seems paradoxical at first glance, it is entirely consistent with the definition of Hard words. One of the components that defines a Hard word as Hard is that it has more phonetically similar neighbors than does an Easy word. Accuracy in this study is based on phonetic similarity. Numerically, a response has a better chance of being a neighbor of a Hard word than it does of being a neighbor of a hard word. Indeed, Kaiser et al. (2000) conducted a neighbor analysis of the present dataset and found that incorrect responses to Hard target words were more often within the neighborhood of the target word than were responses to Easy target words.

Interestingly, however, there was no interaction between Coding and Lexical Discriminability, $F(1,39) < 1$. Thus, relative to the "random" score baselines, the benefit gained by "place" transcription was not biased toward either easy or hard words. In other words, responses for Easy and Hard words did not differ in the extent to which they preserved the place of articulation of the segments in the target word.

However, for the "Zue" transcription, a significant interaction of Coding by Presentation was observed, $F(1,39) = 20.013$, $MSE = 0.058$, $p < 0.001$, $\eta^2 = 0.339$. These results indicate that relative to "random" score baselines, "Zue" broad categorization benefits hard words more than easy words. Therefore, subjects preserved the manner of articulation of the target words more often for Hard words than they did for Easy words. The interaction indicates that for Hard target words, responses were more accurate than responses to Easy words, even when they were phonetically incorrect. Because "Zue" transcription preserves distinctions by manner of articulation, we can conclude that the "more accurate" responses to Hard words were more accurate by virtue of the fact that they preserved the manner of articulation of the segments in the target word.

Again, there was no interaction between these three factors and Hearing Group, indicating that the use of partial information was consistent between groups. It is not surprising that the two hearing groups demonstrated similar effects of Lexical Discriminability: the cochlear implant patients tested were all post-lingually deafened adult speakers of American English. Presumably, for these patients, lexical structure developed normally before their hearing loss. To the extent that their learned knowledge about the similarity patterns among spoken words did not change much during the period between hearing loss and implantation, we can expect that effects of lexical structure on spoken word recognition might not diminish due to cochlear implantation.

Effects of Hearing Group (CI vs. NH)

Although there were no interactions between the factors we tested and Hearing Group, normal-hearing listeners and cochlear implant patients did differ in the overall extent to which they were sensitive to partial information. On average, the responses given by cochlear implant patients benefited more from broad categorization than those of normal-hearing listeners, $F(1, 38) = 4.1$, $MSE = 0.07$, $p = .05$, $\eta^2 = 0.098$. The mean enhancement score for cochlear implant patients was 0.25 ($SE = .02$), indicating that

broad coding raised CI scores about 25% of the amount they could have improved. In contrast, the mean enhancement score for normal-hearing listeners was 0.20, indicating that normal hearing scores improved by 20% of their potential gain due to broad categorization. Thus, when cochlear implant patients made an error in identification, they were closer to the target than when normal-hearing listeners made an error.

This result could have been observed for two reasons. First, it should be noted that adding noise to the auditory signal for normal-hearing listeners is hardly a degradation of the signal that is analogous to that experienced by cochlear implant patients. The main effect of Hearing Group observed here might simply be due to a failure on the part of Kaiser et al.'s manipulation designed to equate relative performance levels across the conditions. Alternatively, the main effect may also be due to a learned ability on the part of cochlear implant patients to make better use of partial information than their normal-hearing counterparts do, due to their experience with degraded inputs.

The fact that the Hearing Group factor did not interact with any of the other factors known to effect normal-hearing listeners' spoken word recognition, however, demonstrates that post-lingually deafened adult cochlear implant patients are sensitive to the partial information available under varying conditions, and use this information in much the same way that normal-hearing listeners do.

Discussion

In this study we examined how normal-hearing listeners and cochlear implant patients differ in their use of partial information in the perception of words. All of the listeners were presented with spoken words under a number of different conditions that manipulated presentation modality, lexical status, and the number of talkers. The target-response pairs were broad coded allowing a more detailed investigation into the perception of spoken words. By examining the structure and patterns of incorrect responses, we were able to draw several conclusions about the partial information used by perceivers with normal hearing and cochlear implants.

Effects of Presentation

We found that responses preserved place of articulation when targets were presented audiovisually more often than when targets were presented audio-alone. This result was probably obtained because information regarding place of articulation is more perceptually robust and better specified visually in audiovisual stimuli than it is in audio-alone presentation (Summerfield, 1987), especially for normal-hearing listeners listening in noise and for cochlear implant patients. Of course, as shown in Figure 1, audio-alone scores also benefited from "place" transcription. Obviously, some information regarding place is contained within the audio-alone signal perceived by the subject. However, information regarding place may either be incomplete or partially degraded during audio-alone presentation relative to the information available concerning place during audiovisual presentation, especially for cochlear implant users and normal-hearing listeners in noise.

In contrast, the availability of perceptual information regarding manner of articulation did not appear to change between audiovisual and audio-alone presentation. Responses were similar in the extent to which they were sensitive to manner under both presentation conditions. This suggests that perceptual information about manner in audio-alone presentation does not differ much from the information available in audiovisual presentation. Apparently, the addition of a visual signal does not affect (to a great extent) the information relevant to the manner of articulation contained within the auditory signal alone for either normal-hearing listeners or cochlear implant patients.

These results also indicate that cochlear implant and normal-hearing subjects do not perceive audiovisual or audio-alone perceptual information differently, demonstrating that basic processes of spoken word recognition may be common among normal-hearing subjects and users of cochlear implants.

Effects of Lexical Discriminability

Responses to both easy and hard target words preserved the place of articulation of the target to the same degree. In contrast, there *were* differences in the extent to which responses to easy and hard words preserved manner of articulation. Specifically, responses to hard target words preserved the manner of articulation of targets more than responses to easy target words. Such a result is consistent with an explanation based on the distribution of lexical distinctions across the lexicon. It seems reasonable to assume that, for any given target word, the proportion of neighbors that differ by a particular phonetic distinction would be constant. This is of course an empirically testable assumption. If it were true, that would imply that the likelihood of a target word being identified as a neighbor by a particular distinction would be calculable. The results presented above would then be consistent with a scenario in which the proportion of neighbors that differ by manner, for each word in the entire lexicon, is higher than the proportion of neighbors that differ by place, but only in dense neighborhoods! In other words, the phonological neighbors of hard target words are more similar in regards to manner than the phonological neighbors of easy target words. An in-depth study of the distribution of neighborhood distinctions could test this hypothesis.

In this study we found that cochlear implant and normal-hearing subjects use partial information similarly in identifying spoken words. This implies that the loss of hearing does not significantly change the partial information used during speech perception or the process by which spoken words are recognized. Thus, cochlear implant devices appear to allow the perception of partial information that is useful to speech perception under a variety of conditions.

References

Auer, E. T., Jr., & Bernstein, L. E. (1997). Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *Journal of the Acoustical Society of America, 102*, 3704 - 3710.

Bernstein, L. E., Demorest, M. E., & Eberhardt, S. P. (1994). A computational approach to analyzing sentential speech perception: Phoneme-to-phoneme stimulus-response alignment. *Journal of the Acoustical Society of America, 95*, 3617-3622.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423 - 424.

Erber, N.P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*, 481-492.

Huttenlocher, D. P., & Zue, V. W. (1984). *A model of lexical access from partial phonetic information.* Paper presented at the The IEEE International Conference on Acoustics, Speech and Signal Processing, San Diego, CA.

Kaiser, A. R., Kirk, K. I., Pisoni, D. B., & Lachs, L. (2000). *Audiovisual speech perception of adult cochlear implant users: Integration, talker and lexical effects*. Manuscript in preparation.

Kirk, K.I. (2000). Challenges in the clinical investigation of cochlear implant outcomes. In J. K. Niparko, K. I. Kirk, N. K. Mellon, A. M. Robbins, D. L. Tucci, & B. S. Wilson (Eds.), *Cochlear Implants: Principles and Practices* (pp. 225 - 258). Philadelphia, PA: Lippincott, Williams and Wilkins.

Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16*, 470 - 481.

Lachs, L. (1999). Use of partial stimulus information in spoken word recognition without auditory stimulation, *Research on Spoken Language Processing No. 23*. Bloomington, IN: Speech Research Laboratory, Indiana University Bloomington.

Lachs, L., & Hernández, L. R. (1998). Update: The Hoosier Audiovisual Multitalker Database, *Research on Spoken Language Processing Progress Report 22* (pp. 377 -388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (in press). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear & Hearing*.

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon, *Research on Speech Perception Technical Report No. 6*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19*, 1 - 36.

Massaro, D. W., & Cohen, M. M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Journal of Speech, Language, and Hearing Research, 42*, 21 - 41.

Middleweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold in noise. *Journal of the Acoustical Society of America, 82*, 2145 - 2147.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338 - 352.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*, 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words, *Research on Speech Perception Progress Report No. 10.* Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355 - 376.

Pisoni, D. B. (1999). *Individual differences in the effectiveness of cochlear implants.* Paper presented at the 138th Meeting of the Acoustical Society of America, Columbus, OH.

Pisoni, D. B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear & Hearing, 21*, 70 - 78.

Pisoni, D. B., Cleary, M., Lachs, L., & Kirk, K. I. (2000, November, 2000). *Individual differences in effectiveness of cochlear implants in prelingually deaf children.* Paper presented at the American Speech-Language Hearing Association Annual Convention, Washington, D. C.

Rosenblum, L.D., & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 318-331.

Sheffert, S. M., Lachs, L., & Hernández, L. R. (1996). The Hoosier Audiovisual Multitalker Database, *Research on Spoken Language Processing No. 21* (pp. 578 - 583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Shipman, D. W., & Zue, V. W. (1982). *Properties of large lexicons: Implications for advanced isolated word recognition systems.* Paper presented at the IEEE 1982 International Conference on Acoustics, Speech and Signal Processing.

Sumby, W. H., & Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212 - 215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 3 - 51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tyler, R. S., Fryauf-Bertschy, H., Gantz, B. J., Kelsay, D. M. R., Tyler, R. S., Woodworth, G. G., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology-Head and Neck Surgery, 117*, 180 -187.

Tyler, R. S., Opie, J. M., Fryauf-Bertschy, H., & Gantz, B. J. (1992). Future directions for cochlear implants. *Journal of Speech-Language Pathology and Audiology, 16*, 151 - 163.

Tyler, R. S., Tye-Murray, N., & Lansing, C. R. (1988). Electrical stimulation as an aid to speechreading. Special Issue: New reflections on speechreading. *Volta Review, 90*, 119-148.

Walden, B.E., Prosek, R.H., Montgomery, A.A., Scherr, C.K., & Jones, C.J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20*, 130 - 145.

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Talker Discrimination by Prelingually Deaf Children with Cochlear Implants: Some Preliminary Results[1]

Miranda Cleary and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIDCD Research Grant DC00111 and NIDCD Training Grant DC00012 to Indiana University. We gratefully acknowledge the kind cooperation of Dr. Ann Geers and C. Brenner in making this project possible. We also thank T. Wood and C. Dillon for their help during data collection.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Talker Discrimination by Prelingually Deaf Children with Cochlear Implants: Some Preliminary Results

Abstract. Forty-four school-age children (ages 8 and 9) each of whom had used their Nucleus-22 device for at least four years were tested to assess their ability to discriminate differences between recorded pairs of female voices uttering sentences. Children were asked to respond "same voice" or "different voice" for each trial. The correct answer was "same" for half of the trials, and "different" on the remaining trials. Two conditions were tested. In one condition, the linguistic content of the sentence was always held constant and only the voice of the talker was permitted to vary from trial to trial. In another condition, the linguistic content of the utterance was also varied so that to correctly respond "same" the child needed to recognize that two different sentences were spoken by the same talker. Data from a group of 21 normal-hearing five-year-old children with the same stimulus materials were used to establish that these tasks were well within the capabilities of children without hearing impairment (mean proportion correct on the varied sentence condition = 89%). For the CI children tested, in the "fixed sentence condition" the mean proportion correct was 67% which although significantly different from the score expected by chance of 50%, suggests that the CI children found this discrimination task rather difficult. In the "varied sentence condition," however, the mean proportion correct was only 57%. Although this value is statistically above chance, the results suggest that the CI users were essentially unable to recognize an unfamiliar talker's voice when the content of the paired sentences differed. These findings are discussed in terms of how cochlear implants transmit the speech signal, the contribution of various acoustic cues to talker identity, and known interactions between the perception of linguistic and indexical properties of spoken words. Correlations between performance on the "fixed sentence" version of the task and other processing measures are also reported for the CI group.

Introduction

A large body of research has shown that normal-hearing listeners are sensitive to properties in the acoustic speech signal that provide information about the speech producer. These properties are sometimes referred to as "indexical" properties of the signal and can convey, though imperfectly, information about the talker's gender, age, regional background, emotional state of mind, etc. (Kramer, 1963; McGehee, 1937; Ptacek & Sanders, 1966). Indexical information is usually conceptualized as contrasting with "linguistic" information about the intended pattern of phonemes/phonemic contrasts (Pisoni, 1997). Since linguistic and indexical information are both simultaneously encoded in the same physical acoustic energy, the primary question of interest to speech researchers is how the parallel extraction of these two types of information takes place, and the degree to which these processes interact with each other.

The ability to use indexical information to perceptually discriminate between the speech of different talkers is often taken for granted in communicative situations. In order for a listener to interpret what is being said in the larger context of a spoken conversation, it is usually important to know who is speaking. In situations involving discussion among a large number of people, a listener must keep track of the current speaker and register a change of speaker when it occurs. The difficulty of this task is increased when associated visual cues are unavailable such as in the case of communicating via telephone, listening to the radio, or if one has momentarily turned away from the speaker's face.

As most normal-hearing persons can attest to, even under ideal listening conditions, confusions between talkers sometimes occur. For a hearing-impaired listener using a cochlear implant, these problems are compounded. Research has shown that perceptual errors and communicative breakdowns are more likely to occur even for generally successful users of cochlear implants when they are faced with having to rapidly decode the speech of multiple talkers (Sommers, Kirk, & Pisoni, 1997). Even though generally identified as a problem, relatively little research has investigated how cochlear implants convey the acoustic properties associated with indexical information or how cochlear implant users perceive and interpret this information. Instead, understandably, the focus has been primarily on the perception of acoustic properties known to contribute towards making phonemic distinctions.

In the current study, we asked pediatric users of cochlear implants to make judgments about whether or not pairs of recorded sentences were spoken by the same talker. Two conditions were tested. In one condition, the linguistic content of the paired utterances was identical (referred to henceforth as the "fixed sentence condition"). In the other condition, the linguistic content of the two sentences always differed (the "varied sentence condition"). In this latter condition it was necessary for the listener to be able to identify two separate utterances as spoken by either the same talker or by two different talkers. Since the talkers used in this study were previously unfamiliar, we reasoned that in order to perform the varied sentence condition it would be necessary for the listeners to form a representation, or expectation, of what the speaker of the first utterance in each pair would sound like in a subsequent, linguistically different utterance. Upon hearing the second utterance, the listener would then be able to make a judgment about whether the two sentences were, in fact, spoken by the same talker or by two different talkers.

We attempted to minimize the potential difficulty of this second task condition through several methodological simplifications. In addition to similarities between talkers, key factors in determining the difficulty of a talker discrimination task are the amount of information provided per talker, and the number of different talkers among which the listener is asked to discriminate (Murray & Cort, 1971; Pollack, Pickett, & Sumby, 1954). We therefore used relatively long sentence-length stimuli and only a very small set of three female talkers.

Our manipulation of the linguistic content of the utterances was motivated in part by recent findings on the interaction in perception between the linguistic and indexical properties of speech. Research has shown that the speed of identification of linguistic information (i.e., the phonemic content of the message) is influenced by the presence of indexical variability, and that the perception of certain types of indexical information is, in turn, influenced by linguistic variability (e.g., Miller, 1978; Mullennix & Pisoni, 1990). It is this second relationship that is being explored in the present study, that is, the effect of linguistic variability on judgments about the indexical properties of speech.

A number of related studies involving both normal-hearing and hearing-impaired children have been conducted. In a series of published papers, Jerger and her colleagues investigated whether children demonstrate the same degree of interaction between linguistic and indexical processing as shown by adults. In one of the earlier studies in this series, Jerger et al. (1993) asked normal-hearing children three to six years of age to decide whether a spondee (e.g., "ice cream") was spoken in a male or a female voice. The experimenters then varied whether the judgment was made under the condition of particular words being *consistently associated* with either the male voice or the female voice, or with no predictable association present. They also included a control condition in which only a single word was used for all male/female voice judgment trials. Jerger et al. found that the presence of unpredictable variability in the association between a particular voice and particular word had a significant effect on reaction times in the task, slowing the decision speed by about 95 ms relative to the control condition. The predictable

variation condition was also slower on average than the control condition, but just barely so, on the order of about 30 ms.

In a subsequent study, Jerger, Martin, Pearson, & Dihn (1995) used a similar task with 40 school-age children diagnosed with mild to severe hearing impairments. All of the children used conventional hearing aids and 90% of the group were believed to have acquired their hearing impairment before age 2. According to Jerger et al., these children were able, when using their hearing aids, to identify the two talkers used in the study as either male or female with very high accuracy. The reaction time results from this study suggested that the hearing aid users, particularly the younger children, found it easier to ignore linguistic information while making judgments about indexical information than did a comparison group of normal-hearing children. That is to say, the hearing-impaired children's speeded judgments regarding talker gender showed less interference from unpredictable variation in the linguistic dimension than did the judgments of normal-hearing children.

Although the methodology used in the present study differs considerably from Jerger et al's, the theoretical issues involved are similar. Specifically, we are interested in assessing how well experienced pediatric users of cochlear implants are able to ignore (or generalize beyond) linguistic variability when asked to discriminate between the voices of different talkers. In addition to this theoretical issue which deals with processing strategies, we are also interested more generally in the perception of similarity between talkers, and in determining which acoustic properties cochlear implant users are best able to use to differentiate between talkers. The present study begins to address some of these larger issues.

It has long been known that even given relatively short samples of speech, listeners can readily form impressions (though not necessarily accurate impressions) of the talker's gender, age, emotional state, and linguistic background (Kreiman, 1997). An early study by Mann, Diamond and Carey (1979), reported that the ability to recognize briefly studied unfamiliar voices continues to improve throughout the school-age years and plateaus around adult levels by adolescence. Despite much research in this area, however, the perceptual factors that children and adults use to distinguish between different talkers are far from completely understood. Speech researchers have identified a number of measurable acoustic dimensions along which talkers differ, including fundamental frequency (range, average value, and irregularities in) and absolute formant frequencies, but the degree to which listeners "perceptually weigh" the importance of each dimension has not yet been resolved (see Kreiman, 1997 for discussion).

Our use of the term "indexical" in this report admittedly encompasses a great many aspects of the speech signal, not all of which necessarily relate to cross-talker variability. An important consideration in research on voice perception is that listeners may use a variety of encoding and processing strategies and/or resources in order to accomplish the task of talker discrimination (Kreiman, 1997). Thus, the acoustic cues that yield indexical information are, to a degree, functionally defined according to whether or not they can be shown to have been used by a perceiver for the particular task. In a laboratory situation, the experimenter's selection of the stimulus set largely constrains what possible strategies for discriminating between talkers can be utilized.

Because of the range of strategies listeners may use to discriminate between talkers, it is not straightforward to determine which prior research on the perceptual skills of cochlear implant users is relevant to the present task. In particular, the degree to which earlier studies of cochlear implant users' perception of "supra-segmental" properties of speech such as intonation are relevant to our results is unclear. For example, the prosodic contrasts involved in word stress and intonation are, in fact, linguistic in nature, and depend more on perception of relative f0 values (among other factors) than on absolute f0 levels such as might provide cues to the identity of the speaker. On the other hand, absolute f0 is fairly well established as a very basic perceptual dimension along which listeners tend to discriminate different

talkers, assuming such variation is present. As such, perception of fundamental frequency even as a linguistic entity may be an important factor to consider. In general, the existing literature indicates that for children with cochlear implants, the large acoustic differences in f0 that distinguish declarative versus WH-question intonation, and male from female speech are fairly easily discriminated, even before phonemic distinctions are readily made (e.g., Osberger et al., 1991). Although much of the available research on f0 perception was carried out with the early implant designs, many of these findings should be generalizable to the current generation of cochlear implants as well.

Expectations regarding the children's performance on our talker discrimination task were based in part on the children's history of CI use and the design of the device itself. The eight- and nine-year old children who participated in this study had all used a Nucleus 22 multi-channel cochlear implant for at least four years, and the majority of children at time of testing were using a coding strategy that is capable of representing fairly detailed spectral information about the speech signal. The Nucleus 22 device and its associated spectral peak coding strategy (SPEAK) use a filter bank of 20 filters with center frequencies ranging between 250 Hz to 10,000 Hz (with variable bandwidths between filters) to continually process the incoming waveform (Loizou, 1998). Depending on the distribution of the spectral information, 5 to 10 of these filters (those that best correspond to amplitude peaks in the spectrum) are selected to pass along information to the internal part of the implant. The selection of the active filters is engineered so that vowel sounds retain more spectral detail, while sounds such as fricatives use a smaller number of spectral peaks. According to Loizou (1998), approximately six maxima are used on average. The rate at which pulses are sent via the individual electrodes is dependent on the number of maxima being conveyed and parameters of the individual patient's mapping, but tends to range between 180-300 cycles per second. There is a tradeoff based on available current such that if more maxima are being conveyed, the rate of stimulation is reduced. Fewer maxima are associated with faster stimulation rates and thus better time resolution of the original acoustic information. The SPEAK strategy uses interleaved pulses, such that about every 4 ms, the selected subset of electrodes are stimulated in descending amplitude order. The amplitude of each pulse is governed by the amplitude envelope of the signal issuing from the particular bandpass filter with which it is associated. Unlike previous coding strategies which tried to provide an independent temporal cue to fundamental frequency via stimulation rate, the SPEAK coding strategy, like most other "state of the art" coding strategies, does not use stimulation rate to code f0 but instead leaves fundamental frequency to be decoded by the listener from patterns in the waveform and spectra (Jones, McDermott, Seligman, & Millar, 1995; Seligman & McDermott, 1995).

Given the design of the device and the children's history of use we judged that at least some of the pediatric CI users would be able to make the simple discriminations presented under the "fixed sentence" condition. Because the linguistic content of the sentence was held constant across all comparisons, inter-talker differences should constitute the primary source of any perceived acoustic variation between sentences. For the "varied sentence" condition it was anticipated that the task would prove more difficult, since a generalizable representation of each talker's voice is presumably necessary to accomplish the task. However, if the children were able to ignore the linguistic variability as directed, for the purpose of the task at hand, we judged that the signal provided by the implant should be sufficient to permit some children to form the necessary representations of the different voices. Data from younger normal-hearing children would help us to judge the overall difficulty of this "varied sentence" condition.

Method

Participants

Normal-hearing (NH) Preschoolers. Twenty-one normal-hearing preschoolers were tested as part of a larger project being conducted at the Indiana University Speech Research Laboratory. Thirteen

female and eight male children participated. The children ranged in age from 5;3 to 5;8, mean age = 5;6 (SD = 0;2 months). The mean PPVT receptive vocabulary standardized score for the group was 115.6 (range = 97-138, SD = 14). This average score is one standard deviation above the expected mean for this age. The results reported here were gathered from 22 consecutively recruited children, with data from one child eliminated from the final analysis due to experimenter error.

Pediatric Cochlear Implant Users. Forty-five hearing-impaired pediatric users of cochlear implants participated in this study. All children were participants in a larger study currently being conducted at the Central Institute for the Deaf (see Geers et al., 1999, for details). One child in this group only completed one of the two conditions and all data from this child were subsequently dropped from the analysis, reducing the sample size to 44. The remaining children ranged in age from 7;11 to 9;11, mean age = 8;9 years (SD = 0;6). As can been seen in Table 1, all pediatric cochlear implant users in this study had lost their hearing before age three, with the majority reported as congenitally deaf. The duration of deafness prior to implantation averaged approximately three years and every child had used his/her implant for at least four years prior to the present testing. The group included children who use auditory/oral language as their primary means of communication as well as some children who use total communication, i.e., who rely on manual signs to supplement spoken language.

N=44	Mean	Minimum	Maximum	Std Deviation
Age at Onset of Deafness, in Months	2.52	0	36	7
Duration of Deafness in Years	2.94	.58	5.17	1.11
Duration of CI Use in Years	5.60	4.09	6.87	.66
Number of Active Electrodes	18.20	8	22	2.82

Table 1. Participant characteristics for the pediatric cochlear implant users.

Although we report results below for both the NH and CI users described above, we do not wish to suggest that a direct comparison is appropriate. In actuality, the NH children whose data are reported here completed this discrimination task as part of another study completed prior to any testing of the CI users. That is to say, the NH children were not recruited as a direct comparison group. Nevertheless, we feel that reporting the NH children's performance here is useful at this time to establish that the procedure used was well within the perceptual and cognitive abilities of normally developing children three to four years younger than the CI users in this study.

Stimulus Materials

The stimuli were selected from the Indiana Multi-Talker Sentence Database (Karl & Pisoni, 1994; Bradow, Torretta, & Pisoni, 1996), a CDRom containing digital recordings of 21 talkers each uttering 100 sentences selected from the Harvard Sentence lists (Egan, 1948; IEEE, 1969). All sound files were sampled at 20 kHz with 16-bit amplitude quantization and normalized such that the average RMS values for all files were equated. For detailed description of the recording procedures see Karl & Pisoni (1994). Eight sentences were used for the practice trials and another twenty-four sentences were selected for use during the test trials. (See Appendix.) Speaking rate #02 (medium rate) from the CDRom was used for all stimuli. The sentences were selected to have roughly similar construction and were all between 1.61 and 2.16 seconds in duration (8 to 11 syllables in length). An effort was made to not select sentences containing vocabulary the children would be unfamiliar with, however, due to the nature of the available database, there remain some words that are probably unfamiliar to hearing-impaired children (e.g., "colt"

"brim" and "reef"). For related future studies we are currently making a new set of recordings of the HINT-C, the sentences of which contain more appropriate vocabulary.

For this preliminary study, tokens from two male talkers were selected for the practice trials and tokens from three female talkers were selected for the test trials. The male talkers used for the practice stimuli were talkers #01 (gravely), and #21 (deeper). (These talkers are referred to as m1 and m9 in Bradlow, Torretta, & Pisoni, 1996.) The three female talkers used for the test stimuli were talkers #06 (smooth, deeper, bit older), #07 (gravely, young, unpleasant), #23 (higher, young, sweet) (talkers f2, f3, and f10 in Bradow, Torretta, & Pisoni, 1996). The three female talkers were judged by the experimenter to differ, at least impressionistically, along the dimensions of age, and roughness of voice. Thus, although better-controlled examination of particular indexical dimensions is something we are working towards, the type of variation represented by the three talkers represented here is itself multidimensional. The recordings from talkers #06, #07, and #23 are, however, similar in that all are clearly produced by female adults with similar speaking rates, similar regional accents, and no marked emotional quality.

Among the reasons for selecting the female talkers over the male talkers as the test stimuli was the fact that the female talkers in this particular database have generally higher speech intelligibility scores than the male talkers (Bradlow, Torretta, & Pisoni, 1996). We reasoned that use of less intelligible stimuli could possibly distract listeners from the primary talker discrimination task in spite of the fact that participants were aware that the linguistic content of the test tokens was irrelevant. Of the ten available female talkers, talkers #06, #07, and #23 all had speech intelligibility scores above the mean for the group (>89.5%)(Bradlow, Torretta, & Pisoni, 1996). In addition, the three talkers were selected such that 1) their recorded tokens were quite close in overall duration for each sentence, on average, 2) there was some separation between the talkers' mean f0 values and 3) the talkers were not strongly idiosyncratic relative to the other talkers (i.e., no extremely strongly evident age or regional dialect separation among the three, unlike the two remaining highly intelligible female talkers). As reported by Bradlow, Torretta, and Pisoni (1996), mean f0 for the talkers over the full set of 100 sentences contained in the original database were as follows: #06 = ~168 Hz, #07 = ~179 Hz, #23 =~237 Hz. Our impression was that even normal-hearing persons might occasionally confuse the three selected talkers if close attention was not paid, thus limiting the possibility of ceiling effects in the simple accuracy measure.

Because a same/different discrimination task was to be used, six trials representing every possible ordered pairing of the three voices were employed for the "different voice" trials. For the six "same voice" trials, each of the three voices was paired with itself twice. This was the case in both the fixed sentence and varied sentence conditions. Within each pair, a one-second silent interval was inserted between the offset of the first sentence and the onset of the second sentence.

Procedure

Normal-hearing Children. Each of the 21 normal-hearing children passed a hearing screening at 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz at a level of 20 dB HL using a portable Maico Hearing Instruments pure tone audiometer (MA27) and TDH-39P headphones. A response at 25 dB HL was accepted for 250 Hz due to ambient room noise. Left and right ears were tested separately. Before the discrimination task was introduced, the children were tested on their understanding of the terms "same" and "different" using picture cards. All of the five-year-olds in this group easily identified a pair of pictures that were the same, and a pair that was different, indicating that they understood the concepts of same and different.

This group of children only received only one condition of the talker discrimination task, namely, the "varied sentence" condition. The discrimination trials were administered using a PC computer using a

control program written in C. After instructing the child about the basic nature of the task, four practice trials were administered via a tabletop loud speaker. All children received the same ordering of practice trials (same, different, same, different) using stimuli from two male talkers. On the first two practice trials the experimenter modeled the task by giving the correct answer after the pair of sentences was played. The child was encouraged to do the last two practice trials on his/her own and feedback was provided. During the practice period, the experimenter explained that if the child wasn't sure about the correct answer, he or she could ask for the (same pair of) sentences to be presented again and this option was demonstrated. Repetition could occur up to two additional times. This option was available for both the practice and test trials. The 12 test trials were presented via headphones (Beyerdynamic, DT100), with the examiner being unable to hear the current trial as it was played. The child was asked to verbally report whether the two talkers were the "Same" or if they were "Different" and was shown how the experimenter would circle the child's answer on a response sheet. Assignment of the 24 different sentences to the 12 test trial pairs, and the order of presentation of the test trials were pseudo-randomized by the computer.

Pediatric Cochlear Implant Users. The pediatric cochlear implant users were tested in a manner very similar to the NH children except that the discrimination task involved an additional condition. The fixed sentence condition was administered first, followed by the varied sentence condition. In the fixed sentence condition, the child heard only one sentence across all twelve trials, as spoken by the three different talkers. Each child heard one of the twenty-four sentences selected for use in the varied sentences condition, and this assignment was balanced such that each sentence used in the varied sentence condition was heard in the fixed sentence condition by approximately two children. A practice block of four trials preceded the running of the test trials. All children received the same four practice trials using a single sentence and two different male voices. In all other aspects, the administration of the practice trials was the same as described for the normal-hearing group

After completing twelve test trials in the fixed sentence condition, the child was given the revised instructions for the varied sentence condition. A practice block of four trials again preceded the running of the test trials. All children received the same four practice trials using eight different sentences and two different male voices. Twelve test trials using the three female talkers and 24 different sentences were then administered.

The pediatric cochlear implant users were tested using a Macintosh portable laptop computer using a Psyscope script written to mimic the C program used with the normal-hearing children. Stimuli were presented via a loudspeaker (Advent AV280, 10 Watts amplifier output power, THD < 1%, frequency response 70 Hz-20 kHz) at approximately 70 dB SPL as judged by a sound level meter placed near the location of the child's head. In some cases the level was adjusted upwards at the request of the child. Presentation of all stimuli was audible to the examiner. Although the practice trials were repeated for a few children in order to get the child on task, no test pairs were repeated. Although this is different from the methodology used with the NH children, the impact of this change is probably small because very few of the NH preschoolers requested any repetitions of the test trials.

Four different pseudo-random assignments of the 24 different sentences to the twelve available test pairs were generated prior to testing and nearly equal numbers of children were tested with each randomization. Presentation order of the same talker/different talker test trials was pseudo-randomized by the computer.

The procedures followed with the cochlear implant users were administered by a clinician experienced in working with hearing-impaired children. This clinician was trained in the task administration by the researcher responsible for gathering the data from the normal-hearing children.

Results and Discussion

Normal-hearing Preschoolers. The normal-hearing children had very little difficulty with the varied sentence condition on which they were tested, scoring 89% correct on average as a group. Most children scored either 12/12 or 11/12 correct. The distribution of scores obtained from the children in the NH group is shown on the far right hand panel of Figure 1. The scores of the children as a group differed significantly from chance performance of 50% ($t(20) = 15.28$, $p < .001$). The few errors that were observed primarily involved children incorrectly responding "same" for different voice pairs involving comparisons between talkers #06 and #07. Very few other errors were obtained.

Pediatric Cochlear Implant Users. The score distributions obtained from the CI users for both the "Fixed" and "Varied conditions are shown in the left and center panels of Figure 1. The mean accuracy for the group in the fixed sentence condition was 67% which is significantly above chance performance of 50% (one-sample t-test, $t(43) = 7.13$, $p < .001$. The mean accuracy for the group in the varied sentence condition was 57% correct, which, although significantly above chance (one sample t-test, $t(43) = 3.10$, $p = .003$), indicates that the pediatric CI users encountered considerable difficulty with this task. A paired-samples t-test between scores in the two conditions showed a significant drop in scores for the varied sentence task over the fixed sentence task ($t(43) = 3.66$, $p = .001$) with an average drop of about 11% (or 1.33 trials) across the two conditions. Scores in the two conditions showed a weak but significant positive correlation ($r = +.30$, $p = .049$).

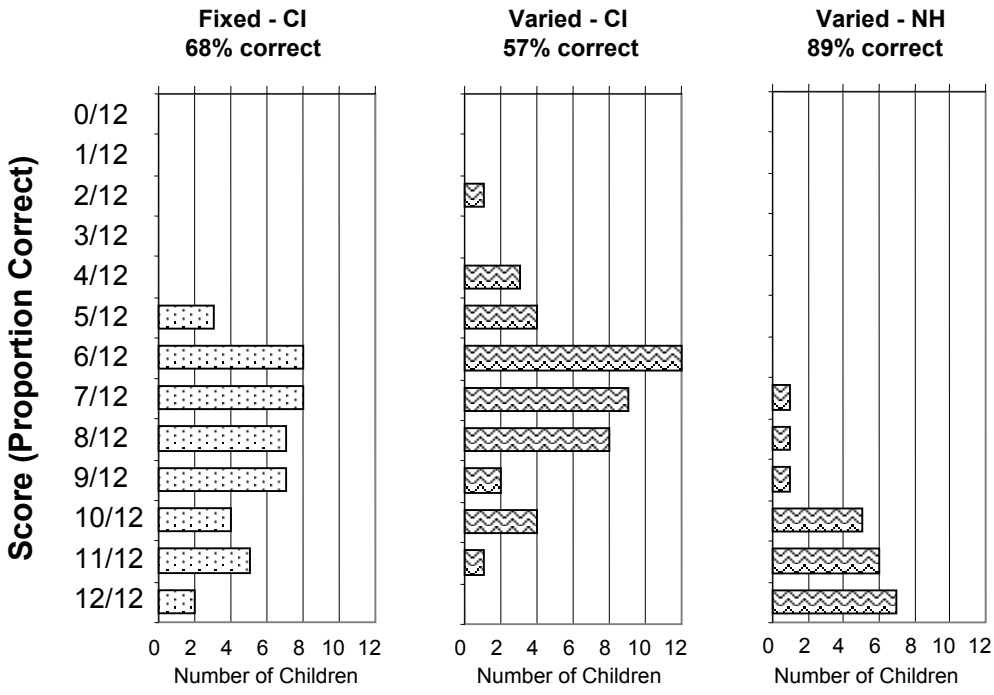


Figure 1. Distribution of scores obtained by the pediatric CI users ($N = 44$) in left and center panels, and normal-hearing five-year olds ($N = 21$), rightmost panel.

The cochlear implant users clearly had much greater difficulty with the varied sentence condition of the talker discrimination task than did the normal-hearing children on the same task. Although we did not test the normal-hearing children on the fixed sentence condition, it is likely that they would have done very well, probably better than the 89% they scored on the varied sentence condition.

Table 2 illustrates the distribution of the two possible error types in each condition for the CI users. One pattern evident in Table 2 is a bias for more often incorrectly responding "different" rather than "same" for pairs tested in the varied sentence condition. No such response bias was observed in the NH children.

	Fixed Sentence Condition		**Varied Sentence Condition**	
Proportion of Possible Errors	**Said "Same" when Different**	**Said "Different" when Same**	**Said "Same" when Different**	**Said "Different" when Same**
	# of children	# of children	# of children	# of children
6/6		1		2
5/6		1	4	8
4/6	3	5	4	9
3/6	9	10	6	9
2/6	14	13	11	8
1/6	12	5	13	6
0/6	6	9	6	2
Mean Proportion of Errors	.30	.35	.34	.52

Table 2. Distribution of errors for the children with cochlear implants. Errors made in the Fixed Sentence condition are shown on the left, errors made in the Varied Sentence condition are shown on the right.

Despite the fact that the talker discrimination scores obtained were not very continuously distributed due to the small number of trials, the variability present in the obtained scores allowed us to calculate correlations between talker discrimination scores and other measures available for these children. Because the pediatric CI users were nearly at chance in the varied sentence condition, meaningful correlations obtained with this measure are unlikely, and although shown in tables below, must be interpreted cautiously.

The results shown in Table 3 indicate that within this sample of cochlear implant users, talker discrimination performance was not significantly correlated in either condition with age at onset of deafness, duration of deafness, or duration of device use. Recall, however, that this group of CI users is relatively homogenous with respect to these factors. In particular, the children were pre-selected to demonstrate relatively little variability along these dimensions, unlike many prior studies involving samples of CI children with greater variation in these traditional predictor variables. For number of active electrodes and degree of exposure to an oral-only communication environment, there is only a slight indication that a greater number of active electrodes and more exposure to oral-only environment are

positively associated with better talker discrimination scores in both conditions. Exposure to oral-only communication was quantified using the communication mode scoring procedure described in Geers et al. (1999), which takes into account the type of communication environment experienced by the child in the year just prior to implantation, each year over the first three years of CI use, and then in the year just prior to the current testing. Table 3 also shows that the two-year spread in the chronological ages of the children was not a factor in performance.

	Proportion Correct Fixed Sentence Condition	Proportion Correct Varied Sentence Condition
Age at Onset of Deafness in Months	.19	.16
Duration of Deafness in Months	-.12	.06
Duration of CI use in Years	-.06	-.19
Number of Active Electrodes	.26	.19
Degree of Exposure to an Oral-Only Communication Environment	.27	.32*
Age in Years	-.13	.06

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 3. Correlations between talker discrimination performance and demographic variables.

	Fixed Sentence Condition	Varied Sentence Condition
BKB – Open Set Sentence Test Key Word Identification	.44**	.16
LNTE – Open Set Spoken Word Identification	.48**	.32*

* $p < 0.05$ (2-tailed), ** $p < .01$

Table 4. Correlations between talker discrimination performance and word recognition measures.

As shown in Table 4, performance on the talker discrimination task was positively correlated with two measures of spoken word recognition. These correlations were moderately large and statistically significant in the case of the fixed sentence condition. In the varied sentence condition, as might be surmised from the proximity of the group mean to chance performance, smaller correlations were observed, with only the correlation with LNTE open-set word recognition reaching statistical significance.

One question that arose in the design of this study was the question of whether the pair-wise comparison task draws on individual differences in short-term memory ability. That is, do the processing demands of the talker discrimination task make use of memory resources given that one stimulus must be kept in working memory for at least one second before the next sentence is played out?

	Fixed Sentence Condition	**Varied Sentence Condition**
WISC Digit Span Forward Points (lip-reading permitted)	.18	.11
Memory Game Span Auditory-Only	.38*	.05
Memory Game Span Auditory Plus Lights	.14	.21
Memory Game Span Lights-Only	~.00	.05

* Correlation is significant at the 0.05 level (2-tailed).

Table 5. Correlations between talker discrimination scores and memory measures obtained from the same sample in a separate study.

As shown in Table 5, performance on the fixed sentence condition of the talker discrimination task was positively correlated with memory span for stimuli presented only in the auditory modality. With the memory span measures that included visual information, however, there was little or no correlation observed with talker discrimination in the fixed sentence condition. This pattern of correlations suggests that either auditory memory plays some small role in the pair-wise discrimination task or that performance on the auditory-only memory span task may in some part reflect basic individual differences in auditory discrimination ability.

General Discussion

The talker discrimination study reported in this paper is very preliminary and we are currently in the process of expanding the scope of this research in a number of directions. Our results do, however, confirm the expectation that hearing-impaired prelingually-deafened children who have acquired language via a multi-channel cochlear implant have more difficulty discriminating between similar-sounding talkers than do normal-hearing children, particularly under conditions where the linguistic content of the message is varied.

There are several aspects of this study that suggest that the results should be interpreted somewhat cautiously. For the pediatric CI users who completed both versions of the task, the “fixed” and “varied” sentence conditions were not counterbalanced in their order of administration. Therefore it is

possible that some type of fatigue effect or order effect related to the task instructions could be responsible for the drop in the cochlear implant users' performance on the "varied sentence" condition. Given the relative brevity of the tasks, however, this is probably unlikely.

Other limitations of the present study include the fact that we do not report data from normal-hearing children using the same-sentence task. As stated previously, we assume they would do very well on this task, however we have not actually tested this. One of the reasons that we initially were hesitant to run a "fixed sentence" version of this task with normal-hearing children was that there are limitations on what we can conclude from any set of results generated using the stimuli described in our method section. The primary problem is that since the Indiana Multi-Talker Sentence Database only contains one recorded token of each talker saying each sentence, the "same talker" trials of the "fixed sentence" condition involve comparing a token with itself. It is not clear that a correct response under these conditions must necessarily reflect encoding of indexical attributes using a representation of a talker's voice. In our future studies we will be including conditions in which different recorded tokens of the same sentence are used in "fixed sentence" same-talker trials.

It may be of interest to note that we have recently tested one post-lingually deafened adult user of a Clarion 8-channel cochlear implant on the same tasks as used with the children discussed in this paper. This individual, "Mr. S", was previously identified as an extremely successful user of his cochlear implant (e.g., see Goh, Pisoni, Kirk, & Remez, 1999; Herman & Clopper, 1999). For this gentleman, in addition to the fixed sentence and varied sentence conditions, we also included an additional "fixed sentence condition" involving 24 additional trials with two male and two female talkers, such that, in 6 of the 12 trials in which the correct response was "different", the voices of the talkers differed by gender. This adult user made no errors at all in this condition, or in the original fixed sentence condition using only female talkers. He responded correctly in 9 of 12 test trials in the "varied sentence" condition, making three incorrect responses of "different" when the talkers were, in fact, the same. Thus, this cochlear implant user's performance resembled the performance of the pediatric CI users on the "varied sentence" condition. On the fixed sentence condition, however, he performed much more consistently than most of the children with cochlear implants.

The limited indexical variability present in our stimulus set required a more difficult discrimination than is typically used in the few studies that have asked hearing-impaired children to make judgments about indexical properties. This selection was intended to help us look at how small differences between talkers are perceived. Since the cochlear implants of today do, in fact, convey a fairly detailed spectral representation of the speech signal, this is not an unreasonable goal.

In future research we intend to further test whether hearing-impaired children find it easier to ignore linguistic information while making judgments about indexical information than do normal-hearing children. This research is related to a previous line of research in which we have found evidence suggesting that children with cochlear implants do not automatically use semantically redundant linguistic information which can potentially help them perform a multi-modal working memory task, despite the fact that these children can be shown to be able to identify the linguistic auditory information when played in isolation (Cleary, Pisoni, & Geers, in press). We plan to further test the hypothesis that hearing-impaired children, due to their atypical history of spoken language acquisition, are less likely than normally-developing children to try to automatically integrate semantically reinforcing/redundant or semantically conflicting information. It is this development of automaticity that makes normal-hearing children susceptible to interference in auditory Stroop tasks and the related Garner tasks as reported on by Jerger and colleagues.

In our new projects we will need to have recorded sentences that are more suitable for use with young hearing-impaired children so as to make the gathering of linguistic judgments possible alongside indexical judgments. As noted earlier, we plan to record multiple tokens from twelve female talkers of the sentences in the HINT-C Sentence Test as well as generate re-synthesized tokens of these recordings. Re-synthesized versions of these recordings with adjustment of acoustic parameters associated with perception of pitch and breathiness will permit us to have more precise experimental control over the acoustic similarities between talkers.

Once the appropriate stimulus materials are available, in addition to testing additional normal-hearing children on talker discrimination judgments under conditions of "listening in the clear," we plan to examine the perceptual judgments of normal-hearing children under simulated conditions of hearing-loss such as those that mimic the signal processing which takes place using the Nucleus 22 device with the SPEAK processing strategy (e.g., Eisenberg et al., 2000). This research should help shed further light on how pediatric cochlear implant users encode and process the indexical variability that is present in the spoken language they encounter in their everyday lives.

References

Bradlow, A.R., Torretta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20,* 255-272.

Cleary, M., Pisoni, D.B., & Geers, A. (in press). Some measures of verbal and spatial working memory in eight- and nine-year old hearing-impaired children with cochlear implants. *Ear and Hearing.*

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope, 58*, 955-991.

Eisenberg, L.S., Shannon, R.V., Martinez, A.S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America, 107,* 2704-2710.

Geers, A.E., Nicholas, J., Tye-Murray, N., Uchanski, R., Brenner, C., Crosson, J., Davidson, L.S., Spehar, B., Torretta, G., Tobey, E.A., Sedey, A., & Strube, M. (1999). Center for Childhood Deafness and Adult Aural Rehabilitation, Current research projects: Cochlear implants and education of the deaf child, second-year results. In *Central Institute for the Deaf Research Periodic Progress Report No. 35* (pp. 5-20). St. Louis, MO: Central Institute for the Deaf.

Goh, W.D., Pisoni, D.B., Kirk, K.I., & Remez, R.E. (1999). Audio-visual perception of sinewave speech in an adult cochlear implant user: A case study. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 201-210). Bloomington, IN: Indiana University.

Herman, R., & Clopper, C. (1999). Perception and production of intonational contrasts in an adult cochlear implant user. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 301-321). Bloomington, IN: Indiana University.

IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Report No. 297.*

Jerger, S., Martin, R., Pearson, D.A., & Dihn, T. (1995). Childhood hearing impairment: Auditory and linguistic interactions during multidimensional speech processing. *Journal of Speech and Hearing Research, 38,* 930-948.

Jerger, S., Pirozzolo, F., Jerger, J., Elizondo, R., Desai, S., Wright, E., & Reynosa, R. (1993). Developmental trends in the interaction between auditory and linguistic processing. *Perception and Psychophysics, 54,* 310-320.

Jones, P.A., McDermott, H.J., Seligman, P.M., & Millar, J.B. (1995). Coding of voice source information in the Nucleus cochlear implant system. *Annals of Otology, Rhinology, & Otolaryngology-Supplement, 166,* 363-365.

Karl, J.R., & Pisoni, D.B. (1994). Effects of stimulus variability on recall of spoken sentences: A first report. In *Research on Spoken Language Processing Progress Report No. 19* (pp. 145-193). Bloomington, IN: Indiana University.

Kramer, E. (1963). Judgement of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin, 60,* 408-420.

Kreiman, J. (1997). Listening to voices: Theory and practice in voice perception research. In K. Johnson & J.W. Mullennix (Eds.) *Talker Variability in Speech Processing* (pp. 85-108). San Diego: Academic Press.

Loizou, P.C. (1998). Introduction to cochlear implants. *IEEE Signal Processing Magazine,* September, 101-130.

Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27,* 153-165.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17,* 269-271.

Miller, J.L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics, 24,* 175-180.

Mullennix, J., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics, 47,* 379-390.

Murray, T., & Cort, S. (1971). Aural identification of children's voices. *Journal of Auditory Research, 11,* 260-262.

Osberger, M.J., Miyamoto, R.T., Zimmerman-Phillips, S., Kemink, J.L., Stroer, B.S., Firszt, J.B., & Novak, M.A. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant. *Ear and Hearing, 12 (Supplement),* 66S-80S.

Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.) *Talker Variability in Speech Processing* (pp. 9-32). San Diego: Academic Press.

Pollack, I., Pickett, J., & Sumby, W. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26,* 403-406.

Ptacek, P., & Sanders, E. (1966). Age recognition from voice. *Journal of Speech and Hearing Research, 9,* 273-277.

Seligman, P., & McDermott, H. (1995). Architecture of the Spectra 22 speech processor. *Annals of Otology, Rhinology, and Laryngology, Supplement, 166,* 139-141.

Sommers, M.S., Kirk, K.I., & Pisoni, D.B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing, 18,* 89-99.

Appendix

The stimuli were selected from the Indiana Multi-Talker Database, a CDRom containing recordings of 100 of the Harvard Sentences as spoken by 21 talkers.

Practice stimuli: Speaking rate #02, Male talkers #01 (gravely), #21(deeper) (~2 seconds long)

1. Glue the sheet to the dark blue background.(02)
2. Kick the ball straight and follow through.(14)
3. Help the woman get back to her feet.(15)
4. Take the winding path to reach the lake.(32)
5. Mend the coat before you go out. (35)
6. March the soldiers past the next hill. (57)
7. Place a rose bush near the porch steps. (59)
8. See the cat glaring at the scared mouse. (82)

Test Stimuli: Speaking rate #02, Female talkers #06 (older), #07 (young, unpleasant), #23 (young, sweet)

1. The juice of lemons makes fine punch. (06)
2. The box was thrown beside the parked truck. (07)
3. The boy was there when the sun rose. (11)
4. The soft cushion broke the man's fall. (18)
5. The salt breeze came across from the sea. (19)
6. The small pup gnawed a hole in the sock. (21)
7. The colt reared and threw the tall rider. (28)
8. The meal was cooked before the bell rang. (39)
9. The ship was torn apart on the sharp reef. (42)
10. The wide road shimmered in the hot sun. (44)
11. The lazy cow lay in the cool grass. (45)
12. The frosty air passed through the coat. (51)
13. The crooked maze failed to fool the mouse. (52)
14. The wagon moved on well-oiled wheels. (56)
15. The set of china hit the floor with a crash. (67)
16. The two met while playing on the sand. (72)
17. The ink stain dried on the finished page. (73)
18. The horn of the car woke the sleeping cop. (77)
19. The pearl was worn in a thin silver ring. (79)
20. The fruit peel was cut in thick slices. (80)
21. The hat brim was wide and too droopy. (84)
22. The slush lay deep along the street. (91)
23. A wisp of cloud hung in the blue air. (92)
24. A pound of sugar costs more than eggs. (93)

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Lexical Neighborhood Properties of the Original and Revised Speech Perception In Noise (SPIN) Tests[1]

Constance M. Clarke[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by a National Science Foundation Graduate Research Fellowship to the author and was conducted at the Speech Research Laboratory, Indiana University. The author would like to thank Dr. David Pisoni for the opportunity to conduct this research in the SRL and Cynthia Clopper for the original idea for this analysis. This report was originally prepared for a course at Indiana University, P747: Spoken Word Recognition, with Dr. David Pisoni.
[2] Department of Psychology, University of Arizona, Tucson, AZ 85721.

Lexical Neighborhood Properties of the Original and Revised Speech Perception In Noise (SPIN) Tests

Abstract. Recent research has shown that the frequency and density of a word's similarity neighborhood have an influence on its intelligibility over and above the influence of the word's frequency. A computational analysis was carried out to examine the lexical neighborhood properties of the original and revised versions of the Speech Perception In Noise (SPIN) Tests. While the SPIN Test was originally created to assess hearing-impairment for speech, it is also used in experimental psycholinguistic studies. Both versions of the SPIN Test contain eight sentence lists that are equivalent on the phonetic content and frequency of the final (target) word and have been found to be equivalent on intelligibility. Based on these equivalencies, any single list can be used for assessment. The purpose of the present study was to evaluate whether the eight lists of each test version are also equivalent on neighborhood characteristics of the words composing the lists. We calculated two measures of the relationship between the frequency of each target word and the frequency and density of the word's similarity neighborhood. For both test versions, the eight lists were found to be equivalent on both of these measures. The importance of these findings for both test usage and word recognition theory is discussed.

Introduction

The Speech Perception In Noise (SPIN) Test (Kalikow, Stevens, & Elliott, 1977) was originally developed as a test of hearing impairment for speech using sentence length materials. Because the test was intended to evaluate performance in relatively realistic listening conditions, the target words were placed within sentential context, and speech babble was used as noise. The sentences contained five to eight words, each ending in a common, monosyllabic word. The test contained two types of sentences: high probability (HP) and low probability (LP). In HP sentences, the final word was predictable from the semantic content of the sentence (e.g., Stir your coffee with a spoon.), whereas, in LP sentences, the final word was not predictable (e.g., We spoke about the knob.). Each final word appeared in one HP and one LP sentence context. Percent correct transcription of the final words was the measure of interest for the test. Ten lists of 50 sentences each were constructed to be statistically equivalent on mean frequency and phonetic content of the final word. Based on intelligibility tests with normal hearing listeners, two lists were eliminated. The remaining eight lists were not statistically different on intelligibility scores, and were therefore considered equivalent such that only one list need be used for testing purposes.

The SPIN Test was later evaluated by Bilger (1984; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) with hearing-impaired listeners over a range of ages, using the original ten lists created by Kalikow et al. (1977). Bilger et al. found that the lists were not equivalent on intelligibility for this population, and then revised the SPIN test based on item analyses (Elliott, 1995). The resulting test, the Revised SPIN Test (Bilger, 1984; 1994), consisted of eight equivalent lists made up of sentences from Kalikow et al.'s original ten lists. As with the original SPIN Test, the eight lists were equivalent on mean frequency and phonetic content of the final word.

Since its development, the SPIN Test has been used both for clinical evaluation of hearing impairment and for experimental psycholinguistic research (Elliott, 1995). Although the Revised SPIN Test improved the balance of lexical characteristics of the eight lists, the revised version has been more difficult to obtain than the original SPIN Test, the sentence materials for which were included in an

appendix of the original Kalikow et al. (1977) report (Elliott, 1995). As a result, both versions of the test continue to be utilized for studies of various kinds (e.g., Clarke, 2000; Clopper et al., 2000).

Recent findings in the spoken word recognition literature indicate that word frequency and phonetic content, the main lexical factors controlled across lists in both versions of the SPIN Test, may not be the only factors that affect the intelligibility of lexical items. Several studies have demonstrated that the characteristics of a word's lexical neighborhood (Landauer & Streeter, 1973), that is, the set of words that are phonetically similar to a target word, can affect performance on word perception tasks (Eukel, 1980; Luce, 1985; Luce & Pisoni, 1998; Meyer & Pisoni, 1999; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). For example, Luce (1985) found that words with higher intelligibility had a larger proportion of neighbors with a lower frequency than the word itself compared to the words with low intelligibility (cited in Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). Put another way, the highly intelligible words that Luce examined had fewer neighbors of higher frequency, while the less intelligible words had many more neighbors of higher frequency.

The idea that perception of a word is affected by the relationship of the word to other words in the lexicon has been incorporated in a recent model of spoken word recognition. The Neighborhood Activation Model (NAM; Luce & Pisoni, 1998) proposes that the process of word recognition involves, first, activation of the lexical representations of the input word as well as its acoustic-phonetic neighbors, followed by competition among the neighbors. The outcome of the competition is based not only on each item's match with the acoustic-phonetic input, but also on its frequency, the number of lexical neighbors, and the frequency of the neighbors. Therefore, the likelihood of accessing the correct lexical representation depends on the input item's frequency in relation to the density and frequency of its neighborhood. In a series of behavioral experiments using several experimental techniques with normal-hearing listeners, Luce and Pisoni (1998) demonstrated NAM's ability to account for word perception data based on these relational factors.

The purpose of the present computational study is to examine the neighborhood properties of the final words of both the original and revised SPIN Tests. While the eight lists making up these tests were balanced on frequency and phonetic content, they were not balanced on neighborhood characteristics of the test words on each list. Given the evidence that neighborhood characteristics may affect word intelligibility, it is important to know whether any of the lists are substantially different from the others on these factors. As noted above, both versions have been tested for intelligibility, and the eight lists were found to be statistically equivalent (though some measures showed nonequivalence for the lists in the original version). However, given that the measures of equivalence were based on null results of analyses of variance (ANOVAs), and that it would be advantageous to account for any remaining variability in intelligibility among the lists, an analysis of the neighborhood properties of the words on these lists is worthwhile. Whether the SPIN Test is used in a clinical or experimental setting, any factor that may undermine the assumption of equivalence among the lists could disrupt test results and interpretations. It should be noted that neighborhood characteristics would be most relevant to the LP items on each list. It is assumed that neighborhood factors would have an attenuated relationship to the intelligibility of a word in a supportive semantic context.

Method

In both versions of the SPIN test there are 400 sentences based on 200 final (target) words, with each word in one HP sentence and one LP sentence. Each of the eight lists consists of 25 HP sentences and 25 LP sentences, and each list is paired with another such that the words in HP sentences in one list are in LP sentences in the other, and vice versa for the other list. Therefore, for this study, only words

taken from the 25 LP sentences in each list were considered. This resulted in 200 words from each test version.

Lexical statistics for the 200 target words from both tests (a total of 241 unique words since 159 words are common to both tests) were obtained from a 20,000 word computerized database based on Webster's Pocket Dictionary (Luce, 1986; Nusbaum, Pisoni, & Davis, 1984; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985).[3] The database contains several pieces of information about each word, including orthography, a phonemic transcription, written frequency, familiarity (Nusbaum, Pisoni, & Davis, 1984), neighborhood density, and neighborhood frequency. The data of central interest for this study were the frequencies and lexical densities of each word. In this database, the frequency of each word is given as the sum of the Kucera & Francis (1967) written frequencies of the word and all of its homophones. The lexical density of a word is the number of English words that can be obtained by substituting, adding, or deleting one phoneme in any position ("Density B"; Greenberg & Jenkins, 1964; Luce & Pisoni, 1998). An entry in the database was located for each target word based on phonological match. If a target word was plural, the entry for its singular form was used.

The words were grouped by list, and two "second-order" statistics (Meyer & Pisoni, 1999) were calculated for each word. The purpose of these statistics was to measure the frequency of each target word in relation to the density and frequency of its neighborhood. The first statistic (Neighborhood Ratio 1) is a ratio of a target word's log frequency and the mean log frequency of its neighbors (Meyer & Pisoni, 1999):

$$\frac{T}{(\Sigma N_i)/n}$$

where T is the log frequency of the target word, N_i is the log frequency of the ith neighbor of the target word, and n is the number of neighbors. This ratio represents the target word's frequency relative to the mean frequency of its neighbors. If the ratio is greater than 1, the target word's frequency is higher than the neighborhood mean; if it is less than 1, its frequency is lower than the neighborhood mean. The second statistic (Neighborhood Ratio 2) is the ratio of a target word's log frequency and the sum of the log frequencies of the target and its neighbors (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985):

$$\frac{T}{T + \Sigma N_i}$$

This ratio represents the frequency of the target word in comparison with the total frequency of the neighborhood. This is slightly different from the first ratio in that it takes into account the number of neighbors in addition to the central tendency of their frequencies. These two ratios are meant to represent, in different ways, how much competition each target word has in the process of discriminating it from similar words during spoken word recognition (see Luce & Pisoni, 1998).

Results

Original SPIN Test

On average, the frequency of the target words in the original SPIN Test was moderate to low (mean frequency = 21.22; mean log frequency = 2.07), ranging from 1 to 269 words per million (Kucera & Francis, 1967). However, overall the words were highly familiar, with a mean familiarity score of 6.92,

[3] One word, *dove* (/d^v/) from list 7 of the original SPIN Test, could not be found in the database.

ranging from 6.08 to 7.00, on a scale of 1 to 7 (1 = word is unknown, 4 = word is recognized but meaning is unknown, 7 = word is recognized and meaning is well known; Nusbaum, Pisoni, & Davis, 1984). The mean frequencies and mean log frequencies of the eight lists are shown in Table 1. A one-way ANOVA on log frequency showed no significant differences among the lists, $F(7, 191) < 1$. This was expected a priori because the lists were constructed to be equivalent on frequency (Kalikow et al., 1977).

Original SPIN Test

Mean	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	All
Frequency	28.64	17.84	18.92	15.64	24.04	21.44	23.21	20.12	21.22
Log Frequency	2.14	2.01	2.08	2.03	2.04	2.08	2.08	2.14	2.07
Neighborhood Density	17.32	11.84	16.96	14.60	15.76	19.12	16.08	17.24	16.12
Neighborhood Ratio 1	1.10	1.11	1.08	1.07	1.04	1.10	1.09	1.10	1.09
Neighborhood Ratio 2	0.083	0.133	0.081	0.095	0.082	0.076	0.091	0.082	0.090

Table 1. Lexical statistics (means) for the eight lists and for all sentence final words of the original SPIN Test (Kalikow, Stevens, & Elliott, 1977). Neighborhood Ratio 1 = (log frequency word)/(mean log frequency neighborhood). Neighborhood Ratio 2 = (log frequency word)/(log frequency word + sum log frequency all neighbors).

The mean number of neighbors for the words in the original SPIN Test was 16.12. As can be seen in Table 1, the mean Neighborhood Ratio 1 for all eight lists is greater than 1, indicating that on average the log frequency of the target words is greater than the mean log frequency of their neighborhoods. A one-way ANOVA on Neighborhood Ratio 1 showed that there was no significant difference among the eight lists, $F(7, 191) < 1$. A final one-way ANOVA on Neighborhood Ratio 2 also showed no difference among the eight lists, $F(7, 191) = 1.36$, $p = 0.22$. A complete set of the target words of the original SPIN Test and their values on several lexical factors can be found in Appendix A.

Revised SPIN Test

The mean frequency of the target words of the revised SPIN Test was also moderate to low (mean frequency = 21.22; mean log frequency = 2.09), ranging from 1 to 269 words per million (Kucera & Francis, 1967). As in the original version, the revised test contained highly familiar words (mean = 6.93, ranging from 6.08 to 7.00, on a scale from 1 to 7). The mean frequencies and mean log frequencies of each list can be found in Table 2. Again as expected, a one-way ANOVA showed no significant differences among the eight lists on log frequency, $F(7, 192) < 1$.

The mean number of neighbors for the target words in the revised SPIN Test was 15.98. Table 2 shows that, as with the original test, the mean Neighborhood Ratio 1 for all eight lists of the revised test is greater than 1, indicating that on average the log frequency of the target words is greater than the mean log frequency of their neighborhoods. A one-way ANOVA showed that there was no significant difference among the eight lists on Neighborhood Ratio 1, $F(7, 192) < 1$. The third one-way ANOVA on Neighborhood Ratio 2 also showed no difference among the eight lists for the revised test, $F(7, 192) < 1$. A complete set of the target words of the revised SPIN Test and their values on several lexical factors can be found in Appendix B.

Revised SPIN Test									
Mean	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	All
Frequency	16.60	15.08	25.76	23.84	20.16	30.72	20.96	16.60	21.22
Log Frequency	2.09	2.00	2.17	2.13	2.05	2.19	2.07	2.02	2.09
Neighborhood Density	15.56	18.36	14.92	17.92	17.28	13.84	15.36	14.56	15.98
Neighborhood Ratio 1	1.15	1.07	1.08	1.09	1.01	1.14	1.09	1.11	1.09
Neighborhood Ratio 2	0.104	0.070	0.084	0.088	0.072	0.101	0.081	0.094	0.087

Table 2. Lexical statistics (means) for the eight lists and for all words of the revised SPIN Test (Bilger, 1984; 1994). Neighborhood Ratio 1 = (log frequency word)/(mean log frequency neighborhood). Neighborhood Ratio 2 = (log frequency word)/(log frequency word + sum log frequency all neighbors).

Discussion

The purpose of this computational analysis was to examine the lexical neighborhood characteristics of both the original and revised versions of the SPIN Test. The relationship of a test word to its phonological neighborhood is relevant for evaluation of test equivalence because it has been shown in recent research to influence word intelligibility, perhaps to a greater degree than word frequency (Luce & Pisoni, 1998). If one of the lists in the SPIN Test is composed of words that "stand out" among their respective similarity neighborhoods (that is, have a relatively high frequency compared to other words in the neighborhood), it will show systematic differences from the other lists in intelligibility scores, all other things being equal.

The aim of this analysis was to assess whether the eight lists in each test were equivalent on two measures that index the relationship between the sentence final words and their similarity neighborhoods. The lists in both the original and revised versions of the SPIN Test were found to be statistically equivalent on both measures (Neighborhood Ratio 1 and Neighborhood Ratio 2). In addition, we verified that the lists were equivalent on word frequency. These results might have been expected since the lists were equated on intelligibility for both versions (Bilger, 1984; 1994; Kalikow et al., 1977). However, it was possible that the lists still could have differed on neighborhood characteristics because these measure relational properties of words to other phonetically similar words in the lexicon. These potential differences could have been responsible for some of the remaining variability among the lists.

The findings of lexical equivalence are indeed reassuring for the claim that the relationship of a word to its lexical neighborhood is an important predictor of intelligibility (Luce & Pisoni, 1998; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). In particular, this claim implies that if the lexical neighborhood characteristics (e.g., Neighborhood Ratios 1 & 2) are not equivalent across word lists, then intelligibility will not be equivalent across word lists. If this statement is true, then a logical consequence is that if intelligibility *is* equivalent across lists, then lexical neighborhood characteristics must be equivalent across lists. Therefore, the findings here that the neighborhood ratios are equivalent, given that the intelligibilities are equivalent, are supportive of the claim.

For practical purposes, knowledge of the lexical neighborhood properties of these items is important for the use of the SPIN materials for both clinical testing and psycholinguistic research. These materials have the potential to be used with a variety of populations and under a variety of conditions. For example, while the authorized version of the SPIN Test (Bilger, 1994) comes with recorded materials and

a multi-talker babble track, researchers may choose to record their own versions of the test sentences and use other methods of stimulus degradation, such as random noise (Elliott, 1995). The findings by Kalikow et al. (1977; original) and Bilger (1984; 1994; revised) of list equivalence on intelligibility are only generalizable to testing situations with similar populations and comparable listening conditions. However, the lexical neighborhood properties measured here are based on properties of the lexical items themselves and their relationship to the rest of the lexicon. Hence, the assumption of the lists' equivalence on the neighborhood measures is valid for a wider variety of uses of the materials. A known exception is signal to noise ratio (SNR). Lexical neighborhood properties seem to lose their predictive power for word intelligibility at very low or very high SNRs (Meyer & Pisoni, 1999).

As noted above, the relevance of neighborhood property measures for the target words is greatest for the LP sentences. However, it is not clear whether the impact of neighborhood factors on intelligibility is eliminated or simply reduced in HP sentences. Investigation of this question would help to illuminate the relationship between lexical and contextual factors in spoken word recognition. As we gain more evidence for the importance of relational factors within the lexicon, future research should turn towards understanding how these relational factors act in concert with other known aspects of spoken word recognition.

References

Bilger, R.C. (1984). Manual for the clinical use of the revised SPIN Test. Champaign, IL: The University of Illinois.

Bilger, R.C. (1994). Authorized Version Revised Spin Test (Revised Speech Perception in Noise Test). Champaign: The University of Illinois Press. (Note: this test includes a compact disk, the test manual, and software for calculating the babble threshold.)

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32-48.

Clarke, C.M. (2000). Perceptual learning of foreign accented English. Unpublished masters thesis, Tucson, AZ: The University of Arizona.

Clopper, C.G., Carter, A.K., Dillon, C.M., Harnsberger, J.D., Herman, R., Clarke, C.M., Pisoni, D.B., & Hernandez, L.R. (2000). A Multi-talker Multi-dialect Corpus of Spoken American English: An Initial Report on Development. *Research on Spoken Language Processing Progress Report No. 24*. Bloomington, IN: Speech Research Laboratory.

Elliot, L.L. (1995). Verbal auditory closure and the Speech Perception In Noise (SPIN) Test. *Journal of Speech and Hearing Research, 38,* 1363-1376.

Eukel, B. (1980). A phonotactic basis for word frequency effects: Implications for automatic speech recognition. *Journal of the Acoustical Society of America, 68*, S33.

Greenberg, J.H., & Jenkins, J.J. (1964). Studies in the psychological correlates of the sound system of American English. *Word, 20*, 157-177.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Kucera, F. & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Landauer, T.D., & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.

Luce. P.A. (1985). Structural distinctions between high and low frequency words in auditory word recognition. Unpublished doctoral dissertation, Indiana University.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing, 19*, 1-36.

Meyer, T.A. & Pisoni, D.B. (1999). Some computational analyses of the PBK Test: Effects of frequency and lexical density on spoken word recognition. *Ear & Hearing, 20*, 363-371.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*, 357-376.

Pisoni, D.B., Nusbaum, H.C., Luce, P.A., & Slowiaczek, L.M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*, 75-95.

Appendix A: Original SPIN Test (LP items)

List	Final word	Transcription	Frequency	Log Freq[a]	N[b] Density[c]	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
1	lap	l@p	19	2.28	30	11.63	1.71	1.33	0.042
1	cake	kek	13	2.11	26	104.27	2.19	0.97	0.036
1	track	tr@k	38	2.58	10	14.90	1.90	1.36	0.120
1	pad	p@d	8	1.90	26	225.12	1.96	0.97	0.036
1	crates	kret	2	1.30	16	61.25	1.65	0.79	0.047
1	herd	hRd	269	3.43	20	461.95	2.32	1.48	0.069
1	mate	met	21	2.32	28	197.61	2.43	0.96	0.033
1	gin	JIn	23	2.36	20	1215.80	2.31	1.02	0.049
1	sand	s@nd	28	2.45	13	2310.46	2.63	0.93	0.067
1	dive	dYv	23	2.36	17	30.71	1.68	1.41	0.077
1	map	m@p	13	2.11	20	75.75	1.88	1.12	0.053
1	van	v@n	32	2.51	12	728.25	2.81	0.89	0.069
1	hive	hYv	2	1.30	15	322.47	2.01	0.65	0.041
1	bomb	bam	36	2.56	13	17.77	1.84	1.39	0.096
1	strips	strIp	30	2.48	6	18.00	1.68	1.47	0.197
1	yell	yEl	9	1.95	19	115.16	2.25	0.87	0.044
1	hug	h^g	3	1.48	21	8.38	1.60	0.92	0.042
1	knife	nYf	76	2.88	8	191.38	2.45	1.17	0.128
1	wax	w@ks	14	2.15	4	51.75	1.89	1.14	0.221
1	lock	lak	23	2.36	31	75.61	1.93	1.22	0.038
1	doll	dal	10	2.00	16	20.31	1.87	1.07	0.063
1	bruise	bruz	3	1.48	10	6.50	1.60	0.92	0.084
1	pine	pYn	14	2.15	30	35.03	1.91	1.13	0.036
1	dent	dEnt	2	1.30	19	56.37	1.95	0.67	0.034
1	crib	krIb	5	1.70	3	1.00	1.00	1.70	0.362
		Mean:	**28.64**	**2.14**	**17.32**	**254.30**	**1.98**	**1.10**	**0.083**
		SD:	*52.64*	*0.52*	*8.15*	*509.19*	*0.38*	*0.27*	*0.076*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
2	growl	grWl	4	1.60	6	3.17	1.28	1.25	0.173
2	sheets	Sit	45	2.65	23	171.39	2.34	1.13	0.047
2	steam	stim	17	2.23	8	62.25	2.47	0.90	0.101
2	net	nEt	34	2.53	26	290.96	2.41	1.05	0.039
2	draft	dr@ft	24	2.38	7	7.14	1.50	1.58	0.185
2	screen	skrin	48	2.68	2	7.00	1.56	1.72	0.463
2	strap	str@p	2	1.30	6	10.67	1.71	0.76	0.112
2	coast	kost	61	2.79	12	158.00	2.67	1.04	0.080
2	swamps	swamp	5	1.70	1	2.00	1.30	1.31	0.566
2	crop	krap	20	2.30	9	10.67	1.53	1.50	0.143
2	bloom	blum	12	2.08	11	22.36	1.77	1.17	0.096
2	cap	k@p	27	2.43	30	81.13	1.91	1.28	0.041
2	fleet	flit	17	2.23	16	27.25	1.55	1.44	0.083
2	mugs	m^g	1	1.00	21	49.24	1.49	0.67	0.031
2	dart	dart	1	1.00	10	113.10	2.43	0.41	0.039
2	wheat	hwit	9	1.95	9	280.78	2.29	0.86	0.087
2	booth	buT	7	1.85	12	79.92	2.04	0.91	0.070
2	scab	sk@b	1	1.00	4	7.25	1.80	0.55	0.122
2	slave	slev	30	2.48	9	10.00	1.46	1.70	0.159
2	hay	he	19	2.28	26	920.85	2.86	0.80	0.030
2	ant	@nt	28	2.45	12	3204.83	2.59	0.94	0.073
2	stamp	st@mp	8	1.90	4	1.25	1.08	1.77	0.307
2	sport	sport	17	2.23	7	53.00	1.90	1.18	0.144
2	geese	gis	3	1.48	8	65.88	2.36	0.63	0.073
2	slot	slat	6	1.78	17	18.29	1.64	1.09	0.060
		Mean:	**17.84**	**2.01**	**11.84**	**226.33**	**1.92**	**1.11**	**0.133**
		SD:	*16.20*	*0.54*	*7.89*	*648.63*	*0.49*	*0.37*	*0.131*

[a] Log Frequency = $\log_{10}$(frequency) + 1
[b] N = neighborhood
[c] based on one-phoneme substitution/addition/deletion

Appendix A: Original (con't)

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
3	cot	kat	1	1.00	35	180.37	2.04	0.49	0.014
3	fee	fi	16	2.20	31	815.13	2.64	0.84	0.026
3	blame	blem	34	2.53	7	21.43	1.94	1.31	0.157
3	jar	Jar	16	2.20	16	345.13	2.04	1.08	0.063
3	gang	g@G	22	2.34	15	15.60	1.65	1.42	0.086
3	sleeves	sliv	11	2.04	7	43.57	1.84	1.11	0.137
3	foam	fom	37	2.57	16	689.38	2.24	1.15	0.067
3	breath	brET	53	2.72	3	17.33	1.98	1.37	0.314
3	barn	barn	29	2.46	11	24.00	1.87	1.31	0.107
3	scare	skEr	6	1.78	11	41.27	2.16	0.82	0.070
3	limb	lIm	5	1.70	20	149.50	1.96	0.87	0.042
3	rope	rop	15	2.18	27	37.48	1.89	1.15	0.041
3	spoon	spun	6	1.78	11	22.82	1.57	1.13	0.093
3	hips	hIp	10	2.00	28	397.68	2.13	0.94	0.032
3	tack	t@k	4	1.60	37	78.49	1.96	0.82	0.022
3	mast	m@st	6	1.78	14	253.07	2.76	0.64	0.044
3	juice	Jus	11	2.04	13	17.31	1.68	1.22	0.086
3	fist	fIst	26	2.42	11	160.27	2.24	1.08	0.089
3	coach	koC	24	2.38	14	26.14	2.00	1.19	0.078
3	crown	krWn	19	2.28	10	25.20	1.68	1.36	0.120
3	pile	pYl	25	2.40	28	21.54	1.90	1.26	0.043
3	swan	swan	3	1.48	11	1.45	1.13	1.31	0.107
3	coin	kOn	10	2.00	14	136.57	1.86	1.07	0.071
3	bar	bar	82	2.91	24	229.46	2.11	1.38	0.054
3	broom	brum	2	1.30	10	43.60	1.87	0.70	0.065
		Mean:	**18.92**	**2.08**	**16.96**	**151.75**	**1.97**	**1.08**	**0.081**
		SD:	*18.27*	*0.45*	*9.23*	*211.58*	*0.32*	*0.25*	*0.060*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
4	thorns	Torn	3	1.48	9	27.89	1.96	0.75	0.077
4	raft	r@ft	4	1.60	14	5.57	1.38	1.16	0.077
4	drain	dren	18	2.26	13	24.77	1.85	1.22	0.086
4	kick	kIk	16	2.20	28	21.11	1.80	1.23	0.042
4	vest	vEst	4	1.60	19	58.16	2.12	0.75	0.038
4	robe	rob	6	1.78	18	41.89	1.94	0.92	0.048
4	hint	hInt	9	1.95	10	15.70	1.66	1.18	0.105
4	bowl	bol	23	2.36	33	59.18	2.06	1.15	0.034
4	blast	bl@st	15	2.18	3	231.00	2.34	0.93	0.236
4	grease	gris	9	1.95	14	26.50	1.91	1.02	0.068
4	ditch	dIC	10	2.00	16	81.88	2.12	0.94	0.056
4	drum	dr^m	11	2.04	10	448.00	1.89	1.08	0.098
4	crash	kr@S	20	2.30	13	6.00	1.37	1.68	0.115
4	deck	dEk	23	2.36	20	40.20	2.00	1.18	0.056
4	mist	mIst	14	2.15	16	169.25	2.27	0.95	0.056
4	crew	kru	36	2.56	19	79.79	1.97	1.30	0.064
4	brook	brUk	3	1.48	6	62.83	2.30	0.64	0.097
4	goal	gol	60	2.78	28	57.64	1.84	1.51	0.051
4	mouse	mWs	10	2.00	14	79.43	1.93	1.04	0.069
4	cruise	kruz	2	1.30	8	9.63	1.66	0.78	0.089
4	grin	grIn	13	2.11	9	26.56	2.03	1.04	0.104
4	ape	ep	3	1.48	17	156.53	2.21	0.67	0.038
4	sponge	sp^nJ	7	1.85	1	16.00	2.20	0.84	0.456
4	truck	tr^k	57	2.76	9	14.11	1.67	1.65	0.155
4	fur	fR	15	2.18	18	203.28	1.78	1.22	0.064
		Mean:	**15.64**	**2.03**	**14.60**	**78.52**	**1.93**	**1.07**	**0.095**
		SD:	*15.16*	*0.39*	*7.58*	*99.04*	*0.26*	*0.28*	*0.087*

Appendix A: Original (con't)

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
5	junk	J^Gk	8	1.90	9	5.78	1.53	1.25	0.122
5	meal	mil	30	2.48	28	77.21	1.92	1.29	0.044
5	prize	prYz	28	2.45	10	41.00	2.36	1.04	0.094
5	mold	mold	45	2.65	18	88.94	2.10	1.27	0.066
5	scream	skrim	13	2.11	6	25.67	2.03	1.04	0.148
5	joints	JOnt	39	2.59	5	92.60	1.88	1.38	0.216
5	fudge	f^J	1	1.00	6	22.17	1.90	0.53	0.081
5	hedge	hEJ	2	1.30	11	71.36	2.09	0.62	0.054
5	plot	plat	37	2.57	12	17.00	1.65	1.56	0.115
5	rent	rEnt	21	2.32	17	69.00	1.91	1.22	0.067
5	bow	bo	17	2.23	32	615.56	2.54	0.88	0.027
5	firm	fRm	109	3.04	13	13.69	1.67	1.82	0.123
5	lid	lId	19	2.28	23	78.91	2.06	1.11	0.046
5	cramp	kr@mp	2	1.30	6	14.17	1.44	0.90	0.131
5	row	ro	36	2.56	38	196.08	2.25	1.14	0.029
5	spool	spul	1	1.00	9	71.33	1.91	0.52	0.055
5	den	dEn	2	1.30	33	130.64	2.18	0.60	0.018
5	bread	brEd	42	2.62	15	37.60	2.09	1.25	0.077
5	brat	br@t	1	1.00	11	35.91	1.79	0.56	0.048
5	sling	slIG	1	1.00	16	12.13	1.84	0.54	0.033
5	trap	tr@p	20	2.30	13	13.85	1.64	1.40	0.097
5	throat	Trot	51	2.71	6	52.00	2.19	1.23	0.171
5	tea	ti	65	2.81	33	1607.21	2.76	1.02	0.030
5	thief	Tif	8	1.90	8	29.25	1.99	0.96	0.107
5	mop	map	3	1.48	16	22.56	1.80	0.82	0.049
		Mean:	**24.04**	**2.04**	**15.76**	**137.66**	**1.98**	**1.04**	**0.082**
		SD:	*25.60*	*0.66*	*9.83*	*329.02*	*0.30*	*0.35*	*0.050*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
6	shed	SEd	11	2.04	17	249.41	2.55	0.80	0.045
6	roar	ror	13	2.11	31	617.58	2.58	0.82	0.026
6	curb	kRb	13	2.11	12	10.00	1.64	1.29	0.097
6	peg	pEg	4	1.60	10	17.00	1.86	0.86	0.079
6	chat	C@t	5	1.70	22	743.36	1.98	0.86	0.037
6	bet	bEt	20	2.30	32	251.59	2.55	0.90	0.027
6	loot	lut	4	1.60	26	53.19	2.04	0.79	0.029
6	wits	wIt	20	2.30	31	618.23	2.18	1.05	0.033
6	rib	rIb	1	1.00	19	14.58	1.77	0.56	0.029
6	slice	slYs	13	2.11	8	11.13	1.62	1.31	0.140
6	clock	klak	20	2.30	15	11.13	1.59	1.45	0.088
6	cheers	CIr	8	1.90	27	87.22	2.13	0.89	0.032
6	film	fIlm	96	2.98	7	9.29	1.42	2.10	0.231
6	gum	g^m	14	2.15	16	161.31	1.93	1.11	0.065
6	trail	trel	31	2.49	13	26.62	1.90	1.31	0.092
6	drug	dr^g	24	2.38	8	7.50	1.60	1.48	0.156
6	dust	d^st	70	2.85	8	239.88	2.22	1.28	0.138
6	fun	f^n	44	2.64	25	190.76	2.06	1.28	0.049
6	lanes	len	34	2.53	33	48.18	2.05	1.23	0.036
6	knob	nab	2	1.30	21	236.67	1.62	0.80	0.037
6	sap	s@p	1	1.00	27	14.96	1.76	0.57	0.021
6	cliff	klIf	11	2.04	5	46.80	1.84	1.11	0.182
6	rim	rIm	5	1.70	26	129.31	1.96	0.87	0.032
6	tin	tIn	12	2.08	30	830.50	2.15	0.96	0.031
6	task	t@sk	60	2.78	9	17.67	1.53	1.81	0.168
		Mean:	**21.44**	**2.08**	**19.12**	**185.75**	**1.94**	**1.10**	**0.076**
		SD:	*23.54*	*0.52*	*9.23*	*248.22*	*0.32*	*0.36*	*0.060*

Appendix A: Original (con't)

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
7	dove	-	-	-	-	-	-	-	-
7	lungs	l^G	16	2.20	18	88.00	2.04	1.08	0.057
7	chunks	C^Gk	2	1.30	10	7.20	1.65	0.79	0.073
7	seeds	sid	41	2.61	27	184.19	2.31	1.13	0.040
7	pole	pol	27	2.43	33	37.36	1.97	1.23	0.036
7	gown	gWn	16	2.20	8	187.25	2.48	0.89	0.100
7	tide	tYd	45	2.65	22	124.18	2.13	1.24	0.053
7	debt	dEt	13	2.11	28	109.29	2.36	0.90	0.031
7	vault	vclt	2	1.30	6	13.50	1.67	0.78	0.115
7	oath	oT	6	1.78	12	495.83	2.25	0.79	0.062
7	flock	flak	10	2.00	12	10.83	1.43	1.40	0.104
7	wheels	hwil	56	2.75	7	100.86	1.76	1.56	0.183
7	clerk	klRk	34	2.53	8	4.13	1.36	1.87	0.189
7	beads	bid	1	1.00	26	299.19	2.22	0.45	0.017
7	splash	spl@S	3	1.48	2	2.00	1.24	1.19	0.374
7	aid	ed	139	3.14	20	102.35	2.20	1.43	0.067
7	feast	fist	3	1.48	10	230.30	2.61	0.57	0.053
7	bark	bark	14	2.15	15	37.13	2.04	1.05	0.065
7	crumbs	kr^m	3	1.48	11	461.64	2.09	0.71	0.060
7	bay	be	63	2.80	35	570.31	2.36	1.19	0.033
7	calf	k@f	11	2.04	19	118.58	1.86	1.10	0.055
7	glue	glu	8	1.90	11	28.18	1.93	0.99	0.082
7	blade	bled	13	2.11	9	27.44	1.79	1.18	0.116
7	cops	kap	15	2.18	30	35.77	1.84	1.18	0.038
7	spray	spre	16	2.20	7	5.57	1.42	1.55	0.181
		Mean:	**23.21**	**2.08**	**16.08**	**136.71**	**1.96**	**1.09**	**0.091**
		SD:	*30.20*	*0.54*	*9.48*	*164.58*	*0.37*	*0.33*	*0.077*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
8	beak	bik	1	1.00	28	298.21	2.06	0.49	0.017
8	bench	bEnC	35	2.54	7	11.43	1.63	1.56	0.182
8	flood	fl^d	19	2.28	5	35.00	2.14	1.06	0.175
8	pie	pY	17	2.23	32	423.63	2.32	0.96	0.029
8	clue	klu	15	2.18	12	29.08	1.82	1.20	0.091
8	hen	hEn	22	2.34	24	229.54	2.22	1.06	0.042
8	tent	tEnt	20	2.30	19	70.79	2.08	1.11	0.055
8	tub	t^b	13	2.11	17	14.00	1.67	1.26	0.069
8	flame	flem	17	2.23	11	21.27	1.68	1.33	0.108
8	pet	pEt	8	1.90	30	96.63	2.18	0.87	0.028
8	ox	aks	5	1.70	6	15.33	1.62	1.05	0.149
8	toll	tol	16	2.20	33	52.73	1.98	1.11	0.033
8	frogs	frcg	1	1.00	4	8.75	1.58	0.63	0.137
8	mat	m@t	8	1.90	30	636.03	2.37	0.80	0.026
8	skirt	skRt	21	2.32	10	5.00	1.30	1.78	0.151
8	logs	lcg	11	2.04	13	107.08	2.28	0.90	0.065
8	cards	kard	26	2.42	15	40.40	1.85	1.31	0.080
8	sheep	Sip	23	2.36	20	180.15	2.18	1.08	0.051
8	beam	bim	21	2.32	16	444.31	2.19	1.06	0.062
8	silk	sIlk	12	2.08	10	14.70	1.65	1.26	0.112
8	host	host	36	2.56	10	136.40	2.20	1.16	0.104
8	pill	pIl	15	2.18	36	88.97	2.12	1.03	0.028
8	notch	naC	6	1.78	7	664.43	1.96	0.91	0.115
8	pool	pul	111	3.05	18	25.28	1.99	1.53	0.078
8	bend	bEnd	24	2.38	18	52.06	2.16	1.10	0.058
		Mean:	**20.12**	**2.14**	**17.24**	**148.05**	**1.97**	**1.10**	**0.082**
		SD:	*20.95*	*0.44*	*9.63*	*194.99*	*0.28*	*0.28*	*0.049*

Appendix B: Revised SPIN Test (LP items)

List	Final word	Transcription	Frequency	Log Freq[a]	N[b] Density[c]	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
1	crib	krIb	5	1.70	3	1.00	1.00	1.70	0.362
1	growl	grWl	4	1.60	6	3.17	1.28	1.25	0.173
1	hut	h^t	13	2.11	27	196.37	2.16	0.98	0.035
1	knob	nab	2	1.30	21	236.67	1.62	0.80	0.037
1	rag	r@g	10	2.00	28	11.25	1.53	1.30	0.045
1	feast	fist	3	1.48	10	230.30	2.61	0.57	0.053
1	splash	spl@S	3	1.48	2	2.00	1.24	1.19	0.374
1	pond	pand	25	2.40	7	13.29	1.67	1.43	0.170
1	hips	hIp	10	2.00	28	397.68	2.13	0.94	0.032
1	lungs	l^G	16	2.20	18	88.00	2.04	1.08	0.057
1	foam	fom	37	2.57	16	689.38	2.24	1.15	0.067
1	drain	dren	18	2.26	13	24.77	1.85	1.22	0.086
1	mist	mIst	14	2.15	16	169.25	2.27	0.95	0.056
1	sleeves	sliv	11	2.04	7	43.57	1.84	1.11	0.137
1	skirt	skRt	21	2.32	10	5.00	1.30	1.78	0.151
1	host	host	36	2.56	10	136.40	2.20	1.16	0.104
1	crew	kru	36	2.56	19	79.79	1.97	1.30	0.064
1	toll	tol	16	2.20	33	52.73	1.98	1.11	0.033
1	cliff	klIf	11	2.04	5	46.80	1.84	1.11	0.182
1	crook	krUk	3	1.48	8	11.25	1.58	0.93	0.105
1	crack	kr@k	21	2.32	18	6.06	1.39	1.67	0.085
1	pile	pYl	25	2.40	28	21.54	1.90	1.26	0.043
1	van	v@n	32	2.51	12	728.25	2.81	0.89	0.069
1	bend	bEnd	24	2.38	18	52.06	2.16	1.10	0.058
1	hay	he	19	2.28	26	920.85	2.86	0.80	0.030
		Mean:	**16.60**	**2.09**	**15.56**	**166.70**	**1.90**	**1.15**	**0.104**
		SD:	*10.95*	*0.38*	*8.96*	*253.08*	*0.48*	*0.29*	*0.092*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
2	risk	rIsk	54	2.73	9	7.67	1.60	1.71	0.159
2	spoon	spun	6	1.78	11	22.82	1.57	1.13	0.093
2	ox	aks	5	1.70	6	15.33	1.62	1.05	0.149
2	steam	stim	17	2.23	8	62.25	2.47	0.90	0.101
2	coin	kOn	10	2.00	14	136.57	1.86	1.07	0.071
2	drug	dr^g	24	2.38	8	7.50	1.60	1.48	0.156
2	lap	l@p	19	2.28	30	11.63	1.71	1.33	0.042
2	bone	bon	33	2.52	30	163.87	2.14	1.18	0.038
2	tanks	t@Gk	12	2.08	16	16.25	1.70	1.22	0.071
2	gin	JIn	23	2.36	20	1215.80	2.31	1.02	0.049
2	oath	oT	6	1.78	12	495.83	2.25	0.79	0.062
2	den	dEn	2	1.30	33	130.64	2.18	0.60	0.018
2	calf	k@f	11	2.04	19	118.58	1.86	1.10	0.055
2	silk	sIlk	12	2.08	10	14.70	1.65	1.26	0.112
2	lanes	len	34	2.53	33	48.18	2.05	1.23	0.036
2	pie	pY	17	2.23	32	423.63	2.32	0.96	0.029
2	mugs	m^g	1	1.00	21	49.24	1.49	0.67	0.031
2	blush	bl^S	2	1.30	7	27.57	1.98	0.66	0.086
2	clock	klak	20	2.30	15	11.13	1.59	1.45	0.088
2	sword	sord	7	1.85	13	55.00	2.21	0.84	0.060
2	braids	bred	1	1.00	18	31.33	2.07	0.48	0.026
2	map	m@p	13	2.11	20	75.75	1.88	1.12	0.053
2	crash	kr@S	20	2.30	13	6.00	1.37	1.68	0.115
2	pet	pEt	8	1.90	30	96.63	2.18	0.87	0.028
2	wits	wIt	20	2.30	31	618.23	2.18	1.05	0.033
		Mean:	**15.08**	**2.00**	**18.36**	**154.49**	**1.91**	**1.07**	**0.070**
		SD:	*12.28*	*0.46*	*9.21*	*273.54*	*0.31*	*0.31*	*0.042*

[a] Log Frequency = $\log_{10}$(frequency) + 1
[b] N = neighborhood
[c] based on one-phoneme substitution/addition/deletion

Appendix B: Revised (con't)

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
3	chest	Cest	53	2.72	13	74.08	2.21	1.23	0.087
3	ditch	dIC	10	2.00	16	81.88	2.12	0.94	0.056
3	swan	swan	3	1.48	11	1.45	1.13	1.31	0.107
3	joints	Jont	39	2.59	5	92.60	1.88	1.38	0.216
3	pole	pol	27	2.43	33	37.36	1.97	1.23	0.036
3	clue	klu	15	2.18	12	29.08	1.82	1.20	0.091
3	cruise	kruz	2	1.30	8	9.63	1.66	0.78	0.089
3	bark	bark	14	2.15	15	37.13	2.04	1.05	0.065
3	pork	pork	10	2.00	9	34.67	2.09	0.96	0.096
3	tea	ti	65	2.81	33	1607.21	2.76	1.02	0.030
3	geese	gis	3	1.48	8	65.88	2.36	0.63	0.073
3	dent	dEnt	2	1.30	19	56.37	1.95	0.67	0.034
3	sheets	Sit	45	2.65	23	171.39	2.34	1.13	0.047
3	coach	koC	24	2.38	14	26.14	2.00	1.19	0.078
3	throat	Trot	51	2.71	6	52.00	2.19	1.23	0.171
3	cap	k@p	27	2.43	30	81.13	1.91	1.28	0.041
3	wheat	hwit	9	1.95	9	280.78	2.29	0.86	0.087
3	bread	brEd	42	2.62	15	37.60	2.09	1.25	0.077
3	logs	lcg	11	2.04	13	107.08	2.28	0.90	0.065
3	roar	ror	13	2.11	31	617.58	2.58	0.82	0.026
3	strap	str@p	2	1.30	6	10.67	1.71	0.76	0.112
3	firm	fRm	109	3.04	13	13.69	1.67	1.82	0.123
3	prize	prYz	28	2.45	10	41.00	2.36	1.04	0.094
3	bomb	bam	36	2.56	13	17.77	1.84	1.39	0.096
3	stripes	strYp	4	1.60	8	14.13	1.79	0.90	0.101
		Mean:	**25.76**	**2.17**	**14.92**	**143.93**	**2.04**	**1.08**	**0.084**
		SD:	*25.32*	*0.52*	*8.56*	*329.99*	*0.34*	*0.27*	*0.043*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
4	spray	spre	16	2.20	7	5.57	1.42	1.55	0.181
4	dime	dYm	4	1.60	20	92.20	1.86	0.86	0.041
4	truck	tr^k	57	2.76	9	14.11	1.67	1.65	0.155
4	screen	skrin	48	2.68	2	7.00	1.56	1.72	0.463
4	scare	skEr	6	1.78	11	41.27	2.16	0.82	0.070
4	crown	krWn	19	2.28	10	25.20	1.68	1.36	0.120
4	broom	brum	2	1.30	10	43.60	1.87	0.70	0.065
4	aid	ed	139	3.14	20	102.35	2.20	1.43	0.067
4	grin	grIn	13	2.11	9	26.56	2.03	1.04	0.104
4	seeds	sid	41	2.61	27	184.19	2.31	1.13	0.040
4	bugs	b^g	4	1.60	26	190.58	1.80	0.89	0.033
4	tack	t@k	4	1.60	37	78.49	1.96	0.82	0.022
4	deck	dEk	23	2.36	20	40.20	2.00	1.18	0.056
4	rope	rop	15	2.18	27	37.48	1.89	1.15	0.041
4	kick	kIk	16	2.20	28	21.11	1.80	1.23	0.042
4	mast	m@st	6	1.78	14	253.07	2.76	0.64	0.044
4	beef	bif	32	2.51	15	460.00	2.18	1.15	0.071
4	rim	rIm	5	1.70	26	129.31	1.96	0.87	0.032
4	ash	@S	11	2.04	17	986.53	2.18	0.94	0.052
4	bowl	bol	23	2.36	33	59.18	2.06	1.15	0.034
4	mate	met	21	2.32	28	197.61	2.43	0.96	0.033
4	mat	m@t	8	1.90	30	636.03	2.37	0.80	0.026
4	frogs	frcg	1	1.00	4	8.75	1.58	0.63	0.137
4	fist	fIst	26	2.42	11	160.27	2.24	1.08	0.089
4	wheels	hwil	56	2.75	7	100.86	1.76	1.56	0.183
		Mean:	**23.84**	**2.13**	**17.92**	**156.06**	**1.99**	**1.09**	**0.088**
		SD:	*29.07*	*0.50*	*9.87*	*227.65*	*0.31*	*0.31*	*0.092*

Appendix B: Revised (con't)

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
5	fun	f^n	44	2.64	25	190.76	2.06	1.28	0.049
5	fee	fi	16	2.20	31	815.13	2.64	0.84	0.026
5	bet	bEt	20	2.30	32	251.59	2.55	0.90	0.027
5	slice	slYs	13	2.11	8	11.13	1.62	1.31	0.140
5	nap	n@p	4	1.60	20	6.75	1.50	1.07	0.051
5	hedge	hEJ	2	1.30	11	71.36	2.09	0.62	0.054
5	slot	slat	6	1.78	17	18.29	1.64	1.09	0.060
5	brook	brUk	3	1.48	6	62.83	2.30	0.64	0.097
5	grief	grif	10	2.00	11	27.55	1.99	1.01	0.084
5	wax	w@ks	14	2.15	4	51.75	1.89	1.14	0.221
5	dart	dart	1	1.00	10	113.10	2.43	0.41	0.039
5	beads	bid	1	1.00	26	299.19	2.22	0.45	0.017
5	fan	f@n	18	2.26	21	430.62	2.33	0.97	0.044
5	crates	kret	2	1.30	16	61.25	1.65	0.79	0.047
5	flame	flem	17	2.23	11	21.27	1.68	1.33	0.108
5	tide	tYd	45	2.65	22	124.18	2.13	1.24	0.053
5	bar	bar	82	2.91	24	229.46	2.11	1.38	0.054
5	ant	@nt	28	2.45	12	3204.83	2.59	0.94	0.073
5	pill	pIl	15	2.18	36	88.97	2.12	1.03	0.028
5	loot	lut	4	1.60	26	53.19	2.04	0.79	0.029
5	dust	d^st	70	2.85	8	239.88	2.22	1.28	0.138
5	trail	trel	31	2.49	13	26.62	1.90	1.31	0.092
5	sand	s@nd	28	2.45	13	2310.46	2.63	0.93	0.067
5	rug	r^g	13	2.11	22	16.64	1.65	1.28	0.055
5	sport	sport	17	2.23	7	53.00	1.90	1.18	0.144
		Mean:	**20.16**	**2.05**	**17.28**	**351.19**	**2.07**	**1.01**	**0.072**
		SD:	*20.87*	*0.54*	*8.89*	*756.50*	*0.34*	*0.28*	*0.048*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
6	lamp	l@mp	18	2.26	11	13.27	1.72	1.31	0.107
6	shed	Sed	11	2.04	17	249.41	2.55	0.80	0.045
6	trap	tr@p	20	2.30	13	13.85	1.64	1.40	0.097
6	dive	dYv	23	2.36	17	30.71	1.68	1.41	0.077
6	scream	skrim	13	2.11	6	25.67	2.03	1.04	0.148
6	sponge	sp^nJ	7	1.85	1	16.00	2.20	0.84	0.456
6	clip	klIp	6	1.78	11	25.73	1.69	1.05	0.087
6	hen	hEn	22	2.34	24	229.54	2.22	1.06	0.042
6	mink	mIGk	5	1.70	13	47.23	1.98	0.86	0.062
6	cave	kev	9	1.95	22	69.00	1.93	1.01	0.044
6	rib	rIb	1	1.00	19	14.58	1.77	0.56	0.029
6	coast	kost	61	2.79	12	158.00	2.67	1.04	0.080
6	bench	bEnC	35	2.54	7	11.43	1.63	1.56	0.182
6	roast	rost	10	2.00	12	145.42	2.50	0.80	0.062
6	flood	fl^d	19	2.28	5	35.00	2.14	1.06	0.175
6	pool	pul	111	3.05	18	25.28	1.99	1.53	0.078
6	gang	g@G	22	2.34	15	15.60	1.65	1.42	0.086
6	thief	Tif	8	1.90	8	29.25	1.99	0.96	0.107
6	wrist	rIst	10	2.00	16	26.94	1.79	1.12	0.065
6	spy	spY	9	1.95	15	12.20	1.69	1.16	0.072
6	herd	hRd	269	3.43	20	461.95	2.32	1.48	0.069
6	clerk	klRk	34	2.53	8	4.13	1.36	1.87	0.189
6	ape	ep	3	1.48	17	156.53	2.21	0.67	0.038
6	jail	Jel	21	2.32	22	22.95	1.83	1.27	0.055
6	rent	rEnt	21	2.32	17	69.00	1.91	1.22	0.067
		Mean:	**30.72**	**2.19**	**13.84**	**76.35**	**1.96**	**1.14**	**0.101**
		SD:	*54.58*	*0.49*	*5.81*	*106.57*	*0.33*	*0.31*	*0.086*

Appendix B: Revised (con't)

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
7	shell	SEl	22	2.34	21	93.00	2.14	1.09	0.049
7	knife	nYf	76	2.88	8	191.38	2.45	1.17	0.128
7	cheers	CIr	8	1.90	27	87.22	2.13	0.89	0.032
7	skunk	sk^Gk	1	1.00	5	2.00	1.16	0.87	0.148
7	peg	pEg	4	1.60	10	17.00	1.86	0.86	0.079
7	fleet	flit	17	2.23	16	27.25	1.55	1.44	0.083
7	gown	gWn	16	2.20	8	187.25	2.48	0.89	0.100
7	hint	hInt	9	1.95	10	15.70	1.66	1.18	0.105
7	row	ro	36	2.56	38	196.08	2.25	1.14	0.029
7	bay	be	63	2.80	35	570.31	2.36	1.19	0.033
7	task	t@sk	60	2.78	9	17.67	1.53	1.81	0.168
7	sheep	Sip	23	2.36	20	180.15	2.18	1.08	0.051
7	brow	brW	6	1.78	7	26.57	1.45	1.23	0.149
7	shock	Sak	31	2.49	21	20.38	1.77	1.40	0.063
7	brat	br@t	1	1.00	11	35.91	1.79	0.56	0.048
7	yell	yEl	9	1.95	19	115.16	2.25	0.87	0.044
7	thorns	Torn	3	1.48	9	27.89	1.96	0.75	0.077
7	cards	kard	26	2.42	15	40.40	1.85	1.31	0.080
7	track	tr@k	38	2.58	10	14.90	1.90	1.36	0.120
7	gum	g^m	14	2.15	16	161.31	1.93	1.11	0.065
7	net	nEt	34	2.53	26	290.96	2.41	1.05	0.039
7	blade	bled	13	2.11	9	27.44	1.79	1.18	0.116
7	bruise	bruz	3	1.48	10	6.50	1.60	0.92	0.084
7	grease	gris	9	1.95	14	26.50	1.91	1.02	0.068
7	chunks	C^Gk	2	1.30	10	7.20	1.65	0.79	0.073
		Mean:	**20.96**	**2.07**	**15.36**	**95.45**	**1.92**	**1.09**	**0.081**
		SD:	*20.62*	*0.53*	*8.70*	*127.23*	*0.34*	*0.26*	*0.039*

List	Final word	Transcription	Frequency	Log Freq	N Density	Mean Freq N	Mean Log Freq N	N Ratio 1	N Ratio 2
8	grain	gren	27	2.43	20	67.25	2.24	1.08	0.051
8	vest	vEst	4	1.60	19	58.16	2.12	0.75	0.038
8	belt	bElt	29	2.46	17	56.59	2.02	1.22	0.067
8	tub	t^b	13	2.11	17	14.00	1.67	1.26	0.069
8	sap	s@p	1	1.00	27	14.96	1.76	0.57	0.021
8	mouse	mWs	10	2.00	14	79.43	1.93	1.04	0.069
8	spool	spul	1	1.00	9	71.33	1.91	0.52	0.055
8	plea	pli	11	2.04	17	46.00	1.81	1.13	0.062
8	fur	fR	15	2.18	18	203.28	1.78	1.22	0.064
8	lid	lId	19	2.28	23	78.91	2.06	1.11	0.046
8	notch	naC	6	1.78	7	664.43	1.96	0.91	0.115
8	jar	Jar	16	2.20	16	345.13	2.04	1.08	0.063
8	aim	em	37	2.57	21	127.38	2.41	1.07	0.048
8	fudge	f^J	1	1.00	6	22.17	1.90	0.53	0.081
8	chip	CIp	17	2.23	24	11.25	1.69	1.32	0.052
8	juice	Jus	11	2.04	13	17.31	1.68	1.22	0.086
8	mice	mYs	10	2.00	22	128.00	2.24	0.89	0.039
8	mold	mold	45	2.65	18	88.94	2.10	1.27	0.066
8	breath	brET	53	2.72	3	17.33	1.98	1.37	0.314
8	slave	slev	30	2.48	9	10.00	1.46	1.70	0.159
8	stamp	st@mp	8	1.90	4	1.25	1.08	1.77	0.307
8	cork	kork	9	1.95	11	84.91	2.24	0.87	0.074
8	strips	strIp	30	2.48	6	18.00	1.68	1.47	0.197
8	junk	J^Gk	8	1.90	9	5.78	1.53	1.25	0.122
8	raft	r@ft	4	1.60	14	5.57	1.38	1.16	0.077
		Mean:	**16.60**	**2.02**	**14.56**	**89.49**	**1.87**	**1.11**	**0.094**
		SD:	*14.00*	*0.49*	*6.65*	*141.77*	*0.31*	*0.32*	*0.076*

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Perceptual Adjustments to Foreign Accented English[1]

Constance M. Clarke[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by a National Science Foundation Graduate Research Fellowship to the author and was conducted at the Speech Research Laboratory, Indiana University. The author would like to thank Dr. David Pisoni for the opportunity to conduct this research in the SRL and for his assistance and guidance. Much gratitude is also given to Luis Hernandez for his technical support, to Darla Sallee for office support and to Dr. Rebecca Herman for assistance with acoustical analyses.
[2] University of Arizona, Department of Psychology, Tucson, AZ.

Perceptual Adjustments to Foreign Accented English

Abstract. Two training tasks were evaluated for a study of perceptual learning of foreign-accented speech. The purpose of the planned perceptual learning study is to examine whether exposure to the speech of several foreign-accented voices will improve perception of a new voice with the same accent. Two important characteristics of the listeners' task during training with the voices are emphasis on similarities among accented voices and the availability of a method for evaluating listener performance during training. The first task examined was a similarity judgment task for pairs of voices. Multidimensional scaling was used to assess changes in perception through the course of the exposure, but this technique proved not to be sensitive enough to subtle changes in perception for these stimuli. The second task examined was the combination of a similarity judgment task and a transcription task. This dual task method was more successful in satisfying the goals for the training task and is a promising technique for use in the perceptual learning study.

Introduction

An important and still unanswered question in the study of speech perception is how the human speech processing system achieves perceptual constancy in the face of enormous variability in the acoustic signal. Productions of the same speech sound by different speakers are acoustically different. Even different productions of the same sound by one talker are not identical. Yet listeners still perceive the same speech sound across such variation. How does the perceptual system so successfully extract a single phoneme when there are few, if any, truly invariant features across different productions of that phoneme?

In the traditional approach to solving this problem, the perceptual system was thought to engage in a process of normalization when processing speech (Shankweiler, Strange, & Verbrugge, 1977). It was believed that this process stripped away and discarded variability that did not directly specify the intended speech segment (e.g., acoustic consequences of vocal tract characteristics, phonetic context, or speaking rate). What remained was invariant information of some kind that would unambiguously specify an abstract linguistic category (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). However, there is evidence that the features of speech that vary from token to token are not simply noise. They may actually be useful in the processing of speech.

Over the past ten years, research by Pisoni and his colleagues has shown that variable features of speech, such as those caused by the differences between talkers, are not lost or discarded but are stored and used later in perception. For example, using a continuous recognition paradigm, Palmeri, Goldinger, and Pisoni (1993) found that subjects were faster at recognizing that a word had been presented previously in a list when it was presented in the same voice, compared to when it was presented in a new voice. In addition, a study by Nygaard, Sommers, and Pisoni (1994) established that familiarity with voices improves intelligibility for those voices when speaking novel words. These pieces of evidence and others (e.g., Greenspan, Nusbaum, & Pisoni, 1988; McGarr, 1983) call into question the traditional view that the characteristics of specific voices are discarded by the speech perception system as it decodes a speaker's intended message. It seems that these "irrelevant" details are, in fact, learned and put to use in processing new input.

If perceptual constancy for speech cannot be explained by normalization, what other mechanism might be responsible? Previous study of the problem of talker variability has yielded two distinct hypotheses about how phonetic categories are structured and how the system deals with variability. The classic model holds that a phonetic category is based on a single, abstract prototype (Posner & Keele, 1968). "Perceptual operations" (Kolers, 1976; Nygaard et al., 1994; Pisoni, 1997) analyze each incoming speech token and match it with the correct category. Experience with a particular type of speech (e.g., with a particular person's voice), allows the system to learn the unique perceptual operations needed to process that speech. Those perceptual operations can be stored and used for later perception. An alternative approach is based on exemplar-based models of categorization (Jacoby & Brooks, 1984; Medin & Shaffer, 1978; Nosofsky, 1986). Applied to speech perception, exemplar-based approaches claim that phonetic categories are made up of episodic traces of speech segments. In a simple version of this model, the token currently being perceived is matched with the most similar acoustic trace held in memory and is assigned the corresponding phonetic category (Goldinger, 1998; Pisoni, 1997).

Thus far, the studies demonstrating retention of variable features of speech do not distinguish between the prototype-with-perceptual-operations model and the exemplar-based model; the results can be explained by both models. For example, the continuous recognition study by Palmeri et al. (1993) described above is intuitively consistent with the exemplar-based model since improved perception was found for the same word produced by the same voice. But perceptual operations used for processing a particular word in a particular voice could also have been responsible for the results. For example, when a listener first hears a word in the experiment, she retains the perceptual operations used to process and identify that word. When the word, spoken by the same voice, is later presented for recognition, the procedures for processing the item will match the stored procedures. Consequently, the overall familiarity will be greater, and the listener will be more likely to correctly report that the word is old. In comparison to the Palmeri et al. (1993) study, the Nygaard et al. (1994) (see also Nygaard & Pisoni, 1998) study showed the perceptual system capable of a higher level of abstraction. There, the words at training and test were different, yet performance was still better with the familiar voices. This suggests that it is not just the particular episodic word tokens that are held in memory, but rather something more general about a talker's speech. This knowledge might be stored in the form of perceptual procedures for matching the acoustic-phonetic input with phonetic categories. However, an exemplar-based model can also explain the results if it is argued that although the words were different at training and test, exemplars of the individual phonetic categories could have been stored.

One way to distinguish between these two models is to investigate whether the perceptual system can learn even more abstract characteristics of the non-linguistic aspects of speech. A certain level of abstraction cannot be explained with an exemplar-based model. A foreign accent, for example, is a source of non-linguistic variability in speech that causes the speech to differ from native speech in abstract, phonetic rule-governed ways. It has been shown that experience with an accented voice helps perception of new words spoken by that voice (Clarke, 2000; Wingstedt & Schulman, 1987). However, if it could be shown that learning an accent aids in the perception of new words spoken by a *new* talker with the same accent, another level of abstraction would be introduced. Perceptual learning of the accent itself would be demonstrated. This would indicate that the speech perception system can learn at a more abstract level than has so far been established through voice training studies. The characteristics of a voice are largely based on a particular vocal tract and glottal source. However, a foreign accent is based on the structure and content of the native language's phonetic system, its phonological rules, and the way it interacts with the target language (Tarone, 1987). These characteristics are overlaid on a voice and are assumed to be more or less consistent across different speakers with the same accent. This consistency is the basis of a listener's ability to identify "what kind of accent" a non-native speaker has. Yet these consistencies across speakers are abstract. They are phonetic, not acoustic, in nature.

Generalization of accent learning to a new talker would call into question a strict exemplar-based view of speech perception. If voice learning is due to the referencing of previously stored acoustic tokens of each voice, as claimed by exemplar-based models, transfer of learning to a new voice with the same accent would not be expected. This is because the acoustic characteristics of the new voice may be quite different from the stored tokens; the only similarities would be abstract, phonetic ones. The type of perceptual learning that transfers to new voices may be better explained by the storing of perceptual operations. Perceptual operations may be more flexible than episodic traces because different levels of analytical rules could be retained, from very specific (i.e., at the level of acoustic characteristics) to very abstract (i.e., at the level of phonological regularities). The abstract rules could be applied to a larger variety of tokens of the original type of speech (e.g., Spanish-accented speech).

In a recent study, Clarke (2000) investigated whether experience with foreign-accented voices improves perception of the speech of a new talker with the same accent, that is, whether an accent itself can be learned. Borrowing the experimental methodology used by Nygaard et al. (1994), Clarke investigated accent learning by giving two groups of listeners three days of accent training. One group was trained with four Spanish-accented voices and four non-accented voices. The other group was trained with four Chinese-accented voices and four non-accented voices. All voices were female speakers producing American English. The listeners' task during training was to learn the name that went with each of the voices. After three days of recognition training, subjects were given a word intelligibility test with new sentences presented in noise. Test sentences included Spanish- and Chinese-accented voices used in training, as well as new Spanish- and Chinese-accented voices. Test sentences also included one new and one old non-accented voice. Clarke found that the Spanish-trained group had an advantage with the old Spanish-accented voice (one of the voices from training), and the Chinese-trained group had an advantage with the old Chinese-accented voice. This replicated the Nygaard et al. findings and showed that voice learning also occurs with foreign-accented voices. However, the listeners' experience with the accented voices did *not* improve their perception of the new accented voices: the Spanish-trained group showed no advantage for the new Spanish-accented voice, nor did the Chinese-trained group for the new Chinese-accented voice. The results suggested that the perceptual learning of speech is voice specific.

It may be, however, that the lack of transfer to new voices was due to a particular aspect of the training methodology. The training task itself may have interfered with the listener's "motivation" for finding similarities among voices of the same accent. For the three days of training, the listeners' goal was to discriminate the voices and match them with the correct name. This is the same procedure Nygaard et al. (1994) used. While the task encouraged close attention to the acoustic and phonetic characteristics of each voice, it effectively required listeners to look for differences among the voices, not similarities. Perhaps a task that emphasized the commonalities among the voices in each accent group would better support accent learning in addition to individual voice learning. An experiment using such a task would be better able to demonstrate whether the perceptual system can learn the abstract phonological characteristics common to all the accented voices and apply them to a new voice.

The purpose of the following two experiments was to find a new training task that can be used in a replication and extension of Clarke (2000). The new training task had to fulfill two goals: first, emphasize the similarities among the voices with the same accent (e.g., among the four Spanish-accented voices); and second, allow for a way to measure the success of training. This second requirement is necessary in order to, for example, determine which subjects were attending to and benefiting from the task. The first experiment assessed the use of a similarity judgment task. The second investigated a task that included both similarity judgments and sentence transcription. The second task was found to be more successful in meeting the goals stated above.

Experiment 1: Similarity Judgments

In the first task investigated, listeners were asked to make similarity judgments between pairs of voices on a seven-point scale from Very Similar to Not Similar At All. Multidimensional scaling[3] was then used to examine whether their similarity spaces for the voices changed from the beginning of the experiment to the end. It was hoped that the similarity judgment task would serve the purpose of encouraging listeners to focus on the similarities among the accented voices, rather than the differences. The multidimensional scaling technique provided a way of measuring whether listeners' perception of the voices was affected by the task demands. For example, one possible change could be a shift from making similarity judgments based solely on the presence or absence of an accent, to judgments based on perceiving and encoding more fine-grained characteristics of the voices.

Method

Subjects

Twenty-four Indiana University undergraduates (20 female, 4 male) participated as listeners in the experiment for partial fulfillment of a course requirement. Eight participants were excluded from the final analysis: one because of a history of hearing disorder, one because of an error in the experimental program, two because of a failure to follow instructions, and four so that the correct counterbalancing of conditions was maintained[4]. The remaining 16 participants (13 female, 3 male) were monolingual, native speakers of American English who reported no history of speech or hearing disorders at the time of testing.

Materials and Stimuli

Two groups of eight participants each listened to eight female voices. For each group, four of the voices were non-accented (NA) when speaking English (native speakers of English), and four had a noticeable accent (non-native speakers of English)[5]. For one group (Spanish/NA) the accented voices had a Spanish accent, and for the other group (Chinese/NA) the accent was Chinese[6]. These twelve voices (four non-accented, four Spanish-accented, and four Chinese-accented) were the same voices used in the training portion of the Clarke (2000) study. The non-accented speakers were native speakers of American English with no obvious regional accent, ranging in age from 19 to 31. The four Spanish-accented speakers were native speakers of Mexican Spanish, all from the region of Sonora, Mexico, who began learning English after the age of 25 (mean age of English acquisition: 33 years; mean age at time of recording: 38 years). The four Chinese-accented speakers were native speakers of Mandarin Chinese, all from Taiwan, ROC, who began learning English after the age of eleven (mean age of English acquisition: 12 years; mean age at time of recording: 24 years). All accented speakers reported using their native language at least thirty percent of the time in their current daily lives. The voices had originally been recorded in the Speech Perception Laboratory at the University of Arizona, Department of Psychology.

[3] Multidimensional scaling (MDS) is a statistical technique for representing similarity among objects. Similarity data specify the location of objects in an n-dimensional space in which distance is inversely related to similarity.

[4] Beyond the counterbalancing requirements, two of the participants who were excluded from further analyses were chosen because they showed the greatest trend toward a bias in their similarity judgments. The other two were excluded because they were the last to participate.

[5] Although the main interest in these experiments was evidence of perceptual learning for the accented voices, the non-accented voices were included in order to keep the voice set identical to that used in the Clarke (2000) study and in the planned follow-up study. The inclusion of non-accented voices in the full studies is important for verifying that the basic voice learning effect can be obtained with the methodology used.

[6] Different groups listened to the Spanish-accented and Chinese-accented voices because in the original study (Clarke, 2000) accent type was a between-subjects variable. There are no experimental comparisons between groups in the present study.

The voices were recorded onto tape and digitized onto a Macintosh PowerPC 8100 at a sampling rate of 22.05 kHz and a resolution of 16 bits. Amplitude was normalized to 90% of maximum for all sentences, and the individual sentence files were converted to WAVE format.

The sentences used in the experiment were taken from the Revised Speech Perception In Noise (SPIN) test (Bilger, 1984; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Elliot, 1995; Kalikow, Stevens, & Elliott, 1977). The Revised SPIN Test comprises a set of phonetically and frequency balanced sentences designed to assess impairment of hearing for speech. It is made up of five- to eight-word sentences, each ending in a common, one-syllable word. All the sentences used in the current experiment were High Predictability (HP) sentences, meaning that the final word in each sentence was highly predictable from the semantic context of the sentence (e.g., Stir your coffee with a spoon.). One hundred four of these sentences were used in the present study.

Among the eight voices that each group heard (four non-accented and four accented), each voice was paired with every other voice, for a total of 28 unique pairings (a voice was never paired with itself). Each of the 28 pairs of voices was presented once in each of six blocks, for a total of 168 trials. Each individual voice was heard seven times per block, 42 times total. The ordering of the voices in each pair was counterbalanced across blocks such that each voice was heard first and second an equal number of times across the experiment. The order of voices within a particular pair was identical for blocks 1, 3, and 5; the mirror order occurred in blocks 2, 4, and 6. In addition, voice order was counterbalanced across subjects. Within each block, the voice pairs were presented in random order using an on-line randomization program. Finally, in each trial, both voices produced the same sentence. Because of constraints stemming from which speakers had originally recorded which sentences, 64 of the 104 unique sentences had to be repeated once during the experiment in order to fill the 168 trials. However, a sentence was never repeated by the same voice and was never repeated in the same block. The sentences spoken by the non-accented voices were identical across the two groups; the sentences spoken by the Spanish-accented voices for the Spanish/NA group were identical to those spoken by the Chinese-accented voices for the Chinese/NA group.

Procedure

Participants were seated in a quiet room in front a computer keyboard and monitor. Up to six participants were run at a time, in separate booths, with separate computers, and at their own pace. Stimuli were presented to each participant over Beyer Dynamic DT100 headphones at approximately 71 dB SPL from a Pentium 133 MHz IBM compatible computer with a Soundblaster 16 AWE 32 sound card. After reading through an instruction sheet, listeners heard each of the eight voices say one sentence each. This phase was simply to familiarize them with the range and type of voices they would be exposed to; no response was required. Then the main portion of the experiment began. In each trial, listeners were alerted with the word “READY” displayed on the computer screen for 1000 ms. The listeners then heard two voices say the same sentence, with a 500 ms inter-stimulus interval (ISI). Five hundred milliseconds after the second voice, listeners were asked to rate how similar the two voices were to one another on a scale from 1 (labeled “Not Similar At All”) to 7 (labeled “Very Similar”) with the prompt, “PLEASE RATE SIMILARITY”. They typed the rating response into a keyboard, and the response appeared on the screen. Participants were allowed to change the response if they wanted to before submitting it by pressing the ENTER key. In the instructions, listeners were asked to use the whole range of the scale during the experiment. After the response was submitted, a 1000 ms inter-trial interval (ITI) occurred before the next trial began. The entire experiment took approximately 25 minutes, and participants were given a break half way through.

Results

Within each block of the experiment, the similarity judgments for each pair of voices were averaged across all subjects in a group. This produced an 8 x 8 matrix of similarity data for each block in which the average similarity score for every combination of two voices was represented. The first block was considered warm-up and was not included in the analysis. For each group (Spanish/NA and Chinese/NA), Blocks 2 through 6 were submitted as separate matrices to a non-metric multidimensional scaling (MDS) analysis (Euclidean distance metric) using the INDSCAL model in the SPSS 10.0 ALSCAL program. This model takes several matrices and finds a multidimensional spatial solution that best fits the data in all the matrices. The model then determines dimension weights for each individual matrix that describe how much emphasis that matrix gives to each dimension relative to the overall solution. Our interest was in the change in these dimension weights from Block 2 (beginning of the experiment) to Block 6 (end of the experiment) for both groups. A change in the importance listeners placed on each dimension due to experience with the voices would indicate that the exposure was having an effect on perception.

The data were analyzed with both a two-dimensional and a three-dimensional solution. The two-dimensional solution was the most appropriate for both groups' data because the fits were extremely good (Spanish/NA group: stress = .10, R^2 = .96; Chinese/NA group: stress = .06, R^2 = .98) and the dimensions were interpretable as 1) accentedness and 2) other voice characteristics. The two-dimensional MDS solutions (across all blocks) for both groups are shown in Figures 1A (Spanish/NA group) and 1B (Chinese/NA group). Each point represents a voice, and the points are labeled as non-accented (NA 1-4) or accented (Spanish/Chinese 1-4). Inspection of this figure shows that Dimension 1 clearly reflects accentedness: all accented voices have positive values on this dimension and all non-accented voices have negative values. Further support for this conclusion comes from the high correlation between rated accentedness (ratings obtained in the Clarke (2000) study) and Dimension 1 coordinate value (r = +.99, p < .001 for both groups). The source of Dimension 2 is less clear, but may reflect other general voice characteristics. One possible candidate is age of the speaker. There was a marginally significant positive correlation between speaker age and Dimension 2 value for the non-accented voices only (Spanish/NA group: r = +.94, p = .06; Chinese/NA group: r = +.95, p = .05; two-tailed; alpha set to .0125 for multiple correlations); the correlation was not significant for the accented voices. Another possible source of Dimension 2 is voice pitch. There was a trend toward a negative correlation between average minimum F0 and Dimension 2 value for non-accented voices only (Spanish/NA group: r = -.93; p = .07; Chinese/NA group: r = -.97, p = .03; two-tailed; alpha set to .0125 for multiple correlations); again, the correlation was not significant for the accented voices[7]. A definitive interpretation of Dimension 2 is not essential, however, for the objectives of this experiment. Of greatest interest is whether there was a systematic shift, over the course of exposure to the voices, in the relative weightings of the voice dimensions, whatever they may be.

The normalized dimension weights for Blocks 2 through 6 for both groups are shown in Figures 2A (Spanish/NA group) and 2B (Chinese/NA group). It can be seen from the dimension scales themselves that, for all blocks, similarity judgments were overwhelmingly based on accentedness. In all but one block across both groups, Dimension 1 (accentedness) commanded over 89% of the weight in similarity judgments. In terms of changes in dimension weightings from Block 2 to Block 6, however, the

[7] Because the voices were not controlled for anything but accentedness, these analyses are post hoc, and the comments based on them are purely speculative. It is noted, however, that the finding that the accented voices are less separated in the similarity space is consistent with other studies of the perception of accented voices. Goggin, Thompson, Strube, and Simental (1991) and Thompson (1987) have found that listeners are worse at learning to discriminate foreign-accented voices than non-accented voices. These findings suggest that it is more difficult to distinguish subtle differences in the voice characteristics of accented voices compared to those of non-accented voices.

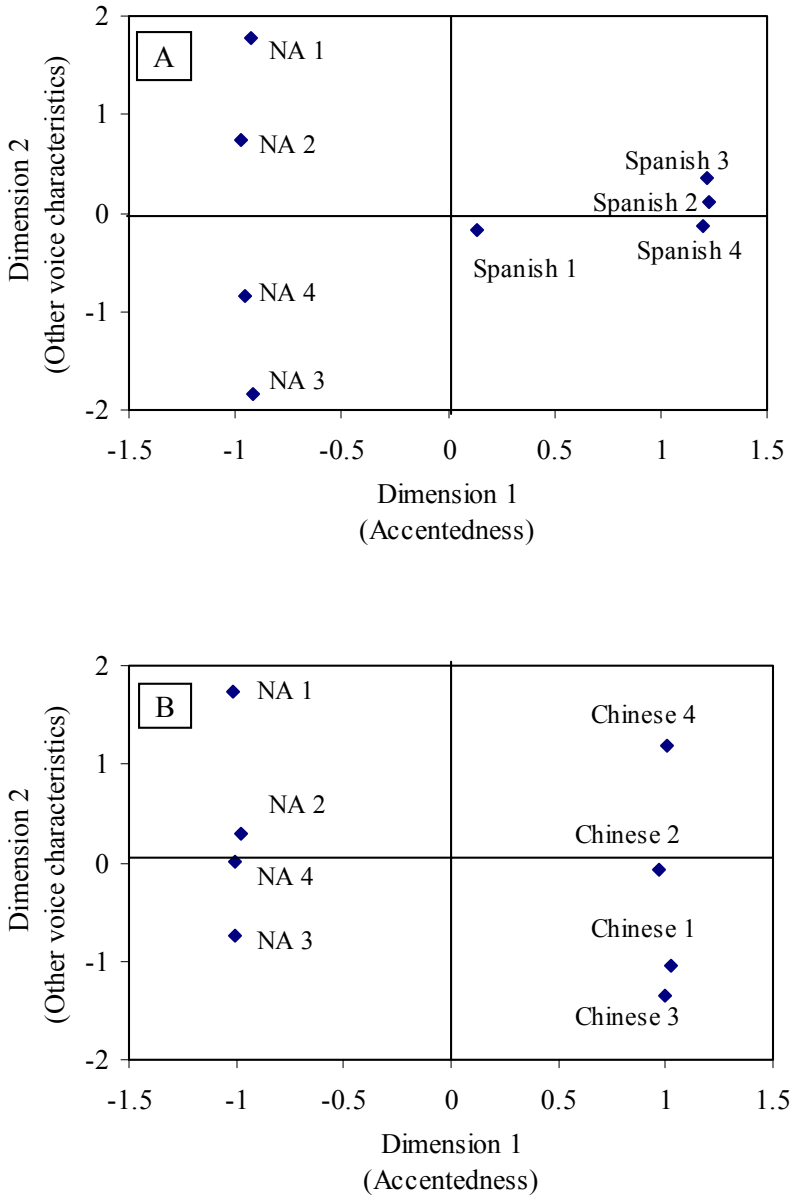


Figure 1. Multidimensional scaling solutions for A) the group listening to four non-accented (NA) and four Spanish-accented voices, and B) the group listening to four non-accented (NA) and four Chinese-accented voices. The solutions are based on similarity judgments between pairs of voices. Each point represents one voice (NA 1-4 were the same for both groups). For both solutions, Dimension 1 was interpreted as Accentedness and Dimension 2 as Other voices characteristics. The dimension scales are arbitrary.

two groups showed different patterns. For the Spanish/NA group, slightly more weight was given to Dimension 2 (other voice characteristics) as the experiment progressed over time. However, for the Chinese/NA group, after Block 2, in which only about 78% of the weight went to the accentedness dimension, almost 100% of the weight went to accentedness. This seems to indicate that after Block 2, the listeners in the Chinese/NA group shifted to a strategy of judging voice similarity almost entirely by

whether the voices matched on accentedness. That is, all accented/accented pairs and non-accented/non-accented pairs were judged as equally similar, and all accented/non-accented pairs were judged as equally dissimilar. This is in contrast with the listeners in the Spanish/NA group, who on the whole seemed to maintain a consistent strategy but gradually became slightly more influenced by the individual voice characteristics.

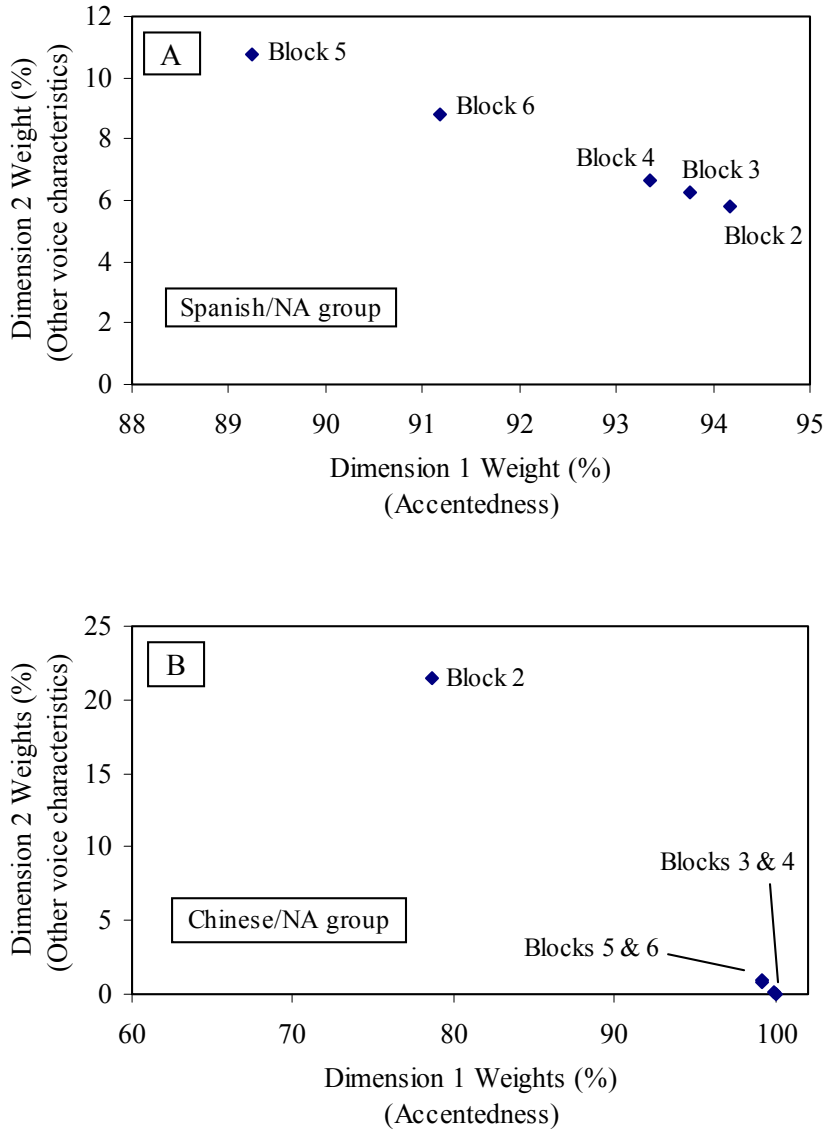


Figure 2. Normalized dimension weights for A) the group listening to non-accented and Spanish-accented voices, and B) the group listening to non-accented and Chinese-accented voices. Each point represents an experimental block of 28 trials. Each block's value on each scale indicates the percentage of weight or importance given to that dimension during that block. Dimension 1 was interpreted as Accentedness and Dimension 2 as Other voices characteristics.

Discussion

The goal of this first experiment was to determine whether a similarity-rating task would be appropriate for use in a study of accent learning. At first glance, the task did meet the initial goal of emphasizing the similarities among the voices. However, the task did not robustly satisfy the second goal: to provide a method of measuring the task's success in affecting perception. A multidimensional scaling analysis was used to look for a change in the subjective similarity space of the voices from the beginning of the experiment, when the voices were unfamiliar to the listeners, to the end, when they were more familiar. To the author's knowledge, the use of MDS in measuring perceptual changes in voice familiarity has not been reported in the literature before. Although MDS seems to be a promising technique for this purpose, it was unsuccessful in the present experiment. First, due to the nature of the voice stimuli, the accentedness dimension of the voice set had almost complete influence on the similarity judgments. This fact likely rendered the similarity measure insensitive to any subtle changes in similarity space that might have been present. Second, the changes that were seen, that is, the change in dimension weightings from the beginning to the end of the experiment, were in opposite directions for the group listening to Spanish and NA voices and the group listening to Chinese and NA voices. The listeners assigned to the Spanish/NA group gave more weight to the "other voice characteristics" as the experiment went on, while the listeners assigned to the Chinese/NA group gave less (and, for most of the experiment, judged solely based on accentedness). Finally, we found that this similarity judgment task was probably too monotonous for a full, three-day training experiment. Therefore, a new task was used in Experiment 2.

Experiment 2: Similarity Judgment and Transcription

The main problem with the first task was that it was difficult to evaluate whether the training was having an effect on listeners' perception of the voices. Therefore, experiment two retained the similarity judgment task, since it was still the best candidate for emphasizing similarities among voices, but added a subsidiary activity: a transcription task. It has been well established that transcription of words or sentences improves with increased voice familiarity (e.g., Greenspan, Nusbaum, & Pisoni, 1988; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Schwab, Nusbaum, & Pisoni, 1985). Transcription is therefore a more reliable and direct way of evaluating whether the experience with the voices is affecting listeners' perception. In this new task, sentence transcription trials were interspersed throughout the experiment along with the similarity rating trials. With this methodology, separate activities served the two objectives of this training task. The similarity trials encouraged listeners to attend to similarities among the voices, and the periodic transcription trials provided a way to track the listeners' ability to understand the voices. We expected that perception would improve with exposure to the voices over time and thus transcription would become more accurate from the beginning of the experiment to the end. As a final note, we also hoped that the inclusion of two different activities would make the task more interesting for the listeners. Moreover, a task that allows for a chance at improvement (the transcription task) might increase participant motivation during the entire procedure.

Method

Subjects

Twenty-eight Indiana University undergraduates (17 female, 11 male) participated as listeners in the experiment for partial fulfillment of a course requirement. Four participants were excluded from the final analysis: two because of exposure to a language other than English at an early age, and two in order

to maintain the correct counterbalancing of conditions[8]. The remaining 24 participants (14 female, 10 male) were monolingual, native speakers of American English who reported no history of speech or hearing disorders at the time of testing.

Materials and Stimuli

Participants were again assigned to two groups. One group (Spanish/NA) listened to four non-accented and four Spanish-accented voices, and the other group (Chinese/NA) listened to four non-accented and four Chinese-accented voices. The voices were the same twelve used in Experiment 1. A new set of the recorded SPIN sentences were used, 64 HP (High Predictability) sentences for the similarity judgment trials, and 24 LP (Low Predictability) sentences for the transcription trials (LP: the final word was not predictable from the semantic context of the sentence, e.g., We spoke about the knob.).

Similarity Judgment Trials. The similarity judgment task was modified slightly from Experiment 1. Instead of hearing two voices and rating how similar they were, listeners heard three voices: a reference voice and two comparison voices. The three voices in a trial were always either all accented or all non-accented. Listeners were asked to choose which comparison voice was more similar to the reference voice. (This is a variant of an XAB task.)

All possible combinations of each reference voice with two other voices from the same accent category were presented twice, once in the first half of the experiment and once in the second half. Thus, there were 96 similarity judgment trials, consisting of 48 non-accented trials and 48 accented trials. Each of the eight voices was the reference voice 12 times total. Within each half of the experiment, the trials were presented in a random order, using on-line randomization. Finally, on a given trial, all three voices always said the same sentence. Because of constraints due to which speakers had originally recorded which sentences, 32 of the 64 unique sentences had to be repeated once during the experiment in order to fill the 96 trials. However, a sentence was never repeated by the same voice and was never repeated in the same half of the experiment. For each reference voice, four of the trials contained non-repeated sentences and eight contained repeated sentences.

Transcription Trials. The transcription trials consisted of one voice saying one sentence (in the clear), followed by the listener typing the entire sentence into the computer keyboard as accurately as possible. There were 24 transcription trials, three trials for each of the eight voices, interspersed among the 96 similarity judgment trials. Only new, LP sentences were presented for transcription. The 24 trials were divided into three blocks, with each block containing one sentence from each of the eight voices. To guard against item effects on transcription accuracy, the presentation order of the blocks was counterbalanced across three groups. The block orders were as follows: Group 1—1, 2, 3; Group 2—2, 3, 1; Group 3—3, 1, 2. Within each block of trials, the eight sentences were presented in random order, using on-line randomization. One transcription trial was presented after every two to six similarity judgment trials (the number between two and six, inclusive, was randomly chosen on-line); hence, the transcription trials were dispersed evenly throughout the experiment, but their occurrence was not predictable.

Procedure

Participants were seated in a quiet room in front of a computer keyboard and monitor. Up to five participants were run at a time, in separate booths, with separate computers, and at their own pace. Stimuli were presented to each participant over Beyer Dynamic DT100 headphones at approximately 71

[8] The two participants who were excluded from further analyses were chosen because they were the last to participate.

dB SPL from a Pentium 133 MHz IBM compatible computer with a Soundblaster 16 AWE 32 sound card. After reading through an instruction sheet, listeners heard each of the eight voices say one sentence each. This phase was simply to familiarize them with the range and type of voices they would be exposed to during the full experiment; no response was required. The main portion of the experiment followed.

On each similarity judgment trial, listeners were alerted to the type of trial coming up with the prompt "SIMILARITY JUDGMENT" displayed on the computer screen for 1500 ms. The sentence they were about to hear was then displayed orthographically for 2000 ms. The words "Reference Voice" were then displayed in the middle of the screen while the reference voice was presented over the headphones. The first comparison voice began 1000 ms after the reference voice finished, and "Comparison 1" was displayed on the lower left side of the screen. "Comparison 1" remained on the screen, and after 500 ms the second comparison voice was played and "Comparison 2" was displayed on the lower right side of the screen. The screen cleared 500 ms after the second comparison voice ended, and a prompt for a response was displayed: "Which is more similar to the reference voice? 1 or 2?" Listeners responded by pressing one of two keys labeled "1" and "2" on the keyboard. After entering their response, participants pressed the ENTER key to move on to the next trial. The ITI was 1000 ms for all trials.

On each transcription trial, listeners were alerted to the type of trial with the prompt "TRANSCRIPTION" which remained on the screen throughout the presentation. After 1000 ms, the sentence to be transcribed was presented over the headphones. Following a pause of 500 ms, listeners were prompted for a response with the words, "Please type the sentence now." They typed what they had heard into the keyboard, and the response appeared on the screen. Participants were allowed to correct mistakes as they typed. When they were finished they pressed the ENTER key to submit their answer. After a 500 ms pause, feedback was provided with the words, "The sentence was:" and the sentence text, displayed for 1500 ms. After the 1000 ms ITI, the next trial began. The entire experiment took approximately 40 minutes, and participants were given a break half way through.

Results

Transcription accuracy was evaluated by scoring predetermined keywords in each of the 24 sentences[9]. The keywords were content words only, including nouns, verbs, adjectives, and adverbs. A keyword was accepted as correct if: it matched the target word exactly, it was an obvious misspelling, it was a homophone, a plural had been added or deleted, or an inflectional affix had been added or deleted. One point was given for each correct keyword, and the score for each sentence was the percentage of correct keywords out of the total possible keywords. Trials in which no response was given were not counted in the total possible score. For each participant, the percent correct score was calculated separately for the first third of the sentences (first eight sentences) and the final third (last eight sentences). Each third included one sentence from each of the voices, and across subjects, each sentence appeared an equal number of times in the first third and in the final third of the experiment.

The mean scores for the first eight and final eight transcription sentences for both groups are shown in Figure 3. For the Spanish/NA group, a Sign Test revealed that a significant number of listeners, 10 out of 12, improved in their transcription accuracy from the first third to the final third of the experiment ($p < .05$). The average improvement for the Spanish/NA group from the first third, M = 77.98%, SD = 5.07, to the final third, M = 82.74%, SD = 8.11, was shown to be marginally significant with a paired t-test, $t(11) = 1.56$, $p = .07$. The Chinese/NA group did not show a significant improvement

[9] Only the results of the transcription task will be reported here. In this experiment the similarity judgment results were of secondary interest and will be analyzed at a later time.

from the first third, M = 81.33%, SD = 10.88, to the final third, M = 85.06%, SD = 5.16, with either a paired t-test ($t(11) = 1.03$, $p = .16$), or a Sign Test (7 out of 12 improved, $p = .39$).

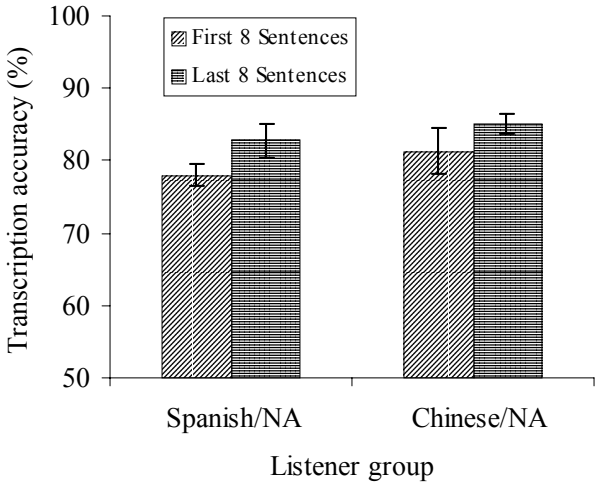


Figure 3. Mean percent accuracy of transcription of keywords in the first 8 and the last 8 sentences of Experiment 2 for the two listener groups. Spanish/NA group: group listening to non-accented and Spanish-accented voices; Chinese/NA group: group listening to non-accented and Chinese-accented voices. Error bars indicate standard error of each mean.

Discussion

The combination similarity judgment and transcription task seems to be promising as a voice-learning task for a full foreign accent learning study. Although the listeners had a relatively short exposure to the voices (40 minutes), the transcription task revealed an improvement in intelligibility of the voices, at least for the Spanish/NA group. The results for the Spanish/NA group are encouraging and suggest the possibility that a significant improvement will also be found for the Chinese/NA group when the training exposure is increased to the planned 2 ½ hours in the full experiment. Thus, we expect that this task will effectively fulfill the second objective of a good training task: to allow for a way to measure the success of training. In addition, the inclusion of similarity judgment trials, and the fact that the voices in each accent category are always grouped within a trial, are likely to fulfill the first goal: to emphasize the similarities among the accented voices. An additional benefit of this task is that the unpredictable transcription trials seem to make it more interesting for the participants and to provide motivation for improvement. This task is therefore an improvement over the similarity-rating task alone, as presented in Experiment 1, and will be the basis for developing the follow-up foreign accent training study.

Conclusion

The original motivation for these experiments was to find a new task that would draw listeners' attention to the similarities among accented voices presented during training, unlike the task used in Clarke (2000). This new training task was to be used in a follow-up study that would re-test whether experience with foreign-accented voices improves perception of a new voice with the same accent. It was argued that a task emphasizing the similarities among the accented voices would provide the best chance

for learning the abstract characteristics of a foreign accent. Transfer of that learning to a new accented voice might then be possible.

The two experiments described here reveal that finding a good voice-training task is not a simple undertaking. The procedure must satisfy several objectives at once, not the least of which is keeping participants interested and motivated so that the training has the desired effect on perception. Through these pilot experiments it was found that two different activities, interweaved throughout the training, provide the best solution for satisfying the two main goals desired for this study. The first goal, emphasizing similarities among the accented voices, is presumably satisfied by the use of a similarity judgment task as the main activity during training. The second goal, providing a way to measure training success, was fulfilled most successfully with a transcription accuracy measure. The transcription task was not completely successful, however, (i.e., for the Chinese/NA group) and may require some modification to make it a more robust measure of perceptual change. For example, transcription sentences could be presented in noise instead of in the clear, as done here. This change would lower overall performance, but may amplify the difference in the listeners' perceptual abilities from the beginning of training to the end. This amplification would be expected since the increased difficulty would demand the use of every perceptual advantage the listeners may have gained during training.

Future directions for this research include testing the similarity judgment/transcription task with the transcription trials in noise to see if this is a stronger measure of perceptual change. Finally, the new task will be applied to a replication and extension of the Clarke (2000) accent training study. It is hoped that this study will provide further insight into the mechanisms involved in the perceptual learning of novel voices and the larger issues of variation and variability in speech and spoken language processing.

References

Bilger, R.C. (1984). Manual for the clinical use of the revised SPIN Test. Champaign, IL: The University of Illinois.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32-48.

Clarke, C.M. (2000). Perceptual learning of foreign accented English. Unpublished masters thesis, Tucson, AZ: The University of Arizona.

Elliot, L.L. (1995). Verbal auditory closure and the Speech Perception In Noise (SPIN) Test. *Journal of Speech and Hearing Research, 38,* 1363-1376.

Goggin, J.P., Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory & Cognition, 19*, 448-458.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Greenspan, S.L., Nusbaum, H.C., & Pisoni, D.B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 421-433.

Jacoby, L.L., & Brooks, L.R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 18* (pp. 1-47). New York: Academic Press.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Kolers, P.A. (1976). Pattern analyzing memory. *Science, 191*, 1280-1281.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

McGarr, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26*, 451-458.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355-376.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309-328.

Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson, & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9-32). San Diego, CA: Academic Press.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Shankweiler, D., Strange, W. & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. Shaw, & J. Bransford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwab, E.C., Nusbaum, H.C., & Pisoni, D.B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395-408.

Tarone, E.E. (1987). The phonology of interlanguage. In G. Ioup, & S.H. Weinberger (Eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System* (pp. 70-85). Cambridge, MA: Newbury House Publishers.

Thompson, C.P. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1*, 121-131.

Wingstedt, M. & Schulman, R. (1987). Comprehension of foreign accents. In W. Dressler, H. Luschutzky, O. Pfeiffer, and J. Rennison (Eds.), *Phonologica 1984* (pp. 339-345). Cambridge: Cambridge, U.P.

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Auditory Learning and Adaptation after Cochlear Implantation: A Preliminary Study of Discrimination and Labeling of Vowel Sounds by Cochlear Implant Users[1]

Mario A. Svirsky,[2,3,4] Alicia Silveira,[3] Hamlet Suarez,[3] Heidi Neuburger,[2] Ted T. Lai,[2] and Peter M. Simmons[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work has been supported by grants from NIH-NIDCD (R01-DC03937, Principal Investigator: M. Svirsky; and training grant T32-DC00012, Principal Investigator: D. Pisoni); the Deafness Research Foundation; the National Organization of Hearing Research; the American Academy of Otolaryngology-Head and Neck Surgery and from BID/CONICYT (Uruguay; Principal Investigators: H. Suarez and M. Svirsky).

[2] Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

[3] Facultad de Medicina y Hospital Maciel, Uruguay.

[4] Department of Biomedical Engineering, Purdue University.

Auditory Learning and Adaptation after Cochlear Implantation: A Preliminary Study of Discrimination and Labeling of Vowel Sounds by Cochlear Implant Users

Abstract. This study examined two possible reasons underlying longitudinal increases in vowel identification by cochlear implant users: improved labeling of vowel sounds and improved electrode discrimination. The Multidimensional Phoneme Identification (MPI) model was used to obtain ceiling estimates of vowel identification for each subject, given his electrode discrimination skills. Vowel identification scores were initially lower than the ceiling estimates, but they gradually approached them over the first few months post-implant. Taken together, the present results suggest that improved labeling is the main mechanism explaining post-implant increases in vowel identification.

Introduction

Cochlear implants (CI's) have been shown to be a safe and effective treatment for profound sensorineural deafness in postlingually deafened adults and in prelingually deaf children. It is not surprising that speech perception scores of the latter continue to increase many years after implantation, because these children must develop an oral language system and develop speech perception and production skills after receiving a CI. However, the full benefit that CI's provide to postlingually deafened adults is not instantaneous either. Normally, asymptotic speech perception performance (i.e., identification of vowels, consonants, words, sentences or running speech) is not apparent for at least several months or weeks after implantation. What are the reasons for this delay, and what is the nature of the underlying process that is indexed by increases in speech perception scores?

There are at least two kinds of reasons for this process. First, CI users may improve their *discrimination* skills over time. To understand speech, CI users must be able to discriminate sounds along relevant acoustic dimensions. For example, vowels are reasonably well characterized by the first two or three formant frequencies (Peterson & Barney, 1952; Hillenbrand, Getty, Clark, & Wheeler, 1995). For CI users, different formant frequencies result in stimulation of different electrodes and thus different cochlear locations. Consequently, the ability to discriminate stimulation delivered to different electrodes may be an important prerequisite to identifying vowels. Unfortunately, the number of intracochlear electrodes is relatively small (22 at most) compared to the number of discriminable steps in a normal hearing cochlea, a factor that limits the spectral information that CI users can receive. Further limitations are due to the fact that even with this relatively small number of electrodes, the ability of CI users to discriminate stimulation delivered to different electrodes is less than perfect. Given that electrode discrimination ability is likely a limiting factor in CI users' reception of spectral information, longitudinal improvement in this ability may underlie the observed improvement in, for example, vowel perception.

Second, CI users may improve their vowel identification over time due to improved *labeling* of speech sounds. The percepts elicited from a cochlear implant are different from normal acoustic hearing, and listeners may be initially unable to label the sounds they hear. In other words, different speech sounds may be distinct from each other (i.e., discrimination may not be the limiting factor in the identification of these sounds), but they may not sound the way listeners *expect* them to sound, leading to identification errors. The possibility of labeling errors is quite plausible, given the fact that cochlear implants do not stimulate the entire neural population of the cochlea but only the most basal 25 mm at best because the electrode array cannot be inserted completely into the cochlea. Therefore, cochlear implants may

stimulate cochlear locations that are more basal and thus elicit higher pitched percepts than normal acoustic stimuli. For example, when the input speech signal has a low frequency peak (e.g. 300 Hz), the most apical electrode is stimulated. The neurons stimulated in response to this signal may have characteristic frequencies of 1000 Hz or even higher. This pronounced modification of the peripheral frequency map might lead to errors in identifying speech sounds, unless the auditory nervous system of CI users is adaptable enough to successfully "re-map" the place frequency code in the cochlea. This adaptation, however, may require weeks or months and may underlie the improvement in speech perception observed after cochlear implantation in postlingually deafened adults. It is important to remember that an individual listener may suffer both labeling and discrimination limitations to speech perception simultaneously.

The goal of this study was to assess the contribution of discrimination and labeling to vowel identification by Spanish-speaking cochlear implant users. To this end, vowel identification and electrode discrimination were assessed longitudinally over several months, starting the day of initial stimulation. The Multidimensional Phoneme Identification (MPI) model (Svirsky, Meyer, Kaiser, Basalo, Silveira, Suarez, Lai, & Simmons, 1999; Svirsky, 2000) was used to obtain "ceiling estimates" of vowel identification, representing the best performance a listener could achieve, given his electrode discrimination skills.

Materials and Methods

The subjects were seven cochlear implant users, all of them native Spanish speakers, implanted by Dr. Suárez in Montevideo, Uruguay. They all had postlingual, profound-to-total deafness. Six of them used the Nucleus-22 device while the remaining one used the Med-El Combi-40 device.

Vowel identification and electrode discrimination were measured up to 6.5 to 32 months after initial stimulation. The first testing session for 6 of the 7 subjects took place immediately after initial stimulation, before they heard any other speech sounds through the implant. Vowel identification testing was repeated at the end of the initial stimulation session for two of the seven subjects. The other subject (subject 4, who was the Med-El user) was tested for the first time 10 days after initial stimulation. Vowel identification testing was done by presenting each one of the five Spanish vowels ten times, in random order, in j-vowel-d context (where "j" indicates the Spanish velar fricative). Each one of the ten repetitions of each vowel was separately uttered, recorded and presented. The speaker was either a male or a female speaker of Uruguayan Spanish, whose utterances were recorded on a personal computer. Results were scored as the percentage of correct responses for the 50 stimuli. In one case (Subject 2, session done three months post-implant), both the male and the female vowels were presented, and the analyses described below were conducted separately for the male speaker and the female speaker datasets.

Electrode discrimination (or, equivalently, discrimination of place of stimulation in the cochlea) was assessed with a pitch-ranking task. Two adjacent electrodes were stimulated in sequence, for 500 ms each with a 500 ms pause in between, and the subject had to say which one of the two sounds was higher pitched. All sounds were presented at maximum comfortable level, and these levels were balanced for equal loudness prior to presentation. Due to the limited available testing time, only nine pairs of adjacent electrodes were tested in Nucleus-22 users: these were pairs 1-2, 2-3 and 3-4 (at the basal end of the electrode array); 9-10, 10-11 and 11-12 (in the middle of the array); and 17-18, 18-19 and 19-20 (at the apical end of the array). Each pair of electrodes was stimulated eight times in random order, with the more basal electrode being stimulated first about 50% of the time. The Med-El user had 11 active electrodes and in her case all pairs of adjacent electrodes were tested. Average d', an index of the ability to correctly pitch rank electrodes was calculated for each subject based on a procedure similar to that

described by Levitt (1972). A d' of 0 indicates no discrimination, d'=1 indicates minimum discrimination ability, and a d' that is greater than 3 indicates near perfect discrimination.

A previous study (Svirsky et al., 1999) provided evidence that the relevant perceptual dimensions for vowel perception in Spanish were A2 (i.e., the amount of energy delivered to electrodes encoding frequencies in the second formant region), F1 and F2 (i.e., the centers of gravity for stimulation pulses delivered to electrodes encoding the F1 and F2 formant frequency ranges, respectively, weighted by pulse amplitude). The MPI model (for a full description see Svirsky, 2000) was used to obtain ceiling estimates of vowel identification, assuming that F1, F2 and A2 were indeed the perceptual dimensions used by all subjects. The MPI model generates a predicted confusion matrix based on a listener's just noticeable differences (JND's) along the relevant perceptual dimensions. In this study, the JND's for the F1 and F2 dimensions were estimated as the inverse of the d' values that were derived from the pitch ranking experiment. Because JND for the A2 dimension was not measured in this study, predictions of vowel perception scores were obtained for each individual using a wide range of JND values for A2. Normally, the MPI model can be used to estimate a listener's maximum possible vowel (or consonant) identification performance, given his JND's for all the relevant perceptual dimensions. Instead of calculating a single ceiling estimate, in this study we obtained a range of values where the actual maximum was expected to fall. These ranges sometimes changed over time, as the listener's pitch ranking skills (and therefore the estimates for F1 and F2 JND's) increased or decreased. The ceiling ranges were compared to each listener's actual vowel identification scores. When the vowel identification scores fall within the ceiling range (which is partly determined by the listener's pitch ranking skills), this suggests that the listener may be labeling vowels in an optimal fashion, and vowel identification is only limited by his ability to discriminate these speech sounds. Conversely, when vowel identification scores are substantially below the ceiling range predicted by the MPI model, this represents strong evidence that the listener is not using all the acoustic information that is available to him, possibly due to limitations in vowel labeling.

Results

Table I shows electrode discrimination scores (d') for all subjects as a function of time. The first four subjects did not show any systematic improvement after the first testing session. In fact, Subject 1 showed a substantial decrease in d' between the day of initial stimulation and a second testing session seven months later. Subjects 5, 6, and 7, on the other hand, showed better discrimination in later testing sessions than they did on the day of initial stimulation.

Although a detailed description exceeds the scope of this manuscript, it should be noted that the MPI model provided good fits to the subjects' confusion matrices that were obtained at least a few months after initial stimulation. In other words, the model was able to predict which vowel pairs would be confused by the subjects as well as which vowel pairs would not be confused. In addition, the model predicted that, for given levels of frequency and amplitude discrimination, vowel identification scores would be higher when the female speaker was used than when the male speaker was used. This is precisely what happened during the three-month post-implant session for Subject 2, the only session where both the male and the female vowels were administered (see the top right panel of Fig. 1). These results provide some validation for the choice of dimensions employed in this study and for the MPI model itself, validating the use of ceiling estimates obtained with it.

Subject	Time after initial stimulation (months)	Average electrode discrimination (d')
1	0	2.3
	7	0.76
2	0	1.2
	3	0.94
	20	0.86
3	0	0.69
	0.1	1.02
	11	0.35
	25	0.64
4	0.3	3.21
	15	2.30
5	0	0.65
	3	2.17
	15	1.94
6	0	0.56
	4	1.75
7	0	0.56
	12	1.18

Table 1. Average electrode discrimination as a function of time. For subjects 1–4, discrimination did not improve after the day of initial stimulation (and it even decreased in some cases), but subjects 5, 6 and 7 did show better electrode discrimination when they were tested a few months after initial stimulation.

Figure 1 shows the vowel identification scores for subjects 1-4, as a function of time. Vertical bars indicate the ceiling ranges for subjects 1-3, as predicted by the MPI model. Predictions were not obtained for subject 4, who used a different device than the other subjects. The day of initial stimulation, subjects 1-3 obtained scores that were substantially lower than the ceiling ranges, but within 2-5 months they all reached vowel scores that were within, or quite close to the ceiling ranges.

Figure 2 shows similar data for subjects 5-7. Because in these cases d' increased with time, the ceiling ranges increased accordingly as a function of time. However, subjects 5 and 6 showed the same pattern as subjects 1-3, failing to reach their ceiling ranges at initial stimulation but reaching those ranges 3-6 months later. Subject 7 was the only one whose initial vowel scores were within his ceiling range at initial stimulation.

Subjects 1 and 5, who were tested at the beginning and at the end of the initial stimulation session, showed a marked increase in vowel identification during the session. Both scored only 28% correct immediately after the implant was turned on, and they increased their scores to 58% (subject 1) and 48% (subject 5) by the end of the two-hour session.

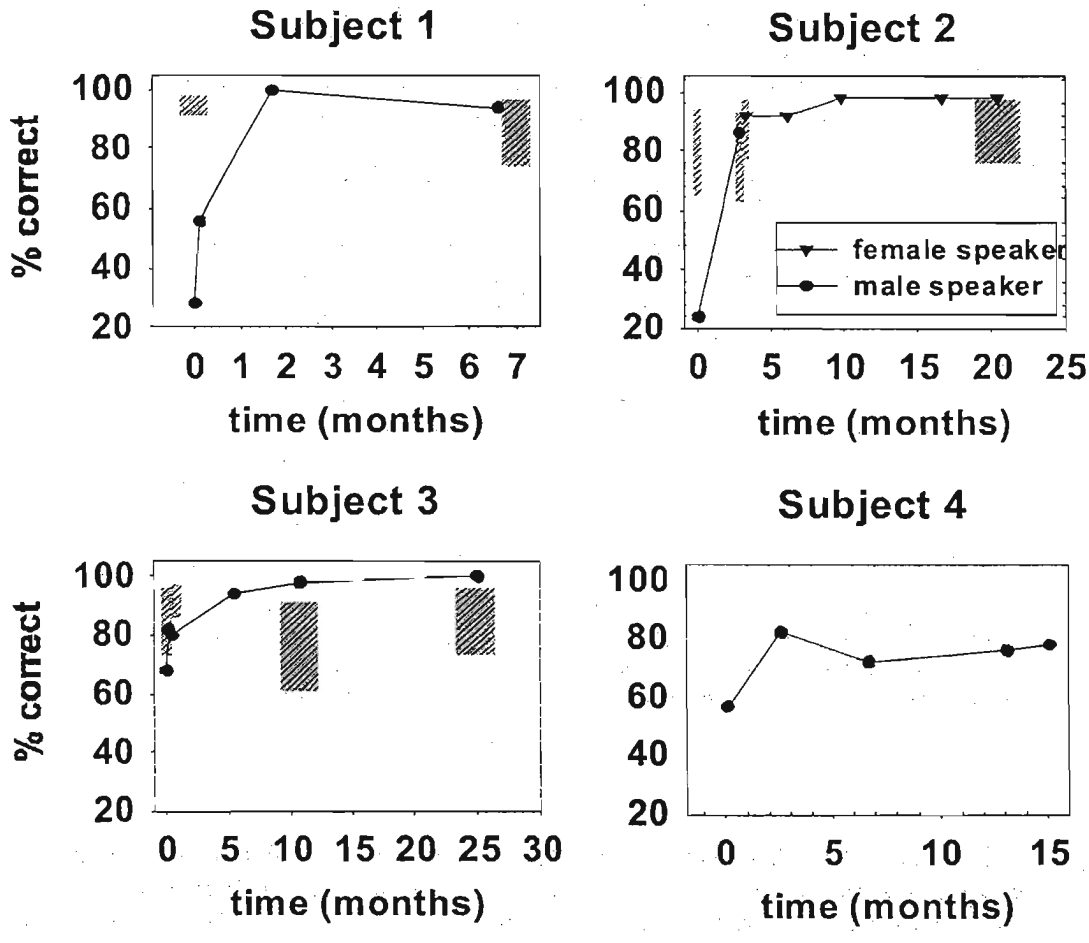


Figure 1. Vowel identification as a function of time for subjects 1-4, whose electrode discrimination skills did not increase post-implant. The vertical bars represent ceiling estimates of vowel identification performance for each individual, as estimated with the MPI model based on the listener's electrode discrimination. Estimates were not obtained for subject 4, who used a different device than the others. Vowel identification by subjects 1, 2 and 3 reaches ceiling estimates by 2-5 months post-implant.

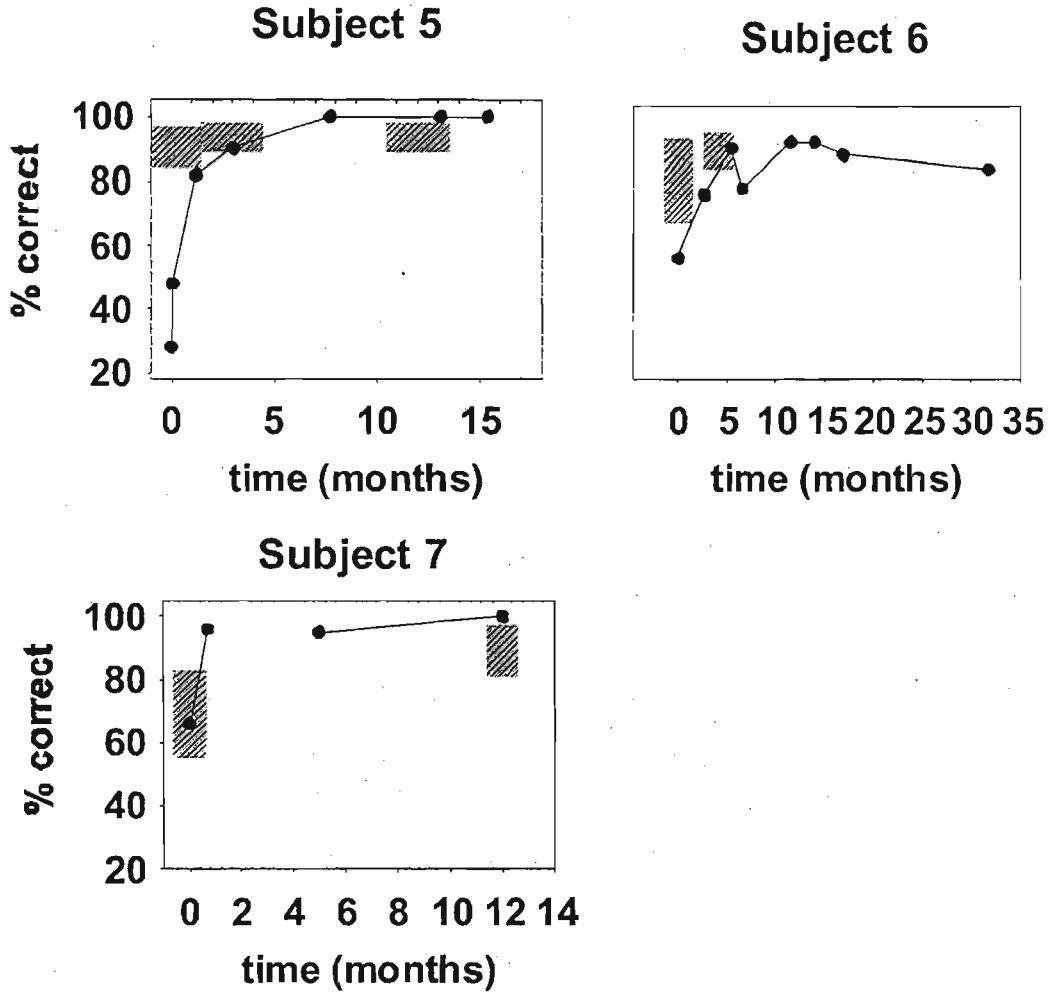


Figure 2. Vowel identification as a function of time for subjects 5-7, whose electrode discrimination skills did increase post-implant. Vowel identification by subjects 5 and 6 reaches ceiling estimates by 3-5 months post-implant, while subject 7 was the only one whose vowel identification scores were within the ceiling estimate range the day of initial stimulation.

Discussion

All subjects showed the expected pattern of improvement in vowel scores over the first few months after initial stimulation. In four of the seven subjects, this change was not accompanied by an increase in electrode discrimination. Taken together, these data suggest that improvement in labeling of vowel sounds was the main mechanism underlying longitudinal increases in vowel identification for these four subjects. The other three subjects presented a more mixed picture, with simultaneous increases in electrode discrimination and in vowel identification. However, predictions obtained with the MPI model suggest that the improvement in electrode discrimination was insufficient to explain the pronounced increases in vowel identification observed in two of these three subjects: in their case, the improvement in vowel identification may have been due to parallel increases in discrimination and labeling skills. These results are consistent with those of Harnsberger et al. (in press), who asked cochlear implant users to select the regions of the F1-F2 plane that sounded like a given vowel. One goal of their study was to determine whether the basalward frequency shift imposed by cochlear implants results in systematic response biases in this task. No such bias was found, indicating that their subjects (who, unlike the subjects in the present study, had used their cochlear implants for at least one year) had learned to label vowels correctly, and their vowel perception was limited mostly by their ability to discriminate formant frequencies. An interesting direction for future research may be to use the method-of-adjustment task employed by Harnsberger et al. longitudinally, starting immediately after implantation, to directly measure changes in the CI users' ability to label vowels correctly. Additionally, it would be informative to image the cochleas of these subjects in order to obtain estimates of electrode location and cochlear length, which in turn would help refine estimates of the amount of basalward shift in these cochlear implant users.

The MPI model was used in this study to tease apart the effect of improved labeling and improved discrimination on speech perception by CI users. This kind of information may be clinically useful because it may suggest areas to be stressed during auditory rehabilitation following cochlear implantation. Subjects whose scores are well below their ceiling estimates may especially benefit from training designed to help them label speech sounds. Conversely, subjects who do reach their ceiling may benefit from training that helps them discriminate better along the acoustic and perceptual dimensions known to be important in speech perception.

References

Harnsberger, J.D., Svirsky, M.A., Kaiser, A.R., Pisoni, D.B., Wright, R., & Meyer, T.A. (in press). Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *Journal of the Acoustical Society of America.*

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English Vowels. *Journal of the Acoustical Society of America,* 97, 3099-3111.

Levitt, H. (1972). Decision Theory, Signal-Detection Theory, and Psychophysics. In E.E. David & P. B. Denes (Eds.), *Human Communication: A Unified View.* New York: McGraw-Hill. Pp. 114-174.

Peterson, G.E. & Barney, H.L. (1952). Control Methods Used in the Study of Vowels. *Journal of the Acoustical Society of America,* 24, 175-184.

Svirsky, M.A. (2000). Mathematical Modeling of Vowel Perception by Users of Analog Multichannel Cochlear Implants: Temporal and Channel-Amplitude Cues. *Journal of the Acoustical Society of America,* 107, 1521-1529.

Svirsky, M.A., Meyer, T.A., Kaiser, A.R., Basalo, S., Silveira, A., Suarez, H., Lai, T.T., Simmons, P.M. (1999). Learning How to Perceive Vowels with a Cochlear Implant: The Role of Discrimination and Labeling. Presented at The Association for Research in Otolaryngology, Twenty-Second Midwinter Research Meeting, Feb. 1999.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Early Word Learning Skills of Hearing-Impaired Children Who Use Cochlear Implants: Development of Procedures and Some Preliminary Findings[1]

Derek M. Houston,[2] Allyson K. Carter, Elizabeth A. Ying,[2] Karen Iler Kirk,[2] and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by NIH-NIDCD Research Grants DC00111, DC00064 and Training Grant DC-00012 to Indiana University. We thank Caitlin Dillon, Cara Lento, Tara O'Neill, and Miranda Cleary for valuable assistance in testing participants. We would also like to thank Beth Jeglum and the staff at the Indiana-University-Purdue-University-Indianapolis Center for Young Children for helping make arrangements with the parents of our normal-hearing control participants and for providing a place for testing.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana.

Early Word Learning Skills of Hearing-Impaired Children Who Use Cochlear Implants: Development of Procedures and Some Preliminary Findings

Abstract. In recent years, cochlear implant (CI) technology has advanced and can now greatly facilitate the spoken language learning of prelingually deafened children. However, there is a great deal of variability in linguistic outcome measures among pediatric CI recipients. Many factors may contribute to this variability in performance, including age of implantation, amount of speech therapy, cognitive factors (such as memory span), and numerous linguistic factors. An important basic linguistic skill that may play a central role in later language development is the ability to map the sound patterns of spoken words onto their referents. This report summarizes and describes the development of a procedure and some preliminary findings of 2- to 5-year-old CI users' and normal-hearing controls' word learning abilities. Each child was presented with either four (2- and 3-year-olds) or eight (4- and 5-year-olds) Beanie Babies™ and labels for their names using interactive play scenarios. Across multiple sessions, the participants were tested for receptive and expressive knowledge of the learned names. This report also describes current plans to test more children and to compare the results of this test with subsequent outcome measures in order to ascertain whether there are any correlations between early object labeling abilities and later language skills.

Introduction

Cochlear implants (CIs) provide profoundly deaf children with the possibility of learning spoken language by allowing them to receive auditory input. However, CIs provide an impoverished signal, and children who receive them have had some amount of prior auditory deprivation. These two factors and others may contribute to the finding that some profoundly deaf children do not succeed in learning spoken language. One of the most interesting and challenging discoveries about pediatric CI users is that there are enormous individual differences in language skills after implantation. Recently, Pisoni, Svirsky, Kirk, and Miyamoto (1997) showed that for individual children with CIs, performance on speech perception, speech production, and language tests were highly correlated with each other. They postulated that the common variance might be attributed to cognitive processing skills, including the phonological encoding, storage, and retrieval of spoken words. In this project, we investigate the early word learning abilities of hearing-impaired children with cochlear implants.

Normal-hearing children begin producing words at approximately 12 months of age. By 18 months, most infants can produce over 50 words and they seem to learn several words each day (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). Most research on word learning focuses on how children learn to correctly associate the sound patterns of words to their referents (e.g., Clark, 1973, 1983; Markman, 1991; Nelson, 1988). Recently, some work has explored children's ability to encode the sound patterns of words. Jusczyk and colleagues have shown that by 8 months, infants can encode the sound pattern of words into memory (Houston & Jusczyk, submitted; Jusczyk & Aslin, 1995; Jusczyk & Hohne, 1997). The ability to encode phonological information into memory enables children to form word representations. Gathercole and Baddeley (1989, 1990) have found a strong relationship between phonological working memory span and vocabulary size. Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991) found a

significant correlation between how often parents use words and their children's acquisition of those words, suggesting that frequency of exposure affects how quickly children learn words. Taken together, all these findings suggest that early word learning may be an important subcomponent or skill that affects later language development.

Children with hearing impairment may be at a disadvantage for encoding phonological information because, to varying degrees, they are unable to discriminate the fine acoustic-phonetic details of speech in their surrounding environment. There is some evidence to suggest that any degree of hearing loss may lead to problems in phonological processing and word learning. For example, in a study of hearing-impaired children who used hearing aids, Gilbertson and Kamhi (1995) assessed children's ability to encode phonological information and to learn words when wearing their hearing aids. The investigators found that hearing-impaired children's unaided level of hearing loss (ranging from mild to moderate) did not correlate significantly with word learning abilities but the ability to encode phonological information did correlate. One group of the hearing-impaired children, when using sensory aids, performed as well as normal-hearing children on language learning tasks, whereas another group had much more difficulty. However, whether any particular hearing-impaired child fell into the normally performing group or the group that had more difficulty did not depend on his/her unaided level of hearing loss. The authors concluded that even a mild hearing loss was a significant risk factor for language impairments characteristic of children with Specific Language Impairment (SLI)[3].

It is possible that difficulty in early word learning for some children may be due to difficulty with a specific aspect or stage of word learning. Susan Carey and colleagues have described the word learning process in terms of two stages (Carey, 1978; Carey & Bartlett, 1978). The first stage, "fast mapping", refers to the initial encoding of the sound pattern of the words and some basic understanding of the meaning. The second stage involves developing a fuller understanding of words by hearing them in several different contexts so that hypotheses about their meaning can be tested. Carey and Bartlett (1978) showed that after a single presentation of a word, preschool children already started to form some basic hypotheses of the meaning of the word when the word was used to name a color term. In another study, Heibeck and Markman (1987) have shown that children as young as two years show fast mapping of shape and texture terms as well as color terms. Hence, while a complete understanding of words may involve a complex and lengthy process, the basic process may begin with an initial "fast mapping" stage of word learning that is immediate and crucial in establishing a solid foundation for later lexical development.

The fast mapping stage of early word learning requires children to encode the phonological information of words very rapidly. Children who have difficulty with phonological encoding may show great difficulty learning words. There may be a high incidence of poor phonological encoding ability among CI users for two reasons. First of all, the auditory information provided by a CI is impoverished when compared to normal hearing. It is possible that this impoverished signal may be a limiting factor in encoding the sound pattern of words. Second, hearing impairment may be a risk factor for SLI (Gilbertson & Kamhi, 1995), and one of the factors of SLI is difficulty with phonological processing skills (Leonard, 1998). Thus, it is possible that children with SLI have difficulty with fast mapping. Support for this possibility has come from a series of studies by Rice and colleagues who have shown that children with SLI

[3] Specific Language Impairment (SLI) is often operationally defined as the presence of language impairments in the absence of other cognitive and sensory impairments, including hearing loss. However, it is possible children could have language impairments associated with SLI (e.g., mapping sound to meaning) in addition to language impairments specifically caused by hearing loss (e.g., encoding phonological information). In this respect, it is reasonable to discuss the possibility of hearing impaired children also having SLI.

have difficulty with "quick and incidental learning," which is similar to fast mapping (Oetting, Rice, & Swank, 1995; Rice, Buhr, & Nemeth, 1990; Rice, Oetting, Marquis, Bode, & Pae, 1994). In one investigation, Rice et al. (1994) found that children with SLI needed many more exposures to words than normal language learners in order to display even a basic understanding of the words. Moreover, children with SLI were particularly prone to forgetting the meaning of words after a short delay, suggesting that their long-term memory representation for words was impoverished. In sum, hearing-impaired children who use CIs may have some difficulty quickly learning novel words because the speech signal they receive is impoverished. Some of these children might have difficulty due to poorly developed phonological processing skills.

It is possible that the ability to quickly encode the sound patterns of words and some basic aspects of meaning may account for the individual variability observed in the language skills of children who use CIs. In the present investigation, we explore the possibility that children who use CIs will demonstrate a high degree of variability in learning novel words after only a few exposures. If this hypothesis is correct, then we would expect there to be a correlation between performance on an early word learning task and other language outcome measures, such as vocabulary knowledge and language development.

The goal of this project was to develop a procedure that assesses young children's ability to learn words after only a few exposures. This project is part of an ongoing investigation to explore how quickly children with cochlear implants can map sound patterns onto referents and to determine the relationship between measures of early word learning and other outcome measures such as spoken word recognition, speech intelligibility and receptive and expressive language abilities. This report describes the results of a preliminary study that was conducted to develop a procedure that can be used to test preschool-aged children's ability to learn words after a brief exposure period. The first part of the project involved selecting names for the Beanie Baby™ stuffed animals that would be taught to the children. This was done by eliciting names from adult participants. Next, we describe the results of a pilot study with several children with CIs who were given variations on our initial word-learning procedure. The results of the pilot testing with the children helped us modify several aspects of the procedure. Finally, we present preliminary data from twenty-four normal-hearing children and two children with CIs.

Pilot Phase

Selection of Stimulus Materials

The stimulus materials used in all of the experiments consisted of a set of sixteen Beanie Baby™ stuffed animals. Each Beanie Baby™ comes with a name assigned by the manufacturer (Ty Corporation®). We did not use these names because some children might already know the names while others might not and because some of the names were related to physical attributes of the stuffed animals while others were not. We decided to elicit names from adult participants that corresponded to salient physical attributes of the Beanie Babies™. This was done to facilitate the association between the names and the Beanie Babies™. Because many of the Beanie Babies™ have several features that could be considered salient, a pilot experiment was conducted to determine the characteristics of each Beanie Baby™ that were most perceptually salient. The goal of this pilot study was to select labels for each Beanie Baby™ that would be used in the experimental study.

Methods

Participants. The participants were 37 IU undergraduates, with no reported history of speech or hearing disorders. Thirty-five of the subjects were native English speakers. All subjects were recruited from the Indiana University community and all received partial credit towards an Introductory Psychology class for their participation. The mean age of the participants was 19.9 years (SD = 1.3).

Materials. Sixteen Beanie Baby™ stuffed animals were used as stimuli. In order to control for possible familiarity effects with the original names, we developed a new set of names for the Beanie Babies™. We decided as a first pass to create the new names in such a way as to give a semantic bootstrap to enable word learning, for example, by using a distinguishing physical characteristic of the animal. The Beanie Babies™ were therefore selected from a larger set of Beanie Babies™ on the basis of whether they had distinguishing characteristics that could be easily named, such as a very long tail, horns, or a bright color.

Procedure. Subjects were tested in three groups in a small experimental classroom. They were given written instructions, in which they were told that the experimenter would hold up each of the sixteen Beanie Babies™ individually, and they would be asked to invent new names for the Beanie Babies™, as if they were teaching the names to a young child. Subjects were asked to re-name the Beanie Babies™ using names that described some physical attributes of the Beanie Babies™. Subjects were instructed to provide up to three new names for each animal, and to use one-word names only. Subjects were provided with answer sheets on which to write the new names. The Beanie Babies™ were presented individually, one at a time, in a random order to the three groups.

Results

For each Beanie Baby™, the responses were recorded and tallied to calculate the frequency of the names generated. A new Beanie Baby™ attribute name was chosen from the response tallies based on two criteria: (1) that the name was the most frequent response among students, and (2) that it reflected a true physical attribute of the animal. For example, the name *"Red"* was the most frequent response and was also an appropriate name for the red bull because it refers to the color of the bull. In contrast, a non-attribute name, *"Teddy,"* was the most frequent response for the brown bear, but was inappropriate for our purposes. The second most common response was *"Fuzzy,"* which we used as it describes an attribute of the bear. The new attribute name was the most frequent response for seven of the Beanie Babies™ (*"Blue," "Red," "Stripes," "Pink," "Spots," "Ears," "Tail"*), the second most frequent response for four of the animals (*"Wings," "Fuzzy," "Legs," "Cottontail"*), the third most frequent response for five of the animals (*"Horns," "Gray," "Teeth," "Bushy"*), and the fourth most frequent response for *"White."* Table 1 lists each original Beanie Baby™ stuffed animal name and description, its new attribute name derived from this procedure, and the percentage of subjects who used the new attribute name.

Original Beanie Baby™ name	"New" attribute name	Frequency of new name response (%)	Original Beanie Baby™ name	"New" attribute name	Frequency of new name response (%)
Crunch the Shark	*Teeth*	11.8	*Dotty* the Dalmatian	*Spots*	46.6
Rocket the Bird	*Blue*	36.6	*Halo* the Angel Bear	*White*	6.1
Batty the Bat	*Wings*	13.7	*Spunky* the Cocker Spaniel	*Ears*	14.9
Kuku the Bird	*Pink*	29.6	*Nuts* the Squirrel	*Bushy*	11.1
Snort the Bull	*Red*	32.4	*Nibbly* the Bunny	*Cottontail*	14.1
Goatee the Goat	*Horns*	8.2	*Spinner* the Spider	*Legs*	14.9
Buster the Bear	*Fuzzy*	9.8	*Prance* the Tabby Cat	*Stripes*	25.6
Tiptoe the Mouse	*Tail*	18.5	*Spike* the Rhino	*Gray*	7.8

Table 1. Original Beanie Baby™ names, new given attribute names, and the frequency with which each new name was generated.

Procedure Development

Once the new names were chosen, a piloting phase was initiated to develop a procedure for assessing word learning in young children with cochlear implants after a brief exposure period. The initial conception of the experiment was as follows. Children would be taught the new names of the Beanie Babies™. Younger children (2;0 – 3;11) would be taught four names and older children (4;0 – 5;11) would be taught eight names. In order to get a baseline measure of how likely it was that the children would spontaneously label the Beanie Babies™ with the target labels, the experiment started with two pretests. In the first pretest, children were presented with each of the Beanie Babies™ they would be taught and were simply asked to give it any name, using a free response format. The second pretest used a forced-choice procedure. The Beanie Babies™ were placed in a row in front of the child, and the child was asked to select the one that might have the name that the experimenter presented to them. For example, the experimenter might say, "Which one do you think is named *Fuzzy*?" The experimenter did this for each of the Beanie Babies™.

Following the pretests, each child was given a sequence of training phases in which they were taught the names of the Beanie Babies™, one at a time, using play scenarios. The experimenter provided the name of each Beanie Baby™ exactly three times. Toys were used to give each Beanie Baby™ some sort of memorable personality. For example, in one scenario, the experimenter would say, "This is *Fuzzy*. *Fuzzy* likes to eat grapes. Can you give the grapes to *Fuzzy*?" After exposure and training with each Beanie Baby™, the children were given tests to assess whether or not they learned to associate the names with the Beanie Babies™. The first test used a forced-choice procedure, exactly like the second pretest. The second test used a cued-recall procedure. The cued-recall test required an expressive response from the child. In this procedure, each Beanie Baby™ was presented one at a time to the child as a cue, and the child was asked to recall its name from memory.

The initial procedure underwent several stages of development during the piloting phase. Six children who use cochlear implants participated in the piloting phase: SHM (4;1), SHZ (6;2), SHS (2;5), SNW (3;2), SMH (5;11), and SOC (4;1). Here, we will summarize our major observations during the piloting phase and describe how these pilot results shaped the final design of the experimental procedures.

• **Children often perseverate on the names they initially choose for the Beanie Babies™.** The initial conception of the procedure involved two pretests. During the procedures, we discovered that the children who were given the pretests (SHM and SHZ) were very resistant to learning new names for the Beanie Babies™. Instead, they tended to perseverate on the names that they initially selected. Hence, the pretests were dropped from the procedure.

• **Children who use cochlear implants need several exposures to words.** In the first stage of the pilot, children received only three exposures to each name before they were tested. With only three exposures per item, two of the participants in the pilot study did not perform above chance. Given the poor performance of the pilot subjects tested under this condition, the number of exposures was increased from three to eight for each name.

• **Imitation is important for word learning.** Another factor that seemed to contribute to the poor performance in the early stages of our pilot testing was that no measure was taken to ensure that the children actually encoded the names they were being taught. To make sure that the children encoded the sound pattern of the words, we asked the children to repeat the names that we produced. It is possible that the act of producing the words helps with children's memory for words because there may be a strong developmental interaction between perception and production (e.g., Vihman, 1993). There are recent data supporting the importance of immediate memory and imitation in novel word learning (Gupta & MacWhinney, 1995).

• **Children often show a preference for new Beanie Babies™.** During one phase of the pilot testing we decided to try teaching the names of the Beanie Babies™ to the children one at a time. Thus, they were first presented with one animal and then were given the forced-choice and cued-recall tests for that animal. If they were correct, they were taught the name of an additional Beanie Baby™. If they were incorrect, they were re-taught the original name. Each time they were correct on both the forced-choice and cued-recall tests, the set size increased by one. The set size increased until the child could no longer respond correctly on three consecutive trials. In carrying out this procedure with SHS, SNW, and SMH, we discovered that as the set size increased, the children showed a novelty preference for the most recent Beanie Babies™ presented. As a result, we decided that during each session, the child would be presented with all of the Beanie Babies™ (four or eight), one at a time, and then tested on all of them.

• **A minimum of one year of cochlear implant experience is necessary.** One of our initial criteria for inclusion in the study was that the child must have had at least six months of implant use. The participant SOC, who had exactly six months of experience, clearly did not have sufficient auditory skills to participate in the study and carry out the tasks. Hence, we increased the criterion to one year of implant use.

Experiment

The piloting phase ended when a procedure was settled on that was simple enough for the pilot participants to complete but did not yield ceiling performance. Children were taught and tested on two sets of Beanie Babies during one session. Their long-term memory of the names

was subsequently assessed in a second session by re-testing them at least two hours later. The final design is described here and some preliminary results from two children with CIs and twenty-four normal-hearing children are reported below.

Methods

Participants. Two groups of children participated in this study. One group of four children was recruited from the population of children with cochlear implants who are routinely followed as part of the ongoing longitudinal studies at Indiana University. The criteria for inclusion was that they were between the ages of 2;0 and 5;11, use oral communication, and had at least one year of cochlear implant experience. Two children who use CIs (ages: 3;2 and 3;10) completed the experiment, but the other two children (ages: 2;4 and 4;2) were unable to complete the experiment due to failure to give any responses and are not included in the Results section. Twenty-four age-matched normal-hearing controls were recruited from the Bloomington, Indiana area and the Center for Young Children daycare center on the Indiana-University-Purdue-University-Indianapolis campus. All 24 completed the initial experiment. Seven of the normal-hearing children and the two children with CIs who completed the initial testing participated in the long-term memory test.

Materials. The stimulus materials consisted of 16 Beanie Babies™ that were assigned names by normal hearing college students (see Stimuli Selection above). Each name corresponds to a salient physical attribute (e.g., "*Red*" is a red bull). The Beanie Babies™ were grouped into sets of four as shown in Table 2. The Beanie Babies™ were selected so that most of the attribute names could describe at least two Beanie Babies™ in the group. For example, "*Wings*", "*Pink*", and "*Blue*" all have wings. This was done so that the children would not be able to completely rely on identifying the attributes in the tasks. For example, when they were asked to identify "*Wings*", three of the four Beanie Babies™ had wings.

Set	Beanie Baby™ Attribute Name	Description
A	Teeth	Shark
	Blue	Blue jay
	Wings	Bat
	Pink	Cockatoo
B	Red	Bull
	Horns	Goat
	Fuzzy	Brown bear
	Tail	Rat
C	Spots	Dalmatian
	White	White bear with halo
	Ears	Cocker spaniel
	Bushy	Squirrel
D	Cotton tail	Rabbit
	Legs	Spider
	Stripes	Cat with stripes
	Gray	Rhino

Table 2. Word set stimuli.

Procedure. Children were taught a set of Beanie Babies™ (Training Phase 1) and then given forced-choice and cued-recall tests for that set (Testing Phase 1). Children were then taught another set of Beanie Babies™ (Training Phase 2) and subsequently given the same tests with the second set (Testing Phase 2). Finally, after at least a two-hour delay, children were given the same tests using the first set of Beanie Babies™ and then using the second set (Long-Term Memory Test).

Training Phase 1. *Children younger than four years.* Each child is exposed to four Beanie Babies™. Before the experiment, the exact order of Beanie Baby™ presentation was randomized and recorded on a form that was then followed during the experiment. One experimenter (Experimenter 1) interacted with the child while a second experimenter (Experimenter 2) assisted Experimenter 1 in following the correct order. Experimenter 2 also recorded the children's responses and the number of times Experimenter 1 produced the name of each Beanie Baby™.

Experimenter 1 presented each Beanie Baby™, one at a time, to the child. A different toy prop was used to create a different play scenario with each Beanie Baby™ in order to keep the task interesting. During the play interaction with each Beanie Baby™, Experimenter 1 used the name of the Beanie Baby™ exactly eight times. During the play scenario, Experimenter 1 tried to elicit three productions of the name from the child. Experimenter 2 recorded how many times the child produced each name. Positive feedback was given when the child produced the correct names. See Appendix for a sample scenario.

Children between four years and six years. The training phase was the same with older children as with the younger children, except that eight Beanie Babies™ were taught instead of four.

Testing Phase 1. *Forced-Choice Test.* The Testing Phase consisted of a forced-choice identification task and a cued-recall test given immediately afterwards. For the forced-choice identification test, all of the Beanie Babies™ (four or eight) were placed in a row in front of each child and hidden from view with a piece of cardboard. Then a toy bus or truck was brought out and placed in front of the child. Experimenter 1 then asked the child to "please put {*one of Beanie Babies™*} into the truck {*or bus*}". The child was encouraged to select one of the Beanie Babies™ but was not given any feedback as to whether the response was the correct choice. For example, the experimenter said "thank you," "good job," or clapped when the child made a selection, regardless of whether or not the response was correct. The Beanie Baby™ was then placed back in the row and the next trial was initiated. Each Beanie Baby™ was requested exactly once.

Cued-Recall Test. For the cued-recall task, Experimenter 1 played a "knock knock" game with the child. One Beanie Baby™ was placed behind a toy doorway. Experimenter 1 and/or the child said "knock knock," the door would open and Experimenter 1 would ask the child, "Who's there?" The child was asked to name the Beanie Baby™, up to three times. Experimenter 2 recorded any response. This procedure was repeated for each Beanie Baby™.

Training Phase 2. This phase was the same as Training Phase 1 except that a different group of four or eight Beanie Babies™ was presented to the child.

Testing Phase 2. This phase was the same as Testing Phase 1 except that the new set of Beanie Babies™ was used.

Long-Term Memory Test. On the same day of testing, but at least two hours after the completion of Testing Phase 2, the child was tested a second time, in order to assess long-term memory for the names. During the long-term memory test, Testing Phase 1 and Testing Phase 2 were repeated again without any retraining or feedback.

Results

The mean accuracy scores for all of the tests are summarized in Table 3 for normal-hearing children, and in Table 4 for hearing-impaired children with CIs. The preliminary data revealed that the normal-hearing children had very high scores for the forced-choice test in both the Immediate and Delay conditions. So far, the children who use CIs have performed comparably on the immediate forced-choice task. However, their performance on the cued-recall test, and both tasks after a delay, were very low.

Immediate Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (12)	0.95	0.08	0.92	0.17
> 4 yrs (12)	0.95	0.10	0.97	0.08

Delay Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (3)	0.88	0.22	0.75	0.25
> 4 yrs (4)	0.94	0.13	0.89	0.14

Table 3. Mean accuracy response for normal-hearing children (N=24).

Immediate Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (2)	0.63	0.18	0.25	0.18

Delay Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (2)	0.13	0	0.19	0.27

Table 4. Mean accuracy response for hearing-impaired children who use cochlear implants (N=2).

Discussion

The procedures that were developed in this project will allow us to assess the word-learning skills of children, which will be valuable in tracking the language development of children who use CIs. The results thus far are very preliminary because only a small number of children who use CIs have been tested. We will test at least 12 children who use CIs from each of the two age groups before analyzing the data and comparing it to the results from the normal-hearing children. Another step in the project is to analyze the results from this test and compare them to the results obtained on several outcome measures. The children who use CIs are routinely given a battery of speech perception, word recognition and language tests up to several years after they receive their CIs. One of our goals is to assess how the variability of children's performance in learning novel words in these tasks is related to the variability of language outcome measures. Measures of early word learning and "fast mapping" in this clinical population may be important new predictors of language development and other language-based outcome measures.

A future direction for this project is to manipulate the phonological properties of the names of the Beanie Babies™. Currently, we are using real names that correspond to salient visual attributes in order to make the learning task as easy as possible. Once these procedures are validated, subsequent experiments will use Beanie Babies™ with nonword names, which will vary in terms of phonological difficulty (e.g., phonotactic probabilities, syllable number or stress). These other projects should provide valuable new information about the ability of children who use CIs to encode phonological information in tasks that require novel word learning skills, imitation, and long-term retention.

References

Carey, S. (1978). The child as word learner. In M. Halle & J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293). Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*. Stanford University.

Clark, E.V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65-110). New York: Academic Press.

Clark, E.V. (1983). Meanings and concepts. In J.H. Flavell & E.M. Markman (Eds.), *Cognitive Development* (Vol. III, pp. 787-840). New York: Wiley.

Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (Serial number 242).

Gathercole, S., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336-360.

Gathercole, S.E., & Baddeley, A.D. (1989). Development of vocabulary in children and short-term phonological memory. *Journal of Memory and Language, 28*, 200-213.

Gilbertson, M., & Kamhi, A.G. (1995). Novel word learning in children with hearing impairment. *Journal of Speech and Hearing Research, 38*, 630-642.

Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language, 34*, 63-88.

Heibeck, T.H., & Markman, E.M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*, 1021-1034.

Houston, D.M., & Jusczyk, P.W. (submitted). Infants' long-term memory for words and voices. manuscript submitted to *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236-248.

Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1-23.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science, 277*, 1984-1986.

Leonard, L.B. (1998). *Children with specific language impairment.* Cambridge, MA: MIT Press.

Markman, E.M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought* (pp. 72-106). Cambridge: Cambridge University Press.

Nelson, K. (1988). Constraints on word learning? *Cognitive Development, 3*, 221-246.

Oetting, J.B., Rice, M.L., & Swank, L.K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research, 38*, 434-445.

Pisoni, D.B., Svirsky, M., Kirk, K.I., & Miyamoto, R.T. (1997). Looking at the "Stars": A first report on the intercorrelations among measures of speech perception, intelligibility and language development in pediatric cochlear implant users. *Progress Report on Spoken Language Processing #21*, Indiana University, Department of Psychology, Bloomington, IN.

Rice, M.L., Buhr, J., & Nemeth, M. (1990). Fast mapping word-learning abilities of language delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*, 33-42.

Rice, M.L., Oetting, J.B., Marquis, J., Bode, J. & Pae, S. (1994). Frequency of input effects on SLI children's word comprehension. *Journal of Speech & Hearing Research, 37*, 106-122.

Vihman, M.M. (1993). The construction of a phonological system. In B. de Boysson-Bardies & S. de Schonen & P. Jusczyk & P. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face perception in the first year of life* (pp. 411-419). Dordrecht: Kluwer.

Appendix

Sample Scenario:

This is *Name.*
Can you say hi to him?
Say "Hi *Name!*"
Now your turn {child says "hi *Name*"}
Name likes to climb the tree.
Can you put him on the tree? {child interacts with BB}
Look – *Name* is on the tree.
Tell him to get down. {child says, "get down *Name*"}
Good. Now, *Name* has to go bye bye.
Say, bye bye *Name.* {child repeats "bye bye *Name*"}

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Early Word Learning Skills of Hearing-Impaired Children Who Use Cochlear Implants: Development of Procedures and Some Preliminary Findings[1]

Derek M. Houston,[2] Allyson K. Carter, Elizabeth A. Ying,[2] Karen Iler Kirk,[2] and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by NIH-NIDCD Research Grants DC00111, DC00064 and Training Grant DC-00012 to Indiana University. We thank Caitlin Dillon, Cara Lento, Tara O'Neill, and Miranda Cleary for valuable assistance in testing participants. We would also like to thank Beth Jeglum and the staff at the Indiana-University-Purdue-University-Indianapolis Center for Young Children for helping make arrangements with the parents of our normal-hearing control participants and for providing a place for testing.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana.

Early Word Learning Skills of Hearing-Impaired Children Who Use Cochlear Implants: Development of Procedures and Some Preliminary Findings

Abstract. In recent years, cochlear implant (CI) technology has advanced and can now greatly facilitate the spoken language learning of prelingually deafened children. However, there is a great deal of variability in linguistic outcome measures among pediatric CI recipients. Many factors may contribute to this variability in performance, including age of implantation, amount of speech therapy, cognitive factors (such as memory span), and numerous linguistic factors. An important basic linguistic skill that may play a central role in later language development is the ability to map the sound patterns of spoken words onto their referents. This report summarizes and describes the development of a procedure and some preliminary findings of 2- to 5-year-old CI users' and normal-hearing controls' word learning abilities. Each child was presented with either four (2- and 3-year-olds) or eight (4- and 5-year-olds) Beanie Babies™ and labels for their names using interactive play scenarios. Across multiple sessions, the participants were tested for receptive and expressive knowledge of the learned names. This report also describes current plans to test more children and to compare the results of this test with subsequent outcome measures in order to ascertain whether there are any correlations between early object labeling abilities and later language skills.

Introduction

Cochlear implants (CIs) provide profoundly deaf children with the possibility of learning spoken language by allowing them to receive auditory input. However, CIs provide an impoverished signal, and children who receive them have had some amount of prior auditory deprivation. These two factors and others may contribute to the finding that some profoundly deaf children do not succeed in learning spoken language. One of the most interesting and challenging discoveries about pediatric CI users is that there are enormous individual differences in language skills after implantation. Recently, Pisoni, Svirsky, Kirk, and Miyamoto (1997) showed that for individual children with CIs, performance on speech perception, speech production, and language tests were highly correlated with each other. They postulated that the common variance might be attributed to cognitive processing skills, including the phonological encoding, storage, and retrieval of spoken words. In this project, we investigate the early word learning abilities of hearing-impaired children with cochlear implants.

Normal-hearing children begin producing words at approximately 12 months of age. By 18 months, most infants can produce over 50 words and they seem to learn several words each day (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). Most research on word learning focuses on how children learn to correctly associate the sound patterns of words to their referents (e.g., Clark, 1973, 1983; Markman, 1991; Nelson, 1988). Recently, some work has explored children's ability to encode the sound patterns of words. Jusczyk and colleagues have shown that by 8 months, infants can encode the sound pattern of words into memory (Houston & Jusczyk, submitted; Jusczyk & Aslin, 1995; Jusczyk & Hohne, 1997). The ability to encode phonological information into memory enables children to form word representations. Gathercole and Baddeley (1989, 1990) have found a strong relationship between phonological working memory span and vocabulary size. Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991) found a

significant correlation between how often parents use words and their children's acquisition of those words, suggesting that frequency of exposure affects how quickly children learn words. Taken together, all these findings suggest that early word learning may be an important subcomponent or skill that affects later language development.

Children with hearing impairment may be at a disadvantage for encoding phonological information because, to varying degrees, they are unable to discriminate the fine acoustic-phonetic details of speech in their surrounding environment. There is some evidence to suggest that any degree of hearing loss may lead to problems in phonological processing and word learning. For example, in a study of hearing-impaired children who used hearing aids, Gilbertson and Kamhi (1995) assessed children's ability to encode phonological information and to learn words when wearing their hearing aids. The investigators found that hearing-impaired children's unaided level of hearing loss (ranging from mild to moderate) did not correlate significantly with word learning abilities but the ability to encode phonological information did correlate. One group of the hearing-impaired children, when using sensory aids, performed as well as normal-hearing children on language learning tasks, whereas another group had much more difficulty. However, whether any particular hearing-impaired child fell into the normally performing group or the group that had more difficulty did not depend on his/her unaided level of hearing loss. The authors concluded that even a mild hearing loss was a significant risk factor for language impairments characteristic of children with Specific Language Impairment (SLI)[3].

It is possible that difficulty in early word learning for some children may be due to difficulty with a specific aspect or stage of word learning. Susan Carey and colleagues have described the word learning process in terms of two stages (Carey, 1978; Carey & Bartlett, 1978). The first stage, "fast mapping", refers to the initial encoding of the sound pattern of the words and some basic understanding of the meaning. The second stage involves developing a fuller understanding of words by hearing them in several different contexts so that hypotheses about their meaning can be tested. Carey and Bartlett (1978) showed that after a single presentation of a word, preschool children already started to form some basic hypotheses of the meaning of the word when the word was used to name a color term. In another study, Heibeck and Markman (1987) have shown that children as young as two years show fast mapping of shape and texture terms as well as color terms. Hence, while a complete understanding of words may involve a complex and lengthy process, the basic process may begin with an initial "fast mapping" stage of word learning that is immediate and crucial in establishing a solid foundation for later lexical development.

The fast mapping stage of early word learning requires children to encode the phonological information of words very rapidly. Children who have difficulty with phonological encoding may show great difficulty learning words. There may be a high incidence of poor phonological encoding ability among CI users for two reasons. First of all, the auditory information provided by a CI is impoverished when compared to normal hearing. It is possible that this impoverished signal may be a limiting factor in encoding the sound pattern of words. Second, hearing impairment may be a risk factor for SLI (Gilbertson & Kamhi, 1995), and one of the factors of SLI is difficulty with phonological processing skills (Leonard, 1998). Thus, it is possible that children with SLI have difficulty with fast mapping. Support for this possibility has come from a series of studies by Rice and colleagues who have shown that children with SLI

[3] Specific Language Impairment (SLI) is often operationally defined as the presence of language impairments in the absence of other cognitive and sensory impairments, including hearing loss. However, it is possible children could have language impairments associated with SLI (e.g., mapping sound to meaning) in addition to language impairments specifically caused by hearing loss (e.g., encoding phonological information). In this respect, it is reasonable to discuss the possibility of hearing impaired children also having SLI.

have difficulty with "quick and incidental learning," which is similar to fast mapping (Oetting, Rice, & Swank, 1995; Rice, Buhr, & Nemeth, 1990; Rice, Oetting, Marquis, Bode, & Pae, 1994). In one investigation, Rice et al. (1994) found that children with SLI needed many more exposures to words than normal language learners in order to display even a basic understanding of the words. Moreover, children with SLI were particularly prone to forgetting the meaning of words after a short delay, suggesting that their long-term memory representation for words was impoverished. In sum, hearing-impaired children who use CIs may have some difficulty quickly learning novel words because the speech signal they receive is impoverished. Some of these children might have difficulty due to poorly developed phonological processing skills.

It is possible that the ability to quickly encode the sound patterns of words and some basic aspects of meaning may account for the individual variability observed in the language skills of children who use CIs. In the present investigation, we explore the possibility that children who use CIs will demonstrate a high degree of variability in learning novel words after only a few exposures. If this hypothesis is correct, then we would expect there to be a correlation between performance on an early word learning task and other language outcome measures, such as vocabulary knowledge and language development.

The goal of this project was to develop a procedure that assesses young children's ability to learn words after only a few exposures. This project is part of an ongoing investigation to explore how quickly children with cochlear implants can map sound patterns onto referents and to determine the relationship between measures of early word learning and other outcome measures such as spoken word recognition, speech intelligibility and receptive and expressive language abilities. This report describes the results of a preliminary study that was conducted to develop a procedure that can be used to test preschool-aged children's ability to learn words after a brief exposure period. The first part of the project involved selecting names for the Beanie Baby™ stuffed animals that would be taught to the children. This was done by eliciting names from adult participants. Next, we describe the results of a pilot study with several children with CIs who were given variations on our initial word-learning procedure. The results of the pilot testing with the children helped us modify several aspects of the procedure. Finally, we present preliminary data from twenty-four normal-hearing children and two children with CIs.

Pilot Phase

Selection of Stimulus Materials

The stimulus materials used in all of the experiments consisted of a set of sixteen Beanie Baby™ stuffed animals. Each Beanie Baby™ comes with a name assigned by the manufacturer (Ty Corporation®). We did not use these names because some children might already know the names while others might not and because some of the names were related to physical attributes of the stuffed animals while others were not. We decided to elicit names from adult participants that corresponded to salient physical attributes of the Beanie Babies™. This was done to facilitate the association between the names and the Beanie Babies™. Because many of the Beanie Babies™ have several features that could be considered salient, a pilot experiment was conducted to determine the characteristics of each Beanie Baby™ that were most perceptually salient. The goal of this pilot study was to select labels for each Beanie Baby™ that would be used in the experimental study.

Methods

Participants. The participants were 37 IU undergraduates, with no reported history of speech or hearing disorders. Thirty-five of the subjects were native English speakers. All subjects were recruited from the Indiana University community and all received partial credit towards an Introductory Psychology class for their participation. The mean age of the participants was 19.9 years (SD = 1.3).

Materials. Sixteen Beanie Baby™ stuffed animals were used as stimuli. In order to control for possible familiarity effects with the original names, we developed a new set of names for the Beanie Babies™. We decided as a first pass to create the new names in such a way as to give a semantic bootstrap to enable word learning, for example, by using a distinguishing physical characteristic of the animal. The Beanie Babies™ were therefore selected from a larger set of Beanie Babies™ on the basis of whether they had distinguishing characteristics that could be easily named, such as a very long tail, horns, or a bright color.

Procedure. Subjects were tested in three groups in a small experimental classroom. They were given written instructions, in which they were told that the experimenter would hold up each of the sixteen Beanie Babies™ individually, and they would be asked to invent new names for the Beanie Babies™, as if they were teaching the names to a young child. Subjects were asked to re-name the Beanie Babies™ using names that described some physical attributes of the Beanie Babies™. Subjects were instructed to provide up to three new names for each animal, and to use one-word names only. Subjects were provided with answer sheets on which to write the new names. The Beanie Babies™ were presented individually, one at a time, in a random order to the three groups.

Results

For each Beanie Baby™, the responses were recorded and tallied to calculate the frequency of the names generated. A new Beanie Baby™ attribute name was chosen from the response tallies based on two criteria: (1) that the name was the most frequent response among students, and (2) that it reflected a true physical attribute of the animal. For example, the name "*Red"* was the most frequent response and was also an appropriate name for the red bull because it refers to the color of the bull. In contrast, a non-attribute name, "*Teddy,*" was the most frequent response for the brown bear, but was inappropriate for our purposes. The second most common response was "*Fuzzy,*" which we used as it describes an attribute of the bear. The new attribute name was the most frequent response for seven of the Beanie Babies™ ("*Blue," "Red," "Stripes," "Pink," "Spots," "Ears," "Tail"*), the second most frequent response for four of the animals ("*Wings," "Fuzzy," "Legs," "Cottontail")*, the third most frequent response for five of the animals ("*Horns," "Gray," "Teeth," "Bushy"*), and the fourth most frequent response for "*White."* Table 1 lists each original Beanie Baby™ stuffed animal name and description, its new attribute name derived from this procedure, and the percentage of subjects who used the new attribute name.

Original Beanie Baby™ name	"New" attribute name	Frequency of new name response (%)	Original Beanie Baby™ name	"New" attribute name	Frequency of new name response (%)
Crunch the Shark	*Teeth*	11.8	*Dotty* the Dalmatian	*Spots*	46.6
Rocket the Bird	*Blue*	36.6	*Halo* the Angel Bear	*White*	6.1
Batty the Bat	*Wings*	13.7	*Spunky* the Cocker Spaniel	*Ears*	14.9
Kuku the Bird	*Pink*	29.6	*Nuts* the Squirrel	*Bushy*	11.1
Snort the Bull	*Red*	32.4	*Nibbly* the Bunny	*Cottontail*	14.1
Goatee the Goat	*Horns*	8.2	*Spinner* the Spider	*Legs*	14.9
Buster the Bear	*Fuzzy*	9.8	*Prance* the Tabby Cat	*Stripes*	25.6
Tiptoe the Mouse	*Tail*	18.5	*Spike* the Rhino	*Gray*	7.8

Table 1. Original Beanie Baby™ names, new given attribute names, and the frequency with which each new name was generated.

Procedure Development

Once the new names were chosen, a piloting phase was initiated to develop a procedure for assessing word learning in young children with cochlear implants after a brief exposure period. The initial conception of the experiment was as follows. Children would be taught the new names of the Beanie Babies™. Younger children (2;0 – 3;11) would be taught four names and older children (4;0 – 5;11) would be taught eight names. In order to get a baseline measure of how likely it was that the children would spontaneously label the Beanie Babies™ with the target labels, the experiment started with two pretests. In the first pretest, children were presented with each of the Beanie Babies™ they would be taught and were simply asked to give it any name, using a free response format. The second pretest used a forced-choice procedure. The Beanie Babies™ were placed in a row in front of the child, and the child was asked to select the one that might have the name that the experimenter presented to them. For example, the experimenter might say, "Which one do you think is named *Fuzzy*?" The experimenter did this for each of the Beanie Babies™.

Following the pretests, each child was given a sequence of training phases in which they were taught the names of the Beanie Babies™, one at a time, using play scenarios. The experimenter provided the name of each Beanie Baby™ exactly three times. Toys were used to give each Beanie Baby™ some sort of memorable personality. For example, in one scenario, the experimenter would say, "This is *Fuzzy*. *Fuzzy* likes to eat grapes. Can you give the grapes to *Fuzzy*?" After exposure and training with each Beanie Baby™, the children were given tests to assess whether or not they learned to associate the names with the Beanie Babies™. The first test used a forced-choice procedure, exactly like the second pretest. The second test used a cued-recall procedure. The cued-recall test required an expressive response from the child. In this procedure, each Beanie Baby™ was presented one at a time to the child as a cue, and the child was asked to recall its name from memory.

The initial procedure underwent several stages of development during the piloting phase. Six children who use cochlear implants participated in the piloting phase: SHM (4;1), SHZ (6;2), SHS (2;5), SNW (3;2), SMH (5;11), and SOC (4;1). Here, we will summarize our major observations during the piloting phase and describe how these pilot results shaped the final design of the experimental procedures.

• **Children often perseverate on the names they initially choose for the Beanie Babies™.** The initial conception of the procedure involved two pretests. During the procedures, we discovered that the children who were given the pretests (SHM and SHZ) were very resistant to learning new names for the Beanie Babies™. Instead, they tended to perseverate on the names that they initially selected. Hence, the pretests were dropped from the procedure.

• **Children who use cochlear implants need several exposures to words.** In the first stage of the pilot, children received only three exposures to each name before they were tested. With only three exposures per item, two of the participants in the pilot study did not perform above chance. Given the poor performance of the pilot subjects tested under this condition, the number of exposures was increased from three to eight for each name.

• **Imitation is important for word learning.** Another factor that seemed to contribute to the poor performance in the early stages of our pilot testing was that no measure was taken to ensure that the children actually encoded the names they were being taught. To make sure that the children encoded the sound pattern of the words, we asked the children to repeat the names that we produced. It is possible that the act of producing the words helps with children's memory for words because there may be a strong developmental interaction between perception and production (e.g., Vihman, 1993). There are recent data supporting the importance of immediate memory and imitation in novel word learning (Gupta & MacWhinney, 1995).

• **Children often show a preference for new Beanie Babies™.** During one phase of the pilot testing we decided to try teaching the names of the Beanie Babies™ to the children one at a time. Thus, they were first presented with one animal and then were given the forced-choice and cued-recall tests for that animal. If they were correct, they were taught the name of an additional Beanie Baby™. If they were incorrect, they were re-taught the original name. Each time they were correct on both the forced-choice and cued-recall tests, the set size increased by one. The set size increased until the child could no longer respond correctly on three consecutive trials. In carrying out this procedure with SHS, SNW, and SMH, we discovered that as the set size increased, the children showed a novelty preference for the most recent Beanie Babies™ presented. As a result, we decided that during each session, the child would be presented with all of the Beanie Babies™ (four or eight), one at a time, and then tested on all of them.

• **A minimum of one year of cochlear implant experience is necessary.** One of our initial criteria for inclusion in the study was that the child must have had at least six months of implant use. The participant SOC, who had exactly six months of experience, clearly did not have sufficient auditory skills to participate in the study and carry out the tasks. Hence, we increased the criterion to one year of implant use.

Experiment

The piloting phase ended when a procedure was settled on that was simple enough for the pilot participants to complete but did not yield ceiling performance. Children were taught and tested on two sets of Beanie Babies during one session. Their long-term memory of the names

was subsequently assessed in a second session by re-testing them at least two hours later. The final design is described here and some preliminary results from two children with CIs and twenty-four normal-hearing children are reported below.

Methods

Participants. Two groups of children participated in this study. One group of four children was recruited from the population of children with cochlear implants who are routinely followed as part of the ongoing longitudinal studies at Indiana University. The criteria for inclusion was that they were between the ages of 2;0 and 5;11, use oral communication, and had at least one year of cochlear implant experience. Two children who use CIs (ages: 3;2 and 3;10) completed the experiment, but the other two children (ages: 2;4 and 4;2) were unable to complete the experiment due to failure to give any responses and are not included in the Results section. Twenty-four age-matched normal-hearing controls were recruited from the Bloomington, Indiana area and the Center for Young Children daycare center on the Indiana-University-Purdue-University-Indianapolis campus. All 24 completed the initial experiment. Seven of the normal-hearing children and the two children with CIs who completed the initial testing participated in the long-term memory test.

Materials. The stimulus materials consisted of 16 Beanie Babies™ that were assigned names by normal hearing college students (see Stimuli Selection above). Each name corresponds to a salient physical attribute (e.g., "*Red*" is a red bull). The Beanie Babies™ were grouped into sets of four as shown in Table 2. The Beanie Babies™ were selected so that most of the attribute names could describe at least two Beanie Babies™ in the group. For example, "*Wings"*, "*Pink"*, and "*Blue"* all have wings. This was done so that the children would not be able to completely rely on identifying the attributes in the tasks. For example, when they were asked to identify "*Wings"*, three of the four Beanie Babies™ had wings.

Set	Beanie Baby™ Attribute Name	Description
A	Teeth	Shark
	Blue	Blue jay
	Wings	Bat
	Pink	Cockatoo
B	Red	Bull
	Horns	Goat
	Fuzzy	Brown bear
	Tail	Rat
C	Spots	Dalmatian
	White	White bear with halo
	Ears	Cocker spaniel
	Bushy	Squirrel
D	Cotton tail	Rabbit
	Legs	Spider
	Stripes	Cat with stripes
	Gray	Rhino

Table 2. Word set stimuli.

Procedure. Children were taught a set of Beanie Babies™ (Training Phase 1) and then given forced-choice and cued-recall tests for that set (Testing Phase 1). Children were then taught another set of Beanie Babies™ (Training Phase 2) and subsequently given the same tests with the second set (Testing Phase 2). Finally, after at least a two-hour delay, children were given the same tests using the first set of Beanie Babies™ and then using the second set (Long-Term Memory Test).

Training Phase 1. *Children younger than four years.* Each child is exposed to four Beanie Babies™. Before the experiment, the exact order of Beanie Baby™ presentation was randomized and recorded on a form that was then followed during the experiment. One experimenter (Experimenter 1) interacted with the child while a second experimenter (Experimenter 2) assisted Experimenter 1 in following the correct order. Experimenter 2 also recorded the children's responses and the number of times Experimenter 1 produced the name of each Beanie Baby™.

Experimenter 1 presented each Beanie Baby™, one at a time, to the child. A different toy prop was used to create a different play scenario with each Beanie Baby™ in order to keep the task interesting. During the play interaction with each Beanie Baby™, Experimenter 1 used the name of the Beanie Baby™ exactly eight times. During the play scenario, Experimenter 1 tried to elicit three productions of the name from the child. Experimenter 2 recorded how many times the child produced each name. Positive feedback was given when the child produced the correct names. See Appendix for a sample scenario.

Children between four years and six years. The training phase was the same with older children as with the younger children, except that eight Beanie Babies™ were taught instead of four.

Testing Phase 1. *Forced-Choice Test.* The Testing Phase consisted of a forced-choice identification task and a cued-recall test given immediately afterwards. For the forced-choice identification test, all of the Beanie Babies™ (four or eight) were placed in a row in front of each child and hidden from view with a piece of cardboard. Then a toy bus or truck was brought out and placed in front of the child. Experimenter 1 then asked the child to "please put {*one of Beanie Babies™*} into the truck {*or bus*}". The child was encouraged to select one of the Beanie Babies™ but was not given any feedback as to whether the response was the correct choice. For example, the experimenter said "thank you," "good job," or clapped when the child made a selection, regardless of whether or not the response was correct. The Beanie Baby™ was then placed back in the row and the next trial was initiated. Each Beanie Baby™ was requested exactly once.

Cued-Recall Test. For the cued-recall task, Experimenter 1 played a "knock knock" game with the child. One Beanie Baby™ was placed behind a toy doorway. Experimenter 1 and/or the child said "knock knock," the door would open and Experimenter 1 would ask the child, "Who's there?" The child was asked to name the Beanie Baby™, up to three times. Experimenter 2 recorded any response. This procedure was repeated for each Beanie Baby™.

Training Phase 2. This phase was the same as Training Phase 1 except that a different group of four or eight Beanie Babies™ was presented to the child.

Testing Phase 2. This phase was the same as Testing Phase 1 except that the new set of Beanie Babies™ was used.

Long-Term Memory Test. On the same day of testing, but at least two hours after the completion of Testing Phase 2, the child was tested a second time, in order to assess long-term memory for the names. During the long-term memory test, Testing Phase 1 and Testing Phase 2 were repeated again without any retraining or feedback.

Results

The mean accuracy scores for all of the tests are summarized in Table 3 for normal-hearing children, and in Table 4 for hearing-impaired children with CIs. The preliminary data revealed that the normal-hearing children had very high scores for the forced-choice test in both the Immediate and Delay conditions. So far, the children who use CIs have performed comparably on the immediate forced-choice task. However, their performance on the cued-recall test, and both tasks after a delay, were very low.

Immediate Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (12)	0.95	0.08	0.92	0.17
> 4 yrs (12)	0.95	0.10	0.97	0.08

Delay Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (3)	0.88	0.22	0.75	0.25
> 4 yrs (4)	0.94	0.13	0.89	0.14

Table 3. Mean accuracy response for normal-hearing children (N=24).

Immediate Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (2)	0.63	0.18	0.25	0.18

Delay Test

	Forced-choice accuracy	standard deviation	Cued-recall accuracy	standard deviation
< 4 yrs (2)	0.13	0	0.19	0.27

Table 4. Mean accuracy response for hearing-impaired children who use cochlear implants (N=2).

Discussion

The procedures that were developed in this project will allow us to assess the word-learning skills of children, which will be valuable in tracking the language development of children who use CIs. The results thus far are very preliminary because only a small number of children who use CIs have been tested. We will test at least 12 children who use CIs from each of the two age groups before analyzing the data and comparing it to the results from the normal-hearing children. Another step in the project is to analyze the results from this test and compare them to the results obtained on several outcome measures. The children who use CIs are routinely given a battery of speech perception, word recognition and language tests up to several years after they receive their CIs. One of our goals is to assess how the variability of children's performance in learning novel words in these tasks is related to the variability of language outcome measures. Measures of early word learning and "fast mapping" in this clinical population may be important new predictors of language development and other language-based outcome measures.

A future direction for this project is to manipulate the phonological properties of the names of the Beanie Babies™. Currently, we are using real names that correspond to salient visual attributes in order to make the learning task as easy as possible. Once these procedures are validated, subsequent experiments will use Beanie Babies™ with nonword names, which will vary in terms of phonological difficulty (e.g., phonotactic probabilities, syllable number or stress). These other projects should provide valuable new information about the ability of children who use CIs to encode phonological information in tasks that require novel word learning skills, imitation, and long-term retention.

References

Carey, S. (1978). The child as word learner. In M. Halle & J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293). Cambridge, MA: MIT Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*. Stanford University.

Clark, E.V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65-110). New York: Academic Press.

Clark, E.V. (1983). Meanings and concepts. In J.H. Flavell & E.M. Markman (Eds.), *Cognitive Development* (Vol. III, pp. 787-840). New York: Wiley.

Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (Serial number 242).

Gathercole, S., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336-360.

Gathercole, S.E., & Baddeley, A.D. (1989). Development of vocabulary in children and short-term phonological memory. *Journal of Memory and Language, 28*, 200-213.

Gilbertson, M., & Kamhi, A.G. (1995). Novel word learning in children with hearing impairment. *Journal of Speech and Hearing Research, 38*, 630-642.

Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language, 34,* 63-88.

Heibeck, T.H., & Markman, E.M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*, 1021-1034.

Houston, D.M., & Jusczyk, P.W. (submitted). Infants' long-term memory for words and voices. manuscript submitted to *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236-248.

Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1-23.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science, 277,* 1984-1986.

Leonard, L.B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Markman, E.M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought* (pp. 72-106). Cambridge: Cambridge University Press.

Nelson, K. (1988). Constraints on word learning? *Cognitive Development, 3*, 221-246.

Oetting, J.B., Rice, M.L., & Swank, L.K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research, 38*, 434-445.

Pisoni, D.B., Svirsky, M., Kirk, K.I., & Miyamoto, R.T. (1997). Looking at the "Stars": A first report on the intercorrelations among measures of speech perception, intelligibility and language development in pediatric cochlear implant users. *Progress Report on Spoken Language Processing #21*, Indiana University, Department of Psychology, Bloomington, IN.

Rice, M.L., Buhr, J., & Nemeth, M. (1990). Fast mapping word-learning abilities of language delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*, 33-42.

Rice, M.L., Oetting, J.B., Marquis, J., Bode, J. & Pae, S. (1994). Frequency of input effects on SLI children's word comprehension. *Journal of Speech & Hearing Research, 37*, 106-122.

Vihman, M.M. (1993). The construction of a phonological system. In B. de Boysson-Bardies & S. de Schonen & P. Jusczyk & P. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face perception in the first year of life* (pp. 411-419). Dordrecht: Kluwer.

Appendix

Sample Scenario:

This is *Name.*
Can you say hi to him?
Say "Hi *Name!*"
Now your turn {child says "hi *Name*"}
Name likes to climb the tree.
Can you put him on the tree? {child interacts with BB}
Look – *Name* is on the tree.
Tell him to get down. {child says, "get down *Name*"}
Good. Now, *Name* has to go bye bye.
Say, bye bye *Name.* {child repeats "bye bye *Name*"}

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Reduced, Citation, and Hyperarticulated Speech in the Laboratory: Some Acoustic Analyses[1]

James D. Harnsberger and Lori A. Goshert

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant DC00012 to Indiana University. We would like to thank Corey Yoquelet for his assistance in the acoustic analysis.

Reduced, Citation, and Hyperarticulated Speech in the Laboratory: Some Acoustic Analyses

Abstract. An acoustic analysis was carried out on a set of sentence stimulus materials varying in speech style (Reduced, Citation, and Hyperarticulated) that was elicited via a technique developed previously in our laboratory by Brink, Wright and Pisoni (1998) and Harnsberger and Pisoni (1999). Sentences recorded from twelve speakers were acoustically analyzed for sentence duration, keyword duration, and F1-F2 vowel space dispersion. The Reduced, Citation, and Hyperarticulated styles varied in terms of articulatory precision, in increasing order. Thus, the styles were predicted to differ significantly in keyword and sentence duration, with longer durations corresponding to more articulatorily precise styles. The styles were also predicted to differ significantly in the extent to which vowels in keywords were centralized, affecting the extent of vowel space dispersion. More disperse spaces were predicted for more articulatorily precise styles. Of the twelve participants, seven produced sentences with either the predicted keyword or sentence duration differences (or both) between all three styles. Eight of the twelve participants also showed the predicted vowel dispersion differences between styles, with greater dispersion corresponding to a larger vowel space. However, for some participants the dispersion differences between the Reduced and Citation styles were quite modest. In addition, all twelve participants produced a Hyperarticulated style that differed in keyword duration, sentence duration, and vowel dispersion from the Reduced and Citation styles, as predicted. Overall, the results demonstrate that it is possible to elicit controlled sentence stimulus materials varying in speech style in a laboratory setting, although the method requires further refinement to elicit these styles more consistently from individual participants.

Introduction

A longstanding problem in studies of speech production and speech perception has concerned the limitations imposed by experimental control and by laboratory settings in the collection of naturalistic speech. Naturalistic, spontaneous speech refers to a speech style commonly employed by talkers and listeners in conversations outside of a laboratory setting. In contrast, the style typically elicited from talkers in a recording session is read speech, sometimes called "lab speech," which differs in numerous ways from more spontaneous styles. These differences can include the duration of the utterance and its constituent words, pausing, and the degree of centralization in the quality of vowels, to name a few (Byrd, 1994; Picheny, Durlach, & Braida, 1989; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Unfortunately, much of what we know about speech production, speech perception, and spoken language processing has relied on a narrow range of speech styles, usually read speech. Theoretical models based on such studies may be severely limited in their capacity to generalize to other speech styles and, most importantly, to the speech styles that listeners encounter most frequently outside of a laboratory environment.

The popularity of read speech in studies of speech perception and spoken language processing has been driven by its numerous advantages to researchers. Read speech can be useful in limiting sources of error in the data collection process, or in avoiding particular confounds that might render the results uninterpretable. Control over the quality and structure of the materials also insures that an experiment can be replicated in other laboratories, a key aspect of any experiment. However, the reliance on studies of read speech elicited in the laboratory has meant that perception of variability that exists among speech styles has not been studied in detail. Other types of "nonlinguistic" variability have been shown to affect

speech perception and spoken word recognition, including talker, rate, and stimulus variability (Bradlow, Nygaard, & Pisoni, 1999; Mullennix & Pisoni, 1990; Nygaard, Sommers, & Pisoni, 1995). These studies suggest that listeners encode in long-term memory significant details and properties of speech signals that they encounter, and that these details influence the subsequent perception and recognition of speech. If listeners are sensitive to detailed properties of speech, then variation in those properties due to speech style differences may also play an important role in perceptual processing, one that has thus far been neglected in studies of speech perception and spoken word recognition.

Ideally, to address the issues of the generalizability of theoretical models to more naturalistic speech and the encoding and use of style-specific detail in spoken word recognition, a method would be needed to elicit different speech styles, particularly more naturalistic ones, while maintaining control over the speech materials elicited. Such a method would constitute a happy compromise between the benefits of experimental control and the benefits of analyzing a more natural, representative sample of speech. Various methods have been developed in prior work to elicit spontaneous speech, including the recording of natural conversation, guided conversations on a particular topic, and narration or map tasks (Hirschberg & Nakatani, 1996; Labov, 1972; Milroy, 1987; Speer, Sokol, & Schafer, 1999; Swerts & Collier, 1992). These methods have proven useful in eliciting specific words, phrases, and discourse units of interest. However, they typically fail to control for phonetic context, and they are often not appropriate for eliciting certain linguistic forms, such as specific sentences.

In our laboratory, we have attempted to develop a method of eliciting different speech styles at the sentence level while controlling for the particular sentence materials used. The range of speech styles we have studied includes a Reduced, or hypoarticulated, style, which should more closely resemble the speech style employed in natural settings than does laboratory read speech. The first version of this method was developed by Brink, Wright, and Pisoni (1998). They attempted to elicit three speaking styles, namely Reduced, or hypoarticulated speech; Citation, or read speech (the style normally used in reading controlled materials in a laboratory setting); and Hyperarticulated speech (i.e., clear speech). Each style was elicited in a separate condition of the experiment. Brink et al. attempted to elicit Reduced speech by having participants read a sentence while engaging in a concurrent processing task, specifically, remembering a digit sequence of five to seven digits that was presented immediately prior to the sentence. After reading the sentence, participants were asked to recall the digit sequence in the same order in which it was presented. The digit span task was considered to be a distractor task, chosen to place the participant under a cognitive load while reading a sentence. The digit span task was chosen as the concurrent task because it was successful in pilot studies in producing the desired speech style while minimizing disfluencies by the participants. Citation speech was elicited by simply having listeners read single sentences presented on a computer screen. Hyperarticulated speech was elicited in an experimental condition similar to the Citation speech condition. Participants were asked to read single sentences presented on a computer screen. During this portion of the experiment, they were prompted in a subset of trials to repeat the sentence "more clearly." After responding to that prompt, participants were given the same prompt a second time, and the second reading was chosen to represent Hyperarticulated speech. This procedure had been used successfully in an earlier study by Johnson, Flemming, and Wright (1993).

Brink et al. tested this methodology with six participants, all native speakers of English, and evaluated its success in a detailed acoustic analysis. They measured several properties of the sentences, as well as keywords in the sentences, including the duration, f0 range, absolute RMS energy, energy range, degree of vowel centralization, and degree of vowel dispersion. The results of the acoustic analysis showed that the method was successful in eliciting a Hyperarticulated speech style that was highly distinct from the Citation style, a result that was found for all six talkers. The duration, vowel centralization, and vowel dispersion measures showed the most consistent differences. However, the method failed to elicit significant differences between the Reduced and Citation sentences for five of the

six talkers. Only one participant produced Reduced speech that was acoustically distinguishable from Citation speech using these measures.[2]

More recently, Harnsberger and Pisoni (1999) extended the work by Brink et al. by testing a variant of the elicitation method for Reduced speech, termed the *calibrated cognitive load method*. Harnsberger and Pisoni calibrated the cognitive load (i.e., the digit span task) to the digit span of individual talkers via an immediate serial recall digit span task administered prior to the speech elicitation task. The cognitive load used by Brink et al. was a fixed load (5 – 7 digits in length), which may have been too easy a concurrent task for some talkers, given that adult digit spans average about 7.7 digits in length (Cavanagh, 1972). The individually calibrated cognitive load proved to be successful in eliciting a Reduced speech style from six of the twelve talkers recorded, a substantial improvement over the method used by Brink et al., though still not ideal. The success of the new method was gauged by a set of perception tests (paired comparison tasks) using phonetically-trained and naïve listeners. The purpose of this study was to evaluate the results of Harnsberger and Pisoni through an acoustic analysis of the sentences produced by talkers in that study. The particular acoustic measures taken were a subset of those used by Brink et al. in an acoustic analysis of their elicited speech materials. The particular subset selected were those that were the most successful in differentiating the three speaking styles elicited from the one talker who produced a consistent Reduced-Citation style contrast.

Methods

Participants

Twelve native speakers of American English (seven females and five males), ranging in age from 18 to 30, participated in this study. Participants received $15 total for participating in two one-hour sessions. None of the participants reported any history of speech or hearing disorders at the time of testing.

Stimulus Materials

The participants read 34 sentences from the 200 sentences comprising the Speech Perception in Noise (SPIN) set (Kalikow, Stevens, & Elliot, 1977). The SPIN sentences are short sentences, five to eight words in length, ending in a high frequency monosyllabic noun. The 34 SPIN sentences selected for this study are listed in Appendix 1. The recordings took place in a sound-attenuated chamber (IAC Audiometric Testing Room, Model 402) using a head-mounted Shure (SM98) microphone positioned one inch away from the participant's chin. The recordings were digitized at 22.05 kHz (16 bit sampling) using a Tucker-Davis Technologies System II and stored on an IBM-PC 486 computer.

Procedures

The participants were all recorded reading the sentences under three different conditions corresponding to three distinct speech styles: (1) Reduced, (2) Citation, and (3) Hyperarticulated. The elicitation procedure consisted of four tasks carried out over two test sessions. In the first session, participants were administered a simple forward digit span task (see Digit Span Task) and were recorded reading sentences in the Reduced condition. In the second session, which took place within seven days of

[2] This speaker's Reduced sentences were also perceptually distinguishable from his/her Citation sentences in a pilot Paired Comparison task with three native speakers of English. These native speakers successfully picked the citation sentences as "more carefully pronounced" in reduced-citation sentence pairs, on an average of 89% of test trials. For a detailed description of the Paired Comparison task, see Experiment 2 in Harnsberger and Pisoni's (1999) study.

the first session, participants were recorded reading sentences in the Citation and Hyperarticulation conditions.

Digit Span Task. In the digit span task, participants were presented with a sequence of single digits (0 - 9) on a computer screen inside of the sound-attenuated chamber, and asked to recall the sequence correctly in the order in which it was presented. The participants' responses were digitized and played via headphones to the experimenter, who sat outside of the booth and scored the responses. The responses themselves were not stored to disk as sound files. The length of the digit sequence that was presented started at four, and then increased or decreased via an adaptive staircase algorithm. The algorithm increased the sequence length by one digit for every two sequences at a given length that were successfully recalled by the participant. Whenever the participant responded to a sequence incorrectly, the sequence length was reduced by one digit on the following trial. Over the course of the 25 trials of the task, the sequence length for individual participants increased until the sequence length began eliciting errors. Thus, by the end of the task, participants were "oscillating" between the sequence length that they could consistently recall, and a longer sequence that induced errors. The longest sequence length that was consistently recalled was taken to be the participant's digit span. This value was then used to calibrate the cognitive load in the Reduced condition.

Reduced Condition. The Reduced condition was similar to the Reduced condition described by Brink et al. and consisted of 136 trials, four trials for each of the 34 SPIN sentences, with a 1 s inter-trial interval. The order of the blocks of four trials varied randomly for each participant. Each trial consisted of four parts: initially, participants were presented with a digit sequence, which remained on the screen for 2 s; then, after a 2.5 s interval, a sentence was displayed on the computer screen for the participant to read; next, the participant's response was recorded over a 6 s window; finally, participants were prompted to recall the digit sequence in the correct order. The length of the digit sequence was based on the participant's digit span as measured in the Digit Span Task. The length of the digit sequence in a given trial was either the same as the span score, or plus/minus one digit. For example, if a participant had a span of seven in the digit span task, he/she would be presented with digit sequences ranging in length from six to eight. The same sentence, embedded in the digit span task, was presented four times, with the fourth reading taken as the reduced sentence for subsequent analysis. Before the recording began for the Reduced condition, participants were told that they would be participating in a short-term memory experiment. Participants were instructed to focus on the digit span task in the Reduced condition, in the hope that they would be less careful in monitoring their production of the test sentences.

Citation and Hyperarticulation Conditions. The Citation and Hyperarticulation conditions were identical to those described earlier by Brink et al. In the Citation condition, participants were prompted to read aloud a sentence that appeared on the computer screen. Each sentence was presented once, for a total of 34 trials, with a 1 s ITI. The order in which the sentences were presented was randomized for each participant. The Hyperarticulation condition was similar to the Citation condition, and consisted of two types of trials. The first trial type, the "citation cycle," was identical to a Citation condition trial. In the second trial type, the "hyperarticulation cycle," participants were also prompted to read aloud a sentence appearing on the computer screen. After reading this sentence, participants were then prompted to "Please read the sentence more clearly." After responding, they were asked again to read the sentence more clearly. Thus, for the hyperarticulation cycle, the same sentence was read three times, with the third reading taken to be the example of the "hyperarticulated" reading of the sentence for subsequent analysis. The 34 sentences each appeared in three citation cycles and one hyperarticulation cycle. The program controlling the experiment was designed to insure that the Hyperarticulation condition began with a citation cycle, and that hyperarticulation cycles were separated by at least two citation cycles.

Acoustic Analysis. The recorded sentences were acoustically analyzed for the duration of the sentences as well as three to four keywords. All of the keywords were content words and commonly appeared in one of three positions within the sentences: (1) near the beginning (usually the participant noun), (2) near the middle (usually the main verb) and (3) in the final position (usually the main object of the verb or of a preposition). Duration was measured directly from the waveforms with accompanying wide band spectrograms for reference using Cool Edit 2000 software.

The keywords in each sentence, in all three styles, were also acoustically analyzed for *vowel dispersion*, defined as the average Euclidian distance in Barks of keyword vowels from the center of an individual's vowel space. Vowel formant measures were made from an overlaid LPC-FFT display. The LPC employed 12-16 coefficients (based on the participant) and a 25 ms frame size. The FFT used a 1024-point window. A wide-band spectrogram was used for reference. The formant measures were made at the point of maximal displacement of F1 and F2. The results of the acoustic analyses were used to examine the differences between the Reduced, Citation, and Hyperarticulated styles of individual participants.

Results

Duration Measures

Figures 1 and 2 display the mean keyword and sentence durations for each participant, respectively. Table 1 shows the differences in the duration measures (in seconds) between the Reduced, Citation, and Hyperarticulated styles for individual participants. The difference scores were computed by subtracting the mean duration measure (i.e., the keyword or sentence duration measure) of the "less precise" style from the "more precise" style. Thus, we predicted positive, significant difference scores in all cases. The mean duration measures for each participant were submitted to separate 3 (Style: Reduced, Citation, Hyperarticulated) X 2 (Unit of Analysis: Keyword, Sentence) repeated measures ANOVAs. For every participant, there were significant main effects of Style and Unit of Analysis, as well as a significant Style by Unit of Analysis interaction. Appendix 2 lists the results of the statistical analysis by participant.

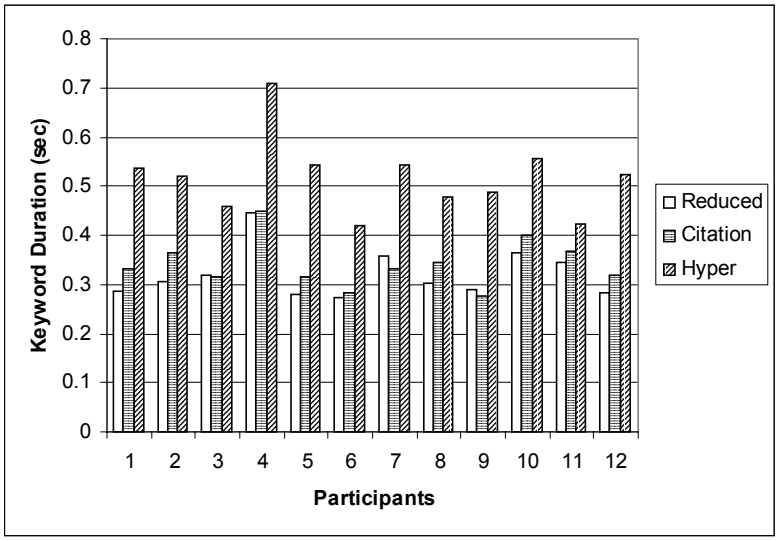


Figure 1. Mean keyword durations of each style for each participant.

Post hoc analyses (Tukey t-tests) showed that seven participants produced positive, significant differences between the Reduced and Citation styles in whole sentences, while only one participant produced positive differences between the two styles in keyword duration. Both the sentences and keywords read in the Hyperarticulated style differed significantly from those read in the Reduced and Citation styles for every participant, as predicted. The differences in duration between the sentences read in the Hyperarticulated style and those in the Reduced and Citation styles were much greater in magnitude than the differences in duration between the Reduced and Citation styles. Overall, seven out of twelve participants differentiated the three styles by manipulating some aspect of the temporal properties of the sentence.

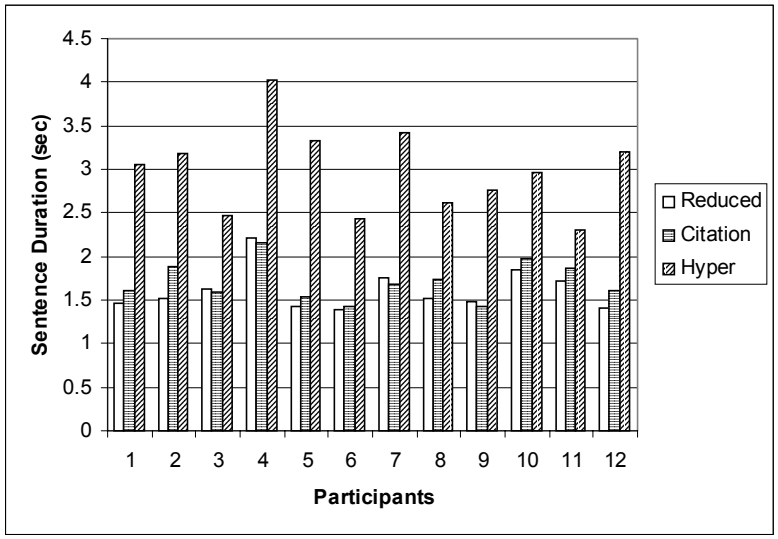


Figure 2. Mean sentence durations of each style for each participant.

	Citation - Reduced		Hyperarticulated - Reduced		Hyperarticulated - Citation	
Subject	Key	Sentence	Key	Sentence	Key	Sentence
1	0.046	0.154**	0.253**	1.587**	0.207**	1.433**
2	0.059*	0.359**	0.215**	1.662**	0.156**	1.303**
3	-0.004	-0.037	0.140**	0.844**	0.144**	0.881**
4	0.003	-0.047	0.261**	1.816**	0.258**	1.863**
5	0.034	0.115*	0.262**	1.895**	0.228**	1.78**
6	0.009	0.041	0.147**	1.047**	0.138**	1.006**
7	-0.026	-0.085	0.184**	1.666**	0.21**	1.751**
8	0.040	0.205**	0.173**	1.089**	0.133**	0.884**
9	-0.012	-0.050	0.197**	1.284**	0.209**	1.334**
10	0.033	0.116*	0.191**	1.115**	0.158**	0.999**
11	0.025	0.133**	0.078**	0.577**	0.053*	0.444**
12	0.034	0.191**	0.239**	1.774**	0.205**	1.583**

Table 1. Mean differences between the "more precise" and "less precise" styles for the duration measures (in seconds). "Key" denotes keyword.
* $p < .05$, ** $p < .01$

Vowel Dispersion

Figure 3 shows the differences in vowel dispersion for each individual participant between the three styles, with greater dispersion corresponding to a larger vowel space. Reduced, Citation, and Hyperarticulated (Hyper) styles were predicted to differ in increasing order in degree of vowel dispersion. An example of vowel spaces differing in degree of dispersion appears in Figure 4, which shows Participant 2's vowel spaces computed from the keyword vowels in each style. Figure 4 demonstrates that, as participants articulate in speaking styles that increase in articulatory precision (i.e., from Reduced to Hyperarticulated, in order of increasing precision), the corresponding vowel spaces expand.

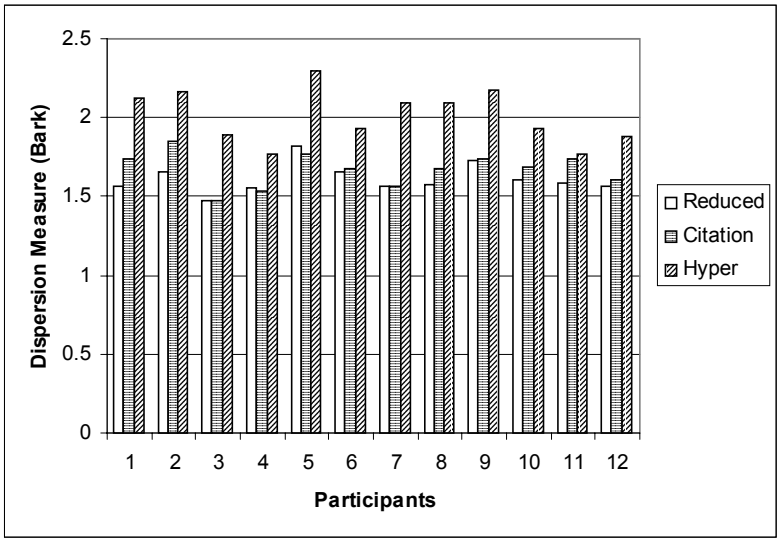


Figure 3. Vowel dispersion measures of each style for each participant.

All twelve participants produced a Hyperarticulated style that differed in vowel dispersion from the Reduced and Citation styles, as predicted. Four participants failed to produce a Reduced style that differed in vowel dispersion from the Citation style in the predicted manner. Eight of the twelve participants showed some degree of vowel dispersion differences in the predicted manner between all three styles. However, for most participants, the differences in vowel dispersion between the Reduced and Citation style were much more modest than those involving the Hyperarticulated style. A statistical analysis of the individual results was not possible given that, for each participant, only one vowel dispersion score could be computed for each style.

Discussion

The goal of this study was to evaluate in an acoustic analysis the success of the calibrated cognitive load method in eliciting three distinct speech styles. The acoustic analysis showed that the revised procedure was successful in eliciting Reduced speech from a majority of the talkers, although large individual differences remain. With a fixed cognitive load, only one of six participants produced reliable differences between the Reduced and Citation styles based on an extensive acoustic analysis of their utterances (Brink et al., 1998). Thus, the results reported by Harnsberger and Pisoni (1999) together with the results of the present study suggest that individually calibrating the cognitive load for the individual participant results in a more consistent elicitation procedure for a Reduced style of speech.

While a success rate of seven out of twelve participants represents a marked improvement over the results reported by Brink et al., the time and effort required by this procedure to elicit the style differences for just 34 different sentences necessitate changes in the experimental procedure to reduce the range of individual differences. First of all, the cognitive load could be increased to one or two digits more than the participant's individual digit span, to insure that the task is sufficiently demanding for the listener as they produce the sentences. A heavier cognitive load may, unfortunately, also have the effect of eliciting more disfluencies. A digit span task with a heavier cognitive load may also prove to be so difficult a task that participants may ignore the span task and simply read the sentence.

Secondly, an adaptive algorithm could be employed throughout the elicitation of Reduced sentences. Currently, a fixed range of loads is used in the elicitation procedure that has been calibrated to the individual participant in an immediate serial recall digit span task (the calibration task). However, due to changes in attention or fatigue, a participant's "effective" digit span could change over the course of the elicitation procedure and, thus, could be higher or lower than that measured in the calibration task. One way to address this possibility would be to adjust the cognitive load adaptively over the course of the elicitation procedure, increasing the load when participants continue to perform well (i.e., recall the digit sequence correctly), and decreasing the load when participants fail to correctly recall a sequence in order.

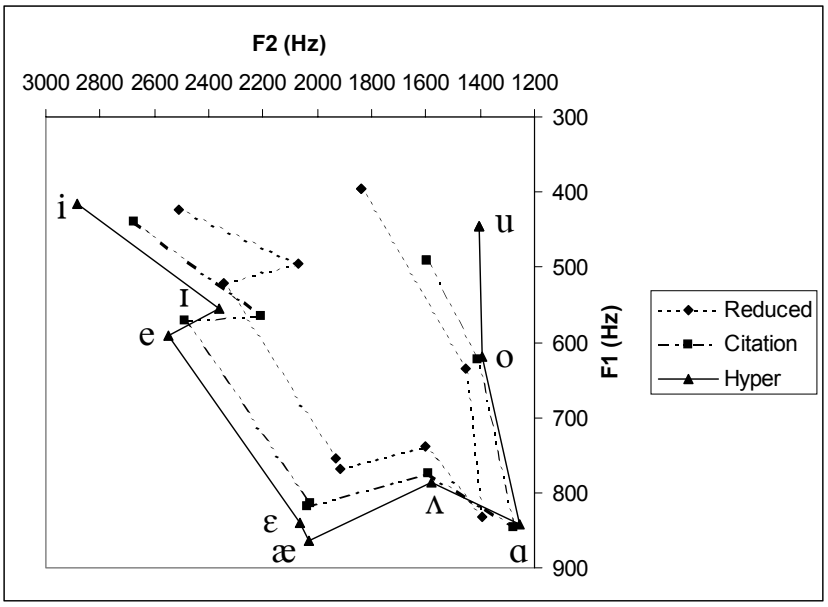


Figure 4. The vowel spaces of participant 2 for keyword vowels in Reduced, Citation, and Hyperarticulated (Hyper) styles.

Finally, the success of the present elicitation procedure, as determined by the acoustic analysis, may be underestimated by the particular measures taken. Our analysis employed the subset of measures used by Brink et al. that were most successful in differentiating the three styles for the single participant who reliably produced them. Brink et al. measured overall sentence energy, word energy, sentence energy range, word energy range, and the pitch range of the sentence as well as sentence and word duration and vowel dispersion. It is possible that, for the stimulus materials measured in this study, additional measures such as those used by Brink et al. would have revealed other acoustic differences between the three speaking styles for listeners.

Summary and Conclusions

In this study, a set of sentences produced in three speech styles by a novel elicitation procedure developed by Brink et al. (1998) and Harnsberger and Pisoni (1999) were acoustically analyzed. In the analysis, the duration and vowel formant frequencies of keywords were measured, as well as the duration of the entire sentence. The results were used to determine the efficacy of the elicitation method. The acoustic analysis showed that six of the twelve participants differentiated the Reduced and Citation styles in both duration and vowel formant frequency. All twelve participants differentiated the Hyperarticulated style from the Reduced and Citation styles. The limited success of the elicitation procedure suggests that further refinement of the procedure is required. Several variants of the elicitation procedure were suggested, including the use of a heavier cognitive load and/or an adaptive load in the elicitation of Reduced sentences. In addition, the limitations of the acoustic analysis were discussed, and greater range of measures was suggested.

References

Bradlow, A.R., Nygaard, L.C., & Pisoni, D.B. (1999). Effects of participant, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics*, *61*, 206-219.

Brink, J., Wright, R., & Pisoni, D.B. (1998). Eliciting speech reduction in the laboratory: Assessment of a new experimental method. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 396-420). Bloomington, IN: Speech Research Laboratory, Indiana University.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39-54.

Cavanagh, J.B. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, *79,* 525-530.

Harnsberger, J.D. & Pisoni, D.B. (1999). Eliciting speech reduction in the laboratory II: Calibrating cognitive loads for individual talkers. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 339-349). Bloomington, IN: Speech Research Laboratory, Indiana University.

Hirschberg, J. & Nakatani, C.H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. ACL-96.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are Hyperarticulated. *Language*, *69*, 505-528.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Milroy, L. (1987). *Observing and analyzing natural language*. Oxford: Basil Blackwell.

Mullennix, J.W. & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *61*, 206-219.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception and Psychophysics*, *57*, 989-1001.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, *32*, 600-603.

Speer, S.R., Sokol, S.B., & Schafer, A.J. (1999). Prosodic disambiguation of syntactic ambiguity in discourse context. *Journal of the Acoustical Society of America, 106,* 2275.

Summers, W., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., & Stokes, M.A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America, 84,* 917-928.

Swerts, M. & Collier, R. (1992). On the controlled elicitation of spontaneous speech. *Speech Communication*, *11*, 463-468.

Appendix 1: Stimulus Sentences (keywords underlined)

The farmer harvested his crop.
His boss made him work like a slave.
He caught the fish in his net.
Close the window to stop the draft.
The beer drinkers raised their mugs.
I made the phone call from a booth.
The cut on his knee formed a scab.
The railroad train ran off the track.
They drank a whole bottle of gin.
The airplane dropped a bomb.
I gave her a kiss and a hug.
The soup was served in a bowl.
The cookies were kept in a jar.
How did your car get that dent?
The baby slept in his crib.
The cop wore a bullet-proof vest.
No one was injured in the crash.
The hockey player scored a goal.
How long can you hold your breath?
At breakfast he drank some juice.
The king wore a golden crown.
He got drunk in the local bar.
The doctor prescribed the drug.
The landlord raised the rent.
Playing checkers can be fun.
Throw out all this useless junk.
Her entry should win first prize.
The stale bread was covered with mold.
I ate a piece of chocolate fudge.
The story had a clever plot.
He's employed by a large firm.
The mouse was caught in the trap.
I've got a cold and a sore throat.

Appendix 2: Statistical tests of individual participant results

Participant	Style	Unit of Analysis	Interaction
1	$F(1, 31) = 5082.37$, $p < .0001$	$F(2, 64) = 635$, $p < .0001$	$F(2, 62) = 342.65$, $p < .0001$
2	$F(1, 30) = 7398.76$, $p < .0001$	$F(2, 60) = 745.2$, $p < .0001$	$F(2, 60) = 447.25$, $p < .0001$
3	$F(1, 30) = 8942.34$, $p < .0001$	$F(2, 66) = 438.19$, $p < .0001$	$F(2, 60) = 225.12$, $p < .0001$
4	$F(1, 32) = 8106.39$, $p < .0001$	$F(2, 64) = 773.4$, $p < .0001$	$F(2, 64) = 437.91$, $p < .0001$
5	$F(1, 31) = 5500.4$, $p < .0001$	$F(2, 66) = 909.39$, $p < .0001$	$F(2, 62) = 531.17$, $p < .0001$
6	$F(1, 32) = 5557.58$, $p < .0001$	$F(2, 66) = 421.69$, $p < .0001$	$F(2, 64) = 240.77$, $p < .0001$
7	$F(1, 31) = 6709.39$, $p < .0001$	$F(2, 64) = 772.32$, $p < .0001$	$F(2, 62) = 484.74$, $p < .0001$
8	$F(1, 27) = 6640.62$, $p < .0001$	$F(2, 66) = 417.87$, $p < .0001$	$F(2, 54) = 222.9$, $p < .0001$
9	$F(1, 23) = 4877.07$, $p < .0001$	$F(2, 64) = 557.29$, $p < .0001$	$F(2, 46) = 297.67$, $p < .0001$
10	$F(1, 29) = 6003.42$, $p < .0001$	$F(2, 66) = 315.57$, $p < .0001$	$F(2, 58) = 161.03$, $p < .0001$
11	$F(1, 26) = 8646.99$, $p < .0001$	$F(2, 52) = 134.32$, $p < .0001$	$F(2, 52) = 79.18$, $p < .0001$
12	$F(1, 20) = 3459.85$, $p < .0001$	$F(2, 51) = 509.1$, $p < .0001$	$F(2, 40) = 299.21$, $p < .0001$

This page left blank intentionally.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Change Deafness: The Inability to Detect Changes in a Talker's Voice[1]

Michael S. Vitevitch

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This research was supported, in part, by NIH-NIDCD Training Grant T32 DC 00012 to Indiana University. I would like to thank Robert Nosofsky, David Pisoni and Holly Storkel for helpful comments and suggestions.

Change Deafness: The Inability to Detect Changes in a Talker's Voice

Abstract. Change blindness is a failure to detect a change in a visual scene. A shadowing task was used to demonstrate an auditory analogue to change blindness—*change deafness*. Participants repeated words varying in lexical difficulty. After a rest-break they heard more words from either the same or a different talker. Answers to explicit questions about the change in talker and implicit measures of behavior (i.e., response latencies) demonstrate that processing is affected by the change, even if participants do not explicitly report a change in talker. Specifically, listeners who did not detect the change in the talker had a greater difference between conditions of lexical difficulty than listeners who noticed the change, or listeners who heard the same talker throughout. These results suggest that failures to detect changes are not limited to the visual domain and that processing at some level may be affected by changes in the environment.

Change blindness is a counterintuitive phenomenon (Levin, Momen, Drivdahl & Simons, in press; Simons & Levin, 1998) in which observers in a variety of paradigms (e.g., Henderson & Hollingworth, 1999; Levin & Simon, 1997; Rensink, O'Regan & Clark, 1997) fail to detect what may be described as obvious changes in the environment. For example, Grimes (1996) found that participants noticed only 30% of the changes in photographs that occurred during an eye movement—even changes as obvious as two heads switching bodies. Simons and Levin (1998) dramatically demonstrated that only 33% of the participants in a real-life interaction noticed that the person asking them for directions was exchanged when a door being carried by confederates momentarily interrupted the discussion.

Is the inability to detect changes in the environment unique to the intense processing demands of the visual system that must encode complex visual-spatial details during very brief eye-fixations, or are there analogous deficits to detecting changes in similarly complex "scenes" in other modalities? In the auditory domain, spoken language may be a comparably complex stimulus. Speech is a complicated auditory signal that simultaneously conveys a conceptual, linguistic message and indexical information to a listener. Indexical information refers to acoustic correlates in the speech signal that provide information on various characteristics of the talker including identity, emotional state, age, dialect, and gender (Pisoni, 1997).

A number of studies have found that changes in the talker producing the stimulus words affects the accuracy with which participants identify those words presented in noise (Mullennix, Pisoni & Martin, 1989; Nygaard, Sommers & Pisoni, 1994), as well as the later recall and recognition of stimulus words (Martin, Mullennix, Pisoni & Sommers, 1989; Palmeri, Goldinger & Pisoni, 1993; Goldinger, Pisoni & Logan, 1991). Furthermore, Goldinger (1998) has shown that words spoken by the same talker in training and test sessions are repeated faster than words spoken by different talkers between training and testing sessions. The results of these studies suggest that changes in the voice of the talker affect processing. Unfortunately, in none of these experiments were participants *explicitly* interrogated to see if they detected the change in the talker that produced the stimuli. Thus, it is unknown if participants explicitly noticed the change in talker, or if they were "deaf" to this change. The results from change blindness studies (e.g., Simons & Levin; 1998) might lead one to predict that most of the participants in the talker-variability studies were unaware of the change in talkers. In contrast, the results from the talker-variability studies clearly demonstrate that participants' responses were affected by the change in the talkers (e.g., Palmeri, Goldinger & Pisoni, 1993).

To reconcile this apparent contradiction the present experiment measured the response latencies of participants to repeat words that were produced by the same talker throughout the experiment or by talkers that were changed halfway through the experiment. Furthermore, participants were *explicitly* asked at the end of the experiment if they noticed the change in the talker. By using both implicit and explicit measures, the present experiment addresses important questions regarding change detection and talker-variability. Specifically, is the inability to detect changes limited to the visual domain, or do participants also exhibit change deafness to changes in the auditory environment? More interestingly, this experiment will allow us to see if behavior is affected *implicitly* by the change in the talker (via differences in reaction times across groups) even if participants do not *explicitly* detect the change (see Chun & Nakayama, in press; Hayhoe, in press; Williams & Simons, in press).

Method

Participants

Twenty-four native speakers of English who reported no history of hearing or speech disorders participated in the experiment for partial fulfillment of an Introductory Psychology research requirement.

Stimuli

One hundred words with a familiarity rating of 6 or higher on a seven-point scale (Nusbaum, Pisoni & Davis, 1984) were selected for this experiment from the Indiana "Easy-Hard" Multi-Talker Speech Database (Torretta, 1995). Fifty words were lexically *easy* and fifty words were lexically *hard*. Lexically easy words have high word frequency and few similar sounding words with a low frequency of occurrence, whereas lexically hard words have low word frequency and many similar sounding words with a high frequency of occurrence (Torretta, 1995). These variables were statistically different in this subset of stimuli. The mean word frequency (based on word counts from Kucera & Francis, 1967) for easy words was 173.02 occurrences per million, and 8.5 occurrences per million for hard words (F (1, 96) = 41.75, $p < .001$). The mean number of similar sounding words, or neighbors, for easy words was 13.36 neighbors, and 27.24 neighbors for hard words (F (1, 96) = 238.71, $p < .001$). The mean frequency of the neighbors for easy words was 34.68 occurrences per million, and 302.19 occurrences per million for hard words (F (1, 96) = 58.42, $p < 001$). The results from a number of different behavioral tasks and participant populations show that, in general, participants respond more quickly and more accurately to lexically easy words than to lexically hard words (e.g., Kirk, Pisoni, Miyamoto, 1997; Luce & Pisoni, 1998; Sommers, 1996).

The same one hundred words were selected from two different male talkers in the database (talkers M0 and M9). These words were pre-tested by ten additional listeners from the same population in an AX "same-different" task to confirm that the selected talkers were perceptually discriminable. These participants heard the same word twice (separated by 50 ms. of silence). The word was spoken either by the same talker or by the two different talkers. When the word was produced by the same talker, participants were 98.6% accurate in responding that the voices were the same. When the word was produced by the two different talkers, participants were 92.2% accurate in responding that the voices were different. These results suggest that the two male voices were highly discriminable perceptually and that any failures to detect the change in the voice were not due to the perceptual similarity of the voices of the talkers.

Procedure

Participants were tested one at a time on a Macintosh Quadra 950 running PsyScope 1.2.2 (Cohen, MacWhinney, Flatt & Provost, 1993) which controlled stimulus randomization and presentation, and collection of response latencies. A headphone-mounted microphone (Beyer-Dynamic DT109) was interfaced to a PsyScope button box that acted as a voice-key. A typical trial proceeded as follows: A stimulus word was presented over the headphones to a participant who had been instructed to repeat the word as quickly and as accurately as possible. Response latency, measured from the beginning of the stimulus, was triggered by the onset of the participant's verbal response. Another trial began 1 s after a response was made. Responses were also recorded on audio-tape for later accuracy analyses.

Each participant received a total of 100 trials. In the first half of the experiment 25 easy words and 25 hard words were presented. In the second half of the experiment the remaining easy and hard words were presented. Each participant was presented with the same words in each half of the experiment, but in a different random order. Halfway through the experiment, participants were given a one-minute rest break. When the experiment resumed, half of the participants heard the same talker present the rest of the stimuli, whereas the other half of the participants heard the other talker present the stimuli. The order of presentation for the talkers was counterbalanced.

When each participant finished the auditory shadowing task, they were asked three questions in the following order:

(1) Did you notice anything unusual about the experiment?
(2) Was the first half of the experiment the same as the second half of the experiment?
(3) Was the voice in the first half of the experiment the same voice that said the words in the second half of the experiment?

These questions were adapted from the naturalistic change blindness experiment by Simons and Levin (1998). Responses to each question were also recorded by the experimenter.

Results

Explicit Measure of Change Deafness

Of the 12 participants who heard the *same* voice in both halves of the experiment, all responded "yes" to question number three, indicating that they had indeed heard the same voice in both halves of the experiment. Of the 12 participants who heard *different* voices in both halves of the experiment, 7 noticed the change in the talker either by stating that the voice was different in response to questions one or two, or by answering "no" to the third question. The remaining 5 participants (42%) did not state that the talker changed when asked questions one and two, and answered "yes" in response to question three, indicating that they failed to detect the change in the talker.

Implicit Measure of Change Deafness

A mixed 2X 3 ANOVA (lexical difficulty as a within factor and talker condition as a between factor) was used to examine the response latencies of the correctly repeated words in the second half of the experiment. Lexical difficulty refers to the easy-hard manipulation among the words. Talker condition was determined by whether a change in talker was presented and if that change was explicitly detected. Listeners who received different talkers in the experiment and failed to explicitly detect the change in the talker are labeled NO in Figure 1. Listeners who received different talkers in the experiment and

explicitly detected the change in the talker are labeled YES in Figure 1. Listeners that received only one talker through out the experiment are labeled SAME in Figure 1.

A main effect of lexical difficulty was found (F (2, 21) = 24.76, $p < .001$) such that easy words (mean = 908 ms) were repeated more quickly than hard words (mean = 934 ms). This result replicates previous studies examining lexical difficulty (e.g., Kirk et al., 1997; Luce & Pisoni, 1998; Sommers, 1996).

The main effect of talker condition was not statistically significant ($F < 1$), but it was in the direction that one might predict based on the results of Goldinger (1998): participants who received the same talker throughout the experiment tended to repeat the words in the second half of the experiment faster than participants who heard different talkers in each half of the experiment. The lack of a significant main effect of talker condition is not unexpected in the present experiment given that *different* easy and hard words were used in each half of the present experiment, whereas Goldinger (1998) used the *same* words in training and test sessions.

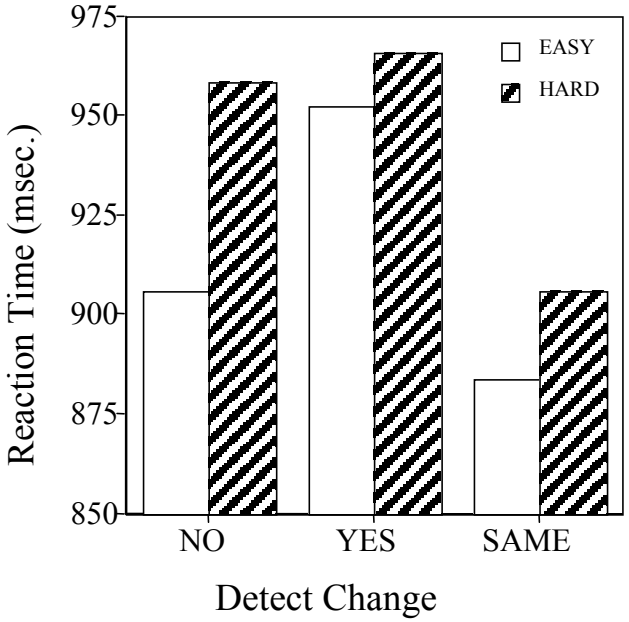


Figure 1. Reaction times from the second half of the experiment to easy and hard words from participants who failed to detect a change in the talker (NO), participants who detected the change in the talker (YES), and participants who received the same talker in both halves of the task (SAME).

Of greatest interest is an interaction between lexical difficulty and talker condition (F (2, 21) = 3.27, $p < .05$). Specifically, the participants that failed to detect the change in the talker had a larger difference between easy and hard words (52 ms) than the participants that detected the change in the talker (13 ms) and the participants that received the same talker throughout the experiment (22 ms). These results suggest that even though participants did not explicitly detect the change in talker, they were implicitly affected by the change in the talker. The results are displayed in Figure 1. No differences in accuracy rates were found (all F's < 1).

Discussion

The results of the present experiment demonstrate that failures to detect changes occur in the auditory modality as well as the visual modality. Forty-two percent of the participants that heard two different talkers failed to report this change when explicitly questioned about it. Thus, participants may experience "deafness" as well as "blindness" to changes in stimuli. The slightly smaller percentage of participants who failed to detect the change in the present study (42%) compared to other studies of change detection (e.g., 70% in Grimes (1996) and 67% in Experiment 2A of Levin & Simons, 1997) could be due to the differences between auditory speech stimuli and visual stimuli. Speech is a stimulus that is distributed through time (e.g., Marslen-Wilson & Tyler, 1980), whereas visual stimuli are not. Alternatively, the speech used in the present experiment may not have been complex *enough* to be equivalent to the visual stimuli employed in some change detection tasks (e.g., Levin & Simons, 1997). The stimuli used in the present experiment were one-syllable words recorded and presented on high-quality audio equipment with minimal background noise. Perhaps if the words were mixed with noise, the "auditory scene" might be comparable in complexity to the stimuli typically used in visual experiments.

More interestingly, the results of the present experiment demonstrate that even when there was not *explicit* evidence that participants detected the change in talker, there was *implicit* evidence that the change affected processing. Specifically, individuals that were deaf to the change had a greater difference between easy and hard words than participants that detected the change in talkers or participants that had the same talker throughout the experiment. The use of implicit and explicit measures in change detection experiments (see Chun & Nakayama, in press; Hayhoe, in press; Williams & Simons, in press) may provide important insights into cognitive and perceptual processing. For example, work by Nosofsky (1987) suggests that certain stimulus characteristics, or dimensions, of representations in memory can be "stretched" to emphasize a salient aspect. In the present experiment, the difference in reaction times as a function of change detection may be due to different individuals stretching different dimensions of the stimulus to varying degrees. The participants that failed to detect the change in talker may have emphasized the lexical difficulty dimension at the expense of the talker dimension of the spoken words. In contrast, the participants that detected the change in the talker may have equally emphasized the dimensions of talker and lexical difficulty.

Although speculative, this attentional-weighting hypothesis is anecdotally supported by a statement from a participant who failed to detect the change in the talker. When the participant was told during the debriefing of the experiment that there were two different talkers, the participant stated that "I was concentrating so much on *what* he was saying I didn't pay attention to the voice." The results of Werner and Thies (in press) also support this hypothesis (see also Shapiro, in press). Werner and Thies found that participants with greater expertise in American football were more likely to detect changes in football images compared to participants with less expertise in American football, suggesting that domain-specific expertise may influence which dimensions of a stimulus are stretched, thereby influencing the detection of changes.

In summary, this experiment demonstrates the existence of *change deafness*. The inability to explicitly detect changes in the auditory domain suggests that change detection is related to attentional demands and is not unique to visual processing. Furthermore, this experiment demonstrates the importance of using *implicit* as well as *explicit* measures of change detection. Although some participants did not explicitly detect the change in the talker, implicit measures of response latency suggest that the change did affect the perceptual and cognitive processing of these participants. Finally, these results support an explanation of change detection based on the distribution of attention-weights and the stretching of stimulus dimensions.

References

Chun, M.M. & Nakayama, K. (in press). On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*.

Cohen J.D., MacWhinney B., Flatt M., and Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257-271.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251-279.

Goldinger, S.D., Pisoni, D.B. & Logan, J.S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 152-162.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (ed.) *Perception: Vancouver Studies in Cognitive Science* (Vol. 2). pp. 89-110. Oxford University Press.

Hayhoe, M. (in press). Vision using routines: A functional account of vision. *Visual Cognition*.

Henderson, J.M. & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*, 438-443.

Kirk, K.I., Pisoni, D.B., Miyamoto, R.C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech and Hearing Research, 40*, 1395-1405.

Kucera, H. & Francis, W.N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Levin, D.T., Momen, N., Drivdahl, S.B. & Simons, D.J. (in press). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition.*

Levin, D.T. & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4*, 501-506.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.

Marslen-Wilson, W.D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-71.

Martin, C.S., Mullennix, J.W., Pisoni, D.B. & Sommers, M. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 676-684.

Mullennix, J.W., Pisoni, D.B. & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nosofsky, R.M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 87-108.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception, Progress Report no. 10.* Speech Research Laboratory. Psychology Department, Indiana University, Bloomington, Indiana.

Nygaard, L.C., Sommers, M. & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309-328.

Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9-32). San Diego: Academic Press.

Rensink, R.A., O'Regan, J.K. & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.

Shapiro, K.L. (in press). Change Blindness: Theory or paradigm? *Visual Cognition*.

Simons, D.J. & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, *5*, 644-649.

Sommers, M.S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, *11*, 333-341.

Torretta, G.M. (1995). The "easy-hard" word multi-talker speech database: An initial report. *Research on Spoken Language Processing, Progress Report No. 20,* Speech Research Laboratory. Psychology Department, Indiana University, Bloomington, Indiana.

Werner, S. & Thies, B. (in press). Is "change blindness" attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Visual Cognition*.

Williams, P. & Simons, D.J. (in press). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition.*

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Speech Perception and Language Skills of Deaf Infants After Cochlear Implantation: A Review of Assessment Procedures and a Research Plan[1]

Derek M. Houston[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by Indiana University Strategic Directions Charter Initiatives Fund and by NIH-NIDCD Training Grant DC00012 and NIH-NIDCD Research Grant DC00064 to Indiana University.
[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana.

Speech Perception and Language Skills of Deaf Infants After Cochlear Implantation: A Review of Assessment Procedures and a Research Plan

Abstract. "Universal Newborn Hearing Screening" law will result in many more infants identified with hearing loss. Thus, many more infants will receive interventions such as cochlear implants at a very young age. In order to evaluate the benefits of receiving cochlear implants during infancy, speech perception and language skills of infants who receive cochlear implants must be assessed. However, most current procedures used to test infant speech perception and language skills have only been used with normal-hearing infants. This paper reviews procedures that have been used to assess sound detection (Behavioral Observation Audiometry (BOA) and Visual Reinforcement Audiometry (VRA)), speech discrimination (VRA, High Amplitude Sucking (HAS), Visual Habituation (VH) Procedure), word learning (VH and Preferential Looking Paradigm (PLP)), and sensitivity to regularities in the ambient language (Headturn Preference Procedure (HPP)). A research plan is outlined and described to adapt the PLP and VH procedures for use with infants who use cochlear implants.

Introduction

Over the past 30 years, technological advances in cochlear implants (CIs) have allowed a growing number of people who are profoundly deaf to perceive sound and understand speech. Hearing-impaired listeners with CIs often show remarkable skills in perceiving and understanding speech and producing spoken language (Dorman, Hannley, Dankowski, Smith, & McCandless, 1989; Miyamoto, Kirk, Robbins, Todd, & Riley, 1996). For instance, many post-lingually deafened adult CI users are able to converse over the telephone without additional sensory aids, and many prelingually deaf children with CIs appear to be able to acquire spoken language normally and enter mainstream school systems. The benefits observed from CI use have led to a broadening of candidacy criteria for receiving a CI. In 1990, the FDA approved CIs for prelingually deaf 2-year-old children. In 1998, this criterion was lowered to 18-month-olds. To accommodate this trend, researchers have developed several new behavioral techniques for evaluating CI benefits that are appropriate for younger and younger children (see Kirk, Diefendorf, Pisoni, & Robbins, 1997). For example, to assess word recognition skills of children who cannot yet read, several behavioral tests require children to point to pictures (as opposed to written words) that correspond to the words they are presented with auditorily (e.g., Geers, 1994).

While steady progress has been made in developing new assessment techniques for speech and language, at the present time, measuring and assessing these skills in children and infants who are too young to follow instructions has been extremely difficult. The only current methods of assessing outcome rely on parent questionnaires (Hayes & Northern, 1996). Having appropriate behavioral performance measures of spoken language skills for young CI users' is critical at this time for several reasons. Last year, the FDA lowered the age criterion for candidacy again -- this time down to 12-months of age. As a result, many infants in the U.S. can receive CIs well before the age at which current behavioral techniques are able to assess their speech perception and language skills. In Europe, infants younger than 6-months have been successfully implanted. Furthermore, position statements and guidelines from the Joint Committee on Infant Hearing (2000) and the American Academy of Pediatrics (1999) are persuading most state lawmakers into implementing "Universal Newborn Hearing Screening" laws, which require hospitals to test the hearing of all newborns. As newborn hearing screening is implemented in hospitals, many more young infants will be identified with a hearing loss and these children will become potential candidates for CIs. Measuring and tracking the perceptual and linguistic development of young prelingually deaf infants who receive CIs will be necessary to assess the possible benefits of performing

the CI surgery at very young ages. However, the current battery of behavioral tests used to measure CI users' spoken language performance was designed for older children (≥ 2 years) and are clearly not appropriate for young infants (≤ 2 years). As the number of young infants who receive CIs increases, it is important that researchers and clinicians develop new methodologies and behavioral techniques to measure the perceptual and linguistic skills of infants and how these processes change over time.

In order to have a full description of the auditory, perceptual, and linguistic progress of infants who use CIs, several speech perception and language skills must be assessed using behavioral techniques. These skills include (but are not limited to): (1) *sound detection*, which is assumed to be a prerequisite for other auditory speech perception abilities (Geers & Moog, 1989); (2) *speech discrimination*, which is the ability to detect differences without necessarily understanding their significance; (3) *word-learning*, which requires the ability to recognize the sound patterns of words and associate them to their referents; and (4*) sensitivity to phonological regularities*, which involves the ability to store in long-term memory language-specific properties and sound patterns in the ambient language.

Over the past 30 years, several methods have been developed to investigate the speech perception and language skills of normal-hearing infants. While audiologists routinely assess hearing-impaired infants' ability to detect simple tones and speech sounds, no procedures have been established to assess hearing-impaired infants' speech discrimination, word learning abilities, and sensitivity to phonological regularities. The sections below briefly review methodologies that have been used to investigate these abilities in normal-hearing infants. A summary of current work-in-progress that employs two of these methodologies to assess speech perception and linguistic skills of profoundly deaf infants who use cochlear implants is provided at the end.

Infant Speech Perception Abilities and Methods of Assessment

Sound Detection

The two primary methods that audiologists use for assessing sound detection and auditory thresholds of infants are *Behavioral Observation Audiometry* and *Visual Reinforcement Audiometry*.

Behavioral Observation Audiometry (BOA). During a BOA procedure, the examiner presents sound out of sight from the infant, using either sound field signals or a manually operated noisemaker, and observes the infant's responses (Northern & Downs, 1991). The type of responses that audiologists look for from an infant depends on the infant's age. For infants between 0 and 4 months of age, an arousal response from sleep (e.g., eye opening, eye blink, movement in arms, legs or body) is the most common observation. Older infants are more likely to be awake during testing and are more likely to show a headturn toward the sound source (Watrous, McConnell, Sitton, & Fleet, 1975). The BOA is very easy to implement. However, many infants will respond to sound only a couple of times and then stop, presumably from boredom and habituation. Hence, it is often difficult for audiologists, using BOA, to assess infants' hearing levels at more than a few frequencies (Northern & Downs, 1991).

Visual Reinforcement Audiometry (VRA). The VRA is a method that relies on conditioning infants to orient to a reinforcer (Primus, 1992). The infant typically is seated with the caregiver in a sound booth while the audiologist is in a control room. In the booth is a reinforcer usually to one side of the infant, although sometimes there may be a reinforcer on both sides. Often, the reinforcer is a mechanical stuffed animal in a Plexiglas box, which can be illuminated and animated to keep the infant's interest (Moore, Thompson, & Thompson, 1975). A loudspeaker sits just below the reinforcer. Also, in the booth is an "attention-getter" directly in front of the infant to draw the infant's attention away from the side. Typically, the attention-getter is either a blinking light, a second mechanical stuffed animal, or an

assistant or a parent entertaining the infant with silent toys (Gravel, 1997). The audiologist, assistant, and caregiver listen to masking music over headphones so they do not know which trials have the change and hence will not bias the infant looking toward the reinforcer.

At the beginning of each trial, the infant's attention is brought to the center by the attention-getter. During the conditioning phase, a tone that is considered likely to be well above the infant's auditory threshold is presented over a loudspeaker in the booth. The reinforcer is presented to the infant if s/he demonstrates an orienting response to the loudspeaker. If there is no orienting response during a trial, then no reinforcement is given. The intensity is increased until the infant orients to the loudspeaker. The idea is to build an association between the tone and the reinforcer (Thompson & Folsom, 1984). Once the infant is conditioned to look towards the reinforcer in response to sound, the audiologist presents lower intensities to assess the infant's auditory thresholds. Because head-turn responses in VRA are conditioned by an interesting reinforcer, infants usually stay engaged in VRA longer than in BOA, allowing audiologists to make more complete assessments (Hayes & Northern, 1996).

Speech Discrimination

Normal-hearing newborns are able to discriminate global, rhythmic properties of speech, and, by the first couple months of life, they are able to make fine-grained discriminations of phonemes and syllables (see Jusczyk, 1997 for a review). The most common behavioral procedures for assessing infants' speech discrimination abilities are the *Conditioned Head Turn Procedure* (a modified version of the VRA), the *High Amplitude Sucking Procedure*, and the *Visual Habituation Procedure*.

Conditioned Head Turn (CHT) Procedure. The CHT is identical to the VRA except that, rather than conditioning the infant to respond to the presence of sound, the experimenter conditions the infant to respond to a change in a sequence of speech stimuli. For example, Kuhl (1979) used this procedure to test infants' ability to discriminate vowels. One vowel was repeated several times (e.g., /i/, /i/, /i/) and then a different vowel was presented (e.g., /o/, /o/, /o/). Infants were rewarded with a visual reinforcer only when they responded to a vowel change. As with the VRA, the experimenters and caregiver listen to masking music over headphones during the procedure.

The CHT can be used with 5- to 18-month-olds, but it is most commonly used with 6- to 10-month-olds (Werker et al., 1998b). The CHT has three stages: (1) A *training stage*, in which the infant is presented with the reinforcer immediately after a new stimulus is presented; (2) a *conditioning stage,* in which the experimenter gradually introduces increasingly longer delay periods between change and reinforcer until the infant performs a criterion number of anticipatory head turns; and (3) a *test phase,* in which the computer randomly presents test and control trials (i.e., the vowel stays the same). Statistical analyses are used to see if infants are more likely to look to the reinforcer during test trials than during control trials. A significant difference in the predicted direction indicates that infants can discriminate the two sounds tested. One disadvantage of this procedure is that many infants do not complete the *conditioning phase* because they never meet the criterion of (usually) three anticipatory looks in a row. The attrition rate for this procedure is in the range from low (~5%) to quite high (~50%) (Werker et al., 1998b).

High Amplitude Sucking (HAS) Procedure. The HAS was originally developed to see if infants would change their sucking behavior in response to changes in visual stimuli (Siqueland & DeLucia, 1969). The procedure was then adapted for use in speech perception. The HAS was the procedure used in the very first infant speech perception experiment, which showed that 2-month-olds could discriminate differences in voice onset time between /ba/ and /pa/ (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The methodology has since been used to investigate young infants' ability to

discriminate many other fine-grained phonetic contrasts (see Jusczyk, 1997 for review). HAS has also been used to show that newborns can discriminate between languages that are rhythmically different from each other (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998). The procedure can be used successfully with normal-hearing newborns up to 4-month-olds.

In the HAS, the infant is given a nonnutritive pacifier that is linked, via a pressure transducer, to a computer, which registers each strong sucking response. The assistant who holds the pacifier listens to masking music over headphones and is unaware of the experimental conditions the infant is assigned to. Presentation of speech stimuli is contingent on a sucking response, giving the infant control of the stimulus presentation rate. The experiment has three phases. During the *baseline phase,* the infant's baseline sucking rate (i.e. number of sucks per minute) is assessed without any speech stimuli. The sensitivity of the equipment is tailored to each infant so that the baseline sucking rate for all infants falls within a pre-established range (typically 15-35 sucks per minute). During the *habituation phase,* one stimulus type is presented until the infant's sucking rate slows and reaches a habituation criterion. The computer keeps track of the number of sucks per minute. In one version of the procedure (see Jusczyk, 1997), the infant must first exhibit a sucking rate above baseline on at least one trial before s/he is allowed to reach the habituation criterion. Habituation is reached as follows. The computer codes the first 1-minute trial above baseline as a "high" trial. The subsequent trial is a new high trial if the sucking rate is at least 75% of the previous trial. If the sucking rate is less than 75% of the previous high trial, then it is considered a "low" trial (i.e., two consecutive trials 75% or less than the previous high trial). The infant reaches the habituation criterion when s/he has two consecutive low trials. At this point, for infants in the experimental group, a new stimulus is presented during the *post-switch phase* for four 1-minute trials. Infants in the control group continue to hear the same stimuli during the *post-switch phase.* The differences in sucking rates between the first two trials of the *post-switch phase* and the last two trials of the *habituation phase* are compared for the experimental and control groups. A significantly greater sucking rate increase in response to the new stimuli after the stimuli are switched is interpreted to mean that infants can discriminate between the two types of stimuli.

The HAS has an even higher attrition rate than the CHT – about 50% or more. The experiment is stopped when infants fall asleep, begin to cry, do not meet the habituation criterion, or simply do not suck on the pacifier. However, because the HAS procedure uses sucking response, which is mastered soon after birth, rather than visual orientation, HAS remains the most common procedure for investigating speech perception of infants younger than 5 months (Jusczyk, 1997).

Visual Habituation (VH) Procedure. The VH procedure is based on the premise that infants will increase their visual fixation times in the presence of a novel stimulus. VH has long been used to investigate infant visual perception (e.g., Cohen, 1969; Kagan & Moss, 1965). In the mid-70s, Horowitz (1975) showed that infants will look longer at a visual display when they are listening to an interesting auditory stimulus. Since then, researchers have used visual habituation to design paradigms that test infant speech perception abilities. For example, VH has been used extensively to show that infants are able to discriminate between native and nonnative phoneme contrasts (e.g., Best, McRoberts, & Sithole, 1988; Polka & Werker, 1994). The basic idea is that over repeated presentations of a single auditory stimulus paired with a visual stimulus, visual fixation to the visual stimulus will eventually decrease. If a novel auditory stimulus is then presented with the same visual stimulus, and infants can discriminate the two auditory stimuli, then visual fixation times should increase (Horowitz, 1975; Werker et al., 1998b).

In the VH procedure, the infant is seated on the caregiver's lap in front of a TV monitor through which the visual and auditory stimuli are presented. There is very little else in the room to distract the infant. One experimenter is in a separate control room and manipulates the presentation of the stimuli. The infant's looks to the monitor are observed and recorded on a computer keyboard or with a button box

in one of two ways. Either the first experimenter watches the infant via a closed-circuit video camera (placed inconspicuously in front of the infant) and monitor or a second experimenter, who is hidden from the infant, watches through peepholes. The experimenter(s) and caregiver listen to masking music over headphones.

The experiment has two main phases: habituation and test. At the beginning of each trial in both phases, the infant's attention is brought to the center with either a blinking light above the monitor or a graphic display presented on the monitor. When the infant looks at the monitor, the experimenter initiates a trial by pushing a button. During the habituation phase, the infant is presented with a simple visual display (e.g., a checkerboard pattern) and an auditory stimulus repeats (e.g., /da/, /da/, /da/...). The experimenter holds down the button as long as the infant continues to look at the monitor. When there is a look away, the experimenter releases the button but pushes it again if and when the infant looks back to the monitor. The stimuli continue until the infant looks away from the monitor for 1 second or more. When the trial ends, the center attention getting stimulus is presented again. The total time that the infant looks at the monitor is summed and recorded for each trial. The habituation trials continue until the infant reaches a habituation criterion. For example, the experimenter may set the habituation criterion to be three consecutive trials where the cumulated looking time for each trial is 50% or less than the average looking time of the first three trials.

In the VH procedure, there are at least two different ways that the experimenter can construct the test phase, depending on the design of the experiment. If the experimenter chooses a between-subjects design, then the trials in the test phase will all consist of either the same visual and auditory stimuli (control group) or the same visual display but different auditory stimuli (e.g., /ba/, /ba/, /ba/…) (experimental group). A difference is then calculated between the average looking time during the test phase and the average looking time during the final trials of the habituation phase. A significantly greater looking time difference between the experimental group and the control group is taken as evidence that infants can discriminate differences between the stimuli. A within-subjects design involves having both old and new stimuli during the test phase for each infant. For a within-subjects design, the order of the old and new stimuli trials must be counter-balanced across infants.

The VH procedure is currently the most commonly used habituation/dishabituation procedure for research on speech perception of infants. It has been used successfully with a wide age range of infants (2- to 14-month-olds). Another advantage of the VH procedure is that the attrition rate is relatively low (~20-25%). The procedure has also been extended to explore word-learning abilities (see below). For a more extensive review of the VH procedure, see Werker et al. (1998b).

Word Learning and Recognition

During the first year of life, infants' speech perception and language skills develop very rapidly. By five months of age, infants can not only discriminate speech sounds, but they can recognize very familiar sound patterns of words, such as their own names (Mandel, Jusczyk, & Pisoni, 1995). Infants also learn to associate the sound patterns of words to their referents. Recent findings have shown that infants begin to associate words to very familiar objects (e.g., their parents) by six months (Tincoff & Jusczyk, 1999), and, by 12-months, infants can identify the meaning of up to 50 words (Fenson et al., 1994). The VH has also been used by some researchers to assess word-learning abilities. Many researchers have used the Preferential Looking Paradigm and its variants to assess word learning and word recognition abilities.

Visual Habituation (VH) Procedure. Recently, a variation of the VH procedure has proven to be successful in exploring infant word learning (Stager & Werker, 1997; Werker, Cohen, Lloyd,

Casasola, & Stager, 1998a). In this variation, infants are habituated to two visual object/auditory label pairs (e.g., V1-A1 and V2-A2). During the test phase, on half of the trials the infants are presented with the same pairings used during the habituation phase. On the other half of the trials, the pairings of visual objects and auditory labels are switched (e.g., V1-A2 and V2-A1). If infants are able to form associations between objects and labels, then they will demonstrate dishabituation (i.e. show longer looking times) to object/label mismatches.

Preferential Looking Paradigm (PLP). Thomas, Campos, Shucard, Ramsay, and Shucard (1981) showed that 1-year-olds will consistently fixate on objects longer when they hear the name of the object than when they hear a nonsense word. Using this basic finding, Golinkoff, Hirsh-Pasek, Cauley, and Gordon (1987) developed a procedure in which infants are presented with two objects side-by-side on TV monitors. At the same time the visual displays are presented, the name of one of the objects is presented several times over loudspeakers. For example, Tincoff & Jusczyk (1999) showed that when presented with an image of their mother on one side and their father on the other side, 6-month-olds will attend longer to the "correct" parent when hearing a synthesized voice repeating either *mommy* or *daddy*.

The standard set-up for the PLP consists of a single plain display wall (approximately 6'x6') with two square holes side-by-side to reveal two monitors and a third hole in the center that allows a video camera to record the infant's looking responses. The infant is seated on the caregiver's lap approximately 5' in front of the display wall. The monitors sit such that they are at about eye level and approximately 30º left and right of center from the perspective of the seated infant. The camera hole is about 5cm and is well above the monitors. There is an attention getting device (typically a blinking light or display of several small lights) centered between the monitors. Behind the display wall are the monitors, camera, and two VCRs, each of which plays stimuli over one of the TV monitors. The experimenter sits either behind the wall or in a control room. The experimenter controls the stimuli with the VCRs and the attention getting light with a switch or button box. The caregiver wears a visor with a piece of cloth hanging from it so that they cannot see the displays on the monitors and potentially influence the infant's looking behavior. The room is dimly lit, and there are no other stimuli in the room that can distract the infant.

During a word recognition experiment, two words are selected along with two visual displays that correspond to the words. For example, the words *apple* and *flower* would be paired with a picture of an apple and a picture of a flower. Verbs and prepositions can be represented with actors performing actions that correspond to the meanings. At the beginning of each trial, the infant's attention is brought to the center with the attention getting light. In an experimenter-controlled version of the PLP, the experimenter observes the infants via the closed circuit video and monitor system and initiates a test trial only after the infant looks to the center. In another version of the PLP, the stimuli just play out straight through the experiment, and the experimenter simply turns on the attention getting lights for a pre-established amount of time (e.g., 2s) in between presentations of the video and audio stimuli.

During the first trial or two of a PLP experiment, infants are presented with both video displays without auditory stimuli in order to get a baseline measure of any bias to look at one display or the other. Following this *saliency phase,* auditory and visual stimuli are presented during the *test phase.* The auditory stimuli are presented via hidden loudspeakers that are either centered or equidistant left and right of the infants. Both visual displays are presented during each test trial after the infants are centered with the attention getting light, but only one word is presented over the loudspeakers. For example, the infants might hear: "Where's the apple? Can you see the apple? Look at the apple. Apple!" when visual displays of both an apple and a flower are presented. The first sentence of the auditory stimuli plays before the visual stimuli begin in order to allow the infants to show an anticipatory response toward the correct side.

The visual objects are always presented on the same sides. Both auditory stimuli are presented several times (about 4 to 8 times each), usually in a random or semi-random order.

Coding of the infants' looking times is computed offline using a VCR and a monitor. A time code must be burnt onto the coding tape either during the testing session while recording the infant or offline. The coder is kept blind to the experimental conditions by muting the volume of the monitor. While coding the videotape, the coder can see when each trial begins and ends by paying attention to the light from the visual stimuli reflecting off of the infants' faces.[3] The coder steps though the trial, frame by frame, and records the looks to the left, right, center, and away. After coding, the data are separated by condition (e.g., *apple* vs. *flower*), and the left and right looks are averaged for each condition. The data can be analyzed several different ways in order to determine if the infants' look more to the correct objects when they are being named. In case infants are more likely to know one object better than the other, researchers often choose to analyze one condition at a time. For example, one might calculate if infants tend to look at the apple more than the flower when *apple* is presented independently of analyzing the reverse case. Often, researchers analyze the looking behavior during the *test phase* and compare it to what was found during the *saliency phase.* Whatever the details of the statistical analyses, the basic idea of the procedure is to see if infants' look more often and longer to one object when it is being named than when the other object is being named – and vice versa.

Recently, Swingley and his colleagues have developed a variant of the PLP to assess infants' speed of word recognition (Swingley, Pinto, & Fernald, 1998). In their procedure, they calculate not only the amount of time infants look toward the "correct" monitor, but also infants' latency to initiate an eye movement toward the correct monitor. This modification allows researchers to explore the time course of lexical retrieval from long-term memory by infants (e.g., Swingley, Pinto, & Fernald, 1999).

The PLP can also be used for word learning with the addition of a *training phase*. The *training phase* is introduced immediately following the *saliency phase*. During the *training phase,* infants are presented with one or more new objects and words. On each trial, only one of the objects (on the left or right side) and a novel word (e.g., *blick*) are presented. For example, infants might see one object on the right side and hear, "Where's the blick? Do you see the blick? Look at the blick. Blick!" After several trials, infants form an association between the visual objects and the spoken words or nonword sound patterns. The *test phase* is the same as in the word recognition design – visual objects that were taught appear together side by side, and, on each trial, only one of the novel words is presented. If the infants can form the correct association, then they will look longer to the direction of an object when they hear its label than when they hear the label of a different object.

The PLP is the most commonly used procedure for investigating spoken word recognition skills and word learning abilities in infants and young children. The procedure has been used successfully with infants ranging in age from 6 months (e.g., Tincoff & Jusczyk, 1999) to 3 years (Naigles, 1998). The attrition rate is also relatively low (about 10-20%).

Variants of the PLP

The Intermodal Preferential Looking Paradigm (IPLP). For some research questions, it is important that the experimenter have an opportunity to interact with the infant during the experimental procedure. For example, if a researcher wants to explore the effects of eye gaze on word learning, then s/he needs to employ a procedure that allows the infant to see where the experimenter is looking during

[3] Alternatively, the experimenter may set up a small light to turn on during each trial of the experiment, which could be placed behind the infants such that it will be recorded by the camera.

the experiment. The IPLP replaces the monitors with a modified Fagan Board. The Fagan Board is a hinged 40 cm x 50 cm flip board that allows the experimenter to Velcro objects to it. The experimenter can flip it back and forth for quick hiding and displaying of the visual stimuli. In the IPLP, the experimenter produces the auditory stimuli using live voice. The first phase of the IPLP is the *exploration phase,* in which the infants are allowed to physically interact with the objects. Next is the *saliency phase*, which is the same as in the PLP – except that the objects are attached by Velcro to the Fagan Board rather than on TV monitors. The next phase is the *labeling phase* – when the experimenter can either look at the object or look away from the object during labeling, depending on the condition. Finally, the *test phase* is identical to the PLP. During this phase, the experimenter hides behind the Fagan board so as to not influence where the infants look.

The Split-Screen PLP. The Split-Screen PLP was developed by Hollich and colleagues in order to facilitate stimuli creation and testing (Hollich, Rocroi, Hirsh-Pasek, & Golinkoff, 1999). In the Split-Screen version of the PLP, a large wide-aspect TV monitor replaces the two individual monitors. Two visual objects appear on different sides of the same monitor rather than on two different monitors. Using a single monitor rather than two separate displays allows for perfect synchronization of the visual stimuli and requires operating only one VCR, rather than two, during testing. The stimuli are made easily by first recording them using a digital camera and then splicing them together using a digital editing program.

Sensitivity to Phonological Regularities

An important aspect of language development is learning language-specific properties. Recent findings suggest that during the first year of life, infants become sensitive to many language-specific properties in the speech signal. For example, 9-month-olds, but not 6-month-olds attend more to lists of words that contain sequences of sounds that are common in the ambient language in their environment than those that are rare or do not occur (Friederici & Wessels, 1993; Jusczyk, Luce, & Charles-Luce, 1994). Findings like these are important for understanding what properties normal-hearing infants are sensitive to during early language development. Infants' sensitivity to regularities in the sound pattern of spoken language reveals not only that they discriminate speech sounds, but that they also encode them into long-term memory and are able to notice common properties. A method that has been extremely helpful in assessing infants' preferences and sensitivities to properties in speech is the Head Turn Preference Procedure.

Head Turn Preference Procedure (HPP). The HPP was first used by Fernald and her colleagues to show that infants prefer infant-directed speech that contains greater pitch peaks and more exaggerated pitch contours than adult-directed speech (Fernald, 1985; Fernald & Kuhl, 1987). In the HPP, the infant is seated on the caregivers' lap in a 3-sided pegboard booth. There is a green light on the front wall and a red light on each of the side walls. A small hole and a video camera are just above the center light. The caregiver and experimenter listen to masking music over headphones so that they cannot influence the outcome of the experiment. The experimenter controls the lights and auditory stimulus, using a button box. The infant's behavior is observed either through holes in the pegboard or in a control room via a closed-circuit video camera and monitor. The experimenter also uses the button box to record the infant's responses online.

In Fernald's version of the HPP, the infant first completes a short training phase in which s/he is presented with one stimulus type at a time on alternating trials. Each stimulus is paired with one of the two blinking lights on the sides. At the beginning of each trial during the test phase, the infant's attention is first brought to the center by the center green blinking light. When the infant looks to the center-light, the light is extinguished and the red side-lights begin blinking. When the infant looks 30º towards one of

the lights, the corresponding stimulus type is presented from behind the light. A preference is indicated if, on average, infants orient significantly more often to one sound pattern than the other.

Recently, Jusczyk and colleagues have used a modified version of the HPP to explore infants' sensitivity to a number of different properties (Jusczyk, 1997; Kemler Nelson et al., 1995). In order to avoid potential side biases, the HPP was modified such that the training period was eliminated and the stimuli were presented randomly to either the left or right side for each trial. Rather than measuring which side the infant orients to, the experimenter measured the average duration of orientation to each stimulus type. At the beginning of each trial, the center-light blinks until the infant looks at it. When the infant is oriented to the center, the experimenter pushes a button on the button box that extinguishes that light and causes one of the side-lights to begin blinking. When the infant orients to the blinking light, the experimenter bushes another button and speech stimuli play from a loudspeaker hidden behind the blinking light. The experimenter responds with button presses each time the infant looks either toward or away from the blinking light. The blinking light and speech continue to play until the infant looks away for two seconds, up to a maximum trial length of about 30 seconds. When the infant orients away from a blinking light but returns within two seconds, the stimuli continue. The amount of time the infant orients to the blinking light is summed for each trial automatically by a computer connected to the button box. For example, in one investigation, Jusczyk, Cutler, and Redanz (1993) showed that 9-month-old English-learning infants orient longer, on average, to lists of words (e.g., *doctor, pliant,* etc.) that follow the predominant stress pattern of English words (i.e., strong/weak) than to lists of words (e.g., *guitar, deride,* etc.) that have the opposite stress pattern (i.e., weak/strong).

Other versions of the HPP have been developed to explore issues of word segmentation and other aspects of language development during the first two years of life (Jusczyk, 1997). In one study, Jusczyk & Aslin (1995) explored infants' ability to recognize the sound pattern of words in sentences. They first familiarized infants with two words by repeating them, one at a time, in citation form. Then they presented sentences, some of which contained the familiarized words and others that contained unfamiliarized target words. By eight months of age, infants attend significantly longer to passages containing familiarized words than to passages containing unfamiliarized target words (Jusczyk & Aslin, 1995). The HPP has also been used to show that infants can recognize familiarized words after delays of one day and longer (Houston & Jusczyk, 2001; Jusczyk & Hohne, 1997). Hence, the HPP is an important methodology that can be used to explore infants' ability to encode speech information into long-term memory. The HPP is successful at exploring the phonological knowledge infants accumulate from exposure to their ambient language (e.g., Jusczyk et al., 1993) as well as testing what phonological information infants can encode during an experiment (e.g., Jusczyk & Aslin, 1995). The procedure has been used successfully with infants from 4.5 months to 2 years, and the attrition rate is about average for infant speech perception measures (approximately 25%).

Current Project

We are now developing an infant speech perception facility in the DeVault Otologic Research Lab in the Riley Children's Hospital ENT clinic. The primary focus of this lab is to make comparisons of speech perception skills of normal-hearing infants and hearing-impaired infants – primarily those who use cochlear implants but also those who use hearing aids for amplification. Very little is currently known about the speech perception skills of hearing-impaired infants, and it is likely that their skills are very limited. Thus, we begin by using the VH and the split-screen PLP to test basic speech perception skills. In the first experiment, we will test infants' ability to discriminate basic speech patterns: a continuous "ahhhh" sound versus a discontinuous "hop hop hop" and rising /i/ versus falling /i/. After the methodologies have proven successful for showing that infants can make these very basic

discriminations, we will explore other speech perception and language skills, such as more subtle phonetic distinctions (e.g., /i/ vs. /u/ and /S/ vs. /m/) and novel word learning.

Issues to be Addressed

Infants' Speech Perception and Language Skills. As described earlier, the VH and the PLP are ideally suited for investigating infant speech perception and language skills. The VH is probably the cognitively simpler of the two procedures. It is designed to measure infants' basic response to a change in auditory information – a startle response or a peak of interest that is shown by increased looking duration. For infants to show learning in the PLP, they must first make associations between the auditory and visual stimuli that are presented to them during the training phase. Infants must discriminate auditory stimuli and then, during the test phase, identify which auditory stimulus they hear in order to match it to the correct visual display in front of them. By using both the VH and the PLP, we hope to be able to measure normal-hearing and hearing-impaired infants' basic auditory discrimination abilities and, hopefully, also their ability to make auditory/visual associations that appear to be important for early word learning.

Validity of the Procedures for Hearing-Impaired Infants. The VH and PLP have never been used before to assess speech perception and language skills of hearing-impaired infants. Also, normal-hearing infants' abilities to discriminate gross pattern differences have not been explored with these procedures. Thus, it will be necessary to assess the validity of the procedures using both normal-hearing and hearing-impaired infants. It has been shown repeatedly over the years that normal-hearing infants can discriminate many subtle phonetic differences (e.g., Best et al., 1988), so it goes without question that they could discriminate gross phonetic-acoustic differences also. By using the VH and PLP to test normal-hearing infants' ability to discriminate "hop hop hop" from "ahhh" and rising /i/ from falling /i/ we will be able to demonstrate that both of these procedures are valid measures of these specific speech perception abilities.

Individual Differences. Another goal in developing new procedures for use in a clinical population is to be able to assess the abilities of individual infants so that speech perception skills can be tracked over time. Research in infant speech perception has been cross-sectional in design, and, as a result, the VH and PLP have not been used to investigate individual infants. In order to evaluate the VH and PLP as possible tools for clinical assessment of speech perception and language abilities in individual infants, both normal-hearing and hearing-impaired infants will be tested repeatedly using the same stimuli, each time they come in for their follow-up clinical appointments (i.e., at 1-month-intervals). If these new procedures are able to measure individual abilities, then we expect to see some consistency in performance over repeated measurements of the same infants.

Inter-Procedure Validity. It is possible that one of the procedures will be useful for testing speech pattern discrimination with normal-hearing and/or hearing-impaired infants, but that the other procedure will not provide a valid measure. To test this possibility the discriminations tested with the two procedures will be switched each month. For example, some infants will be tested on "hop hop hop" versus "ahhh" using the VH and rising /i/ versus falling /i/ using the PLP in one session. And then the next month, they would be tested on "hop hop hop" versus "ahhh" using the PLP and rising /i/ versus falling /i/ using the VH. Thus, both procedures will be used to test both discriminations of all the infants. By taking this approach, the procedures can be used to demonstrate the validity of the methodology. If normal-hearing and/or hearing-impaired infants can demonstrate the ability to make a particular discrimination using one procedure, then we would expect them to show a similar pattern with the other procedure. A pattern of results showing that infants could make a particular discrimination with one procedure but not a second procedure would suggest that the second procedure is not sensitive enough to infants' speech discrimination skills.

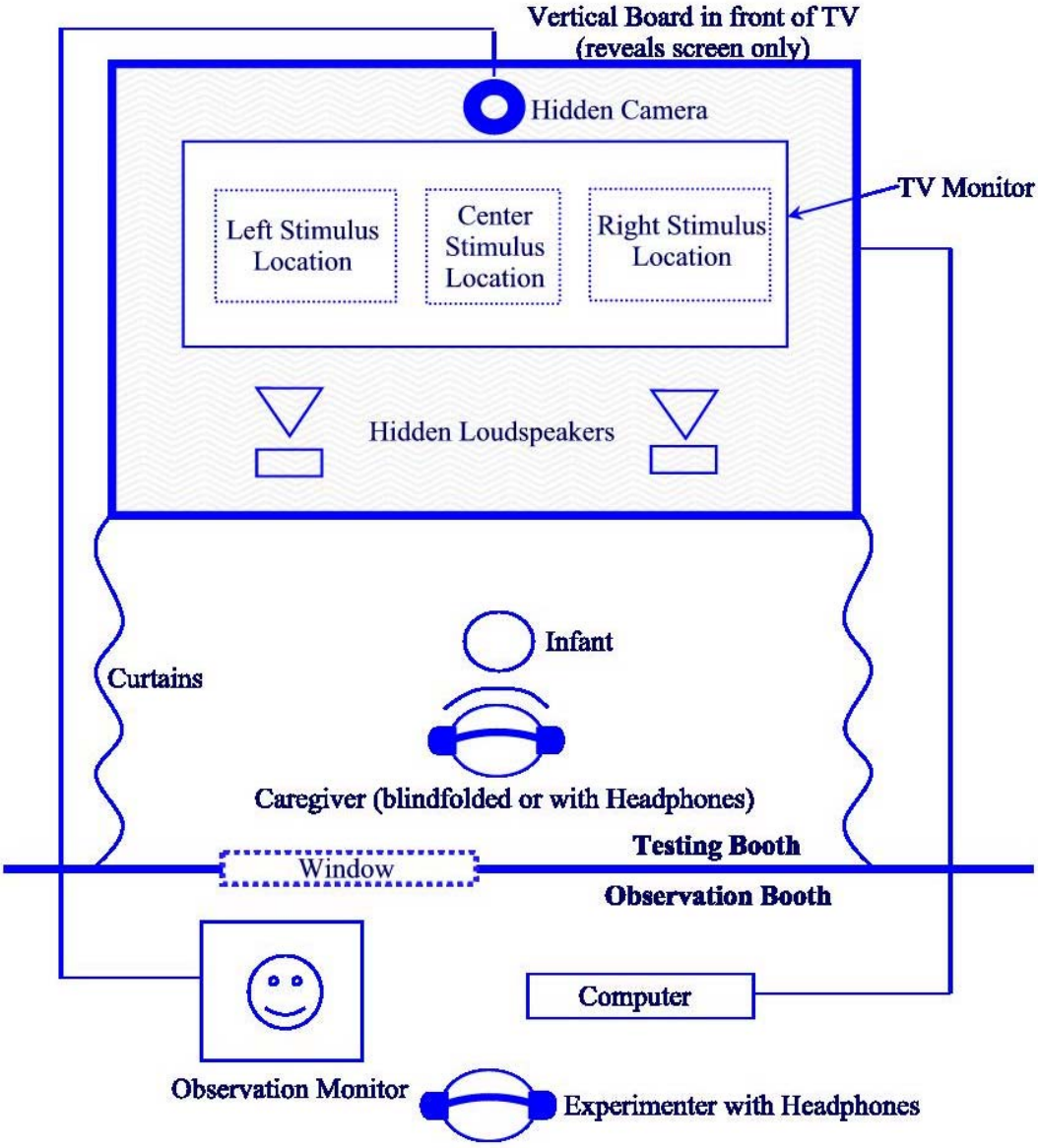


Figure 1. Set up for Preferential Looking Paradigm (PLP) and Visual Habituation (VH) Procedure. During PLP, the caregiver wears a visor as a blindfold, while during the VH, the caregiver wears headphones playing masking music. The Center Stimulus Location is where all the visual stimuli appear during VH and where the graphic of the infant appears in both procedures. The Right and Left Stimulus Locations is where the visual stimuli appear for the PLP only.

Set Up

The VH and PLP will be conducted in a single, soundproof room using the same equipment (see Figure 1). In both procedures, the infant will be seated on the caregiver's lap in front of a 55" wide-aspect monitor. A flat 6-1/2' X 6' wooden structure that is painted black sits in front of the television, revealing

only the monitor of the television so that the infant has nothing else to look at in front. Also, curtains hang from ceiling to floor to the left and right of the infant to prevent distraction by any other objects in the room. The experimenter controls the experiments from a control room adjacent to the soundproof booth. A camera records the infant through a hole in the wooden structure and displays the image onto a monitor in the control room. The experimenter wears headphones playing masking music during both procedures. The caregiver does the same during the VH but wears a visor instead during the PLP, so s/he cannot see which side visual stimuli are presented. During each session, the infant will be tested on his/her ability to discriminate one pair of stimuli ("hop hop hop" vs. "ahhh" or rising /i/ vs. falling /i/) using the VH and the other pair using the PLP. Both procedures will be implemented using Habit software (Cohen, Atkinson, & Chaput, 2000) on a Macintosh G4.

Procedure (VH). The procedure of the VH will be as follows. At the beginning of each trial, a graphic of an infant will appear at the center of the monitor. When the infant looks to the center, the experimenter will push a key on the keyboard, which will extinguish the center attractor and initiate the visual and auditory stimuli. A red and white checkerboard pattern will appear in the center of the monitor, and the infant will hear one of the stimulus items repeat. When the infant looks away, the experimenter will push another key, which will end the trial and begin the next trial. The trials will continue until the average orientation duration of three sequential looks is less than 50% of the average of the initial three looks. The test phase consists of two trials. The 'same' trials are identical to the trials during the habituation phase. The 'switch' trial consists of a novel auditory stimulus with the same visual display. The order of the 'same' and 'switch' trials will be counterbalanced across infants and within infants across sessions.

Procedure (PLP). The visual stimuli used in the PLP will be physically correlated to the speech stimuli. "Hop hop hop" will be paired with video of a toy kangaroo hopping. "Ahhh" will be paired with a video of a toy airplane moving from left to right across the screen. Rising /i/ will be paired with a video of white bubble rising up in a lava lamp. Falling /i/ will be paired with a video of a ball rolling down a plastic spiral ramp. The auditory and visual stimuli will be digitized onto the Macintosh G4, and EditDV™ will be used to create the visual "split-screen" effects and to synchronize the audio and visual stimuli. As with the VH, each trial will begin with a graphic of an infant on the center of the screen as an attention getter, and the test stimuli will be presented once the infant looks to the center. The procedure will consist of: *Saliency Phase* – one trial with both visual stimuli and no auditory stimuli; *Training Phase 1* – eight trials where the two visual/auditory stimulus pairs will be presented one at a time, the first half in alternating order and the second half in random order; *Test Phase 1* – six trials in random order where both visual stimuli are on the screen but only one auditory stimulus is presented; *Training Phase 2* – six more training trials of the same stimuli in random order; and *Test Phase 2* – six more test trials of the same stimuli in random order. The looks of the infants will be coded online by the experimenter but then will be double-checked for reliability by using the videotape that will record the infants during testing.

Clinical and Theoretical Significance

The VH and split-screen PLP methodology that is being developed in this project has important clinical and theoretical significance. From a clinical standpoint, at the present time it is absolutely essential that new behavioral techniques be developed that can be used to assess the benefit of implanting infants with CIs at very young ages. At this time, it is not known if providing CIs at increasingly younger ages will actually provide additional outcome benefits and help promote spoken language development in this population. With new measures of speech perception and novel word learning performance, clinicians will be able to assess the development of speech perception abilities of infant CI users and, as a result, they will become better able to make more informed decisions about the age at which infants should

undergo CI surgery. Being able to track the progress of individual CI users will also allow clinicians to determine when additional interventions may be necessary to improve outcome performance and help these children reach optimal levels of performance with their CIs.

Finally, from a theoretical perspective, it is of interest to compare language development of normally hearing infants to infants who are first deprived of auditory input and then receive it at a later age via a CI. Do these children follow the same developmental course as normal-hearing infants, even though their early auditory experience was radically different? Also, how does the initial absence of auditory information affect an infants' ability to acquire spoken language? Some language development researchers have hypothesized that there is a "sensitive period" in which the capacity to learn languages declines because of decreasing neural plasticity (e.g., Lenneberg, 1967; Newport, 1990). These important theoretical issues in neural and behavioral development can be explored for the first time in a pediatric population by investigating the language development of hearing-impaired infants who are deprived of auditory input during the early part of the sensitive period and then receive a CI. However, this unique research opportunity may be lost without appropriate behavioral techniques that can measure and track the changes in their perceptual skills over time after implantation. Adapting novel techniques like the PLP for use with the young CI population will allow us an unusual opportunity to investigate and measure the effects of early sensory deprivation on speech perception and spoken language acquisition and help us to understand the behavioral and neural basis for the large differences in outcome performance of CI users.

References

American Academy of Pediatrics Task Force on Newborn and Infant Hearing (1999). Newborn and infant hearing loss: Detection and intervention. *Pediatrics, 103*, 527-530.

Best, C.T., McRoberts, G.W., & Sithole, N.M. (1988). Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Cohen, L.B. (1969). Observing responses, visual preferences, and habituation to visual stimuli in infants. *Journal of Experimental Child Psychology, 7*, 419-433.

Cohen, L.B., Atkinson, D.J., & Chaput, H.H. (2000). Habit 2000: A new program for testing infant perception and cognition. (Version 1.0). Austin: the University of Texas.

Dorman, M.F., Hannley, M.T., Dankowski, K., Smith, L., & McCandless, G. (1989). Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant. *Ear & Hearing, 10*, 44-49.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.

Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (Serial number 242).

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*, 181-195.

Fernald, A., & Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*, 279-293.

Friederici, A.D., & Wessels, J.M.I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics, 54*, 287-295.

Geers, A.E. (1994). Techniques for assessing auditory speech perception and lipreading enhancement in young deaf children. In A. E. Geers & J. S. Moog (Eds.), *Effectiveness of cochlear implants and tactile aids for deaf children: The sensory aids study at central institute for the deaf.* (pp. 85-96). Washington, D.C.: Alexander Graham Bell Association for the Deaf.

Geers, A.E., & Moog, J.S. (1989). Evaluating speech perception skills: Tools for measuring benefits of cochlear implants, tactile aids, and hearing aids. In E. Owens & D.K. Kessler (Eds.), *Cochlear implants in young deaf children*. Boston: College-Hill Press.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14*, 23-45.

Gravel, J.S. (1997). Behavioral audiologic testing. In A. Lalwani & K. Grundfast (Eds.), *Pediatric otology and neurotology*. Philadelphia: J. B. Lippincott.

Hayes, D., & Northern, J.L. (1996). *Infants and hearing*. San Diego, CA: Singular.

Hollich, G., Rocroi, C., Hirsh-Pasek, K., & Golinkoff, R. (1999). *Testing language comprehension in infants: Introducing the split-screen preferential looking paradigm.* Paper presented at the Society for Research in Child Development Biennial Meeting, Albuquerque, NM.

Horowitz, F.D. (1975). Infant attention and discrimination: Methodological and substantive issues. *Monographs of the Society for Research in Child Development, 39* (Serial no. 158).

Houston, D.M., & Jusczyk, P.W. (2001). *Infants' long-term memory for words and voices*. Manuscript submitted for publication.

Joint Committee on Infant Hearing: Year 2000 position statement (2000). Principles and guidelines for early hearing detection and intervention programs. *American Journal of Audiology, 9*, 9-29.

Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science, 277*, 1984-1986.

Jusczyk, P.W., Luce, P.A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language, 33*, 630-645.

Kagan, J., & Moss, M. (1965). Studies of attention in the human infant. *Merril-Palmer Quarterly, 11*, 95-127.

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., & Gerken, L.A. (1995). The Headturn Preference Procedure for testing auditory perception. *Infant Behavior & Development, 18*, 111-116.

Kirk, K.I., Diefendorf, A.O., Pisoni, D.B., & Robbins, A.M. (1997). Assessing speech perception in children. In L. Mendel & J. Danhauser (Eds.), *Audiological Evaluation and Management and Speech Perception Training* (pp. 101-132). San Diego: Singular Publishing Group.

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66*, 1668-1679.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Mandel, D.R., Jusczyk, P.W., & Pisoni, D.B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6*, 315-318.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Miyamoto, R.T., Kirk, K.I., Robbins, A.M., Todd, S.L., & Riley, A.I. (1996). Speech perception and speech production skills of children with multichannel cochlear implants. *Acta Oto-Laryngologica, 116*, 240-243.

Moore, J.M., Thompson, G., & Thompson, M. (1975). Auditory localization of infants as a function of reinforcement conditions. *Journal of Speech and Hearing Disorders, 40*, 29-34.

Naigles, L.R. (1998). Developmental changes in the use of structure in verb learning: Evidence from preferential looking. In C. Rovee-Collier, L. P. Lipsitt & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 298-318). Stamford, CT: Ablex.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 756-766.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Northern, J.L., & Downs, M. P. (1991). *Hearing in children*. Baltimore: Williams & Wilkins.

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421-435.

Primus, M. (1992). The role of localization in visual reinforcement audiometry. *Journal of Speech and Hearing Research, 35*, 1137-1141.

Siqueland, E.R., & DeLucia, C.A. (1969). Visual reinforcement of non-nutritive sucking in human infants. *Science, 165*, 1144-1146.

Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381-382.

Swingley, D., Pinto, J.P., & Fernald, A. (1998). Assessing the speed and accuracy of word recognition in infants. In C. Rovee-Collier, L.P. Lipsitt & H. Hayne (Eds.), *Advances in Infancy Research* (Vol. 12, pp. 257-277). Stamford, CT: Ablex.

Swingley, D., Pinto, J.P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71*, 73-108.

Thomas, D.G., Campos, J.J., Shucard, D.W., Ramsay, D.S., & Shucard, J. (1981). Semantic comprehension in infancy: A signal detection analysis. *Child Development, 52*, 798-803.

Thompson, G., & Folsom, R.C. (1984). A comparison of two conditioning procedures in the use of visual reinforcement audiometry (VRA). *Journal of Speech and Hearing Disorders, 49*, 241-245.

Tincoff, R., & Jusczyk, P.W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*, 172-175.

Watrous, B. S., McConnell, F., Sitton, A. B., & Fleet, W. F. (1975). Auditory responses of infants. *Journal of Speech and Hearing Disorders, 40*, 357-366.

Werker, J.F., Cohen, L.B., Lloyd, V.L., Casasola, M., & Stager, C.L. (1998a). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*.

Werker, J.F., Shi, R., Desjardins, R., Pegg, J.E., Polka, L., & Patterson, M. (1998b). Three methods for testing infant speech perception. In A. Slater (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy.* East Sussex, UK: Psychology Press.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Memory Span and Sequence Learning Using Multimodal Stimulus Patterns: Preliminary Findings in Normal-Hearing Adults[1]

Jeff Karpicke and David B. Pisoni

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH Research Grant DC00111 and Training Grant DC00012 to Indiana University. I would like to thank Luis Hernandez and Winston Goh for their technical assistance on this project. Special thanks go to Miranda Cleary and Lorin Lachs for valuable discussion and suggestions.

Memory Span and Sequence Learning Using Multimodal Stimulus Patterns: Preliminary Findings in Normal-Hearing Adults

Abstract. The Simon memory game has been developed in our laboratory as a means of measuring memory span without requiring an explicit verbal response. This report presents a preliminary analysis of memory span data obtained from normal-hearing adults over two sessions using this new methodology. Traditional memory span measures were obtained from two standard tasks, digit span and word span, as well as six versions of the memory game. The memory game required subjects to reproduce sequences of colors by pressing response buttons on a four-alternative response box. In addition to a "memory span" task, in which color sequences of increasing length were generated randomly, a "sequence learning" task was administered using the memory game, in which identical sequences of increasing length were repeated, plus or minus one item, in order to measure longer-term information processing abilities. Color stimuli in each memory game task were presented either visually (a visual-spatial sequence of colored lights), auditorily (a sequence of spoken color words), or audiovisually (a visual-spatial sequence of colored lights and the same sequence of spoken color words presented simultaneously). Results showed that subjects reproduced far longer sequences in the sequence learning task compared to the memory span task, and subjects reproduced longer sequences in the audiovisual condition than in the visual or auditory conditions. Overall, performance was best in conditions where subjects could benefit from sequence repetition and multimodal information redundancy. The results of this study serve as normative, benchmark data for future studies using the memory game with several clinical populations.

Working memory, the system within the human cognitive system responsible for the temporary storage and processing of information, is useful as an explanatory device for both the limitations of high-level cognitive processes and individual differences in information processing abilities (Baddeley, 1992; Richardson, et al., 1996; Engle, Kane, & Tuholski, 1999). The working memory model has three major components: a visual-spatial short-term memory; a verbal short-term memory; and a central executive, an attentional system that controls the flow of information to and from the other components (Baddeley & Hitch, 1974; Kail & Hall, 2001). The verbal short-term memory – often called the phonological or articulatory loop – can be further divided into two subcomponents: a phonological store for verbal material, and mechanisms that enable rehearsal (Baddeley, 1986). The phonological loop has been an influential component of the working memory model, integrating a wide range of data and generating a large body of research (Gupta, 1996; see Baddeley, 1998, for a review of research on the phonological loop).

Verbal short-term memory capabilities are distinguished by two prominent features: a rapid rate of forgetting and a limited capacity. A measure of an individual's short-term memory span is thought to be indicative of that individual's overall information processing capabilities (Miller, 1956). Individual differences in memory span abilities are thus important in the processes of acquiring new knowledge and retrieving stored information from long-term memory (Engle, 1996). Recently, the case has been made that individual differences in short-term memory capabilities are fundamental to individual differences in language-related abilities (Gupta & MacWhinney, 1997; Gupta & Dell, 1999; Gupta, 1996). More specifically, researchers have

hypothesized that the short-term, phonological storage component within the phonological loop is the fundamental mechanism of language learning (Baddeley, Gathercole, & Papagno, 1998).

Traditional methods of measuring verbal short-term memory capacity almost always involve verbal reproduction of presented lists of items. Performance deficits on such traditional memory span measures shown by hearing-impaired individuals, or by other clinical populations who have deficits in speech production, might therefore be the result of problems associated with the hearing impairment itself, rather than a memory deficit. Thus, the verbal response requirement in traditional memory span studies is a potential confound when attempting to obtain short-term memory data from various clinical populations. Recently, the study of individual differences observed in hearing-impaired children with cochlear implants has become a topic of great interest (Pisoni, Cleary, Geers, & Tobey, 2000). It is not clear what the basis is for individual differences in performance among deaf children with cochlear implants on a variety of outcome measures that assess speech perception, language comprehension, speech intelligibility, and reading (Pisoni & Geers, 2000). However, differences in fundamental information processing capabilities may be the foundational factors responsible for individual differences in language processing in children with cochlear implants, as well as in other clinical populations (Pisoni, 2000). It would be advantageous, therefore, to have a means of obtaining memory span measures from clinical populations, especially the hearing impaired, that does not require an explicit verbal response.

The "Simon memory game" has been developed in our laboratory as a means of collecting memory span data without requiring verbal output (Cleary, Pisoni, & Geers, in press; Carlson, Cleary, & Pisoni, 1998). The memory game also allows us to use visual, auditory, and audiovisual stimulus presentation formats. In their first study with children, Cleary, Pisoni, and Geers (in press) found that in all three stimulus presentation formats, deaf children with cochlear implants had shorter memory spans than normal-hearing children. Normal-hearing children were also better than the cochlear implant children at utilizing "multimodal information redundancy" – the added benefit of receiving simultaneous auditory and visual information about the sequence. Cleary, Pisoni, and Geers (in press) concluded that performance differences on the memory game between the normal-hearing children and the cochlear implant children suggests differences in encoding or rehearsal strategies and "atypical" working memory development in deaf children with cochlear implants.

Until now, no study has looked specifically at performance by normal-hearing adults on both the "memory span" and "sequence learning" versions of the memory game. In addition, there are no data on the test-retest reliability of the memory game. Thus, the purpose of this study was to collect normative, benchmark memory span and sequence learning data from normal-hearing, native English-speaking adults using the Simon methodology. The present study involved testing adults over multiple sessions to obtain test-retest reliability measures. The data obtained in this study will be useful in conjunction with data already collected from normal hearing children, pediatric cochlear implant users, deaf children and adults, and other clinical populations (e.g., Sommers & Sawyer, 2001).

Method

Subjects

Forty-eight Indiana University undergraduates participated in Session One. Forty-three of the original forty-eight returned for a second session one week following their first session. Subjects ranged in age from 18 to 24 years, with the mean age of 20.5 years. Subjects were paid $5 for participation in Session One and $10 for participation in Session Two; each session lasted

approximately 45 minutes. All participants were native speakers of English with no speech or hearing disorders and normal or corrected-to-normal vision at the time of testing.

Materials

For the digit span task, tokens of the 10 spoken digits (0 to 9) were obtained from the Texas Instruments 46-Word (TI46) Speaker-Dependent Isolated Word Corpus (Texas Instruments, 1991). For the word span task, tokens of 66 spoken monosyllabic words were drawn from a prerecorded digital database (see Torretta, 1995, for a detailed description). All words were classified as "easy" words: these words are higher in frequency relative to their neighbors and come from a sparsely populated lexical neighborhood (Luce & Pisoni, 1998). Stimuli used in the digit span and word span tasks were presented over high-quality headphones at approximately 75 dB SPL. Subjects made their responses by writing on prepared answer booklets at the end of each trial. After recording their responses, subjects initiated the next trial by pressing the "Enter" key on the keyboard. See Goh and Pisoni (1998) for a more detailed description of the digit span and word span tasks used in the present study.

For the Simon memory game, auditory tokens of the four color words ("red", "yellow", "blue", and "green") were recorded by a single male speaker of American English. The memory game response box was modeled after the commercial product "Simon" by Milton Bradley. It consisted of four colored, back-lit response buttons. Subjects reproduced visual, auditory, or audiovisual sequences of colors by pressing the response buttons on the memory response box. See Cleary, Pisoni, and Geers (in press) for a more detailed description of the Simon memory game.

Procedure

Subjects were tested individually or in groups of three or fewer. All subjects completed the digit span task, then word span task, followed by six versions of the memory game.

In the digit span task, subjects were presented with a list of digits (0-9) over headphones. Once the entire list had been presented, subjects wrote down as many digits from the list as they could remember, in the order in which they were originally presented. The lists of digits began at length 4 and increased to length 10, with two lists presented at each list length for a total of 14 trials (Goh & Pisoni, 1998).

In the word span task, subjects were presented with a list of monosyllabic words, again over headphones. Once the entire list had been presented, subjects wrote down as many words from the list as they could remember, in the order in which they were originally presented. The lists of words began at length 3 and increased to length 8, with two lists presented at each list length for a total of 12 trials. The stimuli in the word span task were non-repeating and without replacement.

The memory game consisted of two different tasks: a "memory span" task and a "sequence learning" task. In the memory span task, subjects were given a sequence of colors and were asked to reproduce the sequence by pressing the response buttons. The sequences of colors were randomly generated, with the stipulation that no item was ever repeated consecutively in a given list. Sequences began at length 1, and subjects were presented with a total of 20 lists. An "adaptive testing procedure" was used to generate the stimulus sequences (Levitt, 1970): if a subject correctly reproduced two consecutive sequences at the same sequence length, the next sequence was increased in length by one item. If a subject made an error in reproducing a

sequence, the next sequence was decreased in length by one item.

In the sequence-learning task, subjects were given a sequence of colors and were asked to reproduce the sequence by pressing the response buttons. In this task, the sequences of colors were repeated. That is, subsequent sequences were exactly the same as the immediate preceding sequences, plus or minus one item. Sequences began at length 3, and subjects were presented with a total of 12 lists. A similar adaptive testing procedure was used in the sequence learning task: if a subject correctly reproduced a given sequence, then the next sequence presented was the identical sequence, increased in length by one item. If a subject made an error in reproducing a sequence, the next sequence was decreased in length by one item (see Cleary & Pisoni, 2001).

For each memory game task, three stimulus presentation formats were used. In the visual (V) condition, subjects saw a visual-spatial sequence of colored lights and heard nothing. In the auditory (A) condition, subjects heard a sequence of spoken color words and saw nothing. In the audiovisual (AV) condition, subjects saw a visual-spatial sequence of colored lights and also heard the same sequence of spoken color words simultaneously. The audiovisual presentation condition involved "multimodal information redundancy*"* (Cleary, Pisoni, & Geers, in press): redundant information about the sequence was presented to the subject simultaneously through both the auditory and visual modalities.

Within each task, the stimulus presentation conditions were counterbalanced, and the tasks themselves were counterbalanced in order of administration. There were 12 different orders of the memory game, and 4 subjects were assigned to each order, making a total of 48 subjects in Session One. Forty-three of the original forty-eight subjects returned for Session Two, 7-10 days following Session One. Subjects who returned for a second session completed all tasks in the same order in which they had completed them in Session One.

Results

Scoring

Two methods were used to score the digit span and word span tasks. The Strict Span score is the longest list length where both trials are perfectly recalled, plus 1/2 point for every subsequent trial also perfectly recalled (Daneman & Carpenter, 1980). The Absolute Span score is the sum of the total number of items in each perfectly recalled trial (LaPointe & Engle, 1990). The Strict Span is an item-based scoring method, while the Absolute Span is a list-based scoring method. Both scoring methods showed the same pattern of results.

Data from the Simon memory game were scored four different ways. The "One-time Score" is the longest sequence length correctly reproduced at least one time (on at least one trial). The "Half-time Score" is the longest sequence length correctly reproduced at least half of the time (on half of all trials). The "All-time Score" is the longest sequence length correctly reproduced one hundred percent of the time (on all trials). Finally, a weighted score was calculated, which is the sum of the proportion of correctly reproduced trials at each sequence length (Cleary, Pisoni, & Geers, in press). All four scoring methods showed the same pattern of results. A summary of all four scoring methods can be found in the Appendix.

Digit Span and Word Span

Data from the digit span and word span tasks for both sessions are shown in Table 1. According to the Strict Span scoring method, subjects averaged a digit span of roughly 7 items and a word span of roughly 5 items. The Absolute scoring method reflects this difference between

the digit span and word span tasks. The results for Session Two were consistent with those for Session One: subjects again averaged a digit span of approximately 7 items and a word span of approximately 5 items. These findings are consistent with earlier findings using similar methods of obtaining digit span and word span scores (Goh & Pisoni, 1998).

	Mean	SD	Min	Max
Session One (n = 48)				
Strict Score				
Digit Span	7.05	1.09	4.5	10
Word Span	5.31	0.70	3.5	7
Absolute Score				
Digit Span	46.85	16.49	13	98
Word Span	28.67	8.27	10	51
Session Two (n = 43)				
Strict Score				
Digit Span	7.30	1.06	4.5	10
Word Span	5.30	0.76	4	7
Absolute Score				
Digit Span	50.98	16.90	13	98
Word Span	28.77	9.10	14	51

Table 1. Digit span and word span scores: both sessions.

Simon Memory Game

Figure 1 shows the weighted scores in the three conditions of the Simon memory game for each session, averaged across subjects. The left panel of Figure 1 displays scores from Session One, while the right panel of Figure 1 displays scores from Session Two. For each session, memory span scores are plotted on the left side, and sequence learning scores are plotted on the right side. The open bar represents scores in the visual only (V) condition; the dotted bar represents scores in the auditory only (A) condition; and the striped bar represents scores in the audiovisual (AV) condition. The means plotted in Figure 1 are also displayed in Table 2.

In both sessions, irrespective of scoring method, the results showed an improvement in performance in the sequence-learning task over the memory span task. For both the memory span and the sequence learning tasks, performance was best in the audiovisual condition, while performance in the auditory only condition was better than performance in the visual only condition. Additionally, the average scores increased slightly from Session One to Session Two.

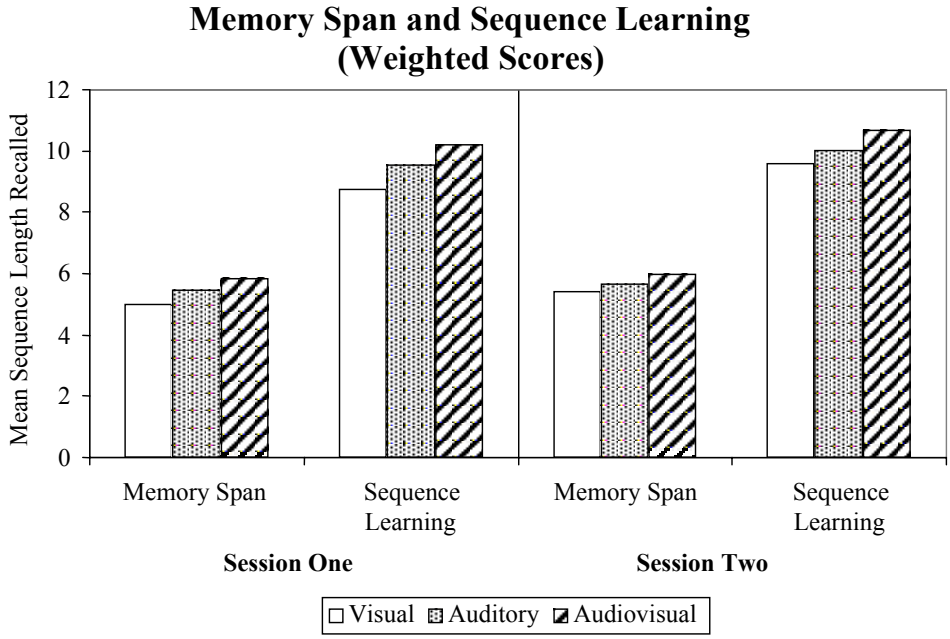


Figure 1. Mean sequence length recalled, Sessions One and Two. Weighted scores.

	Memory Span			Sequence Learning		
	Mean	SD	Range	Mean	SD	Range
Session One						
Visual	5.0	0.89	(3.2, 7.3)	8.7	2.69	(3.9, 14)
Auditory	5.5	0.94	(2.3, 7.4)	9.6	2.88	(4.8, 14)
Audiovisual	5.8	0.78	(4.3, 8.2)	10.2	2.45	(5.6, 14)
Session Two						
Visual	5.4	0.97	(3.7, 8.5)	9.6	2.71	(5.2, 14)
Auditory	5.7	1.00	(4.1, 8.3)	10.0	2.93	(5.0, 14)
Audiovisual	6.0	0.95	(4.3, 8.2)	10.7	2.54	(6.1, 14)

Table 2. Memory Span and Sequence Learning tasks, Sessions One and Two. Weighted scores.

A 2 (Task) x 3 (Stimulus Presentation Format) x 2 (Session) repeated measures analysis of variance (ANOVA) on the weighted scores revealed main effects of Task, $F(1, 42) = 232.348$, $p < .001$, of Stimulus Presentation Format, $F(2, 84) = 15.426$, $p < .001$, and of Session, $F(1, 42) = 5.896$, $p < .05$. No interactions were found among any of these variables.

Table 3 illustrates the main effect of Task in terms of "sequence repetition gain." The difference between a subject's sequence learning score and his or her memory span score can be thought of as a gain in performance due to repetition of the identical sequence. The sequence repetition effect was robust across all three stimulus presentation formats. Performance was increased by approximately four items when the sequence of colors was repeated.

	Session One			Session Two		
	Mean	Proportion	Percent	Mean	Proportion	Percent
Weighted Scores						
Visual	3.7	45/48	94%	4.2	41/43	95%
Auditory	4.1	47/48	98%	4.4	40/43	93%
Audiovisual	4.4	47/48	98%	4.7	43/43	100%

Table 3. Sequence repetition gain.

A post hoc analysis on the main effect of Stimulus Presentation Format, collapsed across Task and across Session, revealed that the difference between the audiovisual and visual only conditions was responsible for this main effect, $t(362) = 3.303$, $p < .001$. Table 4 shows this main effect in terms of a "multimodal redundancy gain." The difference between a subject's score in the audiovisual condition and his or her score in the visual only condition can be thought of as the gain in performance due to redundant auditory information. Multimodal redundancy gain was a robust effect, appearing in a high percentage of all trials in both the memory span and sequence learning tasks.

	Session One			Session Two		
	Mean	Proportion	Percent	Mean	Proportion	Percent
Weighted Scores						
Memory Span	0.84	42/48	88%	0.58	37/43	86%
Sequence Learning	1.46	35/48	73%	1.07	35/43	81%

Table 4. Multimodal redundancy gain.

A second post hoc analysis on the "Stimulus Presentation Format" effect, this time with the memory span and sequence learning tasks analyzed separately, revealed further differences in Stimulus Presentation Format within the memory span task, but not within the sequence-learning task. Table 5 summarizes the findings of this post hoc analysis. Though no Task x Stimulus Presentation Format interaction was found, within the memory span task a significant difference was found between the audiovisual and auditory only conditions, $t(180) = 2.591$, $p < .01$, suggesting that redundant visual information also resulted in a multimodal redundancy gain. Furthermore, a significant difference was found between the auditory and visual conditions, $t(180) = 2.568$, $p < .01$. Thus, auditory information led to a longer memory span than visual information. This is a new finding using the Simon memory game.

Weighted Scores	Memory Span	Sequence Learning
Auditory x Visual	$t(180) = 2.568$, $p < .01$	$t(180) = 1.535$, $p = .126$
Audiovisual x Auditory	$t(180) = 2.591$, $p < .01$	$t(180) = 1.595$, $p = .112$
Audiovisual x Visual	$t(180) = 5.353$, $p < .001$	$t(180) = 3.307$, $p < .001$

Table 5. Post hoc analysis with the memory span and sequence learning tasks analyzed separately.

Test-Retest Reliability

Test-retest reliability for the digit span, word span, Simon memory span, and Simon sequence learning tasks was also assessed in this study. Table 6 summarizes the test-retest reliability of the digit span and word span tasks. Both the list-based and item-based scoring methods show high correlations between scores in Session One and Session Two for the digit span task and the word span task.

Test-Retest Reliability	Digit Span	Word Span
Strict Score	$r = .73$**	$r = .60$**
Absolute Score	$r = .73$**	$r = .59$**

** $p < .01$

Table 6. Test-retest reliability: digit span and word span.

The test-retest reliability of the digit span and word span tasks is useful as a benchmark against which the test-retest reliability scores of the Simon memory game might be compared. The reliability coefficients of each condition of the Simon memory game are shown in Table 7. Overall, moderate positive correlations were obtained in all conditions of the Simon memory game. Notably, the highest correlation was observed in the condition where both sequence repetition and multimodal information redundancy was available ($r = .69$, $p < .01$).

Test-Retest Reliability: Weighted Scores	Memory Span	Sequence Learning
Visual	$r = .40$**	$r = .31$*
Auditory	$r = .46$**	$r = .56$**
Audiovisual	$r = .44$**	$r = .69$**

** $p < .01$
* $p < .05$

Table 7. Test-retest reliability: memory span and sequence learning tasks.

Discussion

Three main effects were found in the present study: Task (memory span vs. sequence learning), Stimulus Presentation Format (Visual vs. Auditory vs. Audiovisual), and Testing Session (One vs. Two). No interactions were found among the three main effects. Subjects were able to reproduce far longer sequences in the sequence-learning task compared with the memory span task. On average, repeating the pattern allowed subjects to reproduce sequences approximately four items longer than patterns presented in the memory span task. This robust effect occurred in all three stimulus presentation conditions. Subjects also reproduced longer sequences in conditions where multimodal information redundancy was available. In particular, redundant auditory information, in addition to the visual-spatial sequence of colors, led to the reproduction of longer sequences in both the memory span and sequence learning tasks. Although no Task x Stimulus Format interaction was evident, the memory span task was more sensitive than the sequence-learning task to the differences between auditory and visual stimulus presentation formats.

Moderate positive test-retest reliability coefficients were found across all stimulus presentation conditions in both the memory span and sequence learning tasks. Equivalent forms of the memory game were administered one week apart, and comparable results were obtained in both sessions. Main effects for Task and Stimulus Presentation Format were found in both Session One and Session Two. The difference in mean scores between the two sessions is probably due to familiarity or practice effects, which could certainly be controlled for in future studies by increasing the duration between testing sessions or allowing a practice session before the testing session.

The present study demonstrates the contribution of sequence repetition and multimodal information redundancy to human memory in a group of young, healthy adults. In normal-hearing adults, using the Simon memory game, sequence repetition and multimodal information redundancy allow subjects to overcome the basic capacity limitations of short-term memory. In certain clinical populations, however, this may not always be the case (Cleary, Pisoni, & Geers, in press; Pisoni, Cleary, Geers, & Tobey, 2000). Individual differences in working memory suggest possible differences in the basic underlying information processing abilities of some clinical populations. It is these differences in central cognitive abilities that appear to be driving individual differences in the ability to acquire and process spoken language (Gupta & MacWhinney, 1997; Pisoni, 2000). The present study of the information processing capabilities

of normal hearing adults, specifically with respect to their ability to use sequence repetition and multimodal information redundancy to overcome short-term memory capacity limitations, provides benchmark data against which other studies of information processing in clinical populations can be compared. Data from our laboratory have already been reported on pediatric cochlear implant users (Cleary, Pisoni, & Geers, in press; Pisoni et al., 2000; Cleary et al., 2000). Other studies of post-lingually deafened adults are underway.

In future studies using the Simon memory game, we intend to investigate the effects of presentation rate on memory span and sequence learning when multimodal information redundancy is present. The rate at which stimuli are recognized and processed has been shown to reveal important sources of individual differences in memory span (Dempster, 1981). Increasing the rate of presentation in the memory game should increase the demands on short-term memory capacity, forcing subjects to rely more heavily on highly automatic processes and thus on redundant multimodal information. We are also currently developing a methodology in our laboratory for presenting "spatially neutral" sequences of colors, in order to block the spatial coding of the visual sequence and force subjects to rely exclusively on verbal coding. Finally, we intend to use the memory game response box format to present multimodal stimuli in implicit learning paradigms, specifically using serial reaction time and artificial grammar tasks (e.g., Nissen & Bullemer, 1987; Reber, 1993).

References

Baddeley, A.D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A.D. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A.D. (1998). *Human memory: Theory and practice*. Boston: Allyn & Bacon.

Baddeley, A.D., Gathercole, S.E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *Recent advances in the psychology of learning and motivation* (Vol. VII, pp. 47-89). New York: Academic Press.

Carlson, J.L., Cleary, M., & Pisoni, D.B. (1998). Performance of normal-hearing children on a new working memory span task. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 251-275). Bloomington, IN: Speech Research Laboratory, Indiana University.

Cleary, M., & Pisoni, D.B. (2001). Sequence learning as a function of presentation modality in children with cochlear implants. Poster presented at *CID New Frontiers Conference*. St. Louis, MO.

Cleary, M., Pisoni, D.B., & Geers, A.E. (in press). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear & Hearing*.

Cleary, M., Pisoni, D.B., Kirk, K.I., Geers, A.E., & Tobey, E. A. (2000). Working memory and language development in children with cochlear implants. Poster presented at *CI2000: The 6th International Cochlear Implant Conference*. Miami, FL.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Dempster, F.N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63-100.

Engle, R.W. (1996). Working memory and retrieval: An inhibition-resource approach. In J.T.E. Richardson (Ed.), *Working memory and human cognition* (pp. 89-119). Oxford: Oxford University Press.

Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). Cambridge, UK: Cambridge Univ. Press.

Goh, W.D., & Pisoni, D.B. (1998). Effects of lexical neighborhoods on immediate memory span for spoken words: A first report. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 195-213). Bloomington, IN: Speech Research Laboratory, Indiana University.

Gupta, P. (1996). Immediate serial memory and language processing: Beyond the articulatory loop. *Technical Report No. CS-96-02*. Urbana: Beckman Institute, Cognitive Science Group.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267-333.

Gupta, P., & Dell, G.S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 447-481). Mahwah, NJ: Lawrence Erlbaum.

Kail, R., & Hall, L.K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition, 29*, 1-9.

LaPointe, L.B., & Engle, R.W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1118-1133.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477.

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19*, 1-36.

Miller, G.A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Nissen, M. J. & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32.

Pisoni, D.B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear & Hearing, 21*, 70-78.

Pisoni, D.B., Cleary, M., Geers, A.E., & Tobey, E.A. (2000). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *The Volta Review, 101*, 111-164.

Pisoni, D.B., & Geers, A.E. (2000). Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language. *Annals of Otology, Rhinology and Laryngology Supplement, 185*, 92-93.

Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.

Richardson, J.T.E., Engle, R.W., Hasher, L., Logie, R.H., Stoltzfus, E.R., & Zacks, R.T. (1996). *Working memory and human cognition*. Oxford: Oxford University Press.

Sommers, M.S., & Sawyer, D. (2001). Predictors of visual enhancement and lipreading in younger and older adults. Poster presented at the meeting of the Acoustical Society of America. Chicago, IL.

Texas Instruments. (1991). TI 46-word speaker-dependent isolated word corpus (CD-ROM). Gaithersburg: NIST.

Torretta, G.M. (1995). The "easy-hard" word multi-talker speech database: An initial report. In *Research on Spoken Language Processing Progress Report No. 20,* (pp. 321-334). Bloomington, IN: Speech Research Laboratory, Indiana University.

Appendix
Simon memory game results scored by four methods

Session One (N = 48)

	Memory Span			Sequence Learning		
	Mean	SD	Range	Mean	SD	Range
One-time Scores						
Visual	5.6	0.82	(4, 7)	9.3	2.46	(5, 14)
Auditory	6.0	0.91	(3, 8)	10.1	2.65	(5, 14)
Audiovisual	6.4	0.79	(5, 9)	10.5	2.28	(6, 14)
Half-time Scores						
Visual	5.4	0.89	(4, 7)	9.1	2.58	(4, 14)
Auditory	5.8	0.97	(3, 8)	10.0	2.76	(5, 14)
Audiovisual	6.2	0.92	(5, 9)	10.5	2.34	(6, 14)
All-time Scores						
Visual	4.0	1.13	(1, 6)	8.5	3.09	(3, 14)
Auditory	4.7	1.22	(3, 7)	9.4	3.21	(4, 14)
Audiovisual	4.8	1.06	(2, 7)	10.1	2.63	(5, 14)
Weighted Scores						
Visual	5.0	0.89	(3.2, 7.3)	8.7	2.69	(3.9, 14)
Auditory	5.5	0.94	(2.3, 7.4)	9.6	2.88	(4.8, 14)
Audiovisual	5.8	0.78	(4.3, 8.2)	10.2	2.45	(5.6, 14)

Session Two (N = 43)

	Memory Span			Sequence Learning		
	Mean	SD	Range	Mean	SD	Range
One-time Scores						
Visual	6.0	1.03	(4, 9)	10.2	2.43	(6, 14)
Auditory	6.2	1.00	(5, 9)	10.5	2.64	(6, 14)
Audiovisual	6.6	0.98	(5, 9)	11.1	2.23	(7, 14)
Half-time Scores						
Visual	5.8	1.07	(4, 9)	10.1	2.48	(5, 14)
Auditory	6.1	1.12	(4, 9)	10.4	2.79	(5, 14)
Audiovisual	6.5	1.06	(5, 9)	11.1	2.28	(6, 14)
All-time Scores						
Visual	4.6	1.37	(2, 6)	9.4	3.13	(4, 14)
Auditory	4.8	1.23	(3, 7)	9.8	3.21	(4, 14)
Audiovisual	5.2	1.19	(3, 7)	10.7	2.66	(5, 14)
Weighted Scores						
Visual	5.4	0.97	(3.7, 8.5)	9.6	2.71	(5.2, 14)
Auditory	5.7	1.00	(4.1, 8.3)	10.0	2.93	(5.0, 14)
Audiovisual	6.0	0.95	(4.3, 8.2)	10.7	2.54	(6.1, 14)

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

A Multi-Talker Dialect Corpus of Spoken American English: An Initial Report on Development[1]

Cynthia G. Clopper, Allyson K. Carter, Caitlin M. Dillon, James D. Harnsberger, Rebecca Herman, Connie M. Clarke,[2] David B. Pisoni and Luis R. Hernandez

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by the NIH-NIDCD R01 Research Grant DC00111 and the NIH-NIDCD T32 Training Grant DC00012 to Indiana University. The authors would like to thank Julie Auger for her assistance in the design of the corpus.
[2] Psychology Department, University of Arizona, Tucson, Arizona

A Multi-Talker Dialect Corpus of Spoken American English: An Initial Report on Development

Abstract. A multi-talker multi-dialect corpus of spoken American English has been designed to provide researchers who are interested in variation and variability with a large number of speech samples from twenty talkers in each of four cities located in phonologically distinct dialect regions of the United States: West (Los Angeles), South (Atlanta), Midland (Indianapolis), and Northern Cities (Chicago). The speech samples to be collected include word-length, sentence-length, and paragraph-length utterances, and have been designed to elicit phonological forms that differentiate the four regions. Once collected, these materials can be used for a range of perceptual and acoustic studies investigating the perception and production of dialect variation in the United States.

Objectives of the Corpus

The purpose of this project is to create a speech corpus containing recordings from a large number of talkers from phonologically distinct dialect regions in the United States for use in a range of perceptual studies and acoustic analyses. Dialect variation, both regional and social in origin, has been an important topic of research in American English since the 1930's when plans for a "Linguistic Atlas of North America" were first discussed (Cassidy, 1993). The first studies were primarily concerned with regional variation, focusing on differences in lexical items produced by older males from rural areas (Chambers, 1993). More recently, dialect research has been extended to include studies on social and ethnic variation, such as African American Vernacular English and Appalachian English (Wolfram & Schilling-Estes, 1998).

Recent research has also begun to focus on phonological variation, particularly on variation and changes in progress that have been documented in the vowel systems of several American English dialects. For example, vowel shifts such as the Northern Cities Vowel Shift found in urban areas surrounding the Great Lakes and the Southern Vowel Shift found in rural areas of the Southern United States have been described in some detail (Wolfram & Schilling-Estes, 1998).

While such phonological variation has been studied via field recordings and transcription, relatively little work has been done to document the acoustic properties of these phenomena or to study their perceptual correlates via playback experiments. While acoustic analysis is a commonly accepted technique for comparing and differentiating the vowel systems of different languages, it is not commonly employed in sociolinguistic research due to Labov's "observer's paradox" (Wolfram & Schilling-Estes, 1998). Simply put, the paradox refers to the effect of the observer's presence (the observer being an experimenter, recording equipment, or any other tool of measurement) on the acoustic properties of speech produced by members of a dialect community of interest. The dialect variation that sociolinguists seek to document is almost always found in forms that appear in speech styles used more frequently in casual conversation, in specific pragmatic or situational contexts, or only with other members of the same dialect community. The intrusion of an experimenter from outside the dialect community and the effect of recording equipment on the formality of the conversational setting are perceived as barriers to the elicitation of the "deepest" form of the dialect in question (Wolfram & Schilling-Estes, 1998). Thus, the most commonly used method to investigate the properties of American English and other dialects involves making audio recordings of spontaneous speech and then phonetically transcribing those interviews.

While such methods are useful in describing relatively gross differences between dialects, they suffer from a number of limitations for researchers interested in the acoustic-phonetic properties of phonological forms of a dialect, and for researchers developing controlled stimulus materials varying in dialect for use in perception experiments. First, the use of spontaneous speech entails a lack of control over the particular stimulus materials elicited. For the experimenter hoping to collect numerous tokens of a particular vowel or word in a common phonetic and prosodic context, it is very difficult to elicit such materials in a natural, spontaneous speech style (cf. Harnsberger & Pisoni, 1999). While certain tasks, such as topically-guided conversations or map tasks, can be used to elicit particular words or prosodic phrases, strict control over the phonetic context of these forms cannot be achieved. Control of phonetic context is crucial for any acoustic analysis, as well as in constructing stimulus materials for use in perception tests.

Given these constraints, and given the purposes of this corpus, we have chosen to elicit speech materials in a read speech style, enabling control over the materials elicited. For the purposes of comparison only, we will also elicit a spontaneous sample from each talker, taking the form of a conversation with the experimenter administering the tests. While eliciting read speech undoubtedly limits the range of phonological variability we will observe between the dialects, we hope to ameliorate this problem by selecting American English dialects that have been shown in prior research to differ robustly from one another in terms of phonological patterns. We are also interested in documenting American English dialects that constitute relatively large communities within the United States. We believe that this will make the corpus as a whole more representative of American English dialectal variation than a corpus that is focused on much smaller dialect communities. We have therefore decided to record twenty talkers from each of four cities, representing four phonologically distinct regions: Atlanta (South), Indianapolis (Midland), Chicago (Northern Cities), and Los Angeles (West). For summary descriptions of each of the regional dialects, and for the rationale behind the selection of the boundaries defining these regions, see Wolfram and Schilling-Estes (1998) and Labov, Ash, and Boberg (1997). While we recognize that these four cities are not representative of all dialects of American English, we expect that they will provide us with some degree of phonological variation that is both acoustically and perceptually prominent, from a relatively large sample of talkers.

The nature of the controlled stimulus materials, the focus on dialect variation, and the large number of talkers we plan to record are the three main features that set this corpus apart from other existing corpora. There are at least three existing spoken language corpora that include speech samples from a large number of talkers from a variety of American English dialects: the Santa Barbara Corpus of Spoken American English (LDC Catalog, 2001c), the CALLFRIEND project (LDC Catalog, 2001a; LDC Catalog, 2001b), and the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Zue, Seneff, & Glass, 1990). The Santa Barbara corpus contains spontaneous speech samples from talkers from a wide range of geographic and socioeconomic backgrounds. The CALLFRIEND project contains recordings of telephone conversations between talkers which are grouped into two broad dialect categories: Southern and Non-Southern. The TIMIT Corpus contains ten read sentences from each of 630 talkers who come from eight defined dialect regions of the United States. The usefulness of the first two corpora in perceptual studies is limited by the lack of common stimulus materials for all talkers. The usefulness of the TIMIT corpus is also limited because of the ten sentences read by each talker, only two of those sentences were read by all 630 talkers.

Spoken language corpora that control for stimulus materials also exist. However, they do not necessarily vary the dialect of the talkers in a systematic fashion. For example, corpora used in our lab such as the “Easy-Hard” Word Multi-Talker Speech Database (Torretta, 1995) and the Talker Variability Sentence Database (Karl & Pisoni, 1994) contain fixed sets of stimuli spoken by 10-20 talkers, but no effort was made to identify or control for dialectal variation in the talkers. The new corpus will combine

the systematic variation in dialect found in the TIMIT corpus with the control over a range of stimulus materials found in the "Easy-Hard" and Talker Variability databases.

Once the corpus has been collected, we plan to use it in our lab for perceptual studies involving dialect identification, categorization, and discrimination by non-native listeners, lexical decision tasks, and voice quality judgement tasks involving dialect manipulations. This corpus will also be used in a series of perceptual learning tasks on dialect intelligibility after laboratory training and dialect manipulations in voice learning. Finally, the corpus will enable us to conduct acoustic-phonetic studies including descriptions of the vowel systems, analyses of diphthongal differences, and investigations into the acoustic correlates of stress across dialects.

Organization of the Corpus

Talkers

Ten males and ten females will be recorded in each of the four cities. Each talker will be a college-aged monolingual native speaker of English, with no history of hearing or speech disorders. In order to obtain a fairly homogenous group of talkers in terms of socioeconomic status, level of education, and linguistic experience, talkers will be recruited from community college campuses and will be asked to complete a lengthy questionnaire. In order to participate, a talker must have lived in the city of interest for his or her entire life and have limited experience with other dialects and languages. Parents of the talkers must also be native English speakers who are local to the area.

Stimulus Materials

The materials list was selected to provide a number of different kinds of speech, including word-length, sentence-length, and paragraph-length materials. The materials themselves were selected with the intent of providing a useful corpus for completing the projects mentioned above.

The word-length materials include CVC's and multisyllabic words and nonwords. The CVC list was designed for this project and consists of 1020 CVC's selected from an online dictionary containing approximately 20,000 entries based on Webster's Pocket Dictionary. The list is composed of all CVC's in the dictionary that received a familiarity rating of 6.0 or greater (on a 7-point scale) by undergraduates (Nusbaum, Pisoni, & Davis, 1984). A small subset of these CVC's was hand-selected for an additional repetition in recording. This subset was selected such that the vowels occurred in consonantal contexts that are expected to reveal systematic differences between the dialects, based on documented shifts and mergers (Callary, 1975; Gordon, 1997; Labov, 1972; Labov, Yeager, & Steiner, 1972; Wolfram & Schilling-Estes, 1998). Additionally, 10 vowels will be recorded in an "hVd" context for use in determining the vowel space of each talker (Hillenbrand, Getty, Clark, & Wheeler, 1995; Hagiwara, 1997). The multisyllabic word list is a subset of the list developed by Carter and Clopper (this volume), and contains 240 words that vary systematically in the number of syllables and the location of primary stress. The multisyllabic nonword list was developed for this project and contains 56 disyllabic forms. These forms have been designed so that half will be realized with primary stress on the first syllable and the other half will be realized with primary stress on the second syllable (Cutler & Carter, 1987; Hammond, 1999; Hayes, 1995; Kelly, 1988; Kelly & Bock, 1988).

The sentence-length materials include high probability, low probability and anomalous sentences. The high probability sentences were taken from all eight of the Speech Perception in Noise (SPIN) lists (Kalikow & Stevens, 1977), with several additional sentences taken from the Hearing in Noise Test (HINT) sentence list (Nilsson, Soli, & Sullivan, 1994) to round out the representation of all English

vowels in the content words in the sentences. In high probability sentences, the final target word is predictable based on the preceding words in the sentence. In low probability sentences, the final target word is not predictable from the rest of the sentence. The low probability sentences were taken from lists 1, 2, 7, and 8 of the SPIN test. The anomalous sentences were created from the SPIN sentences, so that their target words matched those for the low probability sentences that were selected. The remaining words were taken from the high probability sentences in the remaining four lists, using a method similar to Miller and Isard (1963).

The longer materials include a passage and a spontaneous speech sample. The passage selected was the Rainbow Passage (Fairbanks, 1940). This passage has a long history of use in perceptual and acoustic studies, including several involving individual differences and variability (Gelfer & Schofield, 2000; Sapienza, Walton, & Murry, 1999). The short spontaneous speech sample will focus mainly on discussions about the local geographic area and will be used primarily as a reference point for each talker.

Methods

Recording

All recording will be done in sound-attenuated booths located in each city. Materials will be presented visually to the talkers via a portable Macintosh Powerbook G3 computer and the talkers will be asked to read the materials aloud into a head-mounted dynamic unidirectional cardioid microphone (Shure SM10A) as they are presented. Responses will be recorded digitally in real time into individual sound files on the computer and simultaneously on DAT, using a Sony TC8 recorder, as a backup. The nine stimulus sets will be presented in a pseudo-random order, and all stimuli within each set will also be presented randomly.

Future Directions

Collection of the data is expected to begin in the Spring of 2001. We hope to complete the data collection within six months and to have all of the speech available on CD-ROM, with documentation shortly thereafter.

References

Callary, R. (1975). Phonological change and the development of an urban dialect in Illinois. *Language in Society, 4,* 155-169.

Carter, A.K. & Clopper, C.G. (this volume). Prosodic and morphological effects on word reduction in adults: A first report.

Cassidy, F.G. (1993). Area lexicon: the making of DARE. In *American Dialect Research*. Preston, D.R. (ed.) Philadelphia, PA: John Benjamins, 93-106.

Chambers, J.K. (1993). Sociolinguistic dialectology. In *American Dialect Research*. Preston, D.R. (ed.) Philadelphia, PA: John Benjamins, 133-164.

Cutler, A. & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133-142.

Fairbanks, G. (1940). *Voice and Articulation Drillbook*. New York: Harper.

Gelfer, M. & Schofield, K. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice, 14,* 22-33.

Gordon, M.J. (1997). Urban sound change beyond city limits: The spread of the northern cities shift in Michigan. Doctoral dissertation, The University of Michigan.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America, 102*, 655-658.

Hammond, M. (1999). *The Phonology of English*. Oxford: Oxford University Press.

Harnsberger, J.D. & Pisoni, D.B. (1999). Eliciting speech reduction in the laboratory II: Calibrating cognitive loads for individual talkers. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 339-349). Bloomington, IN: Speech Research Laboratory, Indiana University.

Hayes, B. (1995). *Metrical Stress Theory.* Chicago, IL: Chicago University Press.

Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099-3111.

Kalikow, D.N. & Stevens, K.N. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Karl, J.R. & Pisoni, D.B. (1994). Effects of stimulus variability on recall of spoken sentences: A first report. In *Research on Spoken Language Processing Progress Report No. 19* (pp. 145-193). Bloomington, IN: Speech Research Laboratory, Indiana University.

Kelly, M. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27,* 343-358.

Kelly, M. & Bock, J. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, *14,* 389-403.

Labov, W. (1972). The internal evolution of linguistic rules. In *Linguistic Change and Generative Theory*. Stockwell, R.P. & Macaulay, R.K.S. (eds.) Bloomington, IN: Indiana University Press, 101-171.

Labov, W., Ash, S., & Boberg, C. (1997). A National Map of the Regional Dialects of American English. Retrieved June 26, 2000 from the World Wide Web: http://www.ling.upenn.edu/phono_atlas/NationalMap/NationalMap.html.

Labov, W., Yeager, M., & Steiner, R. (1972). A quantitative study of sound change in progress. Philadelphia, PA: U.S. Regional Survey.

LDC Catalog. (2001a). CALLFRIEND American English Non-Southern Dialect. Retrieved January 9, 2001 from the World Wide Web: http://www.ldc.upenn.edu/Catalog/LDC96S46.html.

LDC Catalog. (2001b). CALLFRIEND American English Southern Dialect. Retrieved January 9, 2001 from the World Wide Web: http://www.ldc.upenn.edu/Catalog/LDC96S47.html.

LDC Catalog. (2001c). Santa Barbara Corpus of Spoken American English Part-I. Retrieved January 9, 2001 from the World Wide Web: http://www.ldc.upenn.edu/Catalog/LDC2000S85.html.

Miller, G.A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2,* 217-228.

Nilsson, M., Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95,* 1085-1099.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Sapienza, C., Walton, S., & Murry, T. (1999). Acoustic variations in adductor spasmodic dysphonia as a function of speech task. *Journal of Speech, Language, and Hearing Research, 42,* 127-140.

Torretta, G.M. (1995). The "easy-hard" word multi-talker speech database: An initial report. In *Research on Spoken Language Processing Progress Report No. 20* (pp. 321-334). Bloomington, IN: Speech Research Laboratory, Indiana University.

Wolfram, W. & Schilling-Estes, N. (1998). *American English*. Malden, MA: Blackwell.

Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication, 9*, 351-356.

IV. Publications

IV. Publications

ARTICLES PUBLISHED:

Chin, S.B., Meyer, T.A., Hay-McCutcheon, M., Wright, G.A, & Pisoni, D.B. (2000). Structure of mental lexicons of children who use cochlear implants: Preliminary findings. *Annals of Otology, Rhinology & Laryngology, 109,* 114-116.

Chin, S.B. & Pisoni, D.B. (2000). A phonological system at two years after cochlear implantation. *Clinical Linguistics & Phonetics, 14,* 53-73.

Frisch, S.A., Large, N.R. & Pisoni, D.B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42,* 481-496.

Frisch, A., Meyer, T.A., Pisoni, D.B., Svirksy, M.A. & Kirk, K.I. (2000). Using behavioral data to model open-set word recognition and lexical organization by pediatric cochlear implant users. *Annals of Otology, Rhinology & Laryngology, 109,* 60-62.

Frisch, S.A. & Pisoni, D.B. (2000). Modeling spoken word recognition performance by pediatric cochlear implant users using feature identification. *Ear & Hearing, 21,* 578-589.

Harnsberger, J.D. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Journal of the Acoustical Society of America 108*, 764-83.

Luce, P.A., Goldinger, S.D., Auer, E.T. Jr. & Vitevitch, M.S. (2000). Phonetic priming effects in spoken word shadowing. *Perception & Psychophysics, 62*, 615-625.

Luce, P.A., Goldinger, S.D. & Vitevitch, M.S. (2000). It's good…but is it ART? *Behavioral and Brain Sciences, 23,* 336.

Meyer, T.A., Frisch, S.A., Svirsky, M.A. & Pisoni, D.B. (2000). Modeling phoneme and open-set word recognition by cochlear implant users: A preliminary report. *Annals of Otology, Rhinology & Laryngology, 109,* 68-70.

Pisoni, D.B. (2000). Cognitive factors and cochlear implants: An overview of the role of perception, attention, learning and memory in speech perception. *Ear & Hearing, 21,* 70-78.

Pisoni, D.B., Cleary, M., Geers, A. & Tobey, E.A. (2000). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *Volta Review, 101*, 111-164.

Pisoni, D.B. & Geers, A. (2000). Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing. *Annals of Otology, Rhinology & Laryngology, 109,* 92-93.

Rogers, M.A., & Storkel, H.L. (1999). Planning speech one syllable at a time: The reduced buffer capacity hypothesis in apraxia of speech. *Aphasiology*, *13*, 9-11.

Storkel, H.L. & Rogers, M.A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics, 14*, 407-425.

Svirsky, M.A., Robbins, A.M., Kirk, K.I., Pisoni, D.B. & Miyamoto, R.T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science, 11,* 153-158.

BOOK CHAPTERS PUBLISHED:

Chin, S.B., & Kirk, K.I. (2000). Consonant feature production by children with multichannel cochlear implants, hearing aids, and tactile aids. In S. Waltzman and N. Cohen (Eds.), *Cochlear Implants*. New York: Thieme Medical Publishers. Pp. 309-310.

Kirk, K.I., Pisoni, D.B. & Miyamoto, R.T. (2000). Lexical discrimination by children with cochlear implants: Effects of age at implantation and communication mode. In S. Waltzman and N. Cohen (Eds.), *Cochlear Implants*. New York: Thieme Medical Publishers. Pp. 252-254.

Sehgal, S.T., Kirk, K.I., Pisoni, D.B. & Miyamoto, R.T. (2000). Effect of residual hearing on children's speech perception abilities with a cochlear implant. In S. Waltzman and N. Cohen (Eds.), *Cochlear Implants*. New York: Thieme Medical Publishers. Pp. 219-221.

MANUSCRIPTS ACCEPTED FOR PUBLICATION (IN PRESS):

Carter, A.K. (In press). A phonetic and phonological analysis of weak syllable omissions by children with normally developing language and specific language impairment. *Current Issues in Linguistic Theory.*

Chin, S.B., Finnegan, K.R. & Chung, B.A. (In press). Relationships among types of speech intelligibility in pediatric users of cochlear implants. *Journal of Communication Disorders*.

Cleary, M. & Pisoni, D.B. (In press). Speech perception and spoken word recognition: Research and theory. In B. Goldstein (Ed.), *Handbook of Perception*. Cambridge: Blackwell.

Cleary, M. & Pisoni, D.B. (In press). Speech perception. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*. London, UK: Macmillan.

Cleary, M., Pisoni, D.B. & Geers, A.E. (In press). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear & Hearing.*

Dinnsen, D.A., McGarrity, L.W. & O'Connor, K.M. (In press). Rene Kager: Optimality Theory. *Studies in Second Language Acquisition.*

Dinnsen, D.A., McGarrity, L.W., O'Connor, K.M. & Swanson, K.A. (In press). On the role of sympathy in acquisition. *Language Acquisition*.

Frisch, S.A., Large, N.R., Zawaydeh, B. & Pisoni, D.B. (In press). Emergent phonotactic generalizations in English and Arabic. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Goh, W.D., Pisoni, D.B., Kirk, K.I., & Remez, R.E. (In press). Audio-visual perception of sinewave speech in an adult cochlear implant user: A case study. *Ear & Hearing*.

Harnsberger, J.D., Svirsky, M.A., Kaiser, A.R., Pisoni, D.B., Wright, R. & Meyer, T.A. (In press). Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *Journal of the Acoustical Society of America*.

Herman, R. (In press). Phonetic markers of global discourse structures in English. *Journal of Phonetics.*

Lachs, L., McMichael, K. & Pisoni, D. B. (In press). Speech perception and implicit memory: Evidence for detailed episodic encoding. In J. Bowers & C. Marsolek (Eds.), *Rethinking Implicit Memory*.

Lachs, L., Pisoni, D. B. & Kirk, K. I. (In press). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear & Hearing.*

Meyer, T.A., Pisoni, D.B., Luce, P.A. & Bilger, R.C. (In press). An analysis of the psychometric and lexical neighborhood properties of spondaic words. *Journal of the American Academy of Audiology.*

Miyamoto, R.T., Bichey, B.G., Wynne, M.K. & Kirk, K.I. (In press). Cochlear implantation with large vestibular aqueduct syndrome. *Laryngoscope*.

Vitevitch, M.S. (In press). The influence of onset-density on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance.*

SUBMITTED:

Ferguson, S.H. & Kewley-Port, D. (Submitted). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America.*

Frisch, S., Broe, M. & Pierrehumbert, J. (Submitted). Similarity and phonotactics in Arabic. *Natural Language and Linguistic Theory.*

Goh, W.D., & Pisoni, D.B. (Submitted). Effects of lexical neighborhoods on immediate memory span for spoken words. *Quarterly Journal of Experimental Psychology*.

Harnsberger, J.D. (Submitted). The relationship between the identification and discrimination of non-native nasal contrasts: Evaluating the Perceptual Assimilation Model. *Journal of the Acoustical Society of America.*

Harnsberger, J.D. (Submitted). The effect of linguistic experience on perceptual similarity among nasal consonants: A multidimensional scaling analysis. *Journal of Phonetics.*

Herman, R. (Submitted). Syntactically and pragmatically governed accentuation in Balinese. *Language*.

Herman, R. (Submitted). Utterance-final voice quality variations: Their perceptual structure and acoustic correlates. *Journal of Phonetics*.

Herman, R. & McGory, J.T. (Submitted). The conceptual similarity of intonational tones and its effects on inter-transcriber reliability. *Language and Speech*.

Lachs, L. & Pisoni, D.B. (Submitted). Effects of multi-modal speech cues on recognition memory for spoken words. *Journal of Experimental Psychology: Human Perception and Performance.*

McMichael, K.H. & Pisoni, D.B. (Submitted). Effects of talker-specific encoding on recognition memory for spoken sentences. *Memory & Cognition*.

Meyer, T.A., Pisoni, D.B., Luce, P.A. & Bilger, R.C. (Submitted). Acoustic, psychometric and lexical neighborhood properties of spondaic words: A computational analysis of speech discrimination scores. *Journal of Speech, Language and Hearing Research.*

Pierrehumbert, J. & Herman, R. (Submitted). Intonation. In William Frawley, (ed.) *International Encyclopedia of Linguistics*. 2nd edition. Oxford University Press.

Pisoni, D.B., Svirsky, M.A., Kirk, K.I. & Miyamoto, R.T. (Submitted). Looking at the "Stars": A first report on the interrelations among measures of speech perception, intelligibility and language development in pediatric cochlear implant users. *Journal of Speech, Language, and Hearing Research*.

Sheffert, S.M. & Shiffrin, R.M. (Submitted). Auditory "registration without learning." *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Sheffert, S.M., Pisoni, D.B., Fellowes, J.M. & Remez, R.E. (Under Revision). Learning to recognize talkers from natural, sinewave and reverse speech samples. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Storkel, H.L. (Submitted). Similarity neighborhoods in the developing mental lexicon. *Journal of Child Language.*

Storkel, H.L. (Submitted). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, & Hearing Research.*

Storkel, H.L. & Morrisette, M.L. (Submitted). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in the Schools.*

Vitevitch, M.S. (Submitted). The influence of similarity neighborhoods on phonological speech errors. *Journal of Memory and Language*.

Vitevitch, M.S. (Submitted). Change Deafness: The inability to detect changes in a talker's voice. *Psychological Science.*

Vitevitch, M.S. (Submitted). Influences of the number of phoneme positions that form neighbors on spoken word recognition. *Psychonomic Bulletin & Review*.

Vitevitch, M.S. & Sommers, M. (Submitted). The role of phonological neighbors in the tip-of-the-tongue state. *Journal of Memory and Language*.

CONFERENCE PRESENTATIONS:

Bichey, B.G., Miyamoto, R.T., Wynne, M.K. & Kirk, K.I. (2000). Progressive hearing loss in patients with large vestibular aquaduct syndrome. Paper presented at the annual meeting of the American Auditory Society, April 2000. Scottsdale, AZ.

Carter, A. (2000). The Phonetic manifestation of omitted syllables in children's productions. Paper presented at the Boston University Conference on Language Development, Nov. 3-5, 2000, Boston, MA.

Carter, A. (2000). A phonetic and phonological analysis of weak syllable omissions: Comparing children with normally developing language and specific language impairment. Paper presented at Linguistic Theory, Speech Pathology and Speech Therapy, University of Padova, Italy, August 22-26, 2000.

Cleary, M., Pisoni, D.B., Geers, A.E. & Kirk, K.I. (2000). Individual differences in the effectiveness of cochlear implants in prelingually deaf children: Working memory in the visual and auditory modalities. American Speech-Hearing-Language Association. November 17, 2000, Washington, DC.

Cleary, M., Pisoni, D.B., Kirk, K.I., Geers, A. & Tobey, E. (2000). Working memory and language development in children with cochlear implants. Poster presented at CI2000: The 6th International Cochlear Implant Conference, February 2000, Miami, Florida.

Edwards, J., McGregor, K., Morrisette, M.L., Storkel, H.L. & Windsor, J. (2000). The lexicon in clinical application. American Speech-Language-Hearing Association Convention, November 2000. Washington, DC.

Ferguson, S.H., Kewley-Port, D. & Humes, L.E. (2000). Effects of hearing loss and linear amplification on the relative importance of acoustic cues. Poster presented at the International Hearing Aid Research Conference (IHCON), August 23-28, 2000, Lake Tahoe, CA.

Ferguson, S.H., Kewley-Port, D. & Humes, L.E. (2000). Vowels in clear and conversational speech: Effects of amplification on the relative importance of acoustic cues to vowel intelligibility. Poster presented at the 23rd Midwinter Meeting of the Association for Research in Otolaryngology, February 2000, St. Petersburg Beach, FL.

Harnsberger, J.D., Pisoni, D.B., Svirsky, M., Kaiser, A. & Wright, R. (2000). The 'vowel spaces' of normal-hearing individuals and cochlear-implant users. Presented at the 139th Meeting of the Acoustical Society of America, June 2000, Atlanta, GA.

Harnsberger, J.D., Pisoni, D.B. & Wright, R. (2000). Eliciting speech styles in the laboratory: Assessment of a new experimental method. Presented at the 140th Meeting of the Acoustical Society of America, December 2000, Newport Beach, CA.

Herman, R. (2000). Utterance-final voice quality variations: Their perceptual structure and acoustic correlates. Paper presented at the 140th Meeting of the Acoustical Society of America. December 2000, Newport Beach, CA.

Kaiser, A.R., Kirk, K.I., Pisoni, D.B. & Lachs, L. (2000). Audiovisual speech integration in adults with cochlear implants or normal hearing: Lexical and talker effects. Poster presented at the Twenty-Third Midwinter Research Meeting of the Association for Research in Otolaryngology, February 2000, St. Pete's Beach, FL.

Kaiser, A.R. & Svirsky, M.A. (2000). Using a personal computer to perform real-time signal processing in cochlear implant research. Paper presented at the Ninth IEEE DSP Workshop & First IEEE Workshop on Signal Processing Education, October 2000, Austin, TX.

Kaiser, A.R. & Svirsky, M.A. (2000). Speech perception testing using a real-time PC based speech processor for the Nucleus 22 channel cochlear implant. Twenty-Third Midwinter Research Meeting of the Association for Research in Otolaryngology, February 2000, St. Pete's Beach, FL.

Kaiser, A.R., Svirsky, M.A., Meyer, T.A. & Lento, C.L. (2000). Psychophysical and speech perception performance of post-lingual adults with CIs. CI2000: The 6th International Cochlear Implant Conference, February 2000, Miami, Florida.

Lachs, L., Pisoni, D.B. & Kirk, K.I. (2000). Individual differences in the effectiveness of cochlear implants in prelingually deafened children. Poster presented at the American Speech-Language Hearing Association Annual Convention, November 2000, Washington, D.C.

Lachs, L., Pisoni, D.B., Kirk, K.I. & Miyamoto, R.T. (2000). Some new analyses of the audiovisual integrative abilities of children with cochlear implants: Initial findings and implications. Paper presented at the Twenty-Third Midwinter Research Meeting of the Association for Research in Otolaryngology, February 2000, St. Pete's Beach, FL.

McGarrity, L.W. (2000). To stress or not to stress?: The stress-epenthesis interaction in Yimas. Paper presented at the Sixth Mid-Continental Workshop on Phonology, October 2000, Ohio State University, Columbus, OH.

Meyer, T.A., Svirsky, M.A. & Kaiser, A.R. (2000). Modeling open-set spoken word recognition by cochlear implant users based on psychophysical performance. Twenty-Third Midwinter Research Meeting of the Association for Research in Otolaryngology, February 2000, St. Pete's Beach, FL.

Miyamoto, R.T., Bichey, Bradford G., Wynne, M.K. & Kirk, K.I. (2000). Cochlear implantation with large vestibular aqueduct syndrome. Paper presented at the 2000 Middle Section of the Triological Society, January 2000, Cincinnati, OH.

Miyamoto, R.T., Wynne, M.K., Diefendorf, A.O., Bichey, B.G. & Sorkin, D. (2000). Fluctuating and progressive sensorineural hearing loss in children. Paper presented at the 2000 Convention of the American Speech, Language, and Hearing Association, November 2000, Washington, D.C.

Pisoni, D.B. (2000). Individual differences in effectiveness of cochlear implants in prelingually deaf children. Paper presented at the American Speech-Language Hearing Association Annual Convention, November 2000, Washington, D.C.

Storkel, H.L. (2000). Language interactions in clinical application. With J. A. Edwards, K. K. McGregor, M. L. Morrisette, & J. Windsor in seminar entitled The lexicon in clinical application. American Speech-Language-Hearing Association Convention, November 2000, Washington, D.C.

Storkel, H.L. (2000). The lexicon in development: Childhood. With P.W. Jusczyk, & M.S. Vitevitch, in seminar entitled: The developing mental lexicon. American Speech-Language-Hearing Association Convention, November 2000, Washington, D.C.

Svirsky, M.A., Kaiser, A.R. & Meyer, T.A. (2000). Can cochlear implant users discriminate stimuli that are closer than the distance between adjacent electrodes? Twenty-Third Midwinter Research Meeting of the Association for Research in Otolaryngology, February 2000, St. Pete's Beach, FL.

Svirsky, M.A., Kaiser, A.R. & Meyer, T.A. (2000). How do cochlear implant users understand speech? CI 2000: The 6th International Cochlear Implant Conference, February 2000, Miami, Florida.

Vitevitch, M.S. (2000). Session Presenter in Models of the Lexicon: The Lexicon in Speech Production. Annual Convention of the American Speech-Language-Hearing Association. November 16, 2000, Washington, D.C.

Vitevitch, M.S. (2000). Session Presenter in The Lexicon in Development: The Aging Lexicon. Annual Convention of the American Speech-Language-Hearing Association. November 17, 2000, Washington, D.C.