# RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 26
(2003-2004)

**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. 26 (2003-2004)

## Table of Contents

# INTRODUCTION

This is the twenty-sixth 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.

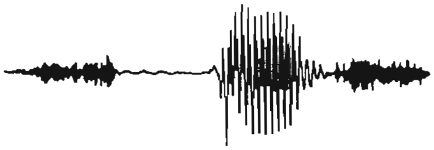
We are distributing progress reports of our research activities because of the ever increasing lag in journal publications and the resulting delay in the dissemination of new information and research findings in the field of spoken language processing. We are, of course, very interested in following 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
USA

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

Copies of this 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, 2003–September 1, 2004)

### 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. | Professor of Otolaryngology–Head and Neck Surgery[3,5] |
| Steven B. Chin, Ph.D | Associate Scientist in Otolaryngology–Head and Neck Surgery[3] |
| Derek Houston, Ph.D | Assistant Professor of Otolaryngology–Head and Neck Surgery[3] |
| Tonya Bergeson, Ph.D | Assistant Professor of Otolaryngology–Head and Neck Surgery[3] |
| Woo Sui Teoh, M.D. | NIH Postdoctoral Trainee[3] |
| David L. Horn, M.D. | NIH Postdoctoral Trainee[3] |
| Rachael F. Holt, Ph.D | NIH Postdoctoral Trainee[3] |
| Stephen J. Winters, Ph.D | NIH Postdoctoral Trainee |
| Ana Schwartz, Ph.D. | NIH Postdoctoral Trainee[6] |
| Rose Burkholder, B.S. | NIH Predoctoral Trainee |
| Cynthia G. Clopper, B.A. | NIH Predoctoral Trainee |
| Brianna Conrey, B.A. | NSF Predoctoral Trainee |
| Caitlin M. Dillon, B.A | NIH Predoctoral Trainee |
| Elena Breiter | Undergraduate Research Assistant |
| Adam Tierney | Undergraduate Research Assistant |
| Sara Phillips | Undergraduate Research Assistant |

### TECHNICAL PERSONNEL

| | |
|---|---|
| Luis R. Hernández, B.A. | Research Associate in Psychology |
| Darla J. Sallee | Administrative Assistant |
| Kuai Hinojosa | Applications Programmer |
| Rebecca A.O. Davis | Research Technician |
| Lanier F. Holt, B.A. | Editorial Assistant |

[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 Professor of Electrical Engineering, Purdue School of Engineering and Technology, Indianapolis, IN.

[6] Now at the Department of Psychology, University of Texas at El Paso, El Paso, TX.

# E-MAIL ADDRESSES

| | |
|---|---|
| Tonya Bergeson | tbergeso@iupui.edu |
| Elena Brieter | ebrieter@indiana.edu |
| Rose Burkholder | rburkhol@indiana.edu |
| Steven B. Chin | schin@iupui.edu |
| Cynthia G. Clopper | cclopper@indiana.edu |
| Rebecca Davis | robryan@iupui.edu |
| Caitlin M. Dillon | cmdillon@indiana.edu |
| Luis R. Hernández | hernande@indiana.edu |
| Rachael F. Holt | raholt@indiana.edu |
| David L. Horn | dlhorn@iupui.edu |
| Derek Houston | dmhousto@indiana.edu |
| Karen I. Kirk | kkirk@iupui.edu |
| Sara Phillips | sacphill@indiana.edu |
| David B. Pisoni | pisoni@indiana.edu |
| Darla J. Sallee | dsallee@indiana.edu |
| Mario A. Svirsky | msvirsky@iupui.edu |
| Woo Sui Teoh | stech@iupui.edu |
| Adam Tierney | attierne@indiana.edu |
| Stephen J. Winters | stwinter@indiana.edu |

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Speech Perception in Deaf Children with Cochlear Implants[1]

**David B. Pisoni**

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

[1] This work was supported by NIH-NIDCD Training Grant T32DC00012, NIH-NIDCD Research Grants R01DC00111, NIH-NIDCD R01DC00064 to Indiana University and NIH Research Grant R01DC03100 to Central Institute for the Deaf. I would like to thank Miranda Cleary for her help and advice on all stages of the research described in this chapter. I would also like to thank Ann Geers, Chris Brenner and Mike Strube of CID for their help and assistance. I am also grateful to Steven Chin for his editorial comments on an earlier version of this paper.

# Speech Perception in Deaf Children with Cochlear Implants

**Abstract.** Cochlear implants work well in many profoundly deaf adults and children. However, despite the success of cochlear implants in many deaf children, large individual differences have been reported on a wide range of speech and language outcome measures. This finding is observed in all research centers around the world. Some children do extremely well with their cochlear implant while others derive only minimal benefits after receiving their implant. Understanding the reasons for the variability in outcomes and the large individual differences following cochlear implantation is one of the most important problems in the field today. This chapter summarizes recent findings on the speech perception skills of deaf children following cochlear implantation. The results of these studies suggest that in addition to several demographic and medical variables, variation in children's success with cochlear implants reflects fundamental differences in rapid phonological coding and verbal rehearsal processes which are used in a wide range of clinical outcome measures used to measure benefit following implantation.

## Introduction

Each week for the last 12 years I have traveled from my home in Bloomington to Riley Hospital for Children at the IU Medical Center in Indianapolis, a distance of some 60 miles each way, to work on an unusual clinical research project. I am part of a multidisciplinary team of basic and clinical researchers who are studying the development of speech perception and language skills of profoundly deaf children who have received cochlear implants. My colleague, Dr. Richard Miyamoto, a pediatric otologist and head and neck surgeon has been providing profoundly deaf adults and children with cochlear implants since the early 1980s when the first single-channel implants were undergoing clinical trials. Since the approval of cochlear implants by the FDA as a treatment for profound deafness, over 60,000 patients have received cochlear implants at centers all over the world (Clarke, 2003).

A cochlear implant is a surgically implanted electronic device that functions as an auditory prosthesis for a patient with a severe to profound sensorineural hearing loss. It provides electrical stimulation to the surviving spiral ganglion cells of the auditory nerve bypassing the damaged hair cells of the inner ear to restore hearing in both deaf adults and children. The device provides them with access to sound and sensory information from the auditory modality. The current generation of multichannel cochlear implants consist of an internal multiple electrode array and an external processing unit (see Figure 1). The external unit consists of a microphone that picks up sound energy from the environment and a signal processor that codes frequency, amplitude and time and compresses the signal to match the narrow dynamic range of the ear. Cochlear implants provide temporal and amplitude information. Depending on the manufacturer, several different place coding techniques are used to represent and transmit frequency information in the signal.

For postlingually profoundly deaf adults, a cochlear implant provides a transformed electrical signal to an already fully developed auditory system and intact mature language processing system. These patients have already acquired spoken language under normal listening conditions so we know their central auditory system and brain are functioning normally. In the case of a congenitally deaf child, however, a cochlear implant provides novel electrical stimulation through the auditory sensory modality and an opportunity to perceive speech sounds and develop spoken language for the first time after a period of auditory deprivation. Congenitally deaf children have not been exposed to speech and do not develop spoken language normally. Although their brain and nervous system continue to develop in the absence of normal auditory stimulation, there is now evidence to suggest that some cortical reorganization has already taken place during the period of sensory deprivation before implantation and that several aspects of speech and language skills after implant may develop in an atypical fashion. Both peripheral and central differences in neural function are likely to be responsible for the wide range of variability observed in outcome and benefit following implantation.

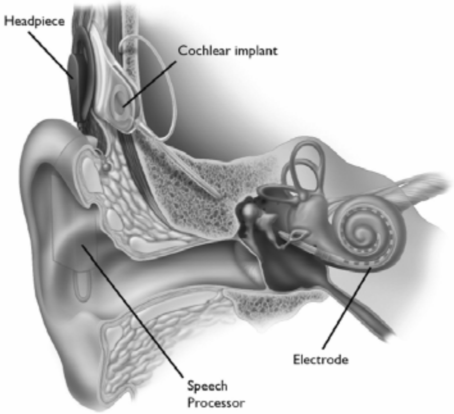


**Figure 1.** Simplified diagram of the internal and external components of a multichannel cochlear implant system. The external speech processor uses a transcutaneous radio-frequency (RF) transmitter to send electrical signals to an internal receiver which is connected directly to the implanted electrode array. (Courtesy of Cochlear Americas).

This chapter is concerned with congenitally deaf children who have received cochlear implants. These children are the most interesting and theoretically important clinical population to study because they have been deprived of sound and auditory stimulation at a very early point in neural and cognitive development. After implantation their hearing is restored with electrical stimulation that is designed to simulate the response of a healthy cochlea to speech and other auditory signals. Aside from the obvious clinical benefits of cochlear implantation as a method of treating profound prelingual deafness in children, this clinical population also provides a unique opportunity to study the effects of auditory deprivation on the development of speech perception and language processing skills. It also permits us to assess the effects of restoration of hearing via artificial electrical stimulation of the nervous system. In some sense, one can think of research on this clinical population as the modern-day analog of the so-called "forbidden experiment" in the field of language acquisition. In this case, after a period of sensory deprivation has occurred, hearing is restored via medical intervention and children receive exposure to sound and stimulation through the auditory modality. Under these conditions, it is possible to study both the consequences of a period of auditory deprivation on speech and language development as well as the effects of restoring hearing using artificial electrical stimulation of the auditory nerve.

While cochlear implants work and appear to work well for many profoundly deaf adults and children, they do not always provide benefits to all patients who receive them. Compared to other behavioral data I have seen in the field of speech perception and spoken word recognition over the years, the audiological outcomes and benefits following cochlear implantation were simply enormous and hard to fully understand at first glance. Some deaf adults and children do extremely well with their cochlear implants and display what initially appears to be near-typical speech perception and language skills on a wide range of traditional clinical speech and language tests when tested under quiet listening conditions in the laboratory. In contrast, other adults and children struggle for long periods of time after they receive their cochlear implant and often never achieve comparable levels of speech and language performance or verbal fluency.

Low-performing patients are unable to talk on the telephone and frequently have great difficulty in noisy environments or situations where more than one person is speaking at the same time. Almost all of these patients do derive some minimal benefits from their cochlear implants because they are able to recognize some nonspeech sounds and have an increased awareness of where they are in their environment in terms of space and time. But they have a great deal of difficulty perceiving speech and understanding spoken language in a robust fashion under a wide range of challenging listening conditions.

I began to wonder why this pattern of results occurred and I became curious about the underlying factors that were responsible for the enormous differences in audiological outcome. Seeing some of these deaf children and talking with them and their parents each week made a big difference in appreciating the magnitude of this problem and the consequences of the wide range of variability in outcome for the children and their families. In addition to the enormous variability observed in the speech and language outcome measures, several other findings have been consistently reported in the clinical literature on cochlear implants in deaf children. An examination of these findings provides some preliminary insights into the possible underlying cognitive and neural basis for the variability in outcome and benefit among deaf children with cochlear implants. A small handful of traditional demographic variables have been found to be strongly associated with outcome and benefit after implantation. When these contributing factors are considered together, it is possible to begin formulating some specific hypotheses about the reasons for the enormous variability in outcome and benefit.

Almost all of the clinical research on cochlear implants has focused on the effects of a small number of demographic variables using traditional outcome measures based on assessment tools developed by clinical audiologists and speech pathologists. Although rarely discussed explicitly in the literature, these behaviorally-based clinical outcome measures of performance are the final product of a large number of complex sensory, perceptual, cognitive and linguistic processes that contribute to the observed variation among cochlear implant users. Until recently, little if any research focused on the underlying information processing mechanisms used to perceive and produce spoken language in this clinical population. Our investigations of these fundamental neurocognitive and linguistic processes have provided some new insights into the basis of individual differences in profoundly deaf children with cochlear implants.

In addition to the enormous individual differences and variation in clinical outcome measures, several other findings have been consistently reported in the literature on cochlear implants in children. Age at implantation has been shown to influence all outcome measures of performance. Children who receive an implant at a young age do much better on a whole range of outcome measures than children who are implanted at an older age. Length of auditory deprivation or length of deafness is also related to outcome and benefit. Children who have been deaf for shorter periods of time before implantation do much better on a variety of clinical measures than children who have been deaf for longer periods of time. Both findings demonstrate the contribution of sensitive periods in sensory, perceptual and linguistic development and serve to emphasize the close links that exist between neural development and behavior, especially, the development of hearing, speech and language (Ball & Hulse, 1998; Konishi, 1985; Konishi & Nottebohm, 1969; Marler & Peters, 1988).

Early sensory and linguistic experience and language processing activities after implantation have also been shown to affect performance on a wide range of outcome measures. Implanted children who are immersed in “Oral-only” communication environments do much better on clinical tests of speech and language development than implanted children who are enrolled in “Total Communication” programs (Kirk, Pisoni, & Miyamoto, 2000). Oral communication approaches emphasize the use of speech and hearing skills and actively encourage children to produce spoken language to achieve optimal benefit from their implants. In contrast, total communication approaches employ the simultaneous use of some form of manual-coded English along with speech to help the child acquire language using both sign and spoken language inputs. The differences in performance between groups of children who are placed in oral communication and total communication education settings are observed most prominently in both receptive and expressive language tasks that involve the use of phonological coding and phonological processing skills such as open-set spoken word recognition, language comprehension and measures of speech production, especially measures of speech intelligibility and expressive language.

Until just recently, clinicians and researchers have been unable to find reliable preimplant predictors of outcome and success with a cochlear implant (see, however, Bergeson & Pisoni, 2004). The absence of preimplant predictors is a theoretically significant finding because it suggests that many complex interactions take place between the newly acquired sensory capabilities of a child after a period of auditory deprivation, properties of the language-learning environment and various interactions with parents and caregivers that the child is exposed to after receiving a cochlear implant. More importantly, however, the lack of preimplant predictors of outcome and benefit makes it difficult for clinicians to identify those children who are doing poorly with their cochlear implant at a time in development when changes can be made to modify and improve their language processing skills.

Finally, when all of the outcome and demographic measures are considered together, the available evidence strongly suggests that the underlying sensory and perceptual abilities for speech and language “emerge” after implantation. Performance with a cochlear implant improves over time for almost all children. Success with a cochlear implant therefore appears to be due, in part, to perceptual learning and exposure to a language model in the environment. Because outcome and benefit with a cochlear implant cannot be predicted reliably from traditional behavioral measures obtained before implantation, any improvements in performance observed after implantation must be due to sensory and cognitive processes that are linked to maturational changes in neural and cognitive development (see Sharma, Dorman & Spahr, 2002).

Our current hypothesis about the source of individual differences in outcome following cochlear implantation is that while some proportion of the variance in performance can be attributed directly to peripheral factors related to audibility and the initial sensory encoding of the speech signal into “information-bearing” sensory channels in the auditory nerve, several additional sources of variance also come from more central cognitive and linguistic factors that are related to psychological processes such as perception, attention, learning, memory and language. How a deaf child uses the initial sensory input from the cochlear implant and the way the environment modulates and shapes language development are fundamental research problems. These problems deal with perceptual encoding, verbal rehearsal, storage and retrieval of phonetic and phonological codes and the transformation and manipulation of phonological and neural representations of the initial sensory input in a range of language processing tasks.

To investigate individual differences and the sources of variation in outcome, we began by analyzing a set of data from a longitudinal project on cochlear implants in children (see Pisoni et al., 1997; 2000). Our first study was designed to study the “exceptionally” good users of cochlear implants—the so-called Stars. These are the children who did extremely well with their cochlear implants after only two years of implant use. The Stars are able to acquire spoken language quickly and easily and appear to be on a developmental trajectory that parallels normal-hearing children although delayed a little in time (see Svirsky et al., 2000). The theoretical motivation for studying the exceptionally good children was based on an extensive body of research on "expertise" and "expert systems" theory (Ericsson & Smith, 1991). Many important new insights have come from studying expert chess players, radiologists and other individuals who have highly developed skills in specific knowledge domains.

## Analysis of the Stars

We analyzed scores obtained from several different outcome measures over a period of six years from the time of implantation to examine changes in speech perception, word recognition and comprehension over time (see Pisoni et al., 2000 for complete report). Before these results are presented, however, we describe how the Stars and a comparison group of lower-performing children were originally selected.

The criterion used to identify the Stars was based on scores obtained from one particular clinical test of speech perception, the Phonetically Balanced Kindergarten (PBK) Words test (Haskins, 1949). This PBK test is an open-set test of spoken word recognition (also see Meyer & Pisoni, 1999) and is very difficult for prelingually deaf children when compared to other closed-set speech perception tests routinely included in the standard clinical assessment battery (Zwolan, Zimmerman-Phillips, Asbaugh, Hieber, Kileny & Telian, 1997). Children who do reasonably well on the PBK test display ceiling levels of performance on other closed-set speech perception tests that measure speech pattern discrimination skills.

Open-set tests like the PBK test measure word recognition and lexical selection processes (Luce & Pisoni, 1998). To perform this test successfully, the child is required to search and retrieve the phonological representation of a test word from lexical memory and repeat it to the examiner. Open-set tests of word recognition are extremely difficult for hearing-impaired children with cochlear implants because the task requires that the child perceive and encode fine phonetic differences based entirely on information present in the speech signal without the aid of any external context or retrieval cues. A child must identify and then discriminate a unique phonological representation from a large number of lexical equivalence classes in memory (see Luce & Pisoni, 1998). It is important to emphasize here that although recognizing isolated spoken words in an open-set test format may seem like a simple task at first glance, it is very difficult for a hearing-impaired child who has a cochlear implant. Typically-developing children with normal hearing routinely display ceiling levels of performance under comparable testing conditions (Kluck et al., 1997).

To learn more about why the Stars do so well on open-set test of word recognition, we analyzed outcome data from children who scored exceptionally well on the PBK test two years after implantation. For comparison, we also obtained PBK scores from a group of low-performing children. The PBK score was used as the "criterial variable" to identify and select two groups of children for subsequent analysis using an "extreme groups" design. The Stars were children who scored in the upper 20% of all children tested on the PBK test two years post-implant. The low-performers consisted of children who scored in the bottom 20% on the PBK test two years post-implant. After the children were sorted into two groups, we examined their performance on a range of other clinical outcome measures that were available as part of large-scale longitudinal study at Indiana University. The speech perception data we discuss here include measures of speech feature discrimination, spoken word recognition and comprehension.

Scores for the two groups were obtained from a longitudinal database containing a variety of demographic and outcome measures from 160 deaf children (see Pisoni et al., 2000). Other measures of vocabulary knowledge, receptive and expressive language and speech intelligibility were also obtained (see Pisoni et al., 1997; 2000 for more details). All of the children in both groups were prelingually deafened. Each child received a cochlear implant because he/she was profoundly deaf and was unable to derive any benefit from conventional hearing aids. All children had used their cochlear implant for two years at the time when these analyses were completed. Using this selection procedure, the two groups turned out to be roughly similar in age at onset of deafness and length of implant use.

### Speech Feature Discrimination

Measures of speech feature discrimination for both consonants and vowels were obtained for both groups of children using the Minimal Pairs Test (Robbins et al., 1988). This clinical test uses a two-alternative forced-choice picture pointing task. The child hears a single word spoken in isolation on each trial using live voice presentation by an examiner and is required to select one of the pictures that correspond to the test item. Examples of two test plates are shown in Figure 2.

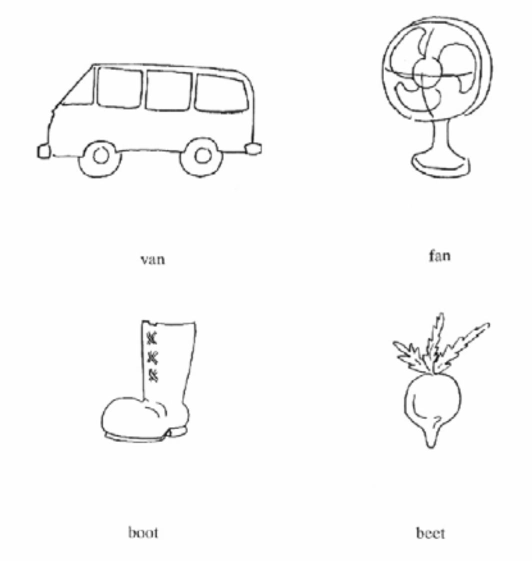


**Figure 2.** Examples of two test plates used to measure speech feature discrimination on the Minimal Pairs Test (from Robbins et al., 1988).

A summary of the consonant discrimination results for both groups of subjects is shown in Figure 3. Percent correct discrimination is displayed separately for manner, voicing and place of articulation as a function of implant use in years. Data for the Stars are shown by the filled bars; data for the low-performers are shown by the open bars in this figure. Chance performance on this task is 50% correct as shown by a horizontal line. A second horizontal line is also displayed in this figure at 70% correct corresponding to scores that were significantly above chance using the binominal distribution.

Examination of the results for the Minimal Pairs Test obtained over a period of six years of implant use reveals several findings. First, performance of the Stars was consistently better than the control group for every comparison across all three consonant features. Second, discrimination performance improved over time with implant use for both groups. The increases were primarily due to improvements in discrimination of manner and voicing by the Stars. At no interval did the mean scores of the comparison group ever exceed chance performance on discrimination of voicing and place features. Although increases in minimal pair discrimination performance were observed over time for the controls, their scores never reached the levels observed with the Stars, even for the manner contrasts that eventually exceeded chance performance in Years 4, 5 and 6.

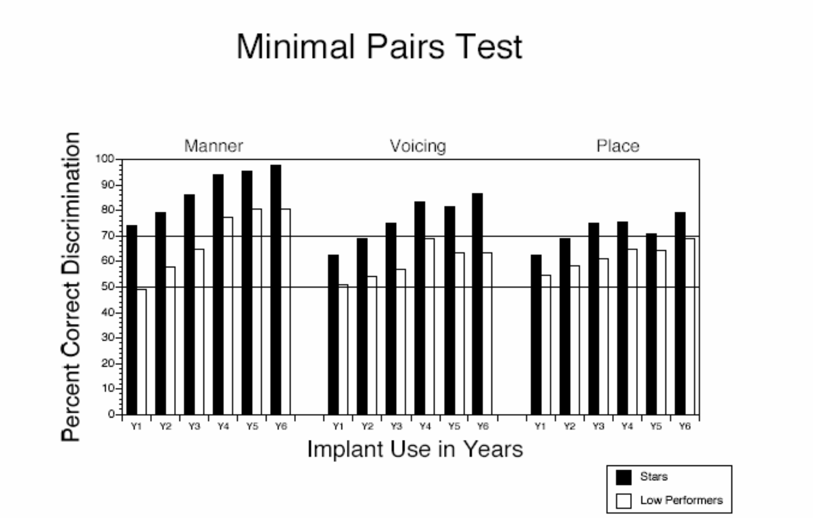


**Figure 3.** Percent correct discrimination on the Minimal Pairs Test (MPT) for manner, voicing and place as a function of implant use. The Stars are shown by filled bars, the Low-Performers are shown by open bars (from Pisoni et al., 2000).

The results of the Minimal Pairs Test demonstrate that both groups of children have difficulty perceiving, encoding and discriminating fine phonetic details of isolated spoken words in a simple two-alternative closed-set testing format. Although the Stars discriminated differences in manner of articulation after one year of implant use and showed consistent improvements in performance over time for both manner and voicing contrasts, they still had a great deal of difficulty reliably discriminating differences in place of articulation, even after five years of experience with their implants. In contrast, the low-performing children were just barely able to discriminate differences in manner of articulation above chance after four years of implant use. The lower-performing children also had a great deal of difficulty discriminating differences in voicing and place of articulation even after five or six years of use.

The pattern of speech feature discrimination results shown in Figure 3 suggests that both groups of children encode spoken words using “coarse” phonological representations. Their representations appear to be “underspecified” and contain much less fine-grained acoustic-phonetic detail than the lexical representations that normal hearing children typically use. The Stars were able to discriminate manner and to some extent voicing much sooner after implantation than the low-performers. In addition, the Stars also displayed consistent improvements in speech feature discrimination over time after implantation.

The speech feature discrimination data reveal several differences in the encoding of sensory information and the phonological representations that are used for subsequent word learning and lexical development. It is likely that if a child cannot reliably discriminate small phonetic differences between pairs of spoken words that are phonetically similar under these relatively easy forced-choice test conditions, they will also have difficulty recognizing words in isolation with no context or retrieving the phonological representations of highly familiar words from memory for use in simple speech production tasks that require immediate repetition. We would also expect them to display a great deal of difficulty in recognizing and imitating nonwords as well which have no lexical representations.

## Spoken Word Recognition

Two additional word recognition tests were used to measure open-set word recognition. Both tests use words that are familiar to preschool age children. The Lexical Neighborhood Test (LNT) contains monosyllabic words; the Multi-syllabic Lexical Neighborhood test (MLNT) contains multisyllabic words (Kirk, Pisoni & Osberger, 1995). Both tests contain two different sets of words that are used to measure lexical discrimination and provide detailed information about how the lexical selection process is carried out. Half of the items in each test are lexically "easy" words and half are lexically "hard" words.

The differences in performance on the easy and hard words provide an index of how well a child is able to make fine phonetic discriminations among acoustically similar words. Differences in performance between the LNT and the MLNT provide a measure of the extent to which the child is able to make use of word length cues to recognize and access words from the lexicon. The test words are presented in isolation one at a time by the examiner using a live-voice auditory-only format. The child is required to imitate and immediately repeat a test word after it is presented by the examiner.

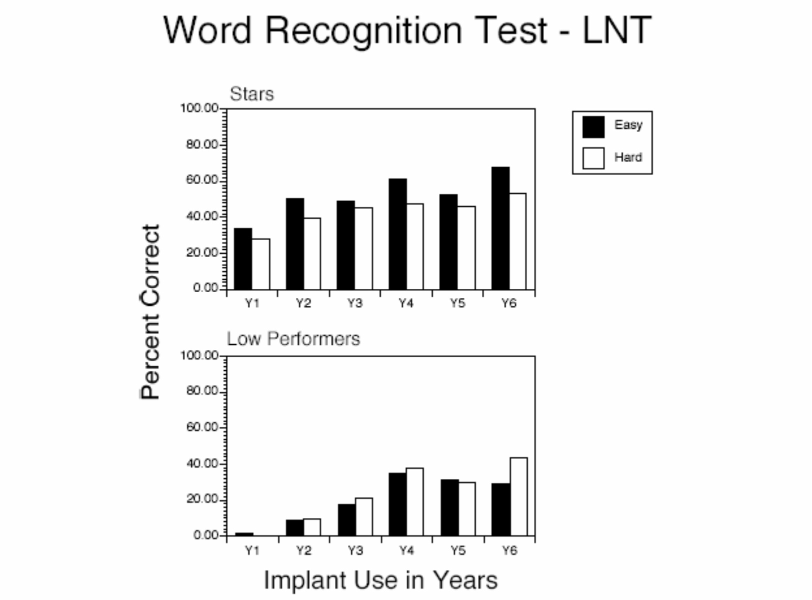


**Figure 4.** Percent correct word recognition performance for the Lexical Neighborhood Test (LNT) monosyllabic word lists as a function of implant use and lexical difficulty (from Pisoni et al., 2000).

Figures 4 and 5 show percent correct word recognition obtained on the LNT and the MLNT for both groups of children as a function of implant use. The data for the Stars are shown in the top panel of each figure; the data for the low-performers are shown in the bottom panels. Scores for the "easy" and "hard" words are shown separately within each panel.

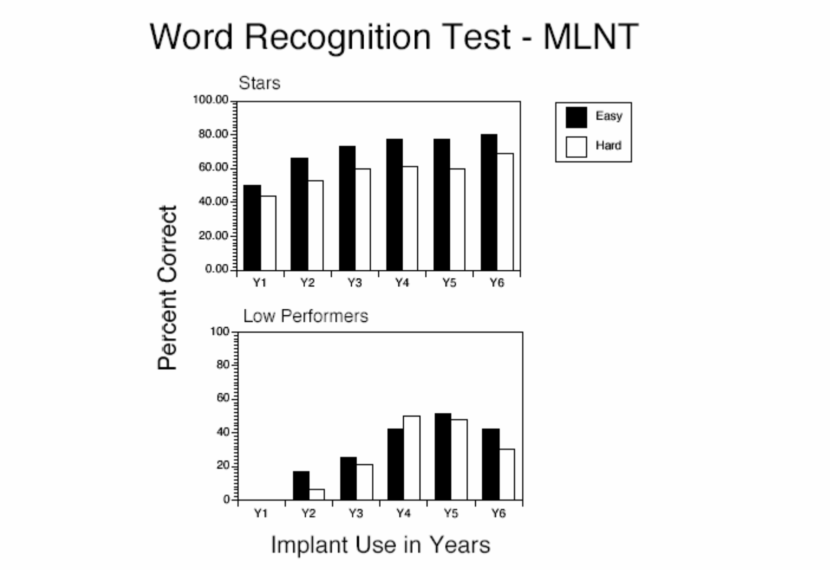


**Figure 5.** Percent correct word recognition performance for the Multi-syllabic Lexical Neighborhood Test (MLNT) word lists as a function of implant use and lexical difficulty (from Pisoni et al., 2000).

Several consistent differences in performance are shown in Figures 4 and 5. The pattern of these differences provides some insights into the task demands and processing operations used in open-set word recognition tests. First, the Stars consistently demonstrate higher levels of word recognition performance on both the LNT and the MLNT than the low-performers. These differences are present across all six years but they are most prominent during the first three years after implantation. Word recognition scores for the low-performers on both the LNT and the MLNT are very low and close to the floor compared to the performance observed for the Stars who are doing moderately well on this test although they never reached ceiling levels of performance on either test even after six years of implant use. Normal-hearing children typically display very high levels of performance on both of these tests by age 4 (Kluck et al., 1997).

The Stars also displayed a word length effect at each testing interval. Recognition was always better for long words on the MLNT than for short words on the LNT. This pattern is not present for the low-performers who were unable to do this open-set task at all during the first three years. The existence of a word length effect for the Stars suggests that these children are recognizing spoken words “relationally” in the context of other words they have in their lexicon (Luce & Pisoni, 1998). If these children were just recognizing words in isolation, either holistically as global temporal patterns or segment-by-segment without reference to the representations of words they already know, we would expect performance to be worse for longer words than shorter words because longer words contain more stimulus information. But this is not what we found.

The pattern of results for the Stars is exactly the opposite of this prediction and parallels earlier results obtained with normal-hearing adults and normal-hearing typically-developing children (Luce & Pisoni, 1998; Kirk et al., 1995; Kluck et al., 1997). Longer words are easier to recognize than shorter words because they are phonologically more distinctive and discriminable and therefore less confusable with other phonetically similar words. The present findings suggest that the Stars are recognizing words based on their knowledge of other words in the language using processing strategies that are similar to those used by normal-hearing listeners.

Additional support for role of the lexicon and the use of phonological knowledge in open-set word recognition is provided by another finding. The Stars also displayed a consistent effect of "lexical discrimination." As shown in Figures 4 and 5, the Stars recognized lexically "easy" words better than lexically "hard" words. The difference in performance between "easy" words and "hard" words is present for both the LNT and the MLNT vocabularies although it is larger and more consistent over time for the MLNT test. Once again, the lower-performing children did not display sensitivity to lexical competition among the test words.

The differences in performance observed between these two groups of children on both open-set word recognition tests were not at all surprising because the two extreme groups were initially created based on their PBK scores, another open-set word recognition test. However, the overall pattern of the results shown in Figures 4 and 5 is theoretically important because the findings demonstrate that the processes used in recognizing isolated spoken words are not specific to the particular test items on the PBK test or the experimental procedures used in open-set tests of word recognition. The differences between the two groups of children readily generalized to two other open-set word recognition tests that use completely different test words.

The pattern of results strongly suggests a common underlying set of linguistic processes that is employed in recognizing and imitating spoken words presented in isolation. Understanding the cognitive and linguistic processing mechanisms that are used in open-set word recognition tasks may provide new insights into the underlying basis of the individual differences observed in outcome measures in children with cochlear implants. It is probably no accident that the PBK test, which is considered the "gold standard" of performance, has had some important diagnostic utility in identifying the exceptionally good users of cochlear implants over the years (see Kirk et al., 1995; Meyer & Pisoni, 1999). The PBK test measures fundamental language processing skills that generalize well beyond the specific word recognition task used in open-set tests. The important conceptual issue is to explain why this happens and identify the underlying cognitive and linguistic processing mechanisms used in open-set word recognition tasks as well as other language processing tasks. We will return to this issue again below.

## Comprehension of Common Phrases

Language comprehension performance was also measured in these two groups of children using the Common Phrases Test (Osberger et al., 1991), an open-set test with three presentation formats: auditory-only (CPA), visual-only (CPV) and combined auditory plus visual (CPAV). Children are asked questions or given directions to follow under these three conditions. The results of the Common Phrases Test are shown in Figure 6 for both groups of subjects as a function of implant use for the three different presentation formats.

Figure 6 shows that the Stars performed consistently better than the low-performers in all three presentation conditions and across all six years of implant use although performance begins to approach ceiling levels for both groups in the CPAV condition after five years of implant use. CPAV conditions were always better than either the CPA or CPV conditions. This pattern was observed for both groups of subjects. In addition, both groups displayed improvements in performance over time in all three presentation conditions. Not surprisingly, the largest differences in performance between the two groups occurred in the CPA conditions. Even after three years of implant use, the lower-performing children were barely able to perform the common phrases task above 25% correct when they had to rely entirely on auditory cues in the speech signal to carry out the task.

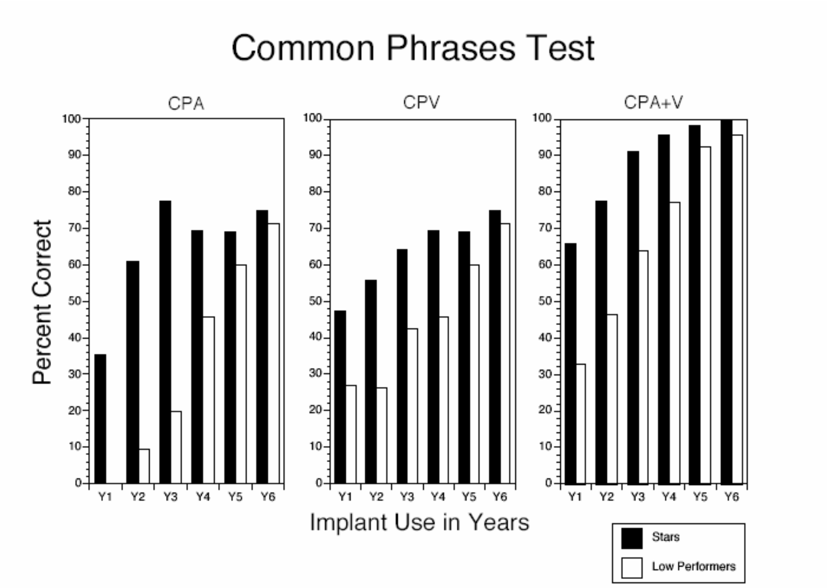


**Figure 6.** Percent correct performance on the Common Phrases Test (CPT) for auditory-only (CPA), visual-only (CPV) and combined auditory plus visual presentation modes (CPAV) as a function of implant use. The Stars are shown by filled bars, the low-performers are shown by open bars (from Pisoni et al., 2000).

## Correlations among Measures of Speech Perception

These descriptive results show that the exceptionally good performers, the Stars, do well on measures of speech feature discrimination, spoken word recognition and language comprehension. They also do well on other tests of receptive and expressive language, vocabulary knowledge and speech intelligibility (see Pisoni et al., 1997; 2000). This pattern of findings suggests that a common source of variance may underlie the exceptionally good performance of the Stars on a range of different speech and language outcome measures. Until our investigation of the exceptionally good children, no one had studied individual differences in outcome in this clinical population or the underlying perceptual, cognitive and linguistic processes. The analyses of speech feature discrimination, spoken word recognition, and spoken language comprehension scores summarized here demonstrate that a child who displays exceptionally good performance on the PBK test also shows good scores on other speech perception tests. Analyses of the other outcome measures revealed a similar pattern of results.

To assess the relations between these different tests, we carried out a series of simple correlations on the speech perception scores and the other outcome measures. We were interested in the following questions: Does a child who performs exceptionally well on the PBK test also perform exceptionally well on other tests of speech feature discrimination, word recognition and comprehension? Is the good performance of the Stars restricted only to open-set word recognition tests or is it possible to identify a common underlying variable or process that can account for the relations observed among the other outcome measures?

Simple bivariate correlations were carried out separately for the Stars and low-performers using the test scores obtained after one year of implant use (see Pisoni et al., 1997; 2000 for the full report). The results of the correlational analyses on the outcome measures revealed a strong and consistent pattern of intercorrelations among all of the test scores for the Stars (see Pisoni et al. 1997, 2000). This pattern was observed for all three of the speech perception tests described here as well as vocabulary knowledge, receptive and expressive language and speech intelligibility. The outcome measures that correlated the most strongly and most consistently with the other tests were the open-set word recognition scores on the LNT and MLNT tests.

The finding that performance on open-set word recognition was strongly correlated with all of the other outcome measures was of special interest to us. The pattern of intercorrelations among all these dependent measures strongly suggests a shared common underlying source of variance. The extremely high correlations with the open-set word recognition scores on the LNT suggests that the common source of variance may be related to the processing of spoken words, specifically to the encoding, storage, retrieval and manipulation of the phonological representations of spoken words. The fundamental cognitive and linguistic processes used to recognize (decompose) and repeat (reassemble) spoken words in an open-set tests like the PBK or LNT are also used in other language processing tasks, such as comprehension and speech production and even nonword repetition, which draw on the same sources of phonological information about spoken words in the lexicon.

The results of the correlational analyses suggest several hypotheses about the source of the differences in performance between the Stars and the low-performers. Some proportion of the variation in outcome appears to be related to how the initial sensory information is processed and used in clinical tests that assess speech feature discrimination, word recognition, language comprehension and speech production. Unfortunately, the data available on these children were based on traditional audiological outcome measures that were collected as part of their annual clinical assessments. All of the scores on these behavioral tests are “endpoint measures” of performance that reflect the final product of perceptual and linguistic analysis. Process measures of performance that assess what a child does with the sensory information provided by his/her cochlear implant were not part of the standard research protocol used in our longitudinal study so it was impossible to examine differences in processing capacity and speed. It is very likely that fundamental differences in both information processing capacity and speed are responsible for the individual differences observed between these two groups of children.

For a variety of theoretical reasons, we refocused our research efforts to study “working memory.” One reason is that working memory plays a central role in human information processing because it serves as the primary interface between sensory input and stored knowledge in long-term memory. Another is that working memory has also been shown to be a major source of individual differences in processing capacity across a wide range of domains from perception to memory to language (Ackerman, Kyllonen & Roberts, 1999; Carpenter, Miyake & Just, 1994; Baddeley, Gathercole, & Papagno, 1998; Gupta & MacWhinney, 1997).

## Measures of Working Memory

To obtain some new measures of working memory capacity from a large group of deaf children following cochlear implantation, we began collaborating with Dr. Ann Geers and her colleagues at Central Institute for the Deaf (CID) in St. Louis where there was a large-scale clinical research project underway. They collected a wide range of different outcome measures of speech, language and reading skills from 8- and 9-year old children who had used their cochlear implants for at least three and one-half years. Thus, in this study, chronological age and length of implant use were controlled.

Using the test lists and procedures from the WISC III (Wechsler 1991), forward and backward auditory digit spans were obtained from 176 deaf children who were tested in separate groups during the summers of 1997, 1998, 1999 and 2000. Forward and backward digit spans were also collected from an additional group of 45 age-matched normal-hearing 8- and 9-year old children who were tested in Bloomington, Indiana, and served as a comparison group (see Pisoni & Cleary, 2003).

The WISC-III memory span task requires the child to repeat a list of digits that is spoken live-voice by an experimenter at a rate of approximately one digit per second (WISC-III Manual, Wechsler 1991). In the "digits-forward" condition, the child is required simply to repeat the list as heard. In the "digits-backward" condition, the child is told to "say the list backward." In both subtests, the lists begin with two items and increase in length until a child gets two lists incorrect at a given length, at which time testing stops. Points are awarded for each list correctly repeated with no partial credit.

A summary of the digit span results for all five groups of children is shown in Figure 7. Forward and backward digit spans are shown separately for each group. The children with cochlear implants are shown in the four panels on the left by year of testing; the normal-hearing children are shown on the right. Each child's digit span in points was calculated by summing the number of lists correctly recalled at each list length.

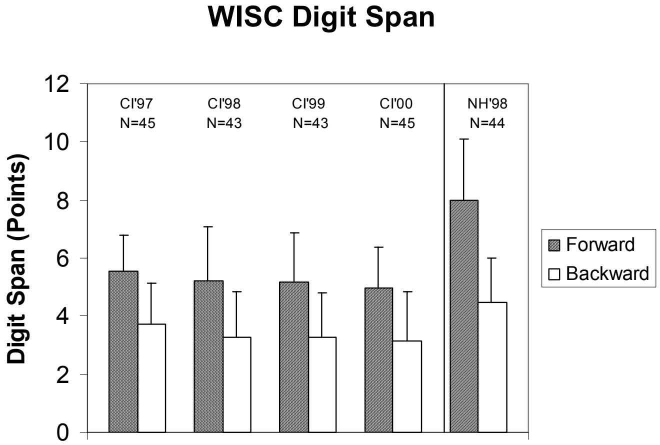


**Figure 7.** WISC digit spans scored by points for the four groups of 8- and 9-year old children with cochlear implants and for a comparison group of 8- and 9-year-old normal-hearing children. Forward digit spans are shown by the shaded bars, backwards digit spans by the open bars. Error bars indicate one standard deviation from the mean (from Pisoni & Cleary, 2003).

The results shown in Figure 7 reveal a systematic pattern of the forward and backward digit spans for the deaf children with cochlear implants. All four groups are quite similar to each other. In each group, the forward digit span is longer than the backward digit span. The pattern is quite stable over the four years of testing despite the fact that these scores were obtained from independent groups of children. The difference in span length between forward and backward report was highly significant for the entire group of 176 deaf children and for each group taken separately ($p < .001$).

The forward and backward digit spans obtained from the 44 age-matched normal-hearing children are shown in the right-hand panel of Figure 7. These results show that the digit spans for the normal-hearing children differ in several ways from the spans obtained from the children with cochlear implants. First, both digit spans are longer than the spans obtained from the children with cochlear implants. Second, the forward digit span for the normal-hearing children is much longer than the forward digit spans obtained from the children with cochlear implants. This latter finding is particularly important because it demonstrates atypical development of the deaf children's short-term memory capacity and suggests several possible differences in the underlying processing mechanisms that are used to encode and maintain sequences of spoken digits in immediate memory.

Numerous studies have suggested that forward digit spans reflect coding strategies related to phonological processing and verbal rehearsal mechanisms used to maintain information in short-term memory for brief periods of time before retrieval and output response. Differences in backward digit spans, on the other hand, are thought to reflect the contribution of controlled attention and operation of higher-level "executive" processes that are used to transform and manipulate verbal information for later processing operations (Rudel & Denckla, 1974; Rosen & Engle, 1997).

The digit spans for the normal-hearing children shown in Figure 7 are age-appropriate and fall within the published norms for the WISC III. However, the forward digit spans obtained from the children with cochlear implants are atypical and suggest possible differences in encoding and/or verbal rehearsal processes used in immediate memory. In particular, the forward digit spans reflect differences in processing capacity of immediate memory between the two groups of children. These differences may cascade and affect other information processing tasks that make use of working memory and verbal rehearsal processes. Because all of the clinical tests that are routinely used to assess speech and language outcomes rely heavily on component processes of working memory and verbal rehearsal, it seems reasonable to assume that these tasks will also reflect variability due to basic differences in immediate memory and processing capacity.

## Correlations with Digit Spans

In order to learn more about the differences in auditory digit span and the limitations in processing capacity, we examined the correlations between forward and backward digit spans and several speech and language outcome measures also obtained from these children at CID (see Pisoni & Cleary, 2003). Of the various demographic measures available, the only one that correlated strongly and significantly with digit span was the child's communication mode. This measure is used to quantify the nature of the child's early sensory and linguistic experience after receiving a cochlear implant in terms of the degree of emphasis on auditory-oral language skills by teachers and therapists in the educational environment.

We found that forward digit span was positively correlated with communication mode ($r = +.34$, $p < .001$). Children who were in language learning environments that primarily emphasized oral skills displayed longer forward digit spans than children who were in total communication (TC) environments. However, the correlation between digit span and communication mode was highly selective in nature because it was restricted only to the forward digit span scores; the backward digit spans were not correlated with communication mode or any of the other demographic variables.

In order to examine the effects of early experience in more detail, a median split was carried out on the communication mode scores to create two subgroups. Figure 8 shows the digit spans plotted separately for the oral and total communication children for each of the four years of testing at CID. The oral group consistently displayed longer forward digit spans than the total communication group. While the differences in forward digit span between oral and total communication children were highly significant, the differences in backward digit span were not. This pattern suggests that the effects of early sensory and linguistic experience on immediate memory is related to coding and verbal rehearsal processes that affect only the forward digit span conditions in this task.

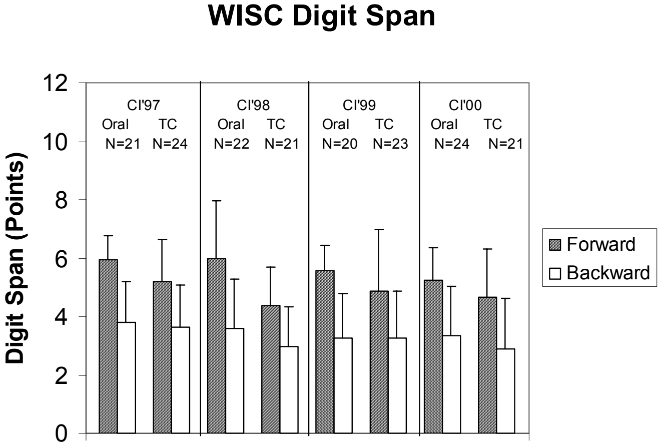


**Figure 8.** WISC digit spans scored by points for the four groups of 8- and 9-year old children with cochlear implants, separated by communication mode. For each year, scores for the oral group are shown to the left of those for the total communication group. Forward digit spans are shown by the shaded bars, backwards digit spans by the open bars. Error bars indicate one standard deviation from the mean (from Pisoni & Cleary, 2003).

The difference in forward digit span between oral and total communication children is present for each of the four groups. These differences could be due to several factors such as more efficient encoding of the initial stimulus patterns into stable phonological representations in working memory, speed and efficiency of the verbal rehearsal processes used to maintain phonological information in working memory or possibly even speed of retrieval and scanning of information in working memory after recognition has taken place. All three factors could influence measures of processing capacity and any one of these could affect the number of digits correctly recalled from immediate memory in this task.

### Digit Spans and Word Recognition

Although these results indicate that early experience in an environment that emphasizes oral language skills is associated with longer forward digit spans and increased information processing capacities of working memory, without additional converging measures of performance, it is difficult to specify precisely what elementary processes and information processing mechanisms are actually affected by early experience and which ones are responsible for the increases in forward digit spans observed in these particular children. Recent studies of normal-hearing children have demonstrated close “links” between working memory and learning to recognize and understand new words (Gupta & MacWhinney, 1997; Gathercole et al., 1997). Other research has found that vocabulary development and several important milestones in speech and language acquisition are also associated with differences in measures of working memory, specifically, measures of digit span, which can be used as estimates of processing capacity of immediate memory (Gathercole & Baddeley, 1990).

To determine if immediate memory capacity is related to spoken word recognition, we correlated the WISC forward and backward digit span scores with three different measures of word recognition. A summary of the correlations between digit span and word recognition scores based on these 176 children is shown in Table I.

The WIPI test (Word Intelligibility by Picture Identification Test) is a closed-set test of word recognition in which the child selects a word from among six alternative pictures (Ross & Lerman, 1979). As described earlier, the LNT is an open-set test of word recognition and lexical discrimination that requires the child to imitate and reproduce an isolated word (Kirk, Pisoni & Osberger, 1995). Finally, the BKB is an open-set word recognition test in which key words are presented in sentences (Bench, Kowal & Bamford, 1979).

**Table I**
**Correlations between WISC digit span and three measures of spoken word recognition (from Pisoni & Cleary, 2003).**

| | Simple Bivariate Correlations | | Partial Correlations[a] | |
|---|---|---|---|---|
| | WISC Forward Digit Span | WISC Backward Digit Span | WISC Forward Digit Span | WISC Backward Digit Span |
| Closed Set Word Recognition (WIPI) | .42*** | .28*** | .25** | .12 |
| Open Set Word Recognition (LNT-E) | .41*** | .20** | .24** | .07 |
| Open Set Word Recognition in Sentences (BKB) | .44*** | .24** | .27*** | .09 |

*** p <.011, ** p<.01

[a]Statistically Controlling for: Communication Mode Score, Age at Onset of Deafness, Duration of Deafness, Duration of Cochlear Implant Use, Number of Active Electrodes, VIDSPAC Total Segments Correct (Speech Feature Perception Measure), Age

Table I displays two sets of correlations. The left-hand portion of the table shows the simple bivariate correlations of the forward and backward digit spans with the three measures of word recognition. The correlations for both the forward and backward spans reveal that children who had longer WISC digit spans also had higher word recognition scores on all three word recognition tests. This finding is present for both forward and backward digit spans. The correlations are all positive and reached statistical significance.

The right-hand portion of Table I shows a summary of the partial correlations among these same measures after we statistically controlled for differences due to chronological age, communication mode, duration of deafness, duration of device use, age at onset of deafness, number of active electrodes and speech feature discrimination. When these "contributing variables" were removed from the correlational analyses, the partial correlations between digit span and word recognition scores became smaller in magnitude overall. However, the correlations of the forward digit span with the three word recognition scores were still positive and statistically significant while the correlations of the backward digit spans were weaker and no longer significant.

These results demonstrate that children who have longer forward WISC digit spans also show higher word recognition scores; this relationship was observed for all three word recognition tests even after the other sources of variance were removed. The present results suggest a common source of variance that is shared between forward digit span and measures of spoken word recognition that is independent of other mediating factors that have been found to contribute to the variation in these outcome measures.

## Digit Spans and Speaking Rate

While the correlations of the digit span scores with communication mode and spoken word recognition suggest fundamental differences in encoding and rehearsal speed which are influenced by the nature of the early experience a child receives, these measures of immediate memory span and estimates of information processing capacity are not sufficient on their own to identify the underlying information processing mechanism responsible for the individual differences. Additional converging measures are needed to pinpoint the locus of these differences more precisely. Fortunately, an additional set of behavioral measures was obtained from these children for a different purpose and made available to us for several new analyses.

As part of the research project at CID, speech production samples were obtained from each child to assess speech intelligibility and measure changes in articulation and phonological development following implantation (see Tobey et al., 2000). The speech samples consisted of three sets of meaningful English sentences that were elicited using the stimulus materials and experimental procedures developed by McGarr (1983). All of the utterances produced by the children were originally recorded and stored digitally for playback to groups of naïve adult listeners who were asked to transcribe what they thought the children had said. In addition to the speech intelligibility scores, we measured the durations of the individual sentences in each set and used these to estimate each child's speaking rate.

The sentence durations provide a quantitative measure of a child's articulation speed which we knew from a large body of earlier research in the memory literature was closely related to speed of subvocal verbal rehearsal (Cowan et al., 1998). Numerous studies over the past 25 years have demonstrated strong relations between speaking rate and memory span for digits and words (for example Baddeley, Thompson & Buchanan, 1975). The results of these studies suggest that measures of an individual's speaking rate reflect articulation speed and this measure can be used as an index of rate of covert verbal rehearsal for phonological information in working memory. Individuals who speak more quickly have been found to have longer memory spans than individuals who speak more slowly.

The forward digit span scores for the 168 children are shown in Figure 9 along with estimates of their speaking rates obtained from measurements of their productions of meaningful English sentences. The digit spans are plotted on the ordinate; the average sentence durations are shown on the abscissa. The top panel shows mean sentence durations; the bottom panel shows the log sentence durations. The pattern of results in both figures is very clear; children who produce sentences with longer durations speak more slowly and, in turn, have shorter forward digit spans. The correlations between forward digit span and both measures of sentence duration were strongly negative and highly significant (*r* = -.63 and *r* = -.70; *p* <.001, respectively). It is important to emphasize once again, that the relations observed here between digit span and speaking rate were selective in nature and were found only for the forward digit spans. There was no correlation at all between backward digit span scores and sentence duration in any of our analyses.

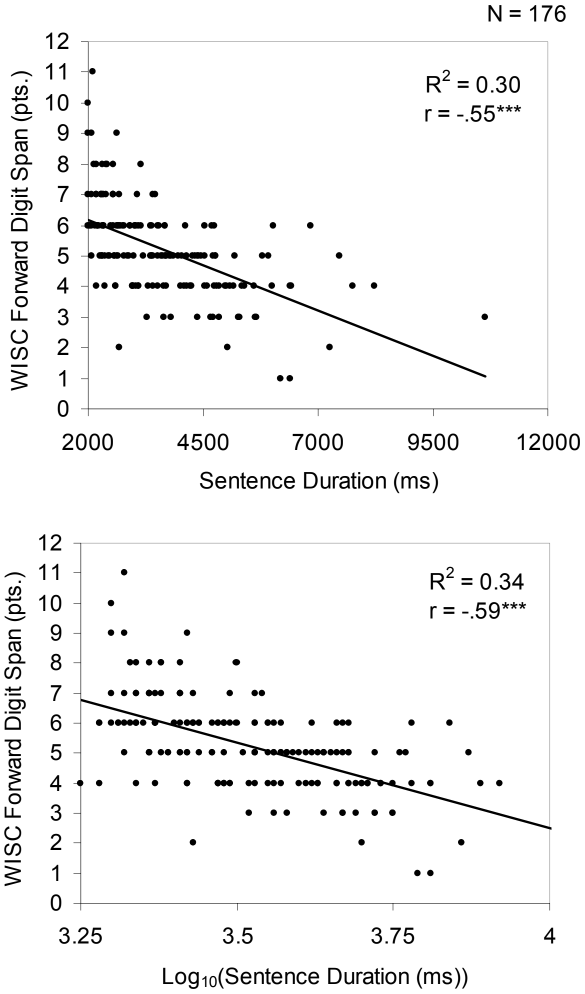


**Figure 9.** Scatterplots illustrating the relationship between average sentence duration for the seven-syllable McGarr Sentences (abscissa) and WISC forward digit span scored by points (ordinate). Each data-point represents an individual child. Non-transformed duration scores are shown in the top panel, log-transformed duration scores in the bottom panel. R-squared values indicate percent of variance accounted for by the linear relation (from Pisoni & Cleary, 2003).

The dissociation between forward and backward digit spans and the correlation of the forward spans with measures of speaking rate suggests that verbal rehearsal speed may be the primary underlying factor that is responsible for the variability and individual differences observed in deaf children with cochlear implants on a range of behavioral speech and language tasks. The common feature of each of these outcome measures is that they all make use of the storage and processing mechanisms of verbal working memory.

## Speaking Rate and Word Recognition

To determine if verbal rehearsal speed is also related to individual differences in word recognition performance, we examined the correlations between sentence duration and the three different measures of spoken word recognition described earlier. All of these correlations are also positive and suggest once again that a common processing mechanism, verbal rehearsal speed, may be the factor that underlies the variability and individual differences observed in these word recognition tasks.

Our analysis of the digit span scores from these deaf children uncovered two important correlations linking forward digit span to both word recognition performance and speaking rate. Both of the correlations with forward digit span suggest a common underlying processing factor that is shared by each of these dependent measures. This factor appears to reflect the speed of verbal rehearsal processes in working memory. If this hypothesis is correct, then word recognition and speaking rate should also be correlated with each other because they make use of the same processing mechanism. This is exactly what we found. As in the earlier analyses, differences due to demographic factors and the contribution of other variables were statistically controlled for by using partial correlation techniques. In all cases, the correlations between speaking rate and word recognition were negative and highly significant. Thus, slower speaking rates are associated with poorer word recognition scores on all three word recognition tests. These findings linking speaking rate and word recognition suggest that all three measures, digit span, speaking rate and word recognition performance are closely related because they share a common underlying source of variance.

To determine if digit span and sentence duration share a common process and the same underlying source of variance which relates them both to word recognition performance, we re-analyzed the intercorrelations between each pair of variables with the same set of the demographic and mediating variables systematically partialled out. When sentence duration was partialled out of the analysis, the correlations between digit span and each of the three measures of word recognition essentially approached zero. However, the negative correlations between sentence duration and word recognition were still present even after digit span was partialled out of the analysis suggesting that processing speed is the common factor that is shared between these two measures.

The results of these analyses confirm that the underlying factor that is shared in common with speaking rate is related to the rate of information processing, specifically, the speed of the verbal rehearsal process in working memory. This processing component of verbal rehearsal could reflect either the articulatory speed used to maintain phonological patterns in working memory or the time to retrieve and scan phonological information already in working memory (see Cowan et al., 1998). In either case, the common factor that links word recognition and speaking rate appears to be related in some way to the speed of information processing operations used to store and maintain phonological representations in working memory (see Pisoni & Cleary, 2003).

## Speech Timing and Working Memory

In addition to our recent studies on verbal rehearsal speed, we have also obtained several new measures of memory scanning during the digit recall task from a group of deaf children with cochlear implants and a group of typically-developing age-matched normal-hearing children (see Burkholder & Pisoni, 2003). Our interest in studying speech timing in these children was motivated by several recent findings reported by Cowan and his colleagues who have carefully measured the response latencies and interword durations during recall tasks in children of different ages.

In one study of immediate recall, Cowan et al. (1994) found that interword pause times provided a reliable measure of the dynamics of the memory scanning and retrieval process during development. Their results showed that children's interword pauses in immediate recall increased as list length increased. This finding supports Cowan's earlier (1992) proposal that serial scanning is carried out during the pauses. Recall of longer lists requires that more items have to be scanned serially, therefore prolonging interword pause time. Additional evidence showing that items in short-term memory are scanned during interword pauses was obtained in another study by Cowan et al. (1998) who found that children with shorter interword pauses also had longer immediate memory spans.

Cowan et al. (1998) also reported that older children have shorter pause durations in immediate recall than younger children. Taken together, their results on speech timing suggest that the memory span increases observed in older children might be associated with both shorter interword pauses during serial recall and faster speaking rates. Shorter interword pauses indicate that the scanning mechanisms used to retrieve items from short-term memory are executed faster and more efficiently in the older children. Combined with increases in articulation speed, this factor may enhance the ability to engage in efficient verbal recall strategies as children develop. These findings on speech timing in immediate memory tasks led Cowan and his colleagues to propose that two processing operations —serial scanning and retrieval of items from short-term memory and subvocal verbal rehearsal of phonological information are used by typically developing children in recall and both of these factors affect measures of working memory capacity (Cowan, 1999; Cowan et al., 1998).

Recently, we obtained several measures of speech-timing during immediate recall from a group of deaf children who use cochlear implants (see Burkholder & Pisoni, 2003). Measures of speaking rate and speech timing were also obtained from an age-matched control group of normal-hearing, typically developing children. Articulation rate and subvocal rehearsal speed were measured from sentence durations elicited using meaningful English sentences. Relations between articulation rate and working memory in each group of children were compared to determine how verbal rehearsal processes might differ between the two populations. To assess differences in speech timing during recall, response latencies, durations of the test items, and interword pauses were measured in both groups of children.

For the analysis of the speech-timing measures during recall, we analyzed only the responses from the digit span forward condition. Analysis of the speech-timing measures obtained during recall revealed no differences in the average duration of articulation of the individual digits or response latencies at any of the list lengths. There was no correlation between the average articulations taken from digit span forward and forward digit span scores when all children were considered together or when the children were evaluated in groups according to hearing ability or communication mode.

However, we found that interword pause durations in recall differed significantly among the groups of children. The average of individual pauses that occurred during recall in the forward condition was significantly longer in the deaf children with cochlear implants than in the normal-hearing children at list lengths three and four.

The results of this study replicated our previous findings showing that profoundly deaf children with cochlear implants have shorter digit spans than their normal-hearing peers. As expected, deaf children with cochlear implants also displayed longer sentence durations than normal-hearing children. Total communication users displayed slower speaking rates and shorter forward digit spans than the oral communication users. In addition to producing longer sentence durations than normal-hearing children, the deaf children with cochlear implants also had much longer interword pause durations during recall. Longer interword pauses are assumed to reflect slower serial scanning processes which may affect the retrieval of phonological information in short-term memory (Cowan, 1992; Cowan et al., 1994). Taken together, the pattern of results indicates that both slower subvocal rehearsal and serial scanning are associated with shorter digit spans in the deaf children with cochlear implants.

The overall pattern of speech-timing results found in both groups of children is quite similar to the findings reported by Cowan et al. (1998) with normal-hearing children. Their findings suggest that covert verbal rehearsal and the speed of serial scanning of items in short-term memory are two factors that affect immediate memory span in normal-hearing children. Cowan et al. also found that children who were faster at subvocal verbal rehearsal and serial scanning displayed longer immediate memory spans than children who executed these processes more slowly. However, his findings were obtained from typically developing normal-hearing children who differed only in chronological age.

Comparable results were observed in our study using children of similar chronological ages but with quite different developmental histories that reflect the absence of sound and early auditory experience during critical periods of perceptual and cognitive development. The effects of early auditory and linguistic experience found by Burkholder and Pisoni (2003) suggest that the development of subvocal verbal rehearsal and serial scanning processes may not only be related to maturationally-based milestones that are cognitively or metacognitively centered, such as the ability to effectively organize and utilize these two processes in tasks requiring immediate recall. Rather, efficient subvocal verbal rehearsal strategies and scanning abilities also appear to be experience- and activity-dependent reflecting the development of neural mechanisms used in speech perception and speech production.

Because the group of deaf children examined in the Burkholder and Pisoni (2003) study fell within a normal range of intelligence, the most likely developmental factor responsible for producing slower verbal rehearsal speeds, scanning rates, and shorter digit spans is an early period of auditory and linguistic deprivation prior to receiving a cochlear implant. Sensory deprivation may result in widespread developmental brain plasticity and neural reorganization, further differentiating deaf children's perceptual and cognitive development from the development of normal-hearing children (Kaas, Merzenich & Killackey, 1983; Shepard & Hardie, 2001). Brain plasticity affects not only the development of the peripheral and central auditory systems but other higher cortical areas as well, both before and after cochlear implantation (Ryugo, Limb, & Redd, 2000; Teoh, Pisoni & Miyamoto, in press a, b).

## Discussion and Conclusions

Our recent findings on speech perception and working memory provide some new insights about the elementary information processing skills of deaf children with cochlear implants and the underlying cognitive and linguistic factors that affect the development of their speech and language skills on a range of outcome measures. These studies were specifically designed to obtain new process measures of performance that assessed the operation of verbal working memory in order to understand the nature of the capacity limitations in encoding and processing phonological information. Several important findings have emerged from our analysis of the memory span data suggesting that working memory capacity, verbal rehearsal speed and scanning processes in short-term memory contribute additional unique sources of variance to the outcome measures obtained with deaf children following cochlear implantation. The pattern of digit span scores, measures of speaking rate and speed of scanning of items in short-term memory clearly demonstrate the presence of atypical development of short-term working memory capacity in these deaf children. It also supports our initial hypothesis that cognitive processing variables contribute to the large individual differences observed in a range of outcome measures used to assess speech and language performance in these children.

The only demographic variable that was correlated with these cognitive processing measures was the child's communication mode. Deaf children who were immersed in oral-only environments displayed longer forward digit spans, faster speaking rates and more efficient scanning of short-term memory than the children who were in total communication environments. The presence of selective effects of early sensory experience on working memory suggests that the stimulus environment and the specific kinds of activities and experiences that children have with their parents and caretakers in the language learning environment operate in a highly selective manner on a specific information processing mechanism and subcomponent of the human memory system that is used for encoding, maintaining and retrieving phonological information in short-term memory. We suspect there may be something unique about the oral environment and the specific experiences and activities that the child engages in on a regular basis that produces selective effects on verbal rehearsal and phonological coding of speech signals.

Because children from total communication environments may simply have less exposure to speech and spoken language in their early linguistic environment after receiving their implant than oral children, they may display problems in both processing and actively rehearsing phonological information in short-term memory. In terms of initial encoding and recognition, the reduced exposure to speech and spoken language may affect the development of automatic attention and specifically the speed with which speech signals can be rapidly identified and encoded into stable phonological representations in short term memory. Thus, total communication children may have fundamental problems in scanning and retrieving phonological information in short-term memory. In terms of verbal rehearsal, total communication children may have slower and less efficient verbal rehearsal processes once information gets into short-term memory simply because they have had less experience than oral children in producing speech and actively generating phonological patterns.

Passive exposure to speech without explicit analysis and conscious manipulation of phonological representations may not be sufficient to develop robust lexical representations of spoken words and fluency in control of speech production. Deaf children who receive cochlear implants may need to be actively engaged in processing spoken language in order to develop automaticity and automatic attention strategies that can be carried out rapidly without conscious effort or processing resources. This may be one direct benefit of auditory-oral education programs. The excellent spoken language skills acquired by children in these programs may reflect the development of highly automatized phonological analysis skills which permit the child to engage in active processing strategies in perception that involve “decomposition” of a speech pattern into a sequence of discrete phonological units and then the “reassembly” of those individual units into sequences of gestures and sensory-motor patterns for use in speech production and articulation.

The development of automatized phonological processing skills may result in increases in the speed and efficiency of constructing phonological and lexical representations of spoken words in working memory. Recovering the internal structure of an input pattern in speech perception as a result of perceptual analysis and then reconstructing the same pattern in speech production may serve to establish permanent links between speech perception and production and may lead to further development of highly efficient sensory-motor articulatory programs for verbal rehearsal and coding of words in working memory. Thus, the development of phonological processing skills may simply be a byproduct of the primary emphasis on speech and oral language skills in oral-only educational environments and may account for why these children consistently display better performance on a wide range of outcome measures of speech and language.

The present set of findings permits us to identify a specific information processing mechanism, the verbal rehearsal process in working memory that is responsible for the limitations on processing capacity. Processing limitations are present in a wide range of clinical tests that make use of verbal rehearsal and phonological processing skills to encode, store, maintain and retrieve spoken words from working memory. These fundamental information processing operations are components of all of the current clinical outcome measures routinely used to assess receptive and expressive language functions. Our findings suggest that the variability in performance on the traditional clinical outcome measures used to assess speech and language processing skills in deaf children after cochlear implantation may simply reflect fundamental differences in the speed of information processing operations such as verbal rehearsal, scanning of items in short-term memory and the rate of encoding phonological and lexical information in working memory.

We believe these new results are clinically and theoretically significant because they suggest a motivated theoretically-based explanation for the enormous variability and individual differences observed in a range of speech and language processing tasks that make use of the same verbal rehearsal processes. As in normal-hearing typically-developing children, the present findings suggest that differences in verbal rehearsal speed may be the primary factor that is responsible for the large individual differences in speech and language development observed in deaf children following cochlear implantation.

## References

Ackerman, P.L., Kyllonen, P.C. & Roberts, R.D. (1999). *Learning and individual differences*. American Psychological Association: Washington, DC.

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

Baddeley, A.D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Ball, G.F. & Hulse, S.H. (1998). Birdsong. *American Psychologist, 53,* 37-58.

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.

Bergeson, T. & Pisoni, D.B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence & B.E. Stein (Eds.), *Handbook of Multisensory Integration,* (pp. 749-772). Cambridge: MIT Press.

Burkholder, R. & Pisoni, D.B. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85,* 63-88.

Carpenter, P.A., Miyake, A. & Just, M.A. (1994). Working memory constraints in comprehension. In M.A. Gernsbacher (Ed.) *Handbook of Psycholinguistics*, (pp. 1075-1122). San Diego: Academic Press.

Clarke, G. (2003). *Cochlear Implants: Fundamentals and Applications*. New York: Springer-Verlag.

Cowan, N. (1992). Verbal memory and the timing of spoken recall. *Journal of Memory and Language*, *31*, 668-684.

Cowan, N. (1999). The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology, 72*, 193-209.

Cowan, N., Keller, T., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language, 33*, 234-250.

Cowan, N., Wood, N.L., Wood, P.K., Keller, T.A., Nugent, L.D. & Keller, C.V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General, 127,* 141-160.

Ericsson, K.A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. New York, NY: Cambridge University Press.

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., Hitch, G.J., Service, E. & Martin, A.J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33,* 966-979.

Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59,* 267-333.

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

Kaas, J. H., Merzenich, M.M., & Killackey, H.P. (1983). The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals. *Annual Review of Neuroscience,* 6, 325-356.

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 Waltzman, S.B., & Cohen, N.L. (Eds.), *Cochlear Implants,* (pp. 252-254). New York: Thieme.

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

Kluck, M., Pisoni, D.B. & Kirk, K.I. (1997). Performance of normal-hearing children on open-set speech perception tests. *Progress Report on Spoken Language Processing #21*, Indiana University, Department of Psychology, Bloomington, IN.

Konishi, M. (1985). Birdsong: From behavior to neuron. *Annual Review of Neuroscience, 8,* 125-170.

Konishi, M., & Nottebohm, R. (1969). Experimental studies in the ontogeny of avian vocalizations. In R.A. Hinde (Ed.), *Bird Vocalizations,* (pp. 29-48). New York: Cambridge University Press.

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

Marler, P., & Peters, S. (1988). Sensitive periods for song acquisition from tape recordings and live tutors in the swamp sparrow, Meiosis georgiana. *Ethology, 77,* 76-84.

McGarr, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26,* 451-458.

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.

Osberger, M.J., Miyamoto, R.T., Zimmerman-Phillips, S., et al. (1991). Independent evaluations of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear Hearing, 12,* 66-80.

Pisoni, D.B. & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear & Hearing, 24,* 106S-120S.

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

Pisoni, D. B., Svirsky, M. A., 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. *Research on Spoken Language Processing Progress Report No. 21*. Bloomington, IN: Speech Research Laboratory, (pp. 51-91).

Robbins, A.M., Renshaw, J.J., Miyamoto, R.T., Osberger, M.J., & Pope, M.L. (1988). *Minimal pairs test.* Indianapolis, IN: Indiana University School of Medicine.

Rosen, V.M. & Engle, R.W. (1997). Forward and backward serial recall. *Intelligence, 25,* 37-47.

Ross, M., & Lerman, J. (1979). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13,* 44-53.

Rudel, R.G. & Denckla, M.B. (1974). Relation of forward and backward digit repetition to neurological impairment in children with learning disability. *Neuropsychologia, 12,* 109-118.

Ryugo, D., Limb, C., & Redd, E. (2000). Brain Plasticity: The impact of the environment on the brain as it relates to hearing and deafness. In J. Niparko (Ed.), *Cochlear Implants, Principles and Practices,* (pp. 33-56). Philadelphia: Lippincott Williams & Wilkins.

Sharma, A, Dorman, M.F. & Spahr, A.J. (2002) A sensitive period for the development of the central auditory system in children with cochlear implants: Implications for age of implantation. *Ear & Hearing, 23*, 532-539.

Shepard, R.K. & Hardie, N. (2001). Deafness-induced changes in the auditory pathway: Implications for cochlear implants. *Audiology and Neuro-Otology, 6*, 305-318.

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.

Teoh, S.W., Pisoni, D.B. & Miyamoto, R.T. (In press, a). Cochlear implantation in adults with prelingual deafness: I. Clinical results. *Laryngoscope.*

Teoh, S.W., Pisoni, D.B. & Miyamoto, R.T. (In press, b). Cochlear implantation in adults with prelingual deafness: II. Underlying constraints that affect audiological outcomes. *Laryngoscope.*

Tobey, E. A., Geers. A. E., Morchower, B., Perrin, J., Skellett, R., Brenner, C., & Torretta, G. (2000). Factors associated with speech intelligibility in children with cochlear implants. *Annals of Otology, Rhinology and Laryngology Supplement, 185,* 28-30.

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

Zwolan, T.A., Zimmerman-Phillips, S., Asbaugh, C.J., Hieber, S.J, Kileny, P.R. & Telian, S.A. (1997). Cochlear implantation of children with minimal open-set speech recognition skills. *Ear & Hearing, 18,* 240-251.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Some New Experiments on Perceptual Categorization of Dialect Variation in American English: Acoustic Analysis and Linguistic Experience[1]

**Cynthia G. Clopper and David B. Pisoni**

*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. We would like to acknowledge the valuable advice that we received on various aspects of this project from Kenneth deJong, Caitlin Dillon, Luis Hernandez, and Robert Nosofsky, as well as the assistance of Jeffrey Reynolds and Adam Tierney with data collection.

# Some New Experiments on Perceptual Categorization of Dialect Variation in American English: Acoustic Analysis and Linguistic Experience

**Abstract.** Traditional methods of research on perceptual dialectology have been limited to tasks that ask participants to draw and label dialect regions on maps or to make attitude judgments about samples of certain linguistic varieties. Only a small handful of perceptual experiments have directly looked at how listeners identify and categorize where talkers are from based on actual speech samples. Our first experiment investigated how well naïve listeners could categorize talkers based on regional dialect of American English using sentence-length utterances. Although listeners performed poorly on this task, a more detailed analysis of their error patterns revealed systematic perceptual confusions. The results suggest that listeners have knowledge of three broad dialect categories: New England, South, and North/West. Since listeners were able to perform the perceptual categorization task at levels reliably above chance, a second experiment was carried out to measure the acoustic-phonetic properties that listeners used to make their categorization judgments. Results of this acoustic analysis study revealed four robust acoustic-phonetic properties that were good predictors of where the talkers were actually from: New England r-lessness and /æ/ backing, North /ou/ offglide centralization, and South Midland /u/ fronting. These properties also appeared to be used by the listeners in making their perceptual judgments. A third experiment explored the effects of residential history on dialect categorization performance using two different groups of listeners with different linguistic experiences. Listeners who had lived in at least three different states comprised the "Army Brat" group; listeners who had lived only in Indiana comprised the "Homebodies" group. The results revealed that the Army Brats performed significantly better than the Homebodies on the perceptual dialect categorization task. This finding suggests that greater exposure to linguistic variation in development leads to better performance on the categorization task. A final experiment investigated the effects of short-term perceptual learning on categorization performance. Using the same talkers and response alternatives as in the first experiment, one group of listeners was trained to categorize one talker from each region and a second group was trained to categorize three talkers from each region. After training, both groups of listeners were then asked to categorize new talkers using the same six dialect regions. Results showed that listeners who were exposed to a range of variability after training with three talkers from each dialect region were better able to generalize to new talkers than the listeners who were trained to identify only a single talker from each region. Taken together, these four experiments provide new evidence for the important role of acoustic-phonetic variation and variability in speech perception and spoken language processing. The implications of these findings for speech perception and spoken language processing are discussed.

## Introduction

Over a decade ago, Klatt (1989) identified five sources of variability in spoken language: ambient conditions, word environment, segment realization, within-speaker variability, and cross-speaker variability. Ambient conditions include such non-speech phenomena as background noise, room reverberation, and the properties of any recording or telephonic equipment involved. During World War II, Miller (1946) and his colleagues developed protocols for testing the effects of noise introduced through communication equipment on speech intelligibility and spoken word recognition. Models of speech perception, however, typically assume ideal listening conditions or a dual-route auditory system in which all ambient noise is filtered out of the speech stream and processed by a separate, general auditory mechanism (Mattingly & Liberman, 1990).

Utterance-specific variability includes word environment and segment realization variability which are both due to the physical properties of the human speech production system. Word environment effects include cross-word coarticulation and durational changes due to stress, focus, and reduction processes in continuous speech. Stress, focus, and reduction also effect segment realization, as do segmental coarticulation and variable production rules (such as releasing a word-final stop). Normalization for these kinds of variability has been one of the major theoretical problems in speech perception research. Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) claimed to have found the invariant cue to consonant place of articulation in the second formant transition in CV syllables. Models of speech perception such as Klatt's (1979) Lexical Access from Spectra (LAFS) model are based on the fundamental assumption that at some level acoustic invariance can be found across utterances.

Finally, within-speaker and cross-speaker sources of variability are related to "indexical" properties of the talker: gender, age, physical size, dialect, social status, emotional state, register, and speaking rate. Emotional state, register, and speaking rate are all sources of within-speaker variability, whereas properties of the talkers such as their gender, age, physical size, dialect, and social status are considered cross-speaker sources of variability. Few researchers working in the mainstream of speech research have focused their attention on the role that these talker-specific sources of variability play in speech perception and spoken language processing and no one has attempted to model how these different sources of variation might be encoded by human listeners.

One exception to this general trend is the research from our laboratory over the last decade. Pisoni and his colleagues have examined the role of talker variability in speech perception and spoken word recognition and have discovered that indexical characteristics of a talker are actually perceived and encoded by listeners along with the meaningful content of the linguistic signal. In one series of studies, Mullennix, Pisoni, and Martin (1989) showed that cross-talker variability affected word recognition performance in noise. Listeners performed more poorly when the talker changed from trial to trial than when the talker remained constant across all trials. In addition, the results of a speeded classification task suggested that indexical properties of the talker are inseparable from the linguistic content of the utterance (Mullennix & Pisoni, 1990).

Perceptual learning studies conducted by Pisoni and his colleagues have also shown that the role of variability in spoken language stimuli in training leads to better generalization performance. For example, in a study on the perceptual learning of novel voices, Nygaard, Sommers, and Pisoni (1994) found that training listeners to identify talkers led to better performance on word recognition in noise when the words were spoken by familiar talkers they had learned to identify in the first part of the experiment than when the words were spoken by unfamiliar talkers. In another study, Logan, Lively, and Pisoni (1991) showed that training Japanese listeners to identify English /r/ and /l/ using highly variable natural stimuli led to better generalization performance to novel talkers and novel utterances than training on less variable materials. These studies provide evidence for the role of variability in speech perception and reveal that perceptual learning can be enhanced through the use of training materials that contain variability and variation.

Variability has been observed in spoken language for many years and is a natural consequence of the speech production process and the physical mechanisms used to control speech articulation. More than 50 years ago, Peterson and Barney (1952) published the results of their well-known pioneering acoustic vowel space study which revealed large amounts of variation both within and across talkers. They plotted the first and second formant values of ten vowels spoken by 76 talkers (including men, women, and children) on an F1 by F2 plane on which they superimposed one ellipse for each of the ten vowels. The ellipses were drawn to include roughly 90% of the tokens for a given vowel. Their results produced a set of 10 relatively large and overlapping ellipses that revealed the large amount of variation in formant frequencies for vowels of the same quality. Even when only those tokens that were correctly identified by listeners 100% of the time were included in the figure, there was still a great deal of overlap between different vowels in the acoustic space. The observed pattern reflects the finding that vowels of different phonemic qualities are often found in exactly the same part of the F1 x F2 vowel space.

Despite this overwhelming evidence for variation and variability in vowel production, however, researchers working on human speech perception and speech synthesis typically focused on the mean formant values reported by Peterson and Barney (1952). For example, Hillenbrand, Getty, Clark, and Wheeler (1995) carried out an extensive study that was designed to replicate Peterson and Barney's vowel spaces using a larger number of talkers and a larger corpus of vowels. Hillenbrand et al. were interested in replicating and extending the mean formant values found in the earlier study. What they found instead was a systematic shift in the low front vowels that reflects the Northern Cities vowel shift in Michigan and other northern states where the new recordings were made. Peterson and Barney had recorded their talkers forty years earlier and had not controlled for regional variety of American English or even for native language. Hagiwara (1997) noted these differences in mean formant frequency between the two studies and pointed out the obvious role that four decades and geographic location played in the differing results. He then replicated the Peterson and Barney study again with talkers from Southern California. He found a shift in the back vowels that reflects the back vowel fronting found in the southern United States and in some parts of California. Hagiwara argued that speech researchers should work to record and measure vowel spaces of talkers in different parts of the country in order to determine the extent of vowel production variability.

Sociolinguists have been documenting precisely this kind of phonological variation in speech for more than thirty years, since Labov revolutionized the field with quantitative methods for collecting spoken language data on variable productions in his famous New York City department store study (Labov, 1972). More recently, Labov, Ash, and Boberg (in press) have collected recordings of over 700 talkers from around the United States and carried out acoustic measurements of the vowels. These measurements have allowed them to determine current dialect boundaries in the United States and to describe current changes in progress such as the Northern Cities vowel shift, the Southern vowel shift, and the /ɑ/ ~ /ɔ/ merger.

In addition to Labov's acoustic measurement work on speech production, other researchers in the field of sociolinguistics have studied how this variation is perceived by naïve listeners. Preston's (1993) work on perceptual dialectology used several unique methods to investigate the kinds of mental representations college students have about dialect variation in the United States. In one study, he gave undergraduate students in Indiana, Hawaii, New York, and Michigan a map of the United States, including state boundaries, and asked them to indicate where people "speak differently." He found that most of his participants indicated some portion of the country as having a southern dialect and identified New York City as having its own unique accent. In addition, he found that participants tended to identify more regional varieties in close geographic proximity to their own hometown than in areas farther away, suggesting a gradient of knowledge about linguistic variation.

Preston (1993) also asked some of the students in this study to complete an attitude judgment task in which they were given a list of the 50 states and were asked to rate each state on the correctness, pleasantness, and intelligibility of the English spoken there. He found an overwhelming tendency for participants to indicate that the most correct English is spoken in northern and western states and that the least correct English is spoken in southern states. Pleasantness ratings tended to be based on where the participants themselves were from, with their home state typically receiving a high pleasantness rating.

Perceptual dialectology studies provide interesting information about what kinds of representations naïve listeners have stored in memory about dialect variation and reveal important differences in these representations based on where the participants are from. However, these kinds of studies do not reveal the underlying psychological and linguistic processes that provide insights into how these listeners actually perceive or encode linguistic variation. Because the participants were not asked to listen to actual speech samples in making their responses, all of the data were based on linguistic representations stored in long-term memory instead of direct behavioral responses to speech stimuli.

A few dialect categorization studies have been conducted that ask naïve listeners to make direct behavioral responses to actual speech stimuli. Purnell, Idsardi, and Baugh (1999) conducted one study of dialect identification using the "matched-guise technique." A single talker left answering machine messages for apartment landlords in various neighborhoods in the San Francisco area using three guises: African American Vernacular English (AAVE), Chicano English (CE), and Standard American English (SAE). Purnell et al. measured dialect identification by recording the number of phone calls that were returned for each guise in each neighborhood. They concluded that the landlords could in fact identify the racial dialect of the talker based on a short answering machine message because the number of returned phone calls for the SAE guise remained constant across all neighborhoods, while the number of returned phone calls for the AAVE and CE guises decreased with the minority population of the neighborhood.

In a perceptual study on regional dialect identification, Preston (1993) played short narratives spoken by nine middle-aged male talkers to naïve listeners in Michigan and Indiana. The talkers were from nine different cities on a north-south continuum from Dothan, Alabama to Saginaw, Michigan. The listeners heard each narrative passage once and were then asked to identify which city they thought each talker was from. In general, the listeners were able to make a coarse distinction between northern and southern talkers, although the boundary between north and south was slightly different for the two groups of listeners.

Finally, Williams, Garrett, and Coupland (1999) conducted a dialect categorization task on the regional varieties of the English spoken in Wales. They recorded two adolescent male talkers from each of six regions of Wales, as well as two adolescent male speakers of Received Pronunciation (RP). Williams et al. played these recordings back to naïve listeners in each of the six regions and asked them to categorize the talkers using an eight-alternative forced-choice task (the six regions in Wales, RP, and "don't know"). Overall categorization performance was about 30% correct. In addition, the listeners were only able to correctly identify 45% of the talkers from their own region. Taken together, the results of these three dialect categorization studies suggest that naïve listeners are able to encode some information about dialect variation and they can use this information to identify where unfamiliar talkers are from. However, their performance is not perfect and in fact seems to be quite effortful.

Recent work in our lab at Indiana University has also explored how naïve listeners categorize unfamiliar talkers by dialect. In particular, Williams et al. (1999), Preston (1993), and Purnell et al. (1999) all suggested that listeners can use their knowledge of variation in their native language to identify where talkers are from. In order to investigate how naïve listeners of American English categorize talkers by dialect, we carried out a perceptual categorization study that was similar to Williams et al.'s earlier investigation using regional varieties of American English and native listeners of American English (Clopper & Pisoni, 2004b). All of the stimuli used in our experiments were drawn from a large corpus of read sentences. Our first research question assessed whether naïve listeners could reliably categorize talkers based on where the talkers were from.

Labov et al.'s (in press) work on dialect variation in the United States has shown the kinds of variation and variability we can expect to find in talkers from different regions of the United States. In order to evaluate which acoustic-phonetic properties listeners might be attending to in making their categorization judgments, we also conducted an acoustic analysis on two sentences that were read by all of our talkers (Clopper & Pisoni, 2004b). Thus, our second research question was designed to assess what acoustic-phonetic properties were available to the listeners and which ones they attended to in making responses in the categorization task.

Preston (1993) found differences between participant groups in all of his perceptual dialectology studies, suggesting that a participant's residential history may have a strong impact on how that person perceives and categorizes language variation. To study this problem, we conducted a second dialect categorization task using two groups of listeners that differed in their personal experience with dialect variation to see how residential history affected their performance on the task (Clopper & Pisoni, 2004a). Thus, our third research question focused on how prior linguistic experience affects performance in the dialect categorization task.

Finally, several recent studies in our laboratory by Nygaard et al. (1994) and Logan et al. (1991) have shown that short-term experience in the laboratory with spoken language variation affects performance on other language processing tasks. Therefore, we also conducted a perceptual learning study using the same perceptual categorization task to determine how categorization performance with unfamiliar talkers would be affected by prior laboratory training in the task.

## Experiment 1: Perceptual Categorization Task

The purpose of the first experiment we conducted was to determine how well a group of naïve listeners can categorize talkers based on regional dialect of American English using a forced-choice categorization task. Utterances from sixty-six white male talkers in their twenties were selected from the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Fisher, Doddington, & Goudie-Marshall, 1986; Zue, Seneff, & Glass, 1990). The TIMIT corpus contains recordings of ten sentences read by 630 different talkers and was originally designed to contain a large degree of variability for use in speech recognition research. For our study, sentences were selected from eleven talkers from each of six different dialect regions in the United States: New England, North, North Midland, South Midland, South, and West.

Eighteen Indiana University undergraduates listened to sentences spoken by the 66 talkers. The listeners were divided post-hoc into three listener groups based on reported residential history: Northern Indiana (N = 7), Southern Indiana (N = 5), and Out-of-State (N = 6). The first two sentences that the listeners heard were those that were spoken by all of the talkers on the TIMIT corpus and are shown in (1). In addition, the listeners also heard each of the talkers reading a different, novel sentence. Examples of these novel sentences are shown in (2).

(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.

(2) a. Beg that guard for one gallon of gas.
b. Barb's gold bracelet was a graduation present.
c. A huge tapestry hung in her hallway.
d. Clasp the screw in your left hand.

In the first phase of the experiment, the listeners heard all 66 talkers reading Sentence (1a) in random order and were asked to categorize the talker by dialect into one of the six geographic regions. The regions were presented on the screen as partial maps of the United States, including state boundaries, and were labeled with the name of the region, as shown in Figure 1. In the second phase of the experiment, the listeners heard all 66 talkers reading Sentence (1b) in random order and were again asked to categorize the talker by dialect. Finally, in the third phase, the listeners heard the 66 talkers reading the novel sentences in random order. No feedback was provided about the accuracy of their responses on this task.

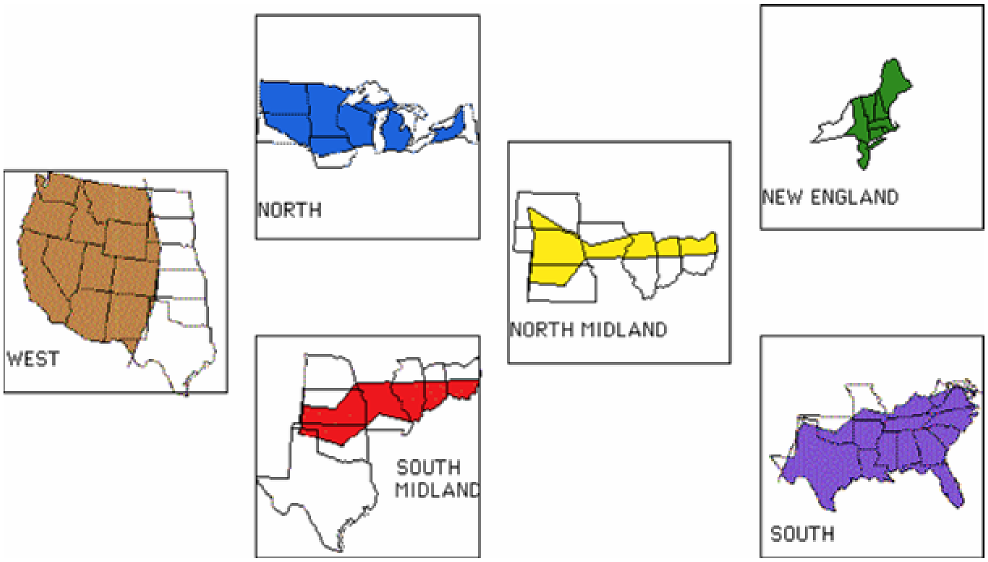


**Figure 1.** The six response alternatives in the categorization task. (From Clopper & Pisoni, 2004b).

The results expressed in terms of overall categorization accuracy are shown in Figure 2. The categorization scores revealed that the listeners performed quite poorly overall, although they were statistically above chance (17%) in all three phases of the experiment.

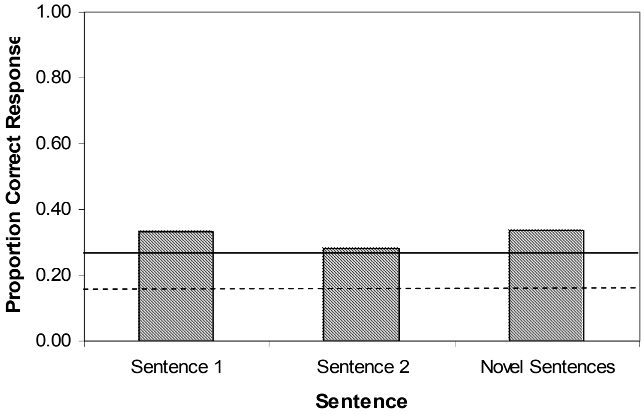


**Figure 2.** Overall proportion correct response in the six-alternative dialect categorization task, collapsed across all listeners and all talkers. Chance performance (17%) is indicated by the horizontal dashed line. Performance significantly above chance (25%), based on a binomial distribution, is indicated by the solid line. (Replotted from Clopper & Pisoni, 2004b).

Figure 3 shows the listeners' performance as a function of the six different dialect regions. The listeners were better able to correctly categorize the New England talkers than any other dialect group. They categorized the Southern talkers more accurately than the Northern or Western talkers. Unlike the talker differences, however, there were no listener group differences in any of the three phases and no individual listener differences within any of the three groups. Overall, the listeners performed consistently, although close to chance, on this task.

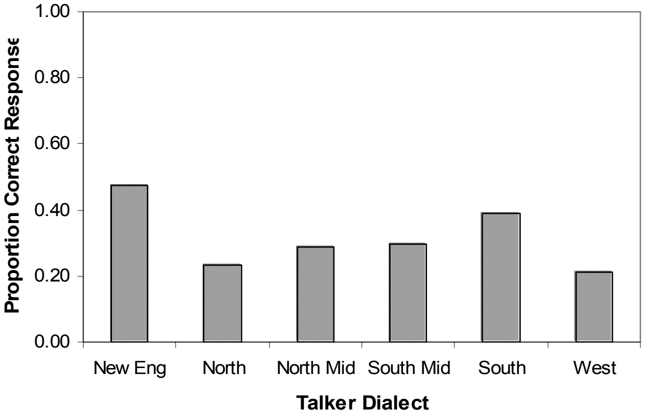


**Figure 3.** Proportion correct categorization responses for each of the six talker dialect regions, collapsed across the three sentence conditions. (Replotted from Clopper & Pisoni, 2004b).

In addition to the analysis of the correct categorization responses, we also conducted a clustering analysis using the confusion matrices of the error responses for each sentence. The confusion matrices were first submitted to the Similarity Choice Model (Nosofsky, 1985) to determine similarity and bias parameters. The similarity parameters indicate the degree of similarity between the dialect regions based on the listener errors. Examination of the bias parameters indicates response bias in selecting one response more often than another. The bias parameters revealed that the listeners' responses were relatively bias-free. The similarity parameters for each sentence were then submitted to ADDTREE (Corter, 1995), which is an iterative additive clustering scheme that selects the most similar pair of cells in the matrix at each iteration to form a cluster and then recalculates the confusion matrix. The resulting clusters from this analysis are shown in Figure 4 for each of the three sentence conditions. In these figures, dissimilarity is indexed in terms of vertical distance in the display, so that the dissimilarity between any two dialect regions is the sum of the lengths of the least number of vertical lines connecting them.

An examination of these clustering solutions revealed three main clusters of the categorization responses for each of the three sentences. For Sentence 1 and the Novel Sentences, the three main clusters were: New England; South and South Midland; and North, North Midland, and West. For Sentence 2, the three main clusters were: New England and North; South and South Midland; and North Midland and West. These results indicate that the confusions made by the listeners were not random, but instead suggest that the listeners were relying on three broad dialect regions in making their categorization judgments: New England, South, and North/West.

Categorization performance improves dramatically if it is measured in terms of the results of the clustering analysis. In particular, the responses on each sentence were rescored so that a response was scored as correct if it fell into the same major cluster as the stimulus item for that sentence. The results of this analysis are shown in Figure 5. This difference between the original measure of performance and the new measure based on the confusions obtained in the categorization task provides additional behavioral evidence that the listeners were systematically making use of three dialect clusters instead of the six regions originally provided by the experimenters.

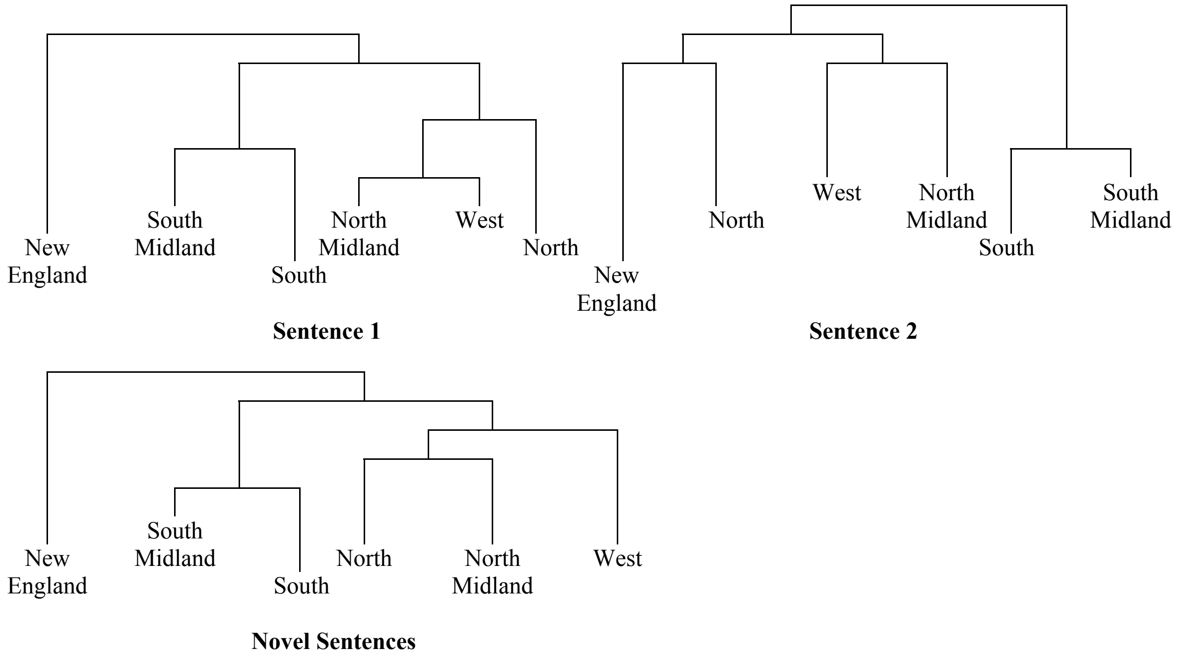


**Figure 4.** Clustering solutions for Sentence 1, Sentence 2, and Novel Sentences, collapsed across all listeners.

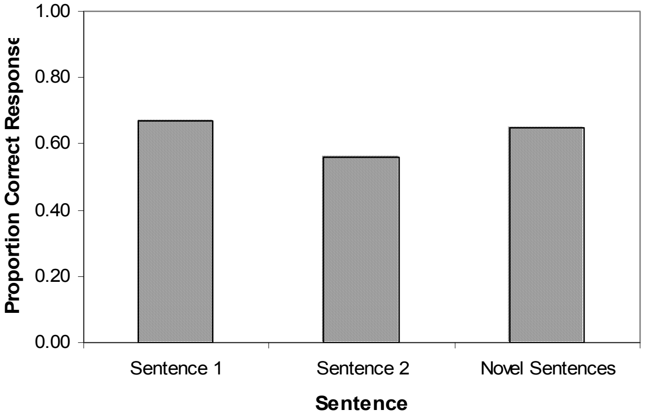


**Figure 5.** Proportion correct response in the categorization task when collapsed across the three main clusters for each of the three sentences. For Sentence 1 and Novel Sentences, collapsed across New England; South and South Midland; North, North Midland and West. For Sentence 2, collapsed across New England and North; South and South Midland; North and West.

Taken together, the results of this first perceptual experiment revealed that listeners are able to categorize talkers by regional dialect at performance levels above chance using a forced-choice task without feedback. In addition, the confusions made by the listeners were systematic in nature and enabled us to investigate patterns in their perception. Specifically, the results of the clustering analysis suggested that listeners make use of three broad perceptual categories instead of six: New England, South, and North/West. Thus, in answer to our first research question, we found that listeners can reliably categorize unfamiliar talkers by dialect using a forced-choice perceptual categorization task, although their performance is not error-free.

## Experiment 2: Acoustic-Phonetic Analysis

Our second experiment was designed to measure several selected acoustic-phonetic properties of these sentences and determine which ones listeners used in making their perceptual categorization judgments. Acoustic measurements were obtained from the first two sentences used in the previous experiment from all of the original 66 male talkers. The eleven acoustic measurements made for each talker are shown in Table 1.

| Word | Segment | Measurement | Acoustic-Phonetic Property |
|---|---|---|---|
| dark | /a/ | change in F3 (midpoint to offset) | r-fulness |
| 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/ | F2 (midpoint) | /u/ backness |
| don’t | /oʊ/ | F2 (midpoint) | /oʊ/ backness |
| | | change in F2 (midpoint to offset) | /oʊ/ diphthongization |
| rag | /æ/ | F2 (midpoint) | /æ/ backness |
| | | change in F2 (onset to offset) | /æ/ diphthongization |
| like | /aɪ/ | change in F2 (midpoint to offset) | /aɪ/ diphthongization |
| oily | /oɪ/ | change in F2 (midpoint to offset) | /oɪ/ diphthongization |

**Table 1.** Acoustic measures selected for comparison between dialect groups.

Each of the acoustic measures was expected to reveal differences between talker groups, based on what is known about phonological variation in American English (e.g., Labov et al., in press; Thomas, 2001). In particular, r-fulness (i.e., rhotic vs. non-rhotic) was predicted to distinguish the New England talkers from the other talkers, reflecting New England r-lessness. Vowel brightness in *wash* was predicted to reveal the *wash* ~ *warsh* alternation and to distinguish South Midland talkers from the others. Fricative voicing and fricative duration in *greasy* were predicted to distinguish the Southern and South Midland talkers from the others. In terms of vowels, /u/ backness was predicted to be lower for the Southern and Western talkers than for the other talkers. The degree of diphthongization of /aɪ/ and /oɪ/ was also predicted to distinguish the South from the other regions. The degree of diphthongization of /oʊ/ as well as /oʊ/ and /æ/ backness were predicted to distinguish the North from the others. Finally, /æ/ diphthongization was predicted to distinguish New England from the others.

A summary of the overall results of the acoustic-phonetic analysis is provided in Table 2. New England talkers were significantly less rhotic than South Midland and Western talkers. Southern talkers had significantly greater voicing in the fricative in *greasy* than New England talkers and a significantly longer fricative in the same word than Northern talkers. South Midland, Southern, and Western talkers had significantly more fronted /u/'s than New England talkers. Northern talkers had a significantly centralized offglide in /oʊ/ relative to the Southern talkers and a significantly fronted /æ/ relative to New England talkers. Thus, several of the selected acoustic-phonetic measures revealed significant differences between the talker groups.

| | **New England** | **North** | **North Midland** | **South Midland** | **South** | **West** |
|---|---|---|---|---|---|---|
| r-fulness (ΔHz) | 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 (ΔHz) | -71 | -148 | -40 | 22 | 37 | -41 |
| /æ/ backness (Hz) | 601 | 399 | 440 | 425 | 494 | 491 |
| /æ/ diphthong (ΔHz) | 256 | 177 | 255 | 280 | 223 | 233 |
| /aɪ/ diphthong (ΔHz) | 452 | 418 | 402 | 278 | 331 | 350 |
| /oɪ/ diphthong (ΔHz) | 301 | 384 | 434 | 250 | 226 | 445 |

**Table 2**. Acoustic measure means by talker dialect group.

In order to determine which acoustic-phonetic properties of the speech signal were good predictors of talker "dialect affiliation" (defined as the dialect labels used to classify the talkers in the TIMIT corpus) we conducted a series of logistic multiple regressions on the acoustic-phonetic measures and dialect affiliation. In each regression, the acoustic-phonetic measures were treated as potential predictor variables of dialect affiliation, which was scored dichotomously ("1" if the talker was from that region and "0" if he was not). The results of these analyses, shown in Table 3, revealed the acoustic-phonetic properties that were good predictors of actual dialect affiliation of the talkers (i.e., good predictors of the TIMIT labels). R-lessness and /æ/ backness were found to be good predictors of the New England talkers. For North dialect affiliation, /oʊ/ offglide centralization and monophthongal /æ/ were found to be good predictors. Fronting of /u/ and backing of /oʊ/ were good predictors of South Midland talkers. Finally, fricative voicing in *greasy* was a good predictor of Southern talkers. None of the acoustic-phonetic measures examined in this study turned out to be good predictors of either North Midland or Western talkers.

| | Significant Variables | Regression Coefficients | Overall $r^2$ |
|---|---|---|---|
| **New England** | r-fulness | -.01 | .33 |
| | /æ/ backness | .02 | |
| **North** | /oʊ/ diphthong | -.01 | .21 |
| | /æ/ diphthong | -.01 | |
| **North Midland** | n/a | | |
| **South Midland** | /u/ backness | -.01 | .19 |
| | /oʊ/ backness | .01 | |
| **South** | fricative voicing | 3.4 | .21 |
| **West** | n/a | | |

**Table 3.** Results of the logistic multiple regression analysis on acoustic-phonetic properties and talker dialect affiliation. For each of the dialect groups, the significant acoustic measures are shown with their regression coefficients and the overall $r^2$ showing model fit. (From Clopper & Pisoni, 2004b).

| | Significant Variables | Regression Coefficients | Overall $r^2$ |
|---|---|---|---|
| **New England** | r-fulness | -.36 | .39 |
| | /æ/ backness | .34 | |
| | /oʊ/ diphthong | -.22 | |
| | vowel brightness | .21 | |
| **North** | /oʊ/ diphthong | -.38 | .27 |
| | /u/ backness | .29 | |
| **North Midland** | /oɪ/ diphthong | .56 | .31 |
| **South Midland** | /u/ backness | -.26 | .38 |
| | vowel brightness | -.34 | |
| | fricative voicing | .33 | |
| **South** | /oɪ/ diphthong | -.39 | .49 |
| | /oʊ/ diphthong | .33 | |
| | /u/ backness | -.33 | |
| | /oʊ/ backness | .31 | |
| | /æ/ diphthong | .20 | |
| **West** | /oɪ/ diphthong | .40 | .16 |

**Table 4.** Results of the linear multiple regression analysis on acoustic-phonetic properties and perceptual categorization. For each of the dialect groups, the significant acoustic measures are shown, along with their regression coefficients and the overall $r^2$ showing model fit. (From Clopper & Pisoni, 2004b).

A second regression analysis was conducted to determine which acoustic-phonetic properties in the signal affected the listeners' categorization behavior in Experiment 1. In this set of linear multiple regressions, the acoustic-phonetic measurements were again treated as predictor variables and the categorization performance of the listener served as the dependent variable. The results of this analysis, summarized in Table 4, reveal those acoustic-phonetic properties that listeners were attending to in making their categorization judgments. In categorizing talkers as New England, listeners were attending to r-lessness, /æ/ backness, centralized /oʊ/ offglides, and vowel brightness in *wash*. For North, listeners were attending to centralized /oʊ/ offglides and backed /u/. Diphthongal /oɪ/ was a good predictor of identification of North Midland and Western talkers for these listeners. Fricative voicing in *greasy*, /u/ fronting, and a dark vowel in *wash* were all good predictors of categorization as South Midland. Finally, /oɪ/ monophthongization, /oʊ/ diphthongization, /u/ fronting, backed /oʊ/, and /æ/ diphthongization were good predictors of categorization as Southern.

Taken together, the results of this acoustic analysis revealed that the dialects of the talkers used in this study could be reliably distinguished based on several robust acoustic-phonetic properties in the speech signal. In addition, seven acoustic-phonetic properties were found to be good predictors of dialect affiliation in the first regression analysis. The second regression analysis revealed that the listeners were attending to 16 acoustic attributes of the talkers in making their categorization judgments. Of the seven predictors of dialect affiliation found in the first regression analysis and the 16 predictors of categorization behavior found in the second regression analysis, four overlapped: New England r-lessness, New England /æ/ backness, North /oʊ/ offglide centralization, and South Midland /u/ fronting. These results suggest that listeners are able to detect and perceive some of the acoustic-phonetic properties that distinguish talkers of different dialects and can use these properties reliably in responding in the categorization task. Thus, in answer to our second research question, we found that listeners used acoustic cues to r-fulness, vowel backness, and vowel diphthongization in making their categorization judgments.

## Experiment 3: Effects of Residential History: Army Brats vs. Homebodies

Due to the post-hoc nature of the listener groups used in Experiment 1 and the small number of participants in each group, we failed to observe any systematic differences between the three groups in their categorization performance. However, given Preston's (1993) earlier findings, we expected to observe differences between listeners based on their past residential history. In addition, an extensive literature on language acquisition suggests that early linguistic experience and activities with many segmental, prosodic, and even indexical contrasts leads to better discrimination of those same contrasts later in life (e.g., Allen, 1983; Peng, Zebrowitz, & Lee, 1993; Polka, 1992; Strange, 1995; Tees & Werker, 1984). Based on these and other findings, early exposure to dialect variation might also be expected to have a lasting influence on a listener's perceptual abilities.

To explore this issue in greater depth, a third experiment was conducted to examine the effects of linguistic experience and residential history on listeners' performance in the dialect categorization task. The same set of sentence materials spoken by the same 66 talkers was used in this experiment as in Experiment 1. The experimental design was also identical to that used in the first experiment. Two new groups of Indiana University undergraduates participated as listeners in this study. The first group, the "Homebodies," consisted of 31 listeners who reported that they had lived exclusively in Indiana. The second group, the "Army Brats," consisted of 30 listeners who reported that they had lived in at least three states (including Indiana).

The perceptual categorization results of this study are shown in Figure 6. The Army Brats were more accurate overall on the six-alternative forced-choice categorization task than the Homebodies. These results suggest that people who have lived in several different states perform better on the categorization task than people who have lived only in one state. Thus, greater linguistic experience and exposure to variation and variability through personal real-life interaction with people from different dialect regions leads to better performance on the dialect categorization task.

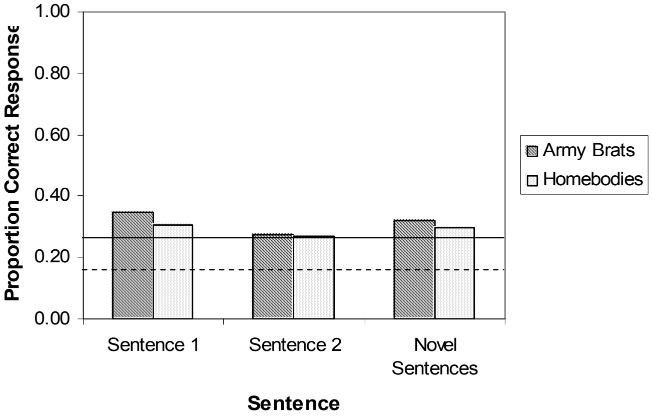


**Figure 6.** Proportion correct responses in each of the three phases of the dialect categorization task for the Army Brat listener group and the Homebodies listener group. Chance performance (17%) is indicated by the dashed line. Performance statistically above chance (25%) is indicated by the solid line. (Replotted from Clopper & Pisoni, 2004a).

## Experiment 4: Some Effects of Perceptual Learning

The third experiment confirmed that real-life linguistic experience affects performance in the dialect categorization task. Experiment 4 was designed to investigate the effects of experience with linguistic variability in a laboratory setting on dialect categorization of unfamiliar talkers. The utterances from the same 66 talkers were used again in this study. Two additional groups of 30 Indiana University undergraduates participated as listeners. The first group, the "one-talker group," was trained to identify one talker from each of the six dialects in three phases of training using the perceptual categorization task with feedback. The second group, the "three-talker group," was trained to categorize three talkers from each of the six dialects in three phases of training. Following three blocks of training, the listeners in both groups were tested on the same talkers they had been trained on in the same categorization task without feedback to ensure that they had actually learned where the talkers were from. After this phase was completed, they participated in a generalization phase in which they heard unfamiliar talkers from each of the six dialect regions reading novel sentences and were asked to categorize them by dialect without feedback.

The results of this perceptual learning experiment are shown in Figure 7. While the group trained to identify one talker in each dialect performed better than the three-talker group on the three training blocks and the final test block using the same set of talkers, the group trained on three talkers actually performed better on the generalization phase with novel talkers than the one-talker group. This "cross-over effect" demonstrates that those listeners who were trained on materials with greater variability initially had more difficulty learning to categorize those materials, but were better able to generalize to new talkers in the final critical generalization phase. These perceptual learning results suggest that short-term exposure to greater stimulus variability even in a highly-controlled laboratory setting leads to better categorization of unfamiliar talkers. Thus, with regard to our third research question, greater linguistic experience through laboratory perceptual learning also leads to better performance on the dialect categorization task.

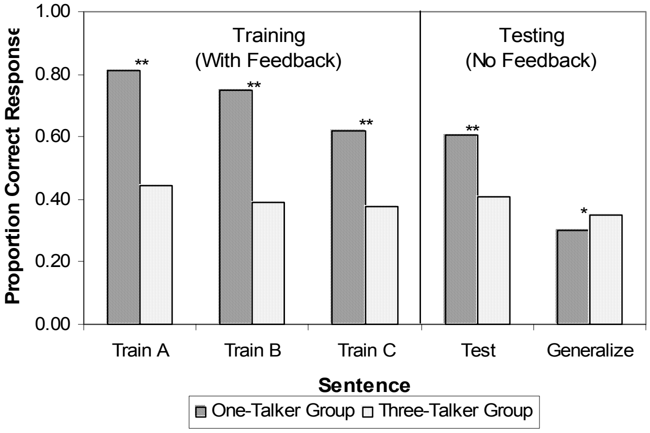


**Figure 7.** Proportion correct responses in each phase of the perceptual learning experiment for each of the listener groups. $**p < .001$, $*p < .05$ based on post-hoc t-tests.

## General Discussion

Experiment 1 revealed that naïve listeners are able to categorize talkers based on regional dialect with above-chance performance using a forced-choice perceptual categorization task. The clustering analysis of the response confusions suggested that listeners make use of three broad dialect categories: New England, South, and North/West. Acoustic analyses carried out in Experiment 2 revealed measurable acoustic-phonetic differences in production between the six dialect regions. In addition, multiple regression analyses suggested that listeners are able to use such stereotyped attributes as r-lessness and /oʊ/ centralization, as well as less stereotyped, but still prominent, attributes such as /u/ fronting and /æ/ backness to categorize talkers by regional dialect. In Experiment 3, we found that real-world exposure to talkers from different dialect regions as a result of living in a number of different geographic locations improves categorization performance. Finally, Experiment 4 revealed that short-term exposure in the laboratory to multiple talkers from each dialect region improves categorization performance on new talkers reading novel sentences, despite higher performance in training phases by the listeners who were exposed to only one talker from each region. Taken together, these results further confirmed that listeners can and do encode dialect variation in normal everyday language perception situations and that they can use the detailed knowledge they gain from these experiences in more formal laboratory tasks such as forced-choice dialect categorization.

The results of our new studies have several important theoretical implications for current models of speech perception and spoken language processing. Proponents of the traditional, abstractionist views of speech and language have stated that "… voice quality, speed of utterance, and other properties directly linked to the unique circumstances surrounding every utterance are discarded in the course of learning a new word" (Halle, 1985) and that "clearly most of the time anyone is listening to English being spoken, he [sic] is listening for the meaning of the message - not to how the message is being pronounced" (Brown, 1990). These traditional views of variation and variability are consistent with the generative linguistics paradigm in which language is described and modeled as being based on a one-to-one mapping between underlying phonological representations and surface phonetic forms. Generative approaches to the study of language have recently even been applied to a theory of language evolution in early humans (Hauser, Chomsky, & Fitch, 2002). The results reported in this chapter, however, support the proposal that variation matters in speech perception and language processing and that listeners have access to fine acoustic-phonetic details and not merely symbolic representations of phonology (Pisoni, 1997).

The claim that variation matters is also supported by previous research on the role of talker variability in language processing tasks. As discussed earlier, Nygaard et al. (1994) showed that speech intelligibility in noise is better when the talkers are familiar than when they are unfamiliar. In addition, Mullennix et al. (1989) showed that listeners were better able to recognize words in noise when all of the words were spoken by a single talker than when the talker changed from trial to trial, suggesting that the listeners were sensitive not only to what was being said, but also to who was saying it. Mullennix and Pisoni (1990) also demonstrated interference in word and voice recognition tasks, revealing that listeners could not entirely ignore either the lexical content of the message or the talker, even when the perceptual task was to attend selectively to only one of the two dimensions. These findings are also consistent with "embodied" approaches to Cognitive Science which place the interactions of the body, mind, and the environment centrally in an understanding of cognition (Clark, 2001).

Taken together, these and other recent findings suggest that linguistic content and talker-specific indexical properties of spoken language are not separated in speech perception, but rather that they are both perceived, processed, and encoded as part of normal language processes. The results of Experiments 3 and 4 also suggest that exposure to linguistic variability, either in real life or in a laboratory setting, improves performance on dialect categorization, suggesting that listeners encode and store detailed acoustic-phonetic information about the dialect of talkers they are exposed to. These sources of information are not lost or discarded by the nervous system. The results of Experiment 2 suggest that listeners are successful in knowing what to listen for and attend to in making their judgments, without any explicit training or feedback. Models and descriptions of language in theoretical linguistics must therefore be able to account for the many-to-one mapping of surface forms to underlying forms that the listeners in these studies could make use of to identify where the talkers were from. In order to perform the categorization task in Experiments 1, 3, and 4 at levels above chance, the listeners had to access and explicitly use detailed phonetic knowledge that they have about variation in English surface forms to categorize the talkers by dialect. Although listeners are able to perform this task above chance, their performance was not perfect and many confusions were observed.

The present set of studies also show that the application of new experimental methodologies in the fields of Cognitive Psychology and Cognitive Science can be fruitfully applied to sociolinguistic issues to further our understanding of linguistic variation and how it is perceived and encoded by naïve listeners and how it is processed by the memory system. Future research on the perception of dialect variation might be able to make use of other methods to measure similarity such as free classification, paired comparison, and similarity scaling tasks. These tasks could provide further converging evidence for the basic underlying psycholinguistic processes used to identify and categorize dialects of English.

In our view, the process of speech perception involves not only the segmentation of the speech signal into meaningful linguistic units (e.g., words, sentences) and the recovery of the structure of the sound patterns, but also the processing and encoding of indexical information about the talker. This talker-specific information is available to the listener and can be used in laboratory tasks, such as categorizing unfamiliar talkers. Certainly, gender differences come to mind as an obvious distinction that we can make based on what we know about how males and females talk. Yet these perceptual abilities have been ignored in theoretical discussions of whether or not talker-specific information is encoded and used in speech perception and spoken word recognition. The results of the present set of experiments demonstrate that even talker-specific characteristics that involve more complex acoustic-phonetic properties and that may be more difficult to perceive or encode, such as regional dialect, are also encoded by naïve listeners, stored in long-term memory, and used in a range of processing tasks.

## References

Allen, G.D. (1983). Linguistic experience modifies lexical stress perception. *Journal of Child Language, 10*, 535-549.

Brown, G. (1990). *Listening to spoken English*. (2nd ed.). New York: Longman.

Clark, A. (2001). *Mindware*. New York: Oxford University Press.

Clopper, C.G., & Pisoni, D.B. (2004a). Homebodies and Army Brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change, 16,* 31-48.

Clopper, C.G., & Pisoni, D.B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32,* 111-140.

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

Fisher, W.M., Doddington, G.R., & Goudie-Marshall, K.M. (1986). The DARPA speech recognition research database: Specifications and status. *Proceedings of the DARPA Speech Recognition Workshop*, 93-99.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America, 102*, 655-658.

Halle, M. (1985). Speculations about the representation of words in memory. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 101-104). The Hague: Mouton.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

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.

Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279-312.

Klatt, D.H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169-226). Cambridge, MA: MIT Press.

Labov, W. (1972). The social stratification of (r) in New York City department stores. In *Sociolinguistic patterns* (pp. 43-69). Philadelphia: University of Pennsylvania Press.

Labov, W., Ash, S., & Boberg, C. (in press). *Atlas of North American English*. Mouton deGruyter.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

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.

Mattingly, I. G., & Liberman, A. M. (1990). Speech and other auditory modules. In G.M. Edelman, W.E. Hall, & W.M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 501-520). New York: Wiley.

Miller, G.A. (1946). Articulation testing methods. In *Transmission and Reception of Sounds Under Combat Conditions*. Summary Technical Report of Division 17, NDRC. Washington, DC. pp. 69-80.

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.

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

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Peng, Y., Zebrowitz, L.A., & Lee, H.K. (1993). The impact of cultural background and cross-cultural experience on impressions of American and Korean male speakers. *Journal of Cross-Cultural Psychology, 24*, 203-220.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

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.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics, 52*, 37-52.

Preston, D.R. (1993). Folk dialectology. In D. R. Preston (Ed.), *American dialect research* (pp. 333-378). Philadelphia: John Benjamins.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18,* 10-30.

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, MD: York Press.

Tees, R.C., & Werker, J.F. (1984). Perceptual flexibility: Maintenance or recovery of the ability of discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*, 579-590.

Thomas, E.R. (2001). *An acoustic analysis of vowel variation in New World English.* Durham, NC: Duke University Press.

Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D.R. Preston (Ed.), *Handbook of perceptual dialectology* (pp. 345-358). Philadelphia: John Benjamins.

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

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Working Memory Capacity, Verbal Rehearsal Speed, and Scanning in Deaf Children with Cochlear Implants[1]

**Rose A. Burkholder 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. We extend thanks to lab members Brianna Conrey and Caitlin Dillon for comments on an earlier version of this manuscript. We are also grateful to Marc Marschark and Patricia Spencer for inviting us to contribute this chapter to their book and providing informative and helpful editorial assistance.

# Working Memory Capacity, Verbal Rehearsal Speed, and Scanning in Deaf Children with Cochlear Implants

**Abstract.** Cochlear implants have been an effective intervention for many profoundly deaf adults and children. Specifically, in prelingually deaf children, cochlear implants provide the first exposure to both environmental sounds and spoken language. After gaining access to sound and spoken language, many children using cochlear implants have been found to develop language with a developmental trajectory that is similar to normally-hearing children. However, several other cognitive skills of deaf children using cochlear implants appear to be atypical and do not develop fully even after several years of cochlear implant use. In this chapter, we review recent findings on cognitive processes in profoundly deaf children who use cochlear implants. Specifically, we discuss the shorter immediate memory spans of deaf children using cochlear implants and explain why measures of their speaking rate and speech timing provide important new clues to their verbal rehearsal and scanning processes. In addition, we also consider recent findings on the nonword repetition skills of deaf children who use cochlear implants. These results suggest that some deaf children with cochlear implants may have difficulties in rapidly encoding, rehearsing, and repeating novel phonological patterns. Our recent findings suggest that fundamental cognitive processes play an important role in the development of speech and language following cochlear implantation.

## Introduction

The development of spoken language and other fundamental cognitive skills is strongly influenced by a variety of early social and sensory experiences. Although both vision and audition contribute to the early sensory experience of typical infants, audition may play a more important role in the earliest stages of perceptual and cognitive development. For instance, even before birth, in the third trimester, sounds and voices that penetrate the womb are easily detectable to most fetuses (Aslin, Jusczyk, & Pisoni, 1998). This auditory prenatal experience has been shown to have an impact on infants' subsequent abilities to recognize speech after birth (DeCasper & Spence, 1986). Thus, early vocal input is highly salient to infants and important for the development of communicative abilities. These findings also suggest that auditory information may provide the basis for some of the earliest memories in infants.

In addition, at birth, although inner ear structures are still developing, a normal infant's hearing is nearly as acute as an adult's (Aslin et al., 1998). However, visual acuity is extremely poor at birth and is typically not fully mature until several months later (Morrone & Burr, 1986). The precocious development of audition in infants suggests that hearing is likely to be the dominant sensory modality that contributes to early language and communication development and may also facilitate the development of cognitive abilities in other domains, including multimodal processing, attention, learning, and memory.

Based on these findings from normally-hearing, typically developing populations, it is important from both a theoretical and clinical perspective to begin investigating the cognitive development of infants who have been deprived of early sensory experience with sound. The lack of sufficient auditory input in humans has been found to have detrimental effects on the development of speech and language in a handful of well known cases of abused and abandoned children such as Genie (Curtis, 1977) and Victor the "Wild Boy of Aveyron" (Lane, 1979). Unfortunately in examinations of such feral children, any experimental outcomes and interpretations are confounded by the social and emotional isolation and abuse that these children may have experienced. However, unlike these rare cases, profoundly deaf children are likely to be unscathed by such severe and tragic circumstances, making them a unique and potentially important clinical population in which to examine the impact of early spoken language deprivation on cognitive and linguistic development. In addition, in recent years, a smaller subset of profoundly deaf children has provided an unusual opportunity to answer what is perhaps an even more provocative question: What happens to the cognitive development of deaf children deprived of early auditory and linguistic experience who later gain exposure to sound and spoken language through a cochlear implant? In this chapter, some preliminary answers to this question are provided.

Due to prior inadequacies in diagnosing hearing problems in newborns and current FDA constraints on implanting deaf children very soon after birth, most profoundly deaf children who receive a cochlear implant have been deprived of auditory input for one or more years (Kirk, 2002). In addition, for many congenitally and prelingually deaf children, cochlear implants provide the first exposure to both environmental sounds and spoken language and, in some cases, may provide the first opportunity for these children to learn any language, spoken or signed. Because of this delayed onset of exposure to spoken language, deaf children using cochlear implants are a unique clinical population in which researchers can examine the ramifications of extended periods of auditory deprivation on the development of speech, language, and other cognitive abilities. In addition to the effects that early auditory deprivation can have on the eventual development of deaf children who have received cochlear implants, the children's behavioral and physiological responses to their new sensory input are important to study as well. Another paramount question to address in these deaf children is whether the sensory input from a cochlear implant can adequately facilitate normal spoken language development and other cognitive abilities that typically rely heavily on auditory and verbal coding skills.

However, some of these questions are not easy to answer because cochlear implants do not simply restore hearing to normal. Rather, they provide listeners with direct electrical stimulation of the auditory nerve and its afferents that must be translated into identifiable auditory percepts and used appropriately for whatever cognitive task is currently required (Rauschecker & Shannon, 2002). Therefore, a child with a cochlear implant must determine what sounds represent and mean and must learn to link these sounds to the visual and auditory events occurring around him or her. This complex perceptual learning task is also important to understand in order to successfully assess the speech, language, attention, learning, and memory skills of profoundly deaf children who have received cochlear implants.

Profoundly deaf children with cochlear implants are also an ideal population in which to study the effects of different *types* of early linguistic experience on language development. The amount and nature of aural-oral experience received after cochlear implantation differs substantially from child to child (Connor, Hieber, Arts, & Zwolan, 2000). The auditory-oral training and educational placement that a deaf child receives after cochlear implantation is typically referred to as his or her communication mode. Communication modes for such children fall along a continuum between exclusively oral communication that uses only speech and total communication that uses a form of signing such as signed exact English or cued speech in addition to spoken language. Although American Sign Language (ASL) does fall within this continuum, it is rarely used by deaf children using cochlear implants.

Differences in each child's communication mode after cochlear implantation allow for comparisons between groups of children based on the richness and robustness of their exposure to spoken language and their experience using primarily oral-aural communication. Such comparisons are informative because they can provide solid behavioral evidence of the degree to which the amount and quality of spoken language exposure received by deaf children with cochlear implants has an effect on the development of their speech, language, and other cognitive abilities. However, one caveat in examining effects of communication mode on speech and language development in deaf children with cochlear implants is that placement into a communication mode or educational program is not random, and children who fail to thrive in oral communication programs are often put into total communication programs.

Although the development of the speech and language skills of deaf children after cochlear implantation has been the primary area of interest to most clinicians who are interested in measuring benefit and outcome in this population, several recent studies have begun to study other cognitive skills of these children such as attention, learning, and memory. However, this new interest in cognitive processes should not be viewed as a divergence from research focused on speech and language in deaf children using cochlear implants. Rather, recent investigations of learning and memory processes in deaf children using cochlear implants may provide new insights into speech and language development and provide principled explanations for the enormous individual differences in outcome and benefit that have been observed in this clinical population (Pisoni, Cleary, Geers, & Tobey, 2000).

An extensive body of literature examining normally-hearing populations has shown that attention, learning, and memory processes are all intertwined and closely related to vocabulary development and language learning. Attention, learning, and memory account for a large amount of variability that is observed in the language skills of normally-hearing adults and children (Baddeley, Gathercole, & Papagno, 1998; Cowan, 1996; Cowan, Nugent, Elliott, Ponomarev, & Sults, 1999; Gupta, 2003). For example, differences in working memory have been found to be closely related to vocabulary knowledge and the development of spoken and written language abilities in normally-hearing adults and children (Cowan, 1996; Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992; Gupta, 2003). In addition, working memory processes have also been linked to language proficiency in deaf children who do not use cochlear implants (Bebko, Bell, Metcalfe-Haggert, & McKinnon, 1998). Thus, in addition to providing vital knowledge about the role of early auditory and linguistic experience in learning and memory, the study of memory processes in deaf children with cochlear implants may also yield new fundamental knowledge about speech and language development.

Direct links between working memory performance and the development of speech and language skills have recently been documented in deaf children using cochlear implants. Pisoni and Cleary (2003) found that immediate memory capacity, measured by forward digit span, was strongly correlated with deaf children's scores on several different word recognition tasks. In addition, serial recall has also been found to be related to the receptive vocabulary of deaf children using cochlear implants (Dawson, Busby, McKay, & Clark, 2002). However, only recently have some of the more intricate aspects of memory processing abilities in deaf children with cochlear implants been explored to uncover how they may influence speech and language development.

The following sections review recent research examining several memory processes that may be intimately connected to speech and language development. Specifically, we summarize findings on working memory capacity, verbal rehearsal, and serial scanning processes in profoundly deaf children using cochlear implants and discuss how they are related to the basic cognitive skills that have been shown to be important to traditional speech and language outcome measures used to assess benefit with a cochlear implant. The usefulness of measuring temporal characteristics of speech, such as speaking rate and interword pause durations, to index the speed of subvocal verbal rehearsal and serial scanning in deaf children using cochlear implants is also discussed. We present the results of these speech-timing studies and consider their implications for the development and the use of subvocal verbal rehearsal and serial scanning in deaf children with cochlear implants and discuss why these two fundamental memory processes contribute to the shorter immediate memory spans observed in these children. Overall, the findings presented here indicate that, in addition to perceptual difficulties related to their hearing impairment and the encoding of degraded auditory input, atypical development of subvocal verbal rehearsal and serial scanning also contribute to the decreased memory spans of deaf children using cochlear implants. Interestingly, these results are similar to what has been found in deaf children who do not use cochlear implants (see Marschark & Mayer, 1998 for a review).

In addition to research on immediate memory and scanning, several recent findings on the nonword repetition skills of deaf children with cochlear implants are described. These new results suggest that some deaf children with cochlear implants have substantial difficulties in rapidly encoding, rehearsing, and repeating novel phonological patterns. Such difficulties indicate that deaf children with cochlear implants have developed atypical phonological processing skills. Finally, we discuss the influence of communication mode and early oral-aural experience on memory and nonword repetition performance. These differences in communication mode suggest that early linguistic experiences and activities after cochlear implantation play a substantial role in perceptual and cognitive development. Taken together, the findings presented here suggest that fundamental linguistic and phonological processing skills used in memory and nonword repetition tasks may play a foundational role in the development of speech and language skills following cochlear implantation. In addition, these findings reveal that basic memory processes such as encoding, subvocal verbal rehearsal, and serial scanning of short-term memory are atypical in deaf children using cochlear implants and appear to be closely related to the nature and amount of early auditory and linguistic exposure received by these children after cochlear implantation.

## Memory Abilities in Deaf Children without Cochlear Implants

Prior to the recent interest in deaf children using cochlear implants and their aural rehabilitation, much research focused on deaf children communicating with manually signed visual-spatial language such as ASL or one of the signing systems created to accompany spoken language. Research concerning the acquisition of signed language and its influence on cognitive and social development encouraged a series of investigations of memory development in this population (e.g., Bebko, 1984; Campbell & Wright, 1990; Liben & Drury, 1977; Marschark & Mayer, 1998). The early work on the memory processes of deaf children who use manual signs and lack a fully developed native spoken language was an important precursor to the current investigations of the memory of deaf children using cochlear implants (Marschark & Mayer, 1998). Several studies have shown that when confronted with a specific language processing task that relies on memory, many deaf children, like their normally-hearing peers, use covert verbal rehearsal as a strategy to maintain items in short-term memory (Bebko, 1984; Liben & Drury, 1977). Covert verbal rehearsal is assumed to involve the repeated cycling of verbally coded memory representations within the phonological loop of working memory in order to prevent memory decay (Baddeley et al., 1975).

One of the strongest pieces of evidence that deaf children use covert verbal rehearsal strategies came from a study by Campbell and Wright (1990). They found that deaf children, like normally-hearing children and adults, are susceptible to the word length effect. Word length effects are observed when the number of lexical items that can be recalled from immediate memory is determined by the length of the words in the list. Word length effects occur because longer words take more time to articulate and subvocally rehearse and cannot be refreshed as quickly and efficiently within the phonological loop. As a result of the decreased rate of subvocal verbal rehearsal, memory spans for lists of longer words will be shorter. Evidence of the word length effect and covert verbal rehearsal strategies in deaf children suggests that they are capable of processing and repeatedly recycling linguistic input within the short-term memory store (Baddeley et al., 1975).

Despite utilizing memory strategies that are similar to their normally-hearing peers, deaf children behave atypically on a wide variety of memory tasks. In particular, phonological memory tasks appear to be the most difficult for deaf children to carry out, especially when they involve encoding and retrieval of sequential information (Banks, Gray, & Fyfe, 1990; Waters & Doehring, 1990). Early onset deafness can also produce substantial differences in performance on memory tasks that require the management and manipulation of phonological or linguistic information. However, what has previously remained unknown is whether, after a prolonged period of auditory and spoken language deprivation, cochlear implantation can ameliorate or even prevent some of these disadvantages and allow deaf children who receive a sensory aid to perform more like their normally-hearing peers on a wide range of language and memory tasks.

## Working Memory Capacity in Deaf Children with Cochlear Implants

Several recent studies have shown that deaf children with cochlear implants have shorter immediate memory spans than their normally-hearing, age-matched peers. The first evidence of shorter memory spans in deaf children using cochlear implants was obtained using the WISC-III auditory digit span task which provided a measure of immediate memory capacity (Pisoni et al., 2000; Pisoni & Geers, 2000). The WISC-III auditory digit span task is administered to children live-voice with lip reading cues available and involves two different recall conditions. In forward digit span recall, children are simply asked to repeat back a sequence of digits in their exact order of presentation. In the backward digit span task, children are required to repeat the digits in the reverse order of their original presentation.

In both the forward and backward digit span tasks, deaf children with cochlear implants performed worse than their normally-hearing peers. Figure 1, adapted from Pisoni and Cleary (2003), displays the forward and backward digit spans obtained from 176 deaf children using cochlear implants obtained over a four year period along with a comparison group of 44 age-matched, normally-hearing children. All children were between 8- and 9-years-old and the deaf children had around 4 to 7 years of experience with their implant (Geers et al., 1999). Subsets of this sample of children were used for all of the subsequent studies discussed in this chapter that were conducted by our lab. The top panel of this figure shows that the digit spans of all four groups of deaf children using cochlear implants were significantly shorter than the digit spans of the normally-hearing children who are shown on the right.

The bottom panel of Figure 1 shows that, in addition to memory span differences found between normally-hearing children and deaf children using cochlear implants, deaf children with cochlear implants who used oral communication methods had longer forward digit spans than children who used total communication. These results provided the first evidence that the quality and quantity of aural and oral exposure can have a systematic effect on immediate memory span capacity for sequential patterns in deaf children using cochlear implants. Specifically, as has been suggested in deaf individuals without cochlear implants, the quality and quantity of oral and aural experience of deaf children with the devices may mediate or influence memory processing strategies such as perceptual and phonological encoding, subvocal verbal rehearsal, and serial scanning (Bebko & Metcalfe-Haggert, 1997).

Given that digit span recall requires the verbal repetition of auditory stimuli, it is reasonable to ask whether perceptual encoding or articulatory difficulties may have substantially contributed to the shorter digit spans observed in the deaf children using cochlear implants, particularly those who used total communication. If a deaf child using a cochlear implant cannot detect and accurately perceive what digit was spoken or has such unintelligible speech that even when the correct response is known it cannot be articulated in any identifiable form, memory capacity may be underestimated. It is also important to consider the role of perceptual or articulatory difficulties in memory performance, because in some memory tasks using only visual stimuli and nonverbal responses, deaf children with cochlear implants perform as well as their normally-hearing peers. For instance, in memory tasks requiring recognition memory for faces or the reproduction of a pattern of visually and spatially arranged dots, deaf children with cochlear implants fall within the normative range of scores obtained for normally-hearing children (Cleary & Pisoni, 2004). Not surprisingly, this result is similar to what has been found in deaf children who do not use cochlear implants (Bellugi, O'Grady, Lillo-Martin, O'Grady-Hynes, Van-Hoek, et al., 1990; Campbell & Wright, 1990; McDaniel, 1980; Olsson & Furth, 1966).

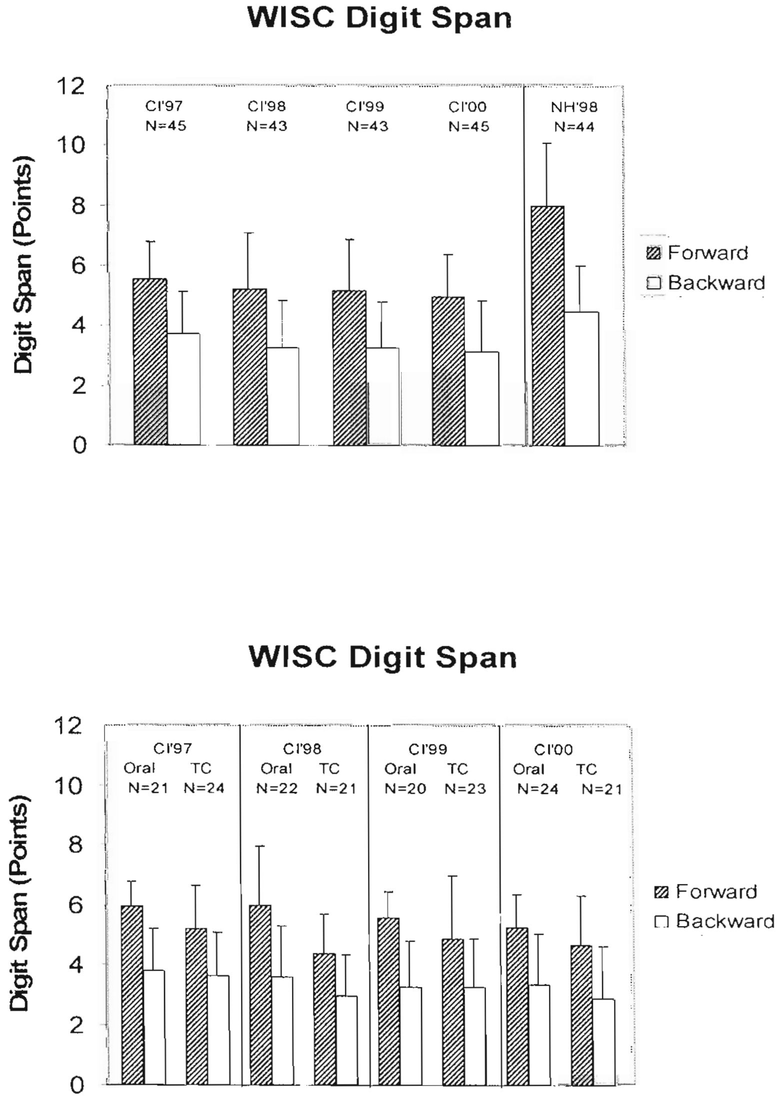


**Figure 1.** Mean WISC-III digit spans from four groups of 8- and 9-year-old deaf children with cochlear implants and normally-hearing children. The top panel shows the digit spans of all deaf and normally-hearing children, and the bottom panel shows the group of deaf children with cochlear implants split according to communication mode. Error bars represent standard error of the mean. (Adapted from Pisoni & Cleary, 2003)

However, deaf children using cochlear implants have been found to have shorter memory spans than their normally-hearing peers in visual memory tasks in which the stimuli are presented sequentially. Using a customized version of the popular memory game "Simon" by Milton Bradley, Cleary, Pisoni, and Geers (2001) reported that deaf children using cochlear implants had shorter reproductive visual memory spans than their normally-hearing peers. Figure 2 shows a version of the Simon apparatus that is nearly identical to the one used to measure memory span in the deaf and normally-hearing children. The Simon memory task used in this study involved presenting randomly generated patterns of colored lights to the children. All patterns were combined using four possible colors (blue, green, red, yellow) and got progressively longer during the task. Children were required to reproduce the patterns by manually pressing the colored and illuminated response buttons on the Simon apparatus which was interfaced to a computer that recorded all responses.

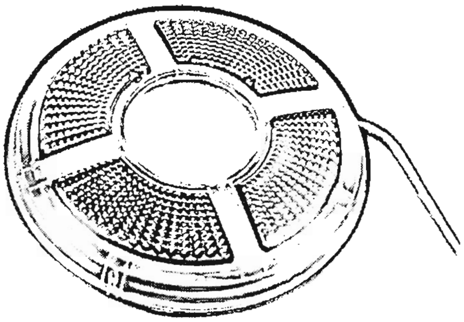


**Figure 2.** Simon memory span game adapted for use with deaf and normally-hearing children.

Cleary et al.'s findings suggest that problems with memory processes other than the early auditory encoding of linguistic input may also contribute to the shorter digit spans of deaf children who use cochlear implants. They reasoned that deaf children using cochlear implants performed poorly even on a visual memory span task because they had difficulty in coding visual sequences verbally and were slower at subvocally rehearsing the verbally coded sequential information in working memory (Cleary et al., 2001). Thus, although the Simon memory task is ostensively based on visual information (the colors of buttons) the most successful strategy to complete the task is to use some form of verbal coding and subvocal verbal rehearsal of color names rather than relying exclusively on visual cues.

Although the deaf children with cochlear implants are able to use subvocal verbal rehearsal, they are at a disadvantage relative to their normally-hearing peers because of their lack of early linguistic experience and aural-oral activities which would ordinarily facilitate rapid execution of this rehearsal strategy. However, the effects of deafness and lack of sensory input on memory performance appear to dissipate when stimuli in memory tasks are not as likely to be verbally encoded (Cleary & Pisoni, 2004; Dawson et al., 2002). For instance, in several serial short-term memory tasks using either tones or hand gestures as stimuli, Dawson and colleagues found that deaf children using cochlear implants performed just as well as their normally-hearing peers.

Taken together, these recent results indicate that verbal encoding problems most likely prevent deaf children with cochlear implants from performing as well as their normally-hearing peers on both auditory and visual memory tasks. However, verbal encoding is not the only underlying memory process required to perform a digit span recall or serial short-term memory task. As mentioned earlier, subvocal verbal rehearsal is also an important component of working memory. To gain a better understanding of how working memory functions in this clinical population, we explored the verbal rehearsal process in much greater detail using several different measures of speech timing.

## Speech Timing and Memory Processes in Deaf Children with Cochlear Implants

In normally-hearing children, several fundamental memory components have been successfully delineated by examining temporal aspects of speech production using speech timing measures. Measures of speech timing during memory tasks completed by normally-hearing children have been used to index both subvocal or covert verbal rehearsal as well as serial scanning of items in short-term memory (Cowan, 1992; Cowan, Wood, Wood, Keller, Nugent et al., 1998). One basic form of speech timing that Cowan and colleagues have measured is overt speaking rate. Measures of overt speaking rate can be used to estimate the rate of subvocal verbal rehearsal in immediate memory. The idea that overt speaking rate is an appropriate measure of subvocal verbal rehearsal is based on a large body of memory research that has found a consistent and strong linear relationship between speaking rate and memory span in both normally-hearing children and adults (Baddeley, Thompson, & Buchanan, 1975; Hitch, Halliday, & Littler, 1989; Hulme & Tordoff, 1989; Kail & Park, 1994; Schweickert, Guentert, & Hersberger, 1990). In general, these studies have found that speakers who articulated faster also had longer digit spans. According to Baddeley and his colleagues (1975), the relationship between speaking rate and immediate memory span occurs because the faster an individual speaks and thus rehearses subvocally, the more frequently items can be refreshed within the phonological loop. A faster rate of rehearsal through this loop will facilitate the recall of more items and ultimately result in a longer memory span.

The finding that speech-timing measures may reflect basic memory processes was explored further by Cowan and his colleagues (1992; 1994) using measures of serial scanning. Scanning is the process by which each item in a list is individually located in short-term memory. This process is carried out by retrieving the items within a list serially during each interword pause taken during the period of recall (Sternberg, 1966). In contrast to measures of overt speaking rate that are frequently derived from sentence repetition or speeded articulation tasks, measures used to estimate serial scanning speed are made during the actual recall process of immediate serial recall tasks. Memory scanning activities can be conveniently indexed by measuring the durations of the interword or inter-item pause durations that occur during digit span recall. Figure 3, adapted from Burkholder and Pisoni (2003b), shows a schematic representation of how interword pauses are measured from speech production samples made during digit span recall.

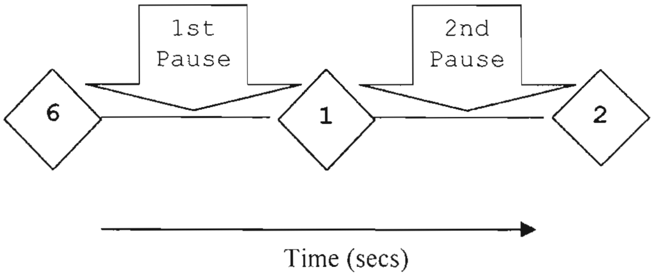


**Figure 3.** Schematic representation of interword pause duration measures made on WISC-III forward digit span responses. Example of a three digit list (6 1 2).

The serial scanning process begins with the onset of the first pause taken during recall and continues until the next item on the list is determined. Thus, during the first pause in serial recall, between the first and second items on the list, scanning occurs until the second item in the list is located and articulated. Similarly, during the pause between the second to last and last items, nearly the entire list has to be scanned until the final item to be recalled is located. Cowan (1992) observed that the pause durations during immediate recall increase as later digits are recalled. This increase in duration occurs because more items from the list must be scanned through in serial order as the final items of the list are recalled.

Cowan and colleagues (1994) also examined maturational effects on speech-timing in normally-hearing children. They found that 8-year-olds spoke significantly faster and had shorter interword pause durations in immediate recall than 4-year-olds. These findings suggested developmental increases in speed of subvocal verbal rehearsal and serial scanning. Increases in subvocal verbal rehearsal speed and serial scanning rates appeared to facilitate immediate memory span recall in the children Cowan examined. In addition to having faster speaking rates and shorter pause times in recall, the 8-year-old children also displayed significantly longer memory spans than the 4-year-old children.

Based on Cowan's findings that speech-timing measures in recall can be used as an index of covert memory processes that influence the immediate memory spans in normally-hearing children, Burkholder and Pisoni (2003b) conducted a speech-timing analysis on the digit span responses of 37 profoundly deaf children using cochlear implants. The children were between 8- and 9-years-old and had 4.5 to 7 years of experience with their cochlear implants. In our study, both the overt speaking rates and pause durations of profoundly deaf children with cochlear implants were compared to a set of measures obtained from a group of 36 age-matched, normally-hearing controls. Speaking rate was obtained from the two groups by measuring the durations of short sentences taken from the McGarr Sentence Intelligibility task (McGarr, 1981).

The McGarr stimulus materials consisted of 36 sentences of 3-, 5-, and 7-syllables, each with 12 sentences at each syllable length. The sentences were elicited by simply asking the children to listen to each sentence as it was read by the clinician or experimenter and then providing them with the written text of the sentence. With the text of the sentence placed in front of them, the children were asked to repeat the sentence at their usual speaking rate. Providing the written text of the sentences reduces the memory load involved in the task and also guards against errors in repetition due to misperception of the spoken sentence. Further assurance that the deaf children with cochlear implants repeated the sentences correctly was achieved by allowing them up to three chances to repeat each sentence. All sentences spoken by both groups of children were digitally recorded and measured using waveform editing software.

Figure 4, adapted from Burkholder and Pisoni (2003b), displays the McGarr sentence durations obtained from the 8- and 9-year-old deaf children with cochlear implants and their aged-matched, normally-hearing peers. The mean durations of the 3-, 5-, and 7-syllable sentences are each shown separately on the abscissa. In addition, the mean duration of all sentences combined together is shown in this figure. The top panel of this figure clearly illustrates that deaf children using cochlear implants had significantly slower speaking rates than their normally-hearing peers on these sentences. In addition, the bottom panel of the figure shows that children who use total communication spoke significantly slower than children who use oral communication methods.

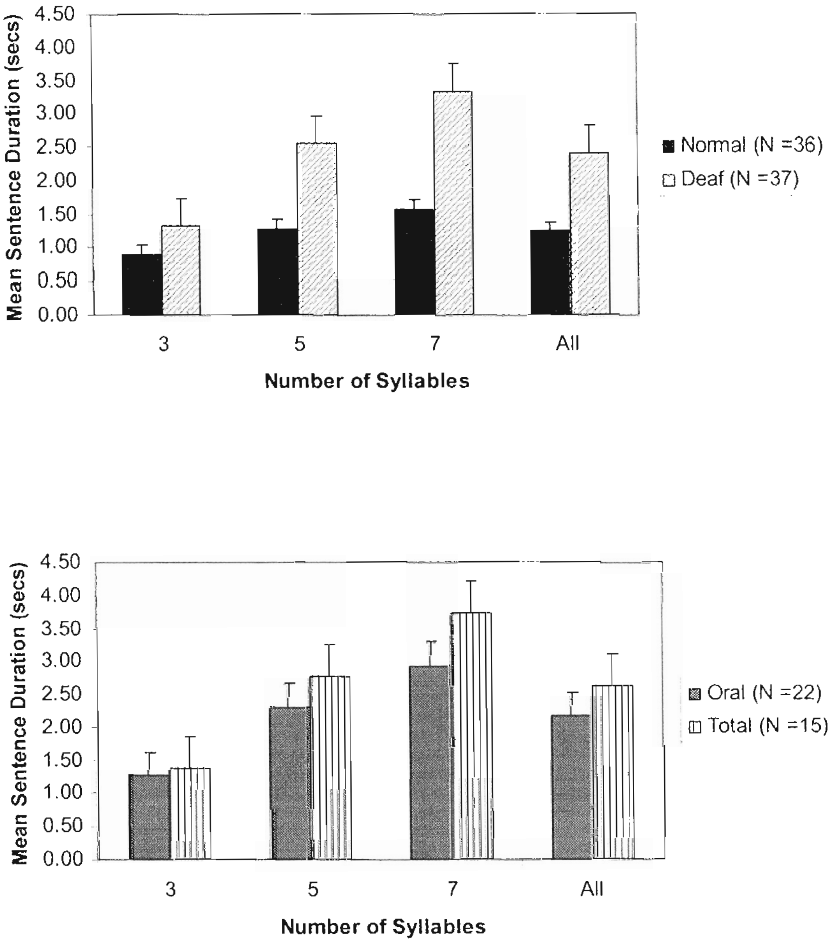


**Figure 4.** Mean sentence durations of normally-hearing children and deaf children using cochlear implants. The top panel shows sentence durations of normally-hearing children and deaf children using cochlear implants, and the bottom panel shows the sentence durations of the deaf children with cochlear implants split according to communication mode. Error bars represent standard error of the mean. (Redrawn from Burkholder & Pisoni, 2003b)

Slower speaking rates have been documented in deaf individuals previously and are attributed in part to the lack of auditory feedback while speaking (Bochner, Barefoot, & Johnson, 1987). However, even with the newly provided auditory feedback from a cochlear implant, deaf children still appear to be unable to produce speaking rates within normal ranges. Based on the findings obtained in previous research examining speaking rate and memory, the pediatric cochlear implant users' inability to overtly articulate at rapid paces may underlie differences in covert verbal rehearsal and result in these children having shorter digit spans than their normally-hearing peers.

The proposal that slower speaking rates and subvocal rehearsal speeds may contribute to the shorter memory spans of the deaf children with cochlear implants was further confirmed through analyses showing a robust correlation between their speaking rates and digit spans (Burkholder & Pisoni, 2003b; Pisoni & Cleary, 2003). Figure 5, adapted from Pisoni and Cleary (2003), displays the correlation between sentence durations and forward digit spans of 176 deaf children using cochlear implants. The log-based transformation of McGarr sentence durations appears on the abscissa and WISC-III digit span points appear on the ordinate.

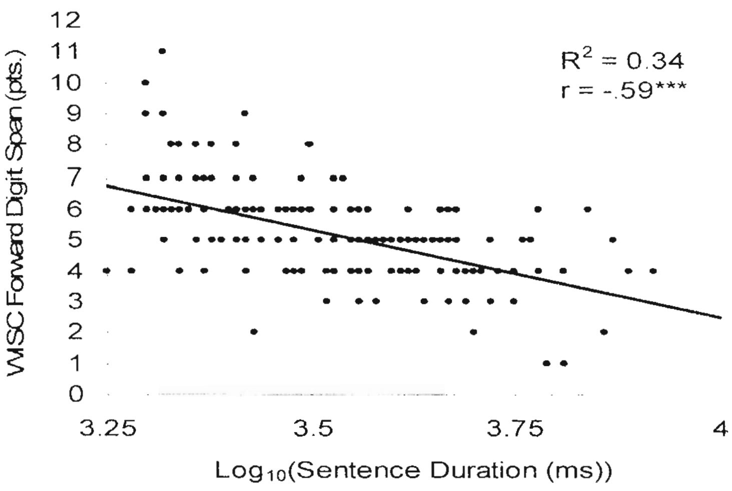


**Figure 5.** Scatterplot illustrating the relationship between average sentence durations for the seven-syllable McGarr Sentences and WISC-III forward digit span scored by points. The sentence durations were log-transformed. R-squared values indicate percent of variance accounted for by the linear relation. (Adapted from Pisoni & Cleary, 2003)

Based on this robust correlation between sentence duration and memory span, it appears that deaf children with cochlear implants and those deaf children without perform worse than their normally-hearing peers on serial recall tasks such as digit span recall because they are not able to subvocally rehearse the digits fast enough. In addition, the correlations between forward digit spans and speaking rate obtained within both the normally-hearing children and deaf children using cochlear implants were of a similar magnitude. These results indicate that basic subvocal verbal rehearsal processes operate similarly in these two different populations and may contribute in comparable ways to immediate memory span for spoken digits.

Although both the pediatric cochlear implant users' shorter forward digit spans and their strong correlation with speaking rate were expected, one other result from this study was unexpected and interesting. Correlations conducted between backward digit span and speaking rate revealed a strong relationship for the group of deaf children with cochlear implants. However, the group of normally-hearing children did not display a correlation between backward digit span and speaking rate (Burkholder & Pisoni, 2003b). This result suggests that the deaf children with cochlear implants may be using the same subvocal verbal rehearsal strategy to complete the backward digit span task that they used in the forward digit span task. This strategy may not be as efficient as the executive planning and organizational strategies that normally-hearing children use to perform the same memory task. Thus, not only may some of the pediatric cochlear implant users' subvocal verbal processing strategies suffer because they are carried out more slowly than normally-hearing children's, but the subvocal verbal rehearsal strategies could also be inappropriately engaged during certain tasks, such as backward digit span recall, in which other planning, rehearsal, and recall strategies would be more useful (Thomas, Milner, & Haberlandt, 2003).

In addition, it is also possible that the shorter digit spans observed in the deaf children with cochlear implants may be due to their inability to scan test items in memory as fast as normally-hearing children. Following the same procedures that Cowan et al. (1994) used, scanning rates during the digit span task were obtained by measuring the interword pause durations taken from digital audio recordings of pediatric cochlear implant users and normally-hearing children completing the actual recall portion of the task. Figure 6, adapted from Burkholder and Pisoni (2003b), displays mean interword pause durations produced by deaf children with cochlear implants and normally-hearing children during WISC-III forward digit span recall. Mean interword pause durations are shown for digit span lists of three and four digits and the longest list recalled by each child which is denoted as the list limit.

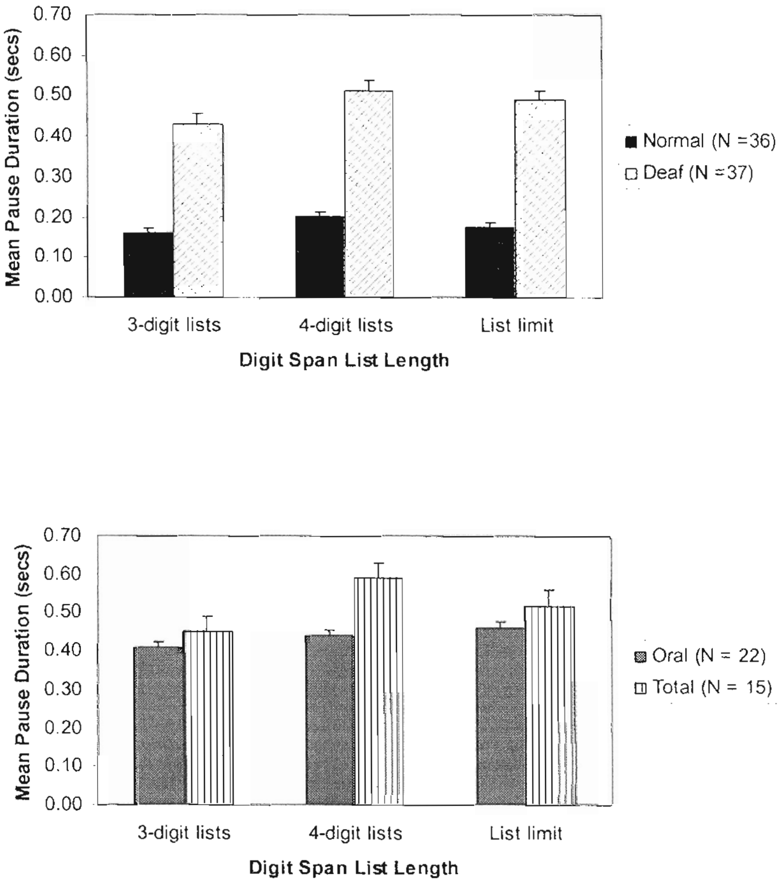


**Figure 6.** Mean interword pause durations taken during WISC-III forward digit span recall for list lengths of 3 and 4 digits and the span limiting list or longest list correctly recalled. The top panel shows the pause durations of both normally-hearing and deaf children using cochlear implants and the bottom panel shows the pause durations of deaf children using cochlear implants split according to communication mode. Error bars represent standard error of the mean. (Redrawn from Burkholder & Pisoni, 2003b)

The top panel of Figure 6 illustrates that the deaf children using cochlear implants were significantly slower than their normally-hearing peers while scanning items during digit span recall. In fact, the pediatric cochlear implant users' interword pause durations taken during digit span recall were nearly twice as long as the normally-hearing children's interword pause durations when recalling lists of three and four digits and in the longest list they were able to correctly recall. These slower serial scanning rates may also be responsible for the shorter memory spans of deaf children using cochlear implants.

Subvocal verbal rehearsal and serial scanning processes both rely on phonological or linguistic encoding. Thus, the lack of early auditory sensory input and linguistic stimulation in the deaf children is likely to be another important factor affecting subvocal verbal rehearsal speed and serial scanning rates resulting in shorter digit spans. When considered together with Cowan et al.'s (1994; 1999) earlier findings on speech timing and memory, the speech-timing study in deaf children using cochlear implants suggests that maturation of subvocal verbal rehearsal and serial scanning not only depends on chronological age but also on the amount and type of early linguistic exposure that children receive. The role that linguistic information plays in these memory processes also explains why the children who used communication methods that stress auditory-oral training performed better than the children who used total communication. The richer auditory-oral exposure and experience that the children using oral communication receive likely facilitates their performance on a wide range of linguistic tasks that use subvocal verbal rehearsal and serial scanning of short-term memory because they are able to rapidly decompose and linguistically represent auditory-verbal sensory information. In addition, children using oral communication methods may have an advantage on auditory memory tasks because they have less difficulty retrieving these items and reassembling them into an intelligible spoken response.

## Memory Recall Errors in Deaf Children Using Cochlear Implants

Recently, additional evidence has been collected suggesting that the pediatric cochlear implant users' shorter auditory digit spans are primarily related to memory processing problems such as subvocal verbal rehearsal and serial scanning. Burkholder and Pisoni (2004) examined and categorized the errors made during digit span recall by deaf children using cochlear implants to determine whether their errors were primarily due to encoding item or order information incorrectly. Each individual error made during spoken recall was classified as one of four types of errors. Errors caused by the recall of digits in an incorrect order were considered to be order errors. Errors caused by the recall of a digit or digits that were not present in the original list were considered to be item errors. Errors in digit span recall caused by the failure to repeat one or more digits were considered to be omissions. Finally, errors that represented both item and order errors were considered to be combination errors.

The two types of errors with the most relevance in this analysis were item and order errors (Conrad, 1965). Order errors result from the loss of temporal order information during encoding or spoken recall. In a serial recall task, encoding sequential order and maintaining this information until recall is a complex process. Therefore, errors in order are most likely related to mistakes in processing due to increased cognitive load. In contrast, item errors result in the replacement of an individual digit in the list with a digit that was not presented in the original list. In a group of deaf children using cochlear implants, item errors are likely indicative of encoding problems rather than slowed or inefficient rehearsal or scanning. Therefore, it is important to dissociate these two types of errors from one another in memory recall.

Figure 7, adapted from Burkholder and Pisoni (2004), shows the proportion of each type of error made in the forward (top panel) and backward (bottom panel) digit span task for both the deaf children using cochlear implants and the group of age-matched, normally-hearing peers. By using this categorization process, Burkholder and Pisoni (2004) found that the proportion of order errors exceeded all other types of errors made by deaf children with cochlear implants during auditory digit span recall.

Although this was an auditory memory task, the performance of the deaf children with cochlear implants appeared to be more influenced by order errors rather than item errors.

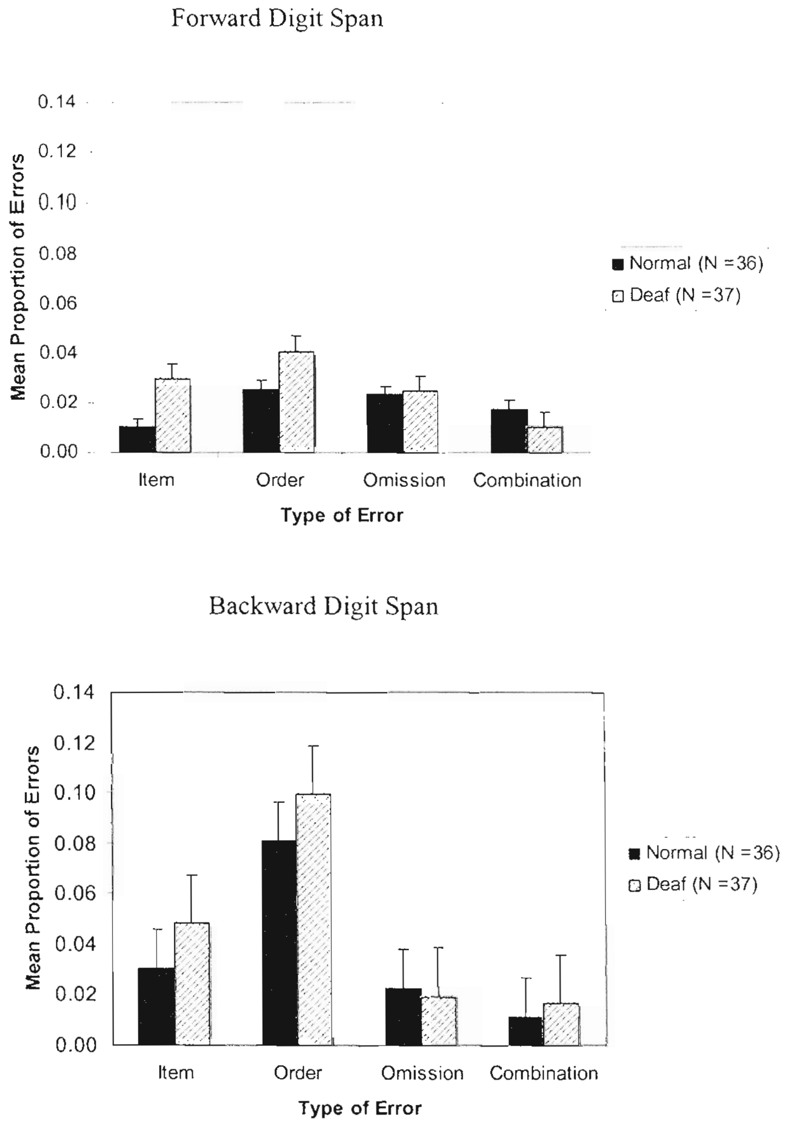


**Figure** 7. Mean proportion of error types made by normally-hearing children and deaf children using cochlear implants during digit span recall in forward (top panel) and backward (bottom panel) recall conditions. Error bars represent standard error of the mean.

The pattern of errors made by the deaf children using cochlear implants was similar to the results found for normally-hearing children. Both the deaf and normally-hearing children committed significantly more order errors than item errors during digit span recall. In addition, order errors were more numerous in backward digit span recall than in forward digit span recall. This result is not surprising because backward digit span is considered to be a more complex and demanding task requiring planning and recall strategies and executive function that may lead to an increased processing load (Li & Lewandowsky, 1995).

Taken together, three converging sets of findings suggest that deaf children with cochlear implants perform more poorly on memory span tasks because they covertly rehearse and scan items in short-term memory more slowly than normally-hearing children. First, deaf children with cochlear implants do poorly even on memory span tasks that do not require encoding of auditory stimuli and spoken responses (Cleary et al., 2001). This result provides support for the proposal that deaf children with cochlear implants not only have difficulty perceiving and encoding auditory stimuli, but they have other memory processing problems that are not tied exclusively to input modalities used in the memory assessments.

Second, Burkholder and Pisoni's (2003b) speech-timing analysis found that the pediatric cochlear implant users' sentence and interword pause durations were significantly longer than the sentence and interword pause durations of their normally-hearing peers. These results suggest that the memory processes reflected in these measures, subvocal verbal rehearsal and serial scanning, also operate more slowly in the deaf children with cochlear implants. Finally, by directly examining the nature of the errors in digit span recall, Burkholder and Pisoni (2004) found a greater proportion of order errors than item errors made in immediate serial recall by deaf children using cochlear implants. This result suggests that deaf children with cochlear implants not only have difficulty in perceiving spoken digits but they frequently fail to encode and maintain the correct serial order of the digits in a test sequence. Therefore, a solid body of evidence suggests that encoding order information may be more difficult for deaf children with cochlear implants than correctly perceiving the degraded stimuli received through their cochlear implant.

## Nonword Repetition in Deaf Children with Cochlear Implants

In addition to measures of immediate memory capacity using digit span, nonword repetition has also been a useful methodology to examine phonological working memory skills in normally-hearing children. Differences in nonword repetition performance have been found to be related to novel word learning in both adults and children (Gathercole & Baddeley, 1989; Gathercole et al., 1992; Gupta, 2003). The relationship between novel word learning and nonword repetition should come as no surprise, because in any word learning task, whether in an experimental or real-world setting, words originally begin as nonwords for those trying to learn them. Therefore, measures of nonword repetition may be useful in assessing the fundamental operation of word learning ability in deaf children using cochlear implants. Unlike spoken word recognition or sentence repetition tasks, nonword repetition is one method that can be used to measure both speech perception and production skills in the absence of higher-level contextual and lexical influences.

However, there are numerous other information processing skills that play a role in successfully completing a nonword repetition task. Although the task of repeating a nonword may intuitively sound as if it is a fairly easy task, successful nonword repetition requires the completion of a complex sequence of sensory, perceptual, and linguistic processes that are executed rapidly in a short period of time. To complete nonword repetition, a listener must encode a novel sound pattern in an auditory-only mode, retain and rehearse the pattern within the phonological loop, and then reassemble the sound pattern into an articulatory motor program for speech. This complex sequence of tasks may be particularly difficult for deaf children who use cochlear implants. Therefore, by examining pediatric cochlear implant users' nonword repetition skills, valuable information can be gathered on several important processes of speech, language, and memory to provide new insights into how these children apply phonological processing skills to novel nonword patterns.

The nonword repetition task utilized in our research was adapted from the stimulus materials developed by Gathercole and her colleagues (1994). The original nonword list included 40 nonwords that sounded like plausible English words. From the original set of 40 nonwords, 20 were selected for use as stimuli for the nonword repetition task conducted with the deaf children who use cochlear implants. These 20 nonwords were selected because of the high degree of variability that was observed when they were repeated by normally-hearing children (Carlson, Cleary, & Pisoni, 1998).

Unlike the McGarr sentence intelligibility task used to elicit small samples of connected speech to measure speaking rate, the nonword repetition task is not administered via live voice and does not involve providing the deaf children with the written text of each nonword. Rather, in the nonword repetition task, the prerecorded stimuli are played back at a comfortable listening level over a loudspeaker placed directly in front of the child. The auditory-only administration method used in the nonword repetition paradigm is significant because no visual cues from the speaker's face are available to the children during the task.

To evaluate the nonword repetition skills of deaf children with cochlear implants, we employed several analysis methods. Detailed and time consuming methods for evaluating the accuracy of the deaf children's nonword repetitions involved the use of both segmental and suprasegmental scoring procedures carried out by several trained transcribers (Carter, Dillon, & Pisoni, 2002; Dillon, Cleary, Pisoni, & Carter, 2004). In these initial analyses, children were scored on their ability to correctly reproduce a number of aspects in each nonword, such as the number of syllables, stress pattern, and phonemes. When scored using traditional segmental and suprasegmental methods, the deaf children correctly produced a nonword only 5% of the time, which indicates a floor performance. Such strict scoring criteria make it difficult to examine any variation in the nonword repetition skills of these deaf children. However, an alternative nonword repetition scoring method using perceptual ratings made by naïve listeners has been useful to quantify the nonword repetition skills of deaf children with cochlear implants.

In the perceptual ratings paradigm that we developed, normally-hearing adult listeners were first presented with one of the same target nonwords, spoken by an adult female, originally presented to the deaf children with cochlear implants. After hearing the nonword target, the listeners were presented with the repetition of that same nonword by one of the deaf children using a cochlear implant. The listeners were told to rate the utterances produced by the children using a scale from 1 to 7 according to how accurate they believed the child's response was, ignoring differences in pitch, when compared to the target pattern that preceded it.

In contrast to the segmental and suprasegmental scoring methods, results from the perceptual ratings task revealed a wide range of variability in nonword repetition skills within the group of deaf children using cochlear implants (Dillon, Burkholder, Cleary, & Pisoni, in press). While most of the children were able to complete the nonword repetition task, a small number of children appeared to be overwhelmed with the task and performed nearly at floor. Communication mode also affected performance on this task. Children who used oral communication received higher nonword ratings than the children who used total communication.

In a sample of 69 deaf children using cochlear implants, Dillon and colleagues (2004) also found that subvocal verbal rehearsal speeds, measured by overt speaking rate, accounted for the most variance in nonword repetition ratings. The results of a regression model fit to the nonword repetition ratings of the deaf children using cochlear implants are shown in Table 1. Along with overt speaking rate, closed-set speech perception, speech intelligibility, and communication mode were also considered in the regression analysis.

The standardized coefficients listed in the far right column indicate the degree of variance in nonword repetition ratings accounted for by each independent variable. The analysis indicated that the speed of subvocal verbal rehearsal, as measured by the durations of overtly articulated sentences, was negatively related to nonword repetition skills in deaf children with cochlear implants. This measure accounted for the most variance in nonword repetition ratings assigned by normally-hearing listeners. This result is consistent with a large body of earlier work on normally-hearing children and adults showing that working memory is strongly correlated with nonword repetition performance (Gathercole & Baddeley, 1989; Gathercole et al., 1992; Gupta, 2003). The present results are particularly interesting because they suggest that processing speed of the subcomponent processes required to complete nonword repetition is the factor that accounts for the most variance in the task when compared to traditional end-point measures of speech perception and production.

| **Factors Contributing to Nonword Repetition Performance** | **Standardized coefficient** |
|---|---|
| **Overt speaking rate:** Subvocal verbal rehearsal speed | |
| Mean sentence duration of McGarr 7-syllable sentences (log, msec) | - .34*** |
| **Closed-set word identification:** Speech Perception | |
| Word Intelligibility by Picture Identification (WIPI) | +.29*** |
| **Speech Production:** Speech Intelligibility | |
| McGarr Sentence Intelligibility | +.28** |
| **Degree of exposure to oral-only communication** | |
| Communication Mode score | +.16* |

*$p < .05$, **$p < .01$, ***$p \leq .001$

**Table 1**. Results of the regression model fit to the nonword repetition ratings of 69 pediatric cochlear implant users.

## Nonword Repetition Duration and Response Latency Measurements

In addition to the data on nonword repetition collected from perceptual ratings and segmental and suprasegmental analyses, Burkholder and Pisoni (2003a) also measured nonword durations and response latencies of the deaf children using cochlear implants. Similar to the speech-timing measures obtained from the deaf children's sentence repetitions and digit span recall responses, the duration and response latency measurements have been informative about the speed of processing in deaf children using cochlear implants. Response latencies provided measures of how long it took each child to plan and initiate his or her response while durations of the nonwords indicated how long it took each child to completely utter a nonword.

The results of this study indicated that the nonword repetition response latencies of the deaf children with cochlear implants were nearly twice as long as the response latencies of a group of age-matched, normally-hearing children. In addition, the durations of the nonword responses of the deaf children with cochlear implants were significantly longer than the durations of the nonwords spoken by normally-hearing children. Taken together, these two results suggest that the deaf children with cochlear implants require significantly more processing time than their normally-hearing peers to encode, rehearse, and articulate novel nonword patterns. Because nonword repetition is frequently conceptualized as a measure of phonological working memory, these results provide additional support concerning the limited working memory capabilities of deaf children who use cochlear implants. Overall, the recent studies on the immediate memory capacity of deaf children with cochlear implants suggested that the lack of early experience with auditory and linguistic input has profound effects not only on speech perception and sensory encoding but also on the children's ability to encode, rehearse, and recall sequential information whether it is presented in the form of visual patterns, highly familiar digits, or novel nonwords.

## Summary and Conclusions

In this chapter, we presented a summary and discussion of several recent studies that assessed the working memory processes and abilities of deaf children with cochlear implants. Although most of the clinical research examining deaf children using cochlear implants has focused on traditional audiological outcome measures of speech and language skills to assess benefit, important new knowledge about speech and language development has come from recent studies on memory processing abilities in this population. These studies have shown that subvocal verbal rehearsal and serial scanning operate much more slowly in deaf children with cochlear implants and contribute to their shorter memory spans. It appears that slower processing speeds may even play a greater role in the deaf children's memory performance than the initial encoding problems related to their current hearing impairment and use of a cochlear implant.

In addition, these studies suggest that the amount and/or nature of the auditory exposure that children receive after implantation can influence their performance on immediate memory tasks that require the encoding, verbal rehearsal, and serial scanning of phonological information in working memory. As in earlier studies, we have found that deaf children with cochlear implants who use oral communication methods consistently performed better than deaf children who use total communication methods on a wide range of tasks (Dillon et al., 2004; Pisoni et al., 2000). However, in these studies, individual differences due to communication mode were reflected in both processing speed and accuracy in tasks assessing immediate memory. Although it might seem reasonable that having a system of visual-manual cues to assist deaf children in communicating would serve as an advantage, the findings presented here suggest that processing additional manual input during communicative exchanges, in addition to having poorer speech skills, may influence the strategies and speed of memory encoding and rehearsal by children using total communication.

This interpretation is consistent with what has been observed in deaf individuals not using cochlear implants. For instance, deaf students who used sign and had poorer speech skills have been found to rely on both verbal and visual encoding strategies during memory tasks (Lichtenstein, 1998). In addition, deaf children and adults using cued speech visually encode hand shape and placement during serial order tasks rather than only verbally encoding the phonological information that cued speech is designed to disambiguate (Leybaert & Lechat, 2001). A similar reliance on both verbal and visual encoding during sequential memory tasks may be a disadvantage to deaf children with cochlear implants who use total communication relative to their orally communicating peers who likely use only verbal or speech-based encoding and rehearsal strategies.

An alternative and perhaps more controversial idea is that manual input may not only influence memory processing strategies used by children using total communication but could create competition for limited resources in auditory-visual modalities used for both hearing and seeing important speech cues during auditory memory span tasks (Bergeson & Pisoni, 2004; Pisoni, 2000). Thus, when a child with a cochlear implant using total communication methods such as signed exact English or cued speech is confronted with manual signs, his or her attention will be drawn to the hand(s) of the speaker in addition to the lips on the speaker's face. It has been well documented in normally-hearing and hearing-impaired adults and children both with and without cochlear implants that speech cues obtained from a speaker's face can provide reliable complementary information about the linguistic content of the speech signal that is equivalent to having a 15 dB increase in the speech signal (MacLeod & Summerfield, 1987; Sumby & Pollack, 1954; Tyler, Fryauf-Bertschy, Kelsay, Gantz, Woodworth, & Parkinson, 1997; Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997).

In deaf children using cochlear implants, the amount of information extracted from visual speech cues and the audio-visual gain demonstrated in speech perception tasks is related to communication mode. Children in total communication programs are less adept at combining auditory and visual sources of speech and do not perform as well in visual-only speech perception conditions relative to their peers who use oral communication (Lachs, Pisoni, & Kirk, 2001; Bergeson, Pisoni, & Davis, 2003). This result suggests that use of signing strategies in addition to speech may prevent children using total communication from fully seeing and processing the articulatory gestures of a speaker's face which provide vital information about the speech signal and may influence the strategies used in the encoding, rehearsal, and recall of verbal information.

In addition to the specific role that linguistic experience has on the development of speech, language, and memory performance in deaf children using cochlear implants, experience-dependent plasticity associated with profound deafness may also contribute to differences in speech and language outcome measures in deaf children using cochlear implants. Neuroplastic effects of auditory deprivation have been well documented in areas ranging from the peripheral auditory system to the cerebral cortex (Shepherd & Hardie, 2001). Recently, it has been suggested that neural plasticity may be associated with postoperative performance with a cochlear implant. Using preoperative PET scan measures, Lee and colleagues (2001) found that deaf children with hypermetabolism within visual pathways performed worse after receiving a cochlear implant than children with reduced metabolism in visual pathways that neighbored auditory cortex. These neural imaging results indicate that the recruitment and reorganization of unused auditory cortex by visual pathways has consequences for the later development of auditory-based language.

Associations between preoperative PET scans in deaf children and their subsequent performance on speech and language tasks after cochlear implantation have also been reported recently by Lee, Oh, Sun, Joo, and Soo (2004). They found that children who go on to be more successful users of cochlear implants had more metabolic activity in cortical areas that are suspected to be active in working memory tasks. This is a theoretically significant finding because it suggests that memory not only affects the abilities and mechanisms of language acquisition directly but also may affect it indirectly by playing a more general neuro-cognitive role in how a deaf child using a cochlear implant learns to encode and derive meaning from the degraded auditory signals provided by the device.

Thus, a continuing problem facing researchers interested in how deaf children using cochlear implants develop speech and language is to more clearly determine how these children learn to use the auditory signals provided to them through an implant in the first place. In addition to relying on working memory processes, the task of encoding and interpreting the new sounds processed by a cochlear implant may also make use of specific perceptual learning abilities, attention, and multimodal audio-visual integration skills. Therefore, in addition to further considering the relationship between memory and language in deaf children using cochlear implants, new research efforts should focus on other important cognitive processes such as perceptual learning, long-term memory, selective attention, and executive function if we are to gain a more detailed picture on both neural and behavioral levels of how deaf children develop speech and language skills while using a cochlear implant.

## References

Aslin, J., Jusczyk, P., & Pisoni, D. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology:, Vol. 2, Cognition, Perception and Language, 5th Ed.* (pp. 147-198). New York: Wiley.

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

Baddeley, A.D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Behavior, 14*(6), 575-589.

Banks, J., Gray, D., & Fyfe, R. (1990). The written recall of primed stories by severely deaf children. *British Journal of Educational Psychology, 60*, 192-206.

Bebko, J. (1984). Memory and rehearsal characteristics of profoundly deaf children. *Journal of Experimental Child Psychology, 38*, 415-428.

Bebko, J.M., & Metcalfe-Haggert, A. (1997). Deafness, language skills, and rehearsal: A model for the development of a memory strategy. *Journal of Deaf Studies & Deaf Education, 2*(3), 131-139.

Bebko, J.M., Bell, M.A., Metcalfe-Haggert, A., & McKinnon, E. (1998). Language proficiency and the prediction of spontaneous rehearsal in children who are deaf. *Journal of Experimental Child Psychology, 68*, 51-69.

Bellugi, U., O'Grady, L., Lillo-Martin, D., O'Grady-Hynes, M., Van-Hoek, K., & Corina, D. (1990). Enhancement of spatial cognition in deaf children. In V. Volterra & C. Erting (Eds.), *From gesture to language in hearing and deaf children.* (pp. 278-299). Berlin: Springer-Verlag.

Bergeson, T.R., Pisoni, D.B., & Davis, R.A.O. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *Volta Review, 103*, 347-370.

Bergeson, T.R., & Pisoni, D.B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence, & B.E. Stein (Eds.), *Handbook of Multisensory Integration.*

Bochner, J.H., Barefoot, S.M., & Johnson, B.A. (1987). Pausing in the speech of deaf young adults, *Journal of Phonetics, 15*, 323-333.

Burkholder, R.A., & Pisoni, D.B. (2003a). *Nonword repetition in children using cochlear implants: Latencies and durations*. Poster session presented at the American Auditory Society Science and Technical Meeting, Scottsdale, AZ.

Burkholder, R.A., & Pisoni, D.B. (2003b). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85*, 63-88.

Burkholder, R.A., & Pisoni, D.B. (2004). *Analysis of Digit Span Recall Errors in Paediatric Cochlear Implant Users.* Poster session presented at the European Symposium of Paediatric Cochlear Implantation, Geneva, Switzerland.

Campbell, R., & Wright, H. (1990). Deafness and immediate memory for pictures: Dissociations between "inner speech" and "inner ear". *Journal of Experimental Child Psychology, 50*, 259-286.

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.

Carter, A., Dillon, C., & Pisoni, D.B. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: suprasegmental analysis. *Clinical Linguistics and Phonetics, 16*, 619-638.

Cleary, M., & Pisoni, D.B. (2004). *Visual and visual-spatial memory measures in children with cochlear implants*. Poster presented at the International Cochlear Implant Conference, Indianapolis, IN.

Cleary, M., Pisoni, D. B., & Geers, A. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing, 22*(5), 395-411.

Connor, C., Hieber, S., Arts, H., & Zwolan, T.A. (2000). Speech, vocabulary, and the education of children using cochlear implants: oral or total communication. *Journal of Speech, Language, and Hearing Research, 43*, 1185-1204.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning & Verbal Behavior, 4*, 161-169.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31*, 668-684.

Cowan, N. (1996). Short-term memory, working memory, and their importance in language processing. *Topics in Language Disorders, 17*, 1-18.

Cowan, N., Keller, T., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language, 33*, 234-250.

Cowan, N., Nugent, L.D., Elliott, E.M., Ponomarev, I., & Sults, J.S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*(5), 1082-1097.

Cowan, N., Wood, N., Wood, P., Keller, T., Nugent, L., & Keller, C. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology, 127*, 141-160.

Curtis, S. (1977). Genie: A Psycholinguistic Study of a Modern Day 'Wild Child.' New York: Academic Press.

Dawson, P., Busby, P., McKay, C., & Clark, G. (2002). Short-term auditory memory in children using cochlear implants and its relevance to receptive language. *Journal of Speech, Language, and Hearing Research, 45*, 789-801.

DeCasper, A.J., & Spence, M.J. (1986). Prenatal maternal speech influences newborns'perception of speech sounds. *Infant Behavior & Development, 9*, 133-150.

Dillon, C.M., Burkholder, R.A., Cleary, M., & Pisoni, D.B. (in press). Perceptual ratings of nonword repetition responses by deaf children after cochlear implantation: Correlations with measures of speech, language, and working memory. *Journal of Speech, Language, and Hearing Research.*

Dillon, C.M., Cleary, M., Pisoni, D.B., & Carter, A.K. (2004). Imitation of nonwords hearing-impaired children with cochlear implants: Segmental analyses. *Clinical Linguistics and Phonetics, 18*, 39-55.

Gathercole, S.E., & Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887-898.

Gathercole, S., Willis, C., Baddeley, A., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*(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.* St. Louis, MO: Central Institute for the Deaf, 5-20.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology, 56A*, 1213-1236.

Hitch, G., Halliday, M., & Littler, J. (1989). Item identification and rehearsal rate as predictors of memory span in children. *Quarterly Journal of Experimental Psychology, 41*, 321-337.

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology, 47*, 72-87.

Kail, R., & Park, Y. (1994). Processing time, articulation time, and memory span. *Journal of Experimental Child Psychology, 57*, 281-291.

Kirk, K. (2002). *Cochlear Implants*. Paper presented at the Cochlear Implant Conference: Practices and research for audiologists and speech language pathologists, Bloomington, IN.

Lachs, L., Pisoni, D.B., & Kirk, K. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing, 22*, 236-251.

Lane, H. (1979). *The Wild Boy of Aveyron.* Boston, MA: Harvard University Press.

Lee, D.S., Lee, J.S., Oh, S.H., Kim, S., Kim, J., Chung, J., Lee, M.C., & Kim, C.S. (2001). Cross-modal plasticity and cochlear implants. *Nature, 409*, 149-150.

Lee, H.J., Oh, S., Sun, K.C., Joo, K.E., & Soo, L.D. (2004). Predicting cochlear implant outcome in a highly variable group of congenitally deaf children: importance of central processing. Poster presented at the *Midwinter meeting of the Association for Research in Otolaryngology*, Daytona Beach, FL.

Leybaert, J., & Lechat, J. (2001). Phonological similarity effects in memory for serial order of cued speech. *Journal of Speech, Language, and Hearing Research, 44*, 949-963.

Li, S.C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 837-847.

Liben, L., & Drury, A. (1977). Short-term memory in deaf and hearing children in relation to stimulus characteristics. *Journal of Experimental Child Psychology, 24*, 60-73.

Lichtenstein, E. (1998). The relationships between reading processes and English skills of deaf college students. *Journal of Deaf Studies and Deaf Education, 1,* 249-262.

McDaniel, E.D. (1980). Visual memory in the deaf. *American Annals of the Deaf, 125,* 17-20.

McGarr, N. (1981). The effect of context on the intelligibility of hearing and deaf children's speech. *Language and Speech, 24*, 255-263.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*, 131-141.

Marschark, M., & Mayer, T.S. (1998). Mental Representation and Memory in Deaf Adults and Children. In M. Marschark &. M.D. Clark (Eds.), *Psychological Perspectives on Deafness* (Vol. 2). Mahwah, NJ: Lawrence Erlbaum Assoc., Inc.

Morrone, M.C., Burr, D.C. (1986) Evidence for the existence and development of visual inhibition in humans. *Nature*, 321, 235-237.

Olsson, J.E., & Furth, H.G. (1966). Visual memory span in the deaf. *American Journal of Psychology, 79*, 480-484.

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. *The Volta Review, 101*, 111-164.

Pisoni, D.B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*, 106S-120S.

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, and Laryngology, 185*, 92-93.

Rauschecker, J., & Shannon, R. (2002). Sending sound to the brain. *Science, 295*, 1025-1029.

Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science, 1*, 74-77.

Shepherd, R.K., & Hardie, N. (2001). Deafness-induced changes in the auditory pathway: Implications for cochlear implants. *Audiology and Neuro-Otology, 6*, 305-318.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652-654.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212-215.

Thomas, J., Milner, H., & Haberlandt, K. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science*, *14*(2), 169-174.

Tyler, R.F., Fryauf-Bertschy, H., Kelsay, D. M., Gantz, B., Woodworth, G., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology Head and Neck Surgery, 117,* 180-187.

Tyler, R.F., Parkinson, A.J., Woodworth, G.G., Lowder, M.W., & Gantz, B.J. (1997). Performance over time of adult patients using the Ineraid or nucleus cochlear implant. *Journal of the Acoustical Society of America, 102,* 508-522.

Waters, G., & Doehring, D. (1990). Reading acquisition in congenitally deaf children who communicate orally: Insights from an analysis of component reading, language, and memory skills. In C.B.A. Levy (Ed.), *Reading and its development.* New York: Academic Press.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals[1]

**Brianna L. Conrey and David B. Pisoni[2]**

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

[1] This work was supported by NIH NIDCD R01 Research Grant DC00111 and NIH NIDCD T32 Training Grant to Indiana University. The first author was also supported by an Indiana University Chancellor's Fellowship and an NSF Graduate Research Fellowship. The authors would like to thank Luis Hernandez for technical support and Sara Phillips and Angelique Horace for help with data collection. We also thank Jim Craig, Jason Gold, Olaf Sporns, and Bryan Donaldson for advice on various aspects of the data analysis. We gratefully acknowledge the audience at the 2003 Audio-Visual Speech Processing workshop for their comments and valuable suggestions on a preliminary version of this work.
[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology—Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

# Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals

**Abstract.** Two experiments were conducted to examine the temporal limitations on the detection of asynchrony in auditory-visual (AV) signals. Each participant made asynchrony judgments about speech and nonspeech signals presented over an 800-ms range of AV onset asynchronies. Consistent with previous findings, all conditions revealed a wide window of several hundred milliseconds over which AV signals were judged to be synchronous. In addition, signals in which the visual component led the auditory component were more likely to be judged as synchronous than signals in which the auditory component led the visual component. In contrast with earlier reports (Dixon & Spitz, 1980; McGrath & Summerfield, 1985), the present results also demonstrated a similar AV synchrony window for speech and nonspeech signals, even when these signals were matched for duration. Visual phonetic characteristics of the speech signals, however, did influence the size and shape of the AV synchrony window. Finally, the onset of the relevant aspects of the stimulus, rather than the duration or offset, was most important for asynchrony judgments for both speech and nonspeech signals. Relationships with recent data on neural mechanisms of multisensory enhancement and convergence are discussed.

## Introduction

The temporal relationships among multimodal stimuli are critical in determining how and whether the stimuli will interact during perceptual processing (Meredith, 2002; Stein & Meredith, 1993). A fundamental question that arises in the study of multisensory perception is the window of time over which multisensory interactions can occur—that is, how close in time do stimuli in two or more sensory modalities need to be in order to interact in processing? One way of studying this issue is to measure the perception of multimodal events that are desynchronized to varying degrees. The present study investigated the detection of auditory-visual (AV) asynchrony for both speech and nonspeech signals.

Because earlier research has demonstrated interactions between audition and vision in human speech perception (Calvert et al., 1997; McGurk & MacDonald, 1976; Sumby & Pollack, 1954), we studied AV asynchrony detection in meaningful speech signals. We also compared detection of AV asynchronies in speech with semantically meaningless nonspeech signals in order to gain insights into the cognitive and neural processes that are involved in this type of crossmodal task.

### Literature Review

Several previous studies have examined AV synchrony detection for speech and nonspeech signals and explored the effects of AV asynchrony on McGurk illusions and speech intelligibility scores. In an early study, Dixon and Spitz (1980) presented adult subjects with films of two AV events, a man reading prose or a hammer hitting a peg. During the experiment, the auditory and visual tracks of the film became gradually out of sync. On half of the trials, the auditory track led the visual track, and on the other half the visual track led the auditory track. Subjects were asked to respond when they noticed that the auditory and visual tracks had become asynchronous. On average, subjects could not reliably detect AV asynchrony for the speech film over a range from when the auditory signal led by 131.1 ms (A131.1V ms) to when the visual signal led by 257.9 ms (V257.9A ms). The "intersensory synchrony window" (Lewkowicz, 1996) for the hammer hitting the nail was much smaller, between A74.8V and V187.5A ms.

McGrath and Summerfield (1985) measured the intelligibility of AV sentences presented in auditory white noise over several visual-leading conditions that ranged up to V160A ms. They reported that although there was a significant effect of delay overall, subjects did not show a significant decrement in speech intelligibility performance until delays of V160A ms were reached. Although the better lipreaders showed a linear trend of decreased performance with increased delay, this pattern was not observed in poor or average lipreaders.

In an additional experiment, McGrath and Summerfield (1985) assessed AV asynchrony detection thresholds using a Lissajou interference pattern as the visual stimulus and a 120-Hz rise/fall gated triangular wave as the auditory stimulus; these patterns were thought to simulate a pair of lips opening to articulate a CV syllable. In an adaptive testing procedure using an XAB forced-choice task, they estimated the 70.7% detection thresholds for auditory-leading and visual-leading asynchronies using these nonspeech patterns. On average, these thresholds were at A78.5V and V137.8A ms. McGrath and Summerfield found a moderate but nonsignificant association between higher nonspeech thresholds and lower lipreading scores. The average auditory-leading and visual-leading thresholds were at A78.5V and V137.8A ms. These thresholds are very close to those reported earlier for nonspeech stimuli by Dixon and Spitz (1980) (A74.8V and V187.5 ms), and Lewkowicz for an adult control group in an infant study (1996) (A65V and V112A ms). They are also similar to the results reported by Bushara, Grafman, and Hallett (2001) for a pilot behavioral task they used in a PET study (A56V and V114A ms).

In addition to Dixon and Spitz (1980), one other study, carried out by Grant, van Wassenhove, and Poeppel (2003), examined discrimination of AV asynchrony in sentences. In a two-interval forced-choice adaptive procedure, they found asynchrony discrimination thresholds of approximately A35V and V160 ms for unfiltered speech and A35V and V225A ms for bandpass-filtered speech. Dixon and Spitz (1980) reported thresholds between A131.1V and V257.9A ms for their speech materials. However, they used a different behavioral task and computed their thresholds with a different procedure.

In addition to the studies of AV synchrony detection and discrimination, several other studies have examined the perception of CV syllables as well as the McGurk illusion over a range of asynchrony levels. In general, AV temporal thresholds for the McGurk illusion are similar to those found in studies of AV synchrony detection and discrimination. For example, Massaro and Cohen (1993) measured AV perception of the syllables /ba/ and /ga/ at asynchronies of V200A, V100A, 0, A100V, and V200A ms. They reported that congruent AV presentations resulted in high accuracy of identification regardless of asynchrony level; however, with visual /ba/ and auditory /da/, /bda/ judgments increased and /da/ judgments decreased with increases in visual lead. In a second experiment, they reported crossmodal influences of the vowels /i/ and /u/ over the same range of asynchronies.

Massaro, Cohen, and Smeele (1996) assessed AV integration for CV syllables over 13 asynchrony levels, from A533V to V533A ms, and concluded that AV integration was not significantly disrupted for asynchronies of up to about A250V or V250A ms. Another study by Munhall, Gribble, Sacco, and Ward (1996), reported that subjects reliably displayed the McGurk effect for visual /aga/ paired with auditory /aba/ for asynchronies between A60V and V240A ms, but did not display the effect reliably when visual /igi/ was paired with auditory /aba/. Similarly, van Wassenhove, Grant, and Poeppel (2002) reported fusion /ta/ responses for visual /pa/ paired with auditory /ka/ between A50V and V200A ms.

Another group of studies explored the effects of AV asynchrony on speech intelligibility of sentences. Pandey, Kunov, and Abel (1986) measured sentence intelligibility when the auditory signal was delayed by 0, 60, 120, 180, 240, and 300 ms, and presented at SNRs of 0 and -10 dB. They reported that AV speech intelligibility was not significantly affected compared with auditory-alone presentation at asynchronies up to V240A ms for the 0 dB SNR, but they did observe a significant decrease in intelligibility by V180A ms for the -10 dB SNR. They also tested a group of normal-hearing experienced lipreaders at auditory delays of 0, 80, 160, and 240 ms, using a SNR of -5 dB, and found that performance declined significantly by V160A ms.

Grant and Seitz (1998) reported that the intelligibility of AV sentences presented in auditory noise was unaffected for hearing-impaired adults until auditory delays of around 200 ms. In another study of integration in asynchronous AV sentences, Grant and Greenberg (2001) found a relatively constant benefit for bandpass-filtered auditory sentences presented audiovisually over asynchronies ranging from A40V to V160-200A ms.

Taken together, most of the behavioral studies—whether measuring detection, discrimination, syllable identification, or sentence intelligibility—have estimated that the synchrony window for AV signals covers a range of several hundred milliseconds. In addition, the studies have reported that this synchrony window is larger when the visual signal leads than when the auditory signal leads. Finally, the studies report a great deal of individual variability in AV asynchrony thresholds among subjects (cf. individual subject data in McGrath & Summerfield (1985), p. 683, Table I, for an example).

Previous studies also suggest that the size of the AV synchrony window may be a function of the specific stimuli used. For example, the overall congruity of the auditory and visual events may influence tasks involving asynchronous AV stimuli (Munhall et al., 1996). In addition, other studies suggest that AV asynchrony may be easier to detect in nonspeech than in speech stimuli. However, speech and nonspeech thresholds have only been estimated using the same task in one study (Dixon & Spitz, 1980), and the speech and nonspeech events chosen in that study were not comparable. The speech event—a man reading prose—contained continuous visual and auditory signals, whereas the nonspeech event—a hammer hitting a nail—was discrete in nature and differed in overall duration from the speech event.

In the present set of experiments, we were interested in how the properties of the auditory and visual signals might affect the AV synchrony window. One of our goals was to compare detection of AV synchrony for speech and nonspeech signals using the same subjects and statistical procedures as well as the same levels of asynchrony for both sets of signals. The speech stimuli were isolated spoken English words rather than samples of connected speech so as to make the speech stimuli consist of single events that were more comparable to the nonspeech stimulus, which was a static circle paired with a simple tone.

In addition, we wanted to examine whether context effects related to the visual properties of the speech might influence judgments of synchrony. To accomplish this, we manipulated the visual characteristics of the speech stimuli in two ways. First, we used one speech condition in which participants viewed point-light displays of a talker's face rather than a fully illuminated face, and second, we used words that had been judged to have either high or low visual-only intelligibility based on results from previous experiments (Bergeson, Reynolds, & Pisoni, 2003; Lachs, 1999; Lachs & Pisoni, in press-a, in press-b). By directly comparing asynchrony judgments of speech and nonspeech, and speech with varying levels of visual information available, we hoped to elucidate some of the variables that influence the size and shape of the AV synchrony window in order to ultimately link our behavioral findings with plausible neural mechanisms.

## Experiment 1

In Experiment 1, we obtained synchrony judgments from each participant under three AV conditions: nonspeech signals (NS), full-face speech (FF), and point-light display speech (PLD). A wide range of AV asynchronies were studied, from A300V to V500A ms.

### Methods

#### Participants

Participants were 15 undergraduate students at Indiana University (5 male and 10 female, mean age of 19.33 years). Eight received partial credit in an introductory Psychology course for their participation; the other seven were paid $10 for their services. All participants were right-handed, monolingual native speakers of American English with no history of hearing or speech disorders and normal or corrected-to-normal vision at the time of testing. The experiment took approximately one hour to complete.

#### Stimuli

The experimental design consisted of three AV conditions: full-face video (FF), point-light display video (PLD), and a nonspeech condition (NS). The NS stimuli were modeled after those used in a PET study by Bushara et al. (2001) that investigated the neural correlates of AV asynchrony processing for several asynchrony levels. The present study used a 4-cm diameter red circle paired with a 2000-Hz tone. As in the earlier Bushara et al. study, both the visual and auditory stimuli were 100 ms in duration.

For the FF condition, 10 familiar English words were chosen from the Hoosier Audiovisual Multitalker Database (Lachs & Hernandez, 1998; Sheffert, Lachs, & Hernandez, 1996), which contains digitized AV movies consisting of single talkers speaking isolated monosyllabic words. The most intelligible of the eight talkers in the database was determined in a previous study, and auditory-only, visual-only, and audiovisual intelligibility data had been collected for all her utterances (Lachs, 1999; Lachs & Pisoni, in press-a, in press-b). In this study, all 10 FF words were spoken by this talker.

All of the FF words used for the present experiment had 100% speech intelligibility scores for both auditory-alone and AV presentation. In order to examine the possible effects of visual-only intelligibility on judgments of AV synchrony, five of the words chosen had high visual intelligibility (VI) and five had low VI. The low VI words all had 0% correct whole-word visual-only intelligibility; these words were back, give, pail, theme, and voice. The high VI words were doubt (10%), fall (80%), knot (40%), loan (30%), and reed (50%). The high VI words had the highest visual-only speech intelligibility scores among the words with 100% intelligibility scores for auditory-alone and AV presentation, and all of the high VI words had higher whole-word visual-only intelligibility scores than the talker's whole-word visual-only intelligibility average of 4.4% (standard error = 1.2%).

The PLD condition also used 10 words, spoken by the most intelligible female talker from a previously recorded audiovisual database of isolated single-syllable English words (Lachs, 2002; Lachs & Pisoni, in press-c). For the PLD movies in this database, the talker had 30 glow-in-the-dark dots glued to the lower half of her face, including her cheeks, jaw, chin, lips, upper and lower teeth, and tongue tip (see (Rosenblum & Saldaña, 1996). The video recordings were made with a black background so that only the movement of the green glow-in-the-dark dots was visible. Whole-word visual-only speech intelligibility was close to 0% correct for all PLD words. However, viseme-confusability matrices for the PLD movies obtained from the most intelligible talker (Bergeson et al., 2003) were used to choose five words that were predicted to have high VI and five others that were predicted to have low VI. The high VI words were boat, site, hope, mouse, and tile, and the low VI words were cod, gain, guide, reach, and thick.

The experimental stimuli used in this study were created using Final Cut Pro 3 (copyright 2003, Apple Computer, Inc.). In all cases, the visual and auditory stimuli were combined beforehand into precompiled movies rather than being assembled "on the fly" during the experiment by the computer. For the asynchronous speech stimuli, the portions of the audio and video tracks that did not overlap with each other were edited from the stimulus movie. (The removed portions did not contain any speech sounds or active articulatory movements.) This was done so that the participants would be unable to rely on any global temporal cues such as the audio track coming on while the screen was blank to determine if the movie was synchronous. Instead, all participants had to make their judgments about synchrony based on whether the presented information was temporally matched across the auditory and visual modalities.

Previous research on AV asynchrony detection (Dixon & Spitz, 1980; Lewkowicz, 1996; Massaro & Cohen, 1993; Massaro et al., 1996; McGrath & Summerfield, 1985; Pandey et al., 1986) and pilot studies in our lab indicated that most normal-hearing young adult subjects should be able to judge AV stimuli as asynchronous with close to 100% accuracy when the auditory signal leads the visual signal by 300 ms (A300V ms) or less and when the visual signal leads the auditory signal by 500 ms (V500A ms) or less. The experimental stimuli used in this study covered this wide range of asynchronies. Because the videos used were recorded at a rate of 30 frames per second, each successive stimulus could differ by 33.33 ms. This resulted in 25 asynchrony levels covering a range of 800 ms, from A300V to V500A. Nine stimuli had auditory leads, one was synchronous, and 15 had visual leads.

### Procedure

The visual stimuli were presented on an Apple Macintosh G4 computer. Auditory stimuli were presented over Beyer Dynamic DT headphones at 70 dB SPL. PsyScope version 1.5.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) was used for stimulus presentation and response collection. All participants were tested in each of the three conditions, NS, FF, and PLD. The conditions were blocked and were always presented in the order NS, FF, and PLD.

The stimuli were presented in a single-interval asynchrony judgment task. On each trial, the participants were asked to judge whether the AV stimulus was synchronous or asynchronous ("in sync" or "not in sync") and were encouraged to respond as quickly and as accurately as possible. Participants were instructed to press one button on a response box if the stimuli were synchronous and another if they were asynchronous. Response hand was counterbalanced across participants but kept constant for each participant on all three conditions of the experiment so as to minimize confusion about the instructions. Before beginning each condition, the participants received instructions and were presented with examples of synchronous and asynchronous movies.

Each of the three conditions consisted of 250 randomized trials, 10 for each of the 25 asynchrony levels. In the NS condition, all trials used the same visual and auditory stimuli, the red circle and the 3000-Hz tone described above. In the FF and PLD conditions, each of the 10 words was presented once at each asynchrony level. At the onset of each trial, a fixation mark ("+") flashed on the computer screen for 200 ms and was followed by 300 ms of blank screen before the test stimulus was presented. The subject's response cued the onset of the next trial.

## Results

Throughout this report, we will refer to synchronous AV stimuli as the 0 condition, for 0-ms delay/lead. Because our figures represent auditory leads to the left side of 0 on the abscissa and visual leads to the right, "lower" will indicate further toward the auditory-leading side of the figure, and "higher" will indicate further toward the visual-leading side of the figure. Similarly, negative numbers will refer to the auditory signal leading the visual signal in time, and positive numbers to the visual signal leading the auditory signal.

The proportion of synchronous responses at each level of asynchrony was determined for each participant. The average proportions are plotted separately for the FF, PLD and NS conditions in Figure 1.

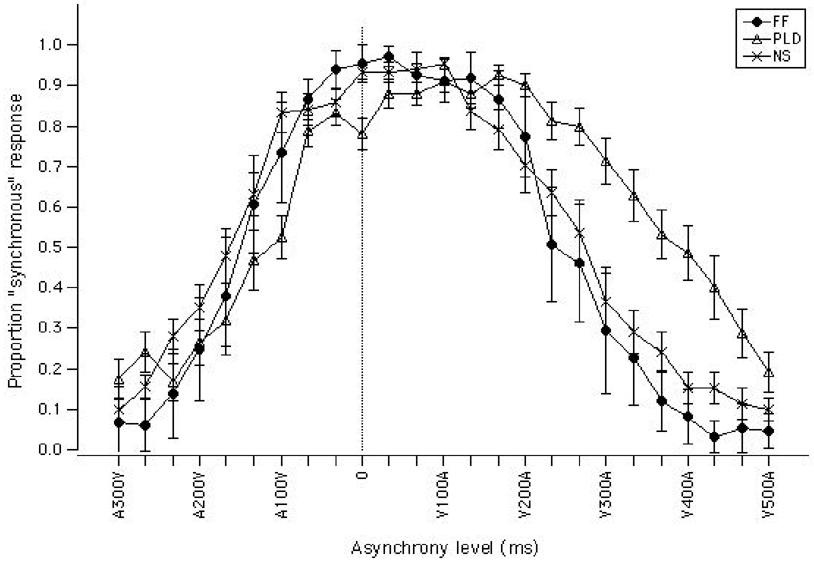


**Figure 1.** Average "in sync" response for all participants in Experiment 1, in FF, NS, and PLD conditions. The dotted vertical line is at 0-ms asynchrony.

In looking at the figure, two major features are apparent. First, the average range of asynchronies identified as synchronous was quite large, on the order of several hundred milliseconds. Second, this range was not centered at 0 ms, the synchronous condition, but was shifted to the right and centered on the visual-leading side of the continuum.

To quantify these findings, we fit each condition from the individual participants' data with a Gaussian curve using Igor Pro 4.05A Carbon (copyright 1988-2002, WaveMetrics, Inc.). This resulted in a total of seven curves for each participant: FF, PLD, NS, and high and low visual-only intelligibility for FF and PLD. From these curves, we obtained estimates of the mean of the AV synchrony window as well as the low and high thresholds for asynchrony detection. For each subject we estimated the mean point of the synchrony window (MPS) as the mean of the Gaussian curve fit to the subject's data. We operationalized the width of the AV synchrony window as the range of asynchronies over which subjects responded that the signals were synchronous more than half the time. For an estimate of this width, we calculated the full width at half maximum (FWHM) of the Gaussian curve. The auditory-leading threshold for synchrony was calculated as the MPS minus the half width at half maximum, and the visual-leading threshold for synchrony was calculated as the MPS plus the half width at half maximum.

Table 1 presents a summary of the curve estimates obtained from a fit of the average response data weighted by the standard error. All statistical tests used estimates from fitting curves to individual subject data. Tukey HSD tests were used for all post-hoc analyses, and the familywise error rate was set to α=.05.

| Condition | MPS | FWHM | A-leading | V-leading |
|---|---|---|---|---|
| FF | 47.405 | 388.39 | -146.79 | 241.60 |
| high VI | 55.33 | 397.68 | -143.51 | 254.17 |
| low VI | 47.749 | 373.42 | -138.96 | 234.46 |
| PLD | 121.59 | 501.36 | -129.09 | 372.27 |
| high VI | 116.42 | 483.63 | -125.40 | 358.24 |
| low VI | 122.53 | 517.35 | -136.14 | 381.20 |
| NS | 52.485 | 421.91 | -158.47 | 263.44 |

**Table 1.** All numbers are in milliseconds. Negative numbers indicate that the auditory signal led the visual signal. VI = visual intelligibility; MPS = mean point of synchrony; FWHM = full width at half maximum; A-leading = auditory-leading threshold; V-leading = visual-leading threshold.

### Mean Point of Synchrony (MPS)

The MPS was significantly larger than 0 in all three conditions of the experiment (FF: *t* (14) = 7.119; PLD: *t* (14) = 15.656; NS: *t* (14) = 4.582; all *p*'s < .001). Most of the participants had an MPS greater than 0 on all three conditions. One participant had an estimated MPS of A13V ms in the FF condition, and two had estimated MPSs of A20V and A24V ms in the NS condition.

A one-way repeated measures ANOVA revealed a significant effect of condition (NS, FF, PLD) on MPS (*F* (2, 28) = 31.326, *p* < .001). Post-hoc Tukey tests indicated that the MPS did not differ significantly for the NS and FF conditions. However, the MPS in the PLD condition was significantly larger than the MPS for either the NS or the FF conditions.

A two-way ANOVA on condition (FF, PLD) and visual-only intelligibility (high, low) revealed no significant overall effect of VI (*F* (1, 14) < 1). However, a significant interaction of Condition x VI (*F* (1,14) = 14.056; *p* < .05) was found. Post-hoc Tukey tests revealed that both high and low VI words in the FF condition had a lower MPS than high or low VI words in the PLD condition. Also, in the FF condition, the high VI words had a significantly higher average MPS than the low VI words. The high VI words had an MPS that was on average 19.12 ms of visual lead higher than the low VI words. Although this difference was small, it was statistically significant and highly consistent across participants. Of the 15 participants, 12 showed the VI effect in the FF condition. Three participants had a lower MPS for the high VI FF words; in those cases the MPSs were 5.54 ms, 9.66 ms, and 12.67 ms smaller for high than for low VI words. The difference between the high and low VI words in the PLD condition was in the opposite direction although it did not reach significance.

### AV Synchrony Window

A one-way repeated measures ANOVA revealed a significant effect of condition on the size of the AV synchrony window (*F* (2, 28) = 4.780, *p* < .05). Post-hoc Tukey HSD tests showed that the synchrony window was larger in the PLD condition than in either the FF or NS conditions. A two-way repeated measures ANOVA on condition and VI revealed no additional significant effect of VI (*F* (1, 14) < 1) and no significant interaction of Condition x VI (*F* (1, 14) < 1).

### Auditory-leading Thresholds

A one-way repeated measures ANOVA revealed no significant effect of condition on the auditory-leading threshold (*F* (2, 28) = 1.850, *p* > .05). Likewise, a two-way ANOVA on condition and VI revealed no significant effects (condition: *F* (1, 14) = 1.233, *p* > .05; VI: *F*(1, 14) < 1; Condition x VI: *F* (1, 14) < 1).

### Visual-leading Threshold

The visual-leading threshold was significantly different across conditions (*F* (2, 28) = 17.550, *p* < .001). Post-hoc Tukey HSD analyses revealed that the PLD condition had a significantly higher visual-leading threshold than either the FF or the NS conditions. The FF and NS conditions did not differ in visual-leading threshold. A two-way ANOVA on condition and VI revealed no additional significant effect of VI (*F* (1, 14) < 1) and no significant interaction of Condition x Intelligibility (*F* (1, 14) = 3.837, *p* > .05).

## Discussion

The present findings are consistent with previous studies and indicate that participants judged AV signals as subjectively synchronous for AV asynchronies that ranged over a window of several hundred milliseconds. In addition, participants judged larger asynchrony levels as subjectively synchronous with visual-leading stimuli than with auditory-leading stimuli. This AV processing asymmetry has also been reported in electrophysiological studies (King & Palmer, 1985; Meredith, 2002; Meredith, Nemitz, & Stein, 1987; B. Stein & Meredith, 1993); in behavioral studies using simple AV asynchronous stimuli (Dixon & Spitz, 1980; Lewald, Ehrenstein, & Guski, 2001; Lewkowicz, 1996); and in behavioral tasks involving AV speech (Dixon & Spitz, 1980; Grant & Greenberg, 2001; Grant & Seitz, 1998; Grant et al., 2003; McGrath & Summerfield, 1985; Pandey et al., 1986).

In contrast with a previous report by Dixon and Spitz (1980), however, the thresholds and means of the AV synchrony window were comparable for nonspeech (NS; average window: A159V to V263A ms; MPS = V52A ms) and full-face (FF) speech signals (average window: A147V to V242A ms; MPS = V47A ms). None of the differences obtained in this study between the NS and FF conditions were significant. However, we did find significant effects due to the visual characteristics of the speech stimuli. The size, MPS, and visual-leading thresholds in the PLD condition were all significantly higher than in the FF or NS conditions. Also, the MPS in the FF high VI condition was significantly higher than the MPS in the FF low VI condition; the effect of VI was small but highly consistent across subjects.

One issue raised by the results obtained in Experiment 1 is whether the duration chosen for the nonspeech signals, 100 ms, could have influenced the characteristics of the AV synchrony window for simple AV signals, and potentially exaggerated the similarity between the windows for speech and nonspeech. For instance, it is possible that 100 ms may be a "special" number for AV interactions. Neurophysiological studies have indicated that multisensory enhancement may not occur if signals from multiple sensory modalities do not occur within 100 ms of each other (King & Palmer, 1985). Similarly, a recent behavioral study in humans has reported multisensory interactions only for AV asynchronies of up to 100 ms (Shams, Kamitani, & Shimojo, 2002). If 100 ms is a "special" duration, then the results for the nonspeech condition in Experiment 1 could be more similar to the results for the FF condition than would be expected if the stimuli used in the NS condition had been shorter or longer in duration.

More generally, Meredith et al. (1987) reported that multisensory neurons in the superior colliculus of the cat show multisensory enhancement when discharge trains from the unimodal stimuli overlap. This multisensory enhancement is greatest during overlap of the peak unimodal discharge trains. In our behavioral task, the AV synchrony window, which we took as a behavioral correlate of multisensory enhancement, could be similarly affected by the overlap of the peak neural response to a stimulus. Using data for NS stimuli of only one duration, it is difficult to assess whether the relevant stimulus information comes from the stimulus onset, offset, or overall duration, or some combination of these. To explore this issue further, we conducted a second experiment to investigate the effects of different durations of NS stimuli on the characteristics of the AV synchrony window.

## Experiment 2

In Experiment 2, we examined the effect of nonspeech signal duration on AV synchrony judgments. We also used the same FF condition as in Experiment 1 as a baseline measure. If signal duration is an important cue in synchrony detection, then subjects might be better at detecting asynchronies in stimuli with shorter durations than in stimuli with longer durations. This might be the case because auditory and visual stimuli that are longer in duration have longer durations of overlap than shorter-duration AV stimuli at the same asynchrony level. For example, suppose we have a stimulus in which the auditory signal leads the visual signal by 300 ms (A300V). If the auditory and visual signals are each 33 ms in duration, then there is a "gap" of 267 ms between the offset of the auditory signal and the onset of the visual signal. However, if the auditory and visual signals are both 500 ms in duration, then the offset of the auditory signal will not occur until 200 ms after the onset of the visual signal. Note that in this example, stimulus offset also varies with stimulus duration. Thus, if the onset asynchrony was more important than the duration and/or offset of the signal for the size of the AV synchrony window, we would not expect to see any significant differences in the asynchrony judgments for nonspeech stimuli of different durations.

### Methods

#### Participants

Participants were 23 undergraduate students at Indiana University (6 male and 17 female, mean age of 19.39 years). All were recruited from the Indiana University subject pool and were paid $10 for their services. All participants were right-handed, monolingual native speakers of American English with no history of hearing or speech disorders and normal or corrected-to-normal vision. The experiment took approximately one hour to complete.

#### Stimuli

Stimuli were created with the same methodology and auditory-visual onset asynchronies used in Experiment 1. The duration of the signals was manipulated so that in the first condition the auditory and visual signals were both 33 ms (NS33); in the second they were 100 ms (as in Experiment 1; here, NS100); and in the third they were 500 ms (NS500). The 500-ms condition was used because the AV words in the FF condition were on average 500-ms long.

The stimuli used in the FF and NS100 conditions were identical to those used in Experiment 1. The NS33 and NS500 conditions used the same red circle and 3000-Hz tone as the NS100 condition, but differed in the duration of the auditory and visual signals, which were both 33 ms in the NS33 condition and both 500 ms in the NS500 condition.

### Procedure

The procedure was identical to that described in Experiment 1, with the following exceptions. All four conditions (FF, NS33, NS100, and NS500) were tested, with 25 asynchrony levels x 10 trials per level = 250 trials per condition. The three NS blocks were tested before the FF block, but the order of the three NS blocks was counterbalanced across participants. Response hand was also counterbalanced across participants.

## Results

Average response data for the FF, NS33, NS100, and NS500 conditions are displayed in Figure 2. As in Experiment 1, individual subject data were fit with Gaussian curves. The mean of the curve was taken as the mean point of the AV synchrony window (MPS), and the auditory- and visual-leading thresholds were the low and high endpoints of the FWHM. Again, all statistical analyses were performed on individual subject data. Table 2 contains a summary of the curve estimates for the average subject data weighted by the standard error.

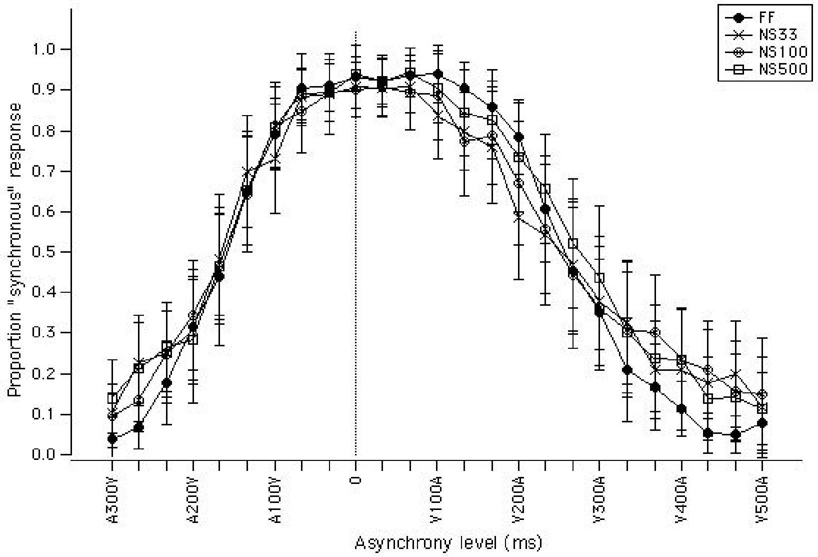


**Figure 2.** Average "in sync" response for all participants in Experiment 2, in FF, NS33, NS100, and NS500 conditions. The dotted vertical line is at 0-ms asynchrony.

| Condition | MPS | FWHM | A-leading | V-leading |
|---|---|---|---|---|
| FF | 53.8748 | 394.42 | -143.34 | 251.09 |
| high VI | 62.8317 | 385.04 | -129.69 | 255.35 |
| low VI | 40.192 | 407.50 | -163.56 | 243.94 |
| NS33 | 39.867 | 410.57 | -165.42 | 245.15 |
| NS100 | 52.757 | 425.02 | -159.75 | 265.27 |
| NS500 | 51.1082 | 448.20 | -172.99 | 275.21 |

**Table 2.** All numbers are in milliseconds. Negative numbers indicate that the auditory signal led the visual signal. VI = visual intelligibility; MPS = mean point of synchrony; FWHM = full width at half maximum; A-leading = auditory-leading threshold; V-leading = visual-leading threshold.

### Mean Point of Synchrony (MPS)

The MPS was significantly larger than zero in all four conditions (FF: *t* (22) = 8.576, *p* < .05; NS33: *t* (22) = 4.517, *p* < .05; NS100: *t* (22) = 2.683, *p* < .05; NS500: *t* (22) = 4.973, *p* < .05). Again, the majority of the participants (15 out of 23) showed an MPS greater than 0 for all conditions. Of the remaining eight participants, four had an MPS greater than 0 in one of the NS conditions only, three had an MPS greater than 0 in two of the NS conditions only, and one had an MPS greater than 0 in the FF condition and two of the NS conditions.

The MPS did not differ significantly across the conditions overall (*F* (3, 66) = 1.049, *p* > .05). However, as observed in Experiment 1, the high VI condition had a significantly higher MPS than the low VI condition (*t* (22) = 4.428, *p* < .05), with an average difference of 21.78 ms. The VI effect on the MPS was found in 20 of the 23 subjects, with the remaining three having low VI MPSs that were 1.33 ms, 3.63 ms, and 39.09 ms higher than their high VI MPSs. This VI effect on the MPS replicated the findings reported in Experiment 1.

### AV Synchrony Window

A one-way ANOVA revealed no significant effect of condition on the size of the AV synchrony window (*F* (3, 66) = 1.785, *p* > .05). The size of the AV synchrony window was not significantly different for high versus low VI FF words (*t* (22) = .730, *p* > .05).

### Auditory-leading Thresholds

The auditory-leading threshold did not differ significantly overall across conditions (*F* (3, 66) = 1.171, *p* > .05). The auditory-leading thresholds for the high and low VI FF words were not significantly different (*t*(22) = .412, *p* > .05).

### Visual-leading Thresholds

A one-way ANOVA revealed no significant effect of condition on the visual-leading threshold (*F* (3, 66) = 1.881, *p* > .05). In addition, high and low VI FF words did not differ significantly for visual-leading threshold (*t* (22) = 1.529, *p* > .05).

### Discussion

The results of Experiment 2 revealed no significant effects of the duration of the NS signals on the AV synchrony window for AV stimuli. In addition, the NS conditions did not differ overall from the FF condition, replicating the results found earlier in Experiment 1. Finally, the MPS for the FF high VI words was significantly higher than the MPS for the FF low VI words by about 20 ms, replicating the VI finding from Experiment 1.

The failure to find any effect of signal duration of the NS stimuli on AV synchrony detection in Experiment 2 suggests that the detection of asynchrony relies on processes related to stimulus onset rather than stimulus offset or duration. Of course, the case may differ for very short or very long stimuli; this is an empirical question that could be addressed in future research. Another explanation of our results is that the subjects were attending only to stimulus onset and ignoring other stimulus properties. Although this account cannot be completely ruled out based on our current data, the subjects were explicitly instructed to respond that the stimuli were synchronous only if they overlapped exactly. Further investigations, manipulating subjects' attentional strategies or response criteria, are needed to resolve this issue more definitively.

To begin to address these issues through a converging approach, we examined the individual words from the FF condition, in which significant effects of VI appeared in both Experiment 1 and Experiment 2. We reasoned that a more detailed analysis of the data for individual words might allow us to pinpoint what features of the particular utterances the subjects were relying on in making their asynchrony judgments. The speaker's face was visible throughout each speech movie, and we edited the movies so that global cues to asynchrony could not be used effectively (see Experiment 1, Methods). As a consequence, the physical onset of the auditory and visual stimuli would be a less reliable cue to asynchrony in the FF condition than in the NS conditions. However, word-internal articulatory events might have some influence on when a particular word was judged to be synchronous.

## Word Item Analysis

The FF condition data obtained from 50 participants (15 from Experiment 1, 23 from Experiment 2, and 12 additional participants) were analyzed separately by word. Figure 3 shows the range of asynchronies over which 50% or more of participants responded "synchronous" for each word. Table 3 shows the VI of the word, size of the window over which 50% or more of participants responded "synchronous" for each word, the auditory-leading and visual-leading limits of that window, and the auditory duration of each word. The number of video frames presented per word varied according to synchrony level as described in the "Methods" section of Experiment 1, so video duration is not included in the table.

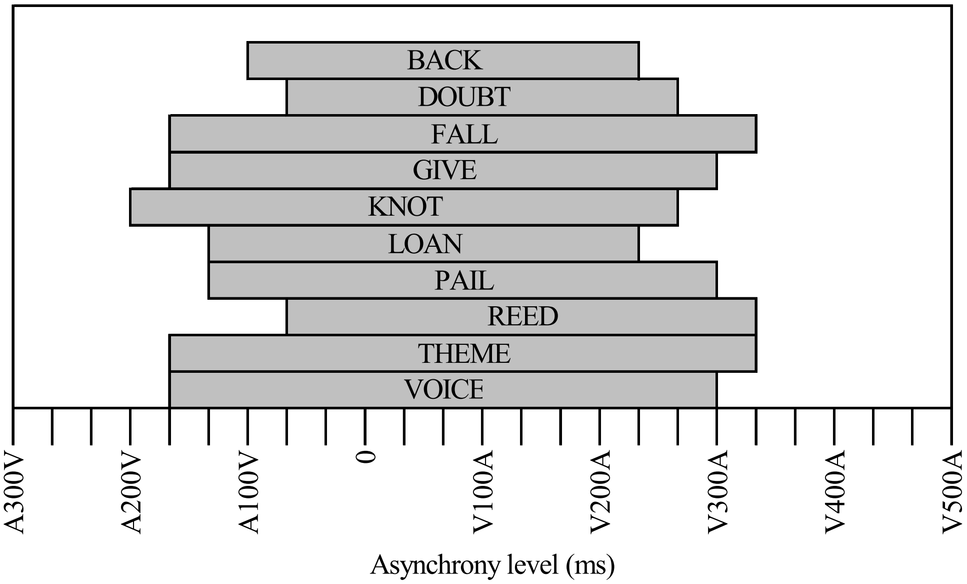


**Figure 3.** Asynchronies for which more than 50% (≥25/50) participants responded "in sync," broken down by word.

| Word | VI | Window | AV limit | VA limit | Dur(A) |
|---|---|---|---|---|---|
| BACK | low | 300 | A100V | V200A | 397 |
| DOUBT | high | 300 | A67V | V233A | 467 |
| LOAN | high | 333 | A133V | V200A | 475 |
| REED | high | 367 | A67V | V300A | 522 |
| PAIL | low | 400 | A133V | V267A | 473 |
| GIVE | low | 433 | A167V | V267A | 388 |
| KNOT | high | 433 | A200V | V233A | 580 |
| VOICE | low | 433 | A167V | V267A | 647 |
| FALL | high | 500 | A167V | V300A | 490 |
| THEME | low | 500 | A167V | V300A | 656 |

**Table 3.** All numbers are in milliseconds. VI = visual intelligibility classification; Window = size of window including asynchrony levels for which 50% or more of participants responded "in sync"; AV limit = auditory-leading limit of window; VA limit = visual-leading limit of window; Dur(A) = duration of the auditory word.

The words in Table 3 are listed in order from those with the smallest asynchrony window to those with the largest window. Smaller windows were taken to indicate greater overall accuracy in judging when the words were synchronous. Table 3 indicates that the overall statistical difference observed in MPS between words defined as "high" or "low" VI based on whole-word scores may not provide the best description of the word item data. Likewise, the auditory duration does not seem to completely explain accuracy. However, the specific phonetic characteristics of the word, particularly the articulatory properties of the initial consonant, do appear to influence accuracy. The two words with the smallest synchrony windows, *back* and *doubt*, both begin with voiced stops articulated in the front of the mouth. The words with the largest synchrony windows, *fall* and *theme*, both begin with voiceless fricatives. The mid-range words begin with liquids (*loan* and *reed*), a voiceless bilabial stop (*pail*), a voiced velar stop (*give*), a nasal (*knot*), and a voiced fricative (*voice*).

The vowels and final consonants of the individual words appear to have a less regular relationship to AV synchrony judgments. For example, the same vowels are found in words with small and large windows (e.g., *reed* and *theme*) and short and long vowels are found at both ends of the spectrum as well. Also, *doubt* and *knot* have voiceless alveolar stops in final position, and *loan* and *theme* both have nasals in final position.

Taken together, this pattern of results suggests that the articulatory properties of the initial phonetic segment influence the accuracy with which words can be identified as synchronous or not. Stops that are voiced and articulated at the front of the mouth were identified most readily as synchronous. Segments articulated at the front of the mouth are easier to see than those at the back (e.g., velar stops; Summerfield, 1987), and stops also provide a discrete and relatively well-defined auditory and visual boundary. The shorter voice-onset time in voiced stops may also be linked more closely to the relevant aspects of the visual articulation than the longer voice-onset time in voiceless stops. Of course, further research will be necessary to determine whether the initial consonant is of primary importance in AV synchrony detection for a larger set of words that are specifically controlled for phonological contrasts, but the results of this initial analysis suggest that the phonetic properties of the initial consonant affect AV asynchrony judgments in this task.

## General Discussion

Although the results of Experiments 1 and 2 are consistent with previous findings reported in the literature, they also provide several new insights into AV asynchrony detection. Both experiments demonstrated a similar AV synchrony window for speech and nonspeech sounds, in contrast to previous reports that suggested a larger window for speech sounds (Dixon & Spitz, 1980; McGrath & Summerfield, 1985). On the other hand, the PLD stimuli resulted in an AV synchrony window that was larger on the visual-leading side than the FF or NS windows. Finally, the onset of the relevant aspects of the stimulus, rather than the duration or offset of the stimulus, seemed to be important for judgments about asynchrony in both speech and nonspeech stimuli.

### The Size and Shape of the AV Synchrony Window

The width of the AV synchrony window may reflect general information processing constraints (Munhall et al., 1996). For example, Guski and Troje (2003) have argued that events are linked at a perceptual level when they occur within around 200 ms of each other; they point out that the window for visual iconic processing is generally held to be around 250 to 300 ms and that the window for auditory echoic memory is around 250 ms. A multisensory interaction window that is several hundred milliseconds long is also consistent with the estimates of the temporal window for multisensory enhancement and/or depression reported in electrophysiological studies in animals (King & Palmer, 1985; Meredith, 2002; Meredith et al., 1987; Stein & Meredith, 1993).

Several possible explanations have been proposed for the auditory-visual asymmetry observed in the intersensory temporal synchrony window. Some researchers have suggested that visual-leading asynchronies are tolerated more easily because they reflect long-term perceptual learning (Dixon & Spitz, 1980; McGrath & Summerfield, 1985). Specifically, perceivers might be able to more easily accommodate multimodal events in which the visual component begins before the auditory component because this type of event is common in their experience of the natural world (e.g., lightning preceding thunder). By contrast, events in which the auditory component comes before the visual component would not be expected based on prior experience and learning.

Another proposal is that the first modality to occur determines the timecourse of processing for asynchronous AV speech (Grant & Greenberg, 2001; Grant et al., 2003). This explanation is based on the hypothesis that visual speech cues from jaw and lip movements provide syllabic information relevant to the perception of place of articulation, whereas auditory speech cues provide information about voicing and manner of articulation (Summerfield, 1987). Visual syllabic information on the order of 200 to 250 ms is taken to be complementary to auditory information, which is hypothesized to be more important for phonological analysis and is conveyed at a faster rate of around 40 to 120 ms or less (Grant & Greenberg, 2001; Grant et al., 2003; van Wassenhove, Grant, & Poeppel, 2003). This explanation of AV interactions predicts a smaller integration window when the auditory speech signal leads than when the visual signal leads. However, this account is not consistent with the present results because we did not obtain any significant differences in asynchrony judgments between full-face speech and simple nonspeech signals, for which phonemic or syllabic considerations do not apply. In addition, recent findings indicate that phonetic information can be used in visual-only tasks, suggesting that meaningful visual information can be conveyed in speech below the syllabic level (Bernstein, Demorest, & Tucker, 2000; Lachs, 1999; Lachs & Pisoni, in press-a, in press-b; Mattys, Bernstein, & Auer, 2002).

Finally, another explanation for the auditory-visual asymmetry involves the timing of auditory and visual signals in the nervous system (Lewald et al., 2001). As Lewald and his colleagues (2001) and Schroeder and Foxe (2002) have noted, there are both physical and physiological differences in the transmission of light and sound. Light travels faster than sound in air, but stimulus transduction takes longer in the retina than in the cochlea (Lewald et al., 2001). Also, in discussing results from an AV asynchrony detection task in infants, Lewkowicz (1996) pointed out that the latency of the earliest evoked potentials are about 30-40 ms faster for auditory than visual signals.

In addition to these considerations, arrival time of auditory and visual signals to different subcortical and cortical regions differs as a function of which region is under consideration and the physical distance of the observer from the stimulus (Lewald & Guski, 2003; Schroeder & Foxe, 2002). For example, at distances of a meter or less, comparable to the observer's distance from the visual stimulus in our experiments (the participants were wearing headphones, so the distance of the auditory stimulus was essentially 0), auditory signals would be predicted to arrive around 40 ms before the visual signals in auditory association cortex. However, at the same distance, superior temporal polysensory areas would receive auditory and visual inputs at approximately the same latency of around 23-25 ms (Schroeder & Foxe, 2002). At further distances of around 40 feet, AV inputs to auditory association cortex would become synchronous while superior temporal polysensory areas would receive auditory and visual inputs asynchronously.

In a recent behavioral study, Sugita and Suzuki (2003) reported that the estimated time of arrival of an auditory stimulus increases with the viewing distance to the visual stimulus. In their study, stimuli were judged as synchronous at a distance of 1 m when the auditory stimulus lagged by 5 ms, and at a distance of 20 m when the auditory stimulus lagged by 50 ms. The authors suggested that up until about 10 m of viewing distance, this increase was consistent with the brain's compensating for the slower velocity of sound. Such results point to the need for further experiments to clarify the role of viewing distance in auditory-visual interactions and perception of AV synchrony.

Although it is unclear at this time whether the relevant differences in auditory and visual processing times lie in transduction or occur later in neural information processing, the present findings and those reviewed from other studies suggest that at least under these presentation conditions auditory information is processed more quickly by the nervous system than visual information. The average MPS was around V50A ms for all but the PLD conditions, suggesting that visual leads of around 50 ms were most likely to be perceived as synchronous and that auditory stimuli may have been processed about 50 ms faster than visual stimuli. If we assume that the AV synchrony window is centered around V50A ms and extend the window 200 ms in either direction, as suggested by general information processing constraint explanations reviewed earlier (Guski & Troje, 2003; Munhall et al., 1996), we obtain a predicted AV synchrony window that extends from A150V to V250A ms. This is quite similar to the results we obtained for the FF and all the NS conditions in the present series of experiments.

## Relationship to Neural Data

The average MPSs for the FF and the three NS conditions ranged from approximately V40A to V60A ms, indicating that the likelihood of a synchronous judgment was maximal when visual input led auditory input by approximately that time interval. Other recent behavioral studies of AV processing indicated that optimal performance on several perceptual tasks occurred with auditory delays of between 50 and 100 ms (Guski & Troje, 2003; Lewald & Guski, 2003). In general, these behavioral results imply that AV interactions relevant for the perception of synchrony occur early in neural processing.

Several neuroimaging studies have reported modulated activity in primary auditory and/or visual cortex during AV perception. For speech stimuli presented visual-only, enhanced auditory cortex activity has been reported in MEG (Sams et al., 1991) and fMRI studies (Calvert et al., 1997; Calvert & Campbell, 2003; MacSweeney et al., 2000) (but see Bernstein et al., 2002, for an exception). An EEG independent-components analysis of AV speech perception also suggested that enhanced auditory cortex activity was an important locus for the visual enhancement effect obtained for AV over auditory-alone presentation (Callan, Callan, Kroos, & Vatikiotis-Bateson, 2001).

Interestingly, Schroeder and Foxe (2002) reported that in the macaque, visual feedback reaches posterior auditory cortex at 50 ms poststimulus, while auditory feedforward input arrives at around 11 ms poststimulus. They suggest two possible origins for the visual signal—superior temporal polysensory areas and prefrontal cortex. In either case, they hypothesized that the earlier arriving auditory input could modulate responsiveness to the later arriving visual input. Based on the estimates of Schroeder and Foxe, visual and auditory inputs could be expected to arrive simultaneously at posterior auditory cortex in the macaque if the visual stimulus occurs approximately 40 ms before the auditory stimulus.

Recent data from event-related potential studies in humans suggest that the earliest audiovisual interaction in cortex can be detected over posterior cortex 40 to 50 ms after the presentation of a synchronous audiovisual stimulus (Giard & Peronnet, 1999; Molholm et al., 2002; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002; van Wassenhove et al., 2003). However, because the ERP methodology has poor spatial resolution, the neural substrates of these early AV components remain unclear, and suggestions vary as to whether they are due to enhanced neural activity in auditory or in visual cortex. One recent study of nonspeech stimuli by Teder-Sälejärvi et al. (2002) concluded that the early AV interaction was due to subjects' anticipation of stimulus presentation. Some researchers, especially those who used nonspeech stimuli, have suggested that the early interaction might be due to visual cortex activation (Fort, Delpuech, Pernier, & Giard, 2002; Giard & Peronnet, 1999), either from recently discovered feedforward projections from auditory cortex to early visual cortex or from auditory feedback from multisensory areas to early visual cortex (Fort et al., 2002; Molholm et al., 2002).

Researchers who used speech stimuli have argued instead for an early visual modulation of auditory processing (Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000; van Wassenhove et al., 2003). Pourtois and colleagues (2000) found that presenting static expressive faces paired with voices resulted in modulation of auditory N1 by 90 to 130 ms poststimulus, the earliest significant interaction they observed. Interestingly, Giard and Peronnet (1999) reported an effect of sensory dominance of their participants on whether early enhancements occurred in visual or auditory cortex. Using nonspeech stimuli, they found that although visually dominant participants showed more enhanced early activity over auditory cortex, auditory-dominant participants showed more enhanced early activity over visual cortex. Further investigation is needed to clarify these issues.

Two recent imaging studies have specifically examined brain response to asynchronous AV stimuli and both have suggested subcortical rather than cortical involvement as the crucial factor. In a PET study, Bushara and colleagues (2001) measured detection of three auditory-leading and three visual-leading asynchronies using a circle paired with a tone. They reported that rCBF responses in the right insula increased with shorter asynchronies and that these responses were positively correlated with responses in the superior colliculus region of the posterior midbrain; the right posterior thalamus, precuneus, and prefrontal cortex; and the left insula. No correlations of right insular activity with superior temporal regions were observed.

In addition, an fMRI study by Olson, Gatenby, and Gore (2002) assessed the McGurk effect in synchronously and asynchronously presented speech, although only one asynchronous condition was used, in which the auditory signal was delayed by 1 second relative to the visual signal. They reported that although superior temporal regions were involved in both synchronous and asynchronous presentations, only the claustrum showed differential (increased) activity during the synchronous presentation condition. Olson et al. suggested that subcortical regions are more highly sensitive to timing of crossmodal stimuli than superior temporal regions. This hypothesis was also proposed in a review article by Calvert (2001), in which she suggested that the superior temporal sulcus is involved in stimulus identification and that the insula and superior colliculus are involved in stimulus timing. However, Calvert suggested that the left claustrum might be preferentially involved in crossmodal matching tasks. Because the Olson et al. study used McGurk stimuli that were incongruent across auditory and visual modalities, the mismatched nature of the stimuli could be responsible for the claustrum activity that was reported in their study.

Further neuroimaging research on auditory-visual asynchrony in both speech and nonspeech signals will be necessary to clarify whether the two types of signals show similar patterns of neural activation. In addition, future neuroimaging studies should be conducted with a wider range of AV asynchronies in order to examine more explicitly the neural responses to stimuli that fall both within and outside the AV synchrony window.

### Context Effects: PLDs

In Experiment 1, we found that the AV synchrony window for PLD stimuli was centered on about V120A and was about 50 ms wider on both the auditory-leading and the visual-leading sides than the FF and NS windows. The PLDs had the same auditory-leading threshold as the FF and NS conditions, but the visual-leading threshold was higher. There are several possible explanations for these results. One suggestion is that presenting the auditory signal first made matching to the unfamiliar PLD visual signal easier compared to presenting the PLD visual signal before the auditory signal. Recent findings by Lachs and Pisoni (in press-c) on crossmodal matching using isolated auditory words and PLDs do not support this idea. Participants in these studies were equally successful at matching auditory to visual and visual to auditory presentations. However, it is possible that other familiarity or "top-down" processing effects may have played a role in the PLD results. In addition, the physical characteristics of the PLD visual signals, such as low luminance and dispersion across the screen, might affect visual processing of these stimuli. Preliminary work in our lab has begun to examine this issue in more detail.

### Context Effects on VI: Word Item Analysis

In both Experiments 1 and 2, the FF high VI words consistently had a higher MPS than the FF low VI words. However, the word-item analysis suggested that the VI results could be explained by participants' sensitivity to the phonetic properties of the initial consonants rather than vowels or final consonants. Initial voiced stops articulated near the front of the mouth were identified as asynchronous most easily, whereas voiceless fricatives were the most difficult. These findings are consistent with the results of Experiment 2, which showed that nonspeech signals of different durations produced similar asynchrony detection results. Taken together, the results suggest that the onset of the two signals rather than the duration or offset is the critical factor controlling the perception of AV synchrony.

The importance of phonetic information in tasks involving AV speech has been reported previously (Bernstein et al., 2000; Lachs, 1999; Lachs & Pisoni, in press-a, in press-b; Mattys et al., 2002; Smeele, Sittig, & van Heuven, 1992). Smeele and her colleagues (1992) reported that for bimodal Dutch nonsense CVC words presented asynchronously in noise, the initial consonant was identified significantly more accurately with the auditory signal leading; conversely, the vowel and final consonant were identified more accurately with visual leads. Further research is needed using a larger inventory of words explicitly controlled for phonological inventory to determine whether the detection of AV asynchrony in speech is also affected by the phonetic and articulatory properties of the vowel and final consonant segments or only by the initial consonant segment as the results of the present study suggest.

### Individual Differences in the AV Synchrony Window

The present study reports results on the perception of AV speech and nonspeech sounds in normal-hearing, typically developing adults. Additional findings have been reported suggesting that other atypical populations may have difficulties processing timing for unimodal and/or crossmodal stimuli. For example, children and adults with dyslexia have difficulties making judgments about the timing of crossmodal stimuli (Laasonen, Service, & Virsu, 2002; Laasonen, Tomma-Halme, Lahti-Nuutila, Service, & Virsu, 2000). It has been suggested that dyslexic individuals may have auditory, visual, and other sensory and motor deficits in processing transient stimuli that change quickly over time (J. Stein & Walsh, 1997).

In our own lab, we had an opportunity to examine the sensitivity to AV asynchrony in a postlingually deafened adult ("Mr. S") who received a cochlear implant after two years of deafness (Goh, Pisoni, Kirk, & Remez, 2001). At the time of testing, our patient had used his implant for nine years. Mr. S has performed visual-only lipreading tasks at a consistently high level, with scores on the CUNY sentences presented visual-only of about 80% of words correct (baseline for cochlear implant patients who participated in another study (Goh et al., 2001) was around 24%). On our AV asynchrony judgment task, Mr. S was more accurate overall at detecting AV asynchrony than all but one of the 50 normal-hearing subjects tested, displaying a smaller AV synchrony window for all conditions and auditory- and visual-leading thresholds closer to 0 for all but the NS100 condition. Figure 4 shows Mr. S's response data along with the data from the normal-hearing subject who performed most similarly to Mr. S. Data from two representative normal-hearing young adult subjects from Experiment 1 are also included for comparison.

In light of Mr. S's impressive lipreading abilities and previous reports that good lipreaders may be better at detecting AV asynchrony (McGrath & Summerfield, 1985; Pandey et al., 1986; but see Grant & Seitz, 1998), further investigation into the potential relationship between lipreading skills and sensitivity to temporal asynchrony between auditory and visual stimuli seems warranted and is currently underway in our lab (see Conrey, 2004, this volume).

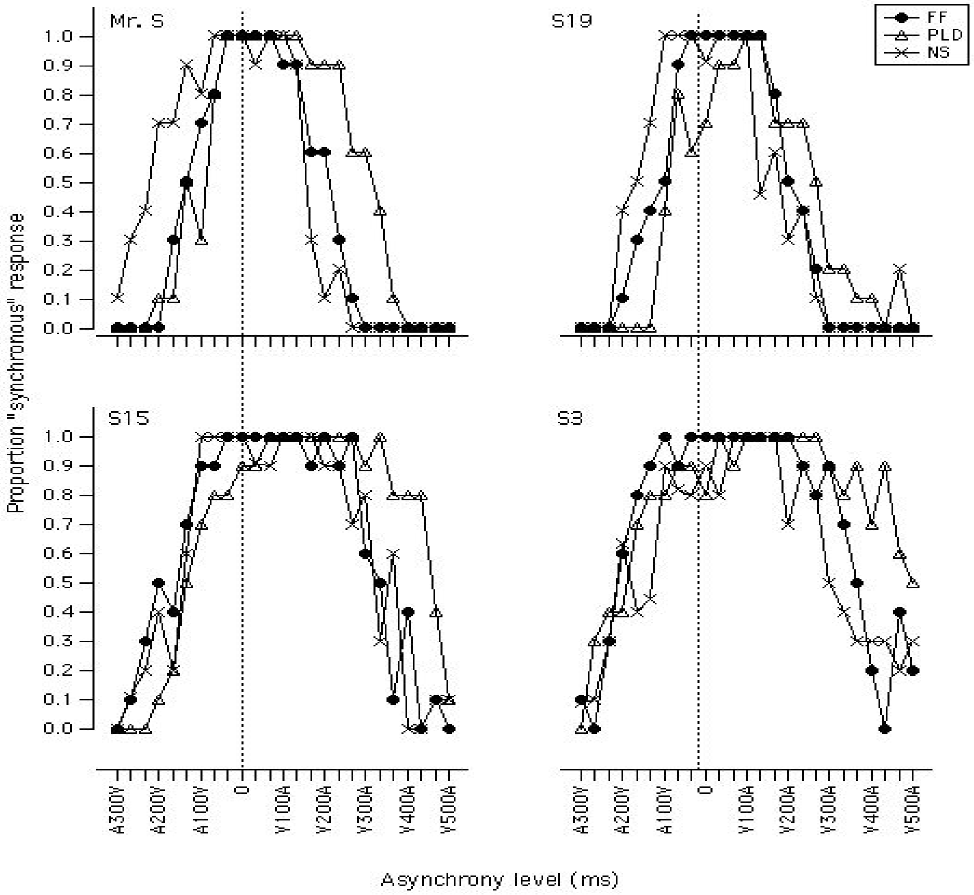


**Figure 4.** "In sync" response data for Mr. S, a cochlear implant patient and exceptionally good lipreader, and for three normal-hearing subjects from Experiment 1. Mr. S's performance on AV asynchrony detection was superior to that of 49 of 50 of our normal-hearing subjects. Top left panel: Mr. S. Top right panel: Subject 19, who performed comparably to Mr. S. Bottom panels: Subject 15 (left) and Subject 3 (right), whose performance was typical for normal-hearing subjects. The dotted vertical lines are at 0-ms asynchrony.

## Conclusions

In summary, the results of the present experiments suggest a window of AV asynchronies several hundred milliseconds wide over which participants were unable to detect asynchronies above chance. FF speech signals and simple nonspeech signals did not differ statistically in terms of the mean, width, or auditory- or visual-leading thresholds of the AV synchrony window. Further research is needed on the characteristics of visual signals, such as PLDs, that significantly affect the size and shape of the AV synchrony window. In addition, the importance of the onset versus the duration or offset of the AV signal should be investigated with longer and shorter speech and nonspeech signals, with manipulations of subjects' attentional strategies, and with phonetically balanced word lists. Finally, future EEG and neuroimaging work should build on earlier studies such as Bushara et al. (2001) and Olson et al. (2002) in order to elucidate the neural mechanisms involved in the perception and detection of synchrony between auditory and visual signals. Such investigations should provide further insights into the timecourse and underlying neural mechanisms involved in auditory-visual multimodal processing.

## References

Bergeson, T.R., Reynolds, J.T., & Pisoni, D.B. (2003). Perception of point light displays of speech by normal-hearing adults and deaf adults with cochlear implants. Paper presented at the *AVSP 2003 International Conference on Auditory-Visual Speech Processing.*

Bernstein, L.E., Auer, E.T., Jr., Moore, J.K., Ponton, C.W., Don, M., & Singh, M. (2002). Visual speech perception without primary auditory cortex activation. *NeuroReport, 13*, 311-315.

Bernstein, L.E., Demorest, M.E., & Tucker, P.E. (2000). Speech perception without hearing. *Perception & Psychophysics, 62*, 233-252.

Bushara, K.O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience, 21*, 300-304.

Callan, D.E., Callan, A.M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. *Cognitive Brain Research, 10*, 349-353.

Calvert, G. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110-1123.

Calvert, G., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C.R., McGuire, P.K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Calvert, G., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience, 15*, 57-70.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Conrey, B.L. (2004). Multimodal sentence intelligibility and the detection of auditory-visual asynchrony in speech and nonspeech signals: A first report. In *Research on Spoken Language Processing Report No. 26* (pp. 345-356). Bloomington, IN: Speech Research Laboratory, Indiana University.

Dixon, N., & Spitz, L. (1980). The detection of audiovisual desynchrony. *Perception, 9*, 719-721.

Fort, A., Delpuech, C., Pernier, J., & Giard, M.-H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research, 14*, 20-30.

Giard, M.H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*, 473-490.

Goh, W.D., Pisoni, D.B., Kirk, K.I., & Remez, R.E. (2001). Audio-visual perception of sinewave speech in an adult cochlear implant user: A case study. *Ear & Hearing, 22*, 412-419.

Grant, K.W., & Greenberg, S. (2001). Speech intelligibility derived from asynchronous processing of auditory-visual information. Paper presented at the *AVSP International Conference on Auditory-Visual Speech Processing*.

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., van Wassenhove, V., & Poeppel, D. (2003). Discrimination of auditory-visual synchrony. Paper presented at the *AVSP 2003 International Conference on Auditory-Visual Speech Processing.*

Guski, R., & Troje, N. (2003). Audio-visual phenomenal causality. *Perception & Psychophysics, 65*, 789-800.

King, A.J., & Palmer, A.R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research, 60*, 492-500.

Laasonen, M., Service, E., & Virsu, V. (2002). Crossmodal temporal order and processing acuity in developmentally dyslexic young adults. *Brain and Language, 80*, 340-354.

Laasonen, M., Tomma-Halme, J., Lahti-Nuutila, P., Service, E., & Virsu, V. (2000). Rate of information segregation in developmentally dyslexic children. *Brain and Language, 75*, 66-81.

Lachs, L. (1999). Use of partial stimulus information in spoken word recognition without auditory stimulation. In *Research on Spoken Language Processing Report No. 25* (pp. 82-114). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L. (2002). Vocal tract kinematics and crossmodal speech information (*Research on Spoken Language Processing Technical Report No. 10*). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., & Hernandez, L.R. (1998). Update: The Hoosier Audiovisual Multitalker Database. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 377-388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., & Pisoni, D.B. (in press-a). Crossmodal source identification in speech perception. *Ecological Psychology*.

Lachs, L., & Pisoni, D.B. (in press-b). Crossmodal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*.

Lachs, L., & Pisoni, D.B. (in press-c). Specification of crossmodal source information in isolated kinematic displays of speech. *Journal of the Acoustical Society of America*.

Lewald, J., Ehrenstein, W.H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research, 121*, 69-79.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*, 468-478.

Lewkowicz, D.J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1094-1106.

MacSweeney, M., Amaro, E., Calvert, G., Campbell, R., David, A.S., McGuire, P.K., et al. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *NeuroReport, 11*, 1729-1733.

Massaro, D., & Cohen, M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication, 13*, 127-134.

Massaro, D., Cohen, M.M., & Smeele, P.M.T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America, 100*, 1777-1786.

Mattys, S.L., Bernstein, L.E., & Auer, E.T., Jr. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics, 64*, 667-679.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America, 77*, 678-684.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Meredith, M.A. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research, 14*, 31-40.

Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience, 7*, 3215-3229.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research, 14*, 115-128.

Munhall, K.G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*, 351-362.

Olson, I.R., Gatenby, J.C., & Gore, J.C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research, 14*, 129-138.

Pandey, C.P., Kunov, H., & Abel, M.S. (1986). Disruptive effects of auditory signal delay on speech perception with lip-reading. *The Journal of Auditory Research, 26*, 27-41.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *NeuroReport, 11*, 1329-1333.

Rosenblum, L.D., & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 318-331.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O.V., Lu, S.-T., et al. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141-145.

Schroeder, C.E., & Foxe, J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research, 14*, 187-198.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*, 147-152.

Sheffert, S.M., Lachs, L., & Hernandez, L.R. (1996). The Hoosier audiovisual multitalker database. In *Research on Spoken Language Processing No. 21* (pp. 578-583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Smeele, P.M.T., Sittig, A.C., & van Heuven, V.J. (1992). Intelligibility of audio-visually desynchronised speech: Asymmetrical effect of phoneme position. Paper presented at the *International Conference on Spoken Language Processing.*

Stein, B., & Meredith, M.A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stein, J., & Walsh, V. (1997). To see but not to read: The magnocellular theory of dyslexia. *Trends in Neurosciences, 20*, 147-152.

Sugita, Y., & Suzuki, Y. (2003). Implicit estimation of sound-arrival time. *Nature, 421*, 911.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to 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*.

Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., & Hillyard, S.A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*, 106-114.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2002). Temporal integration in the McGurk effect. Paper presented at the Poster presented at the annual meeting of the *Society for Cognitive Neuroscience*, San Francisco.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2003). Electrophysiology of auditory-visual speech integration. Paper presented at the *AVSP 2003 International Conference on Auditory-Visual Speech Processing.*

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Perception and Comprehension of Synthetic Speech[1]

**Stephen J. Winters and David B. Pisoni**

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

[1] This research was supported by the NIH-NIDCD Research Grant R01 DC-00111 and NIH-NIDCD T32 Training Grant DC-00012 to Indiana University. We would like to thank Horabail Venkatagiri and Jan van Santen for their helpful comments and suggestions during the preparation of this paper.

# Perception and Comprehension of Synthetic Speech

**Abstract.** An extensive body of **r**esearch on the perception of synthetic speech carried out over the past 30 years has established that listeners have much more difficulty perceiving synthetic speech than natural speech. Differences in perceptual processing have been found in a variety of behavioral tasks, including assessments of segmental intelligibility, word recall, lexical decision, sentence transcription, and comprehension of spoken passages of connected text. Alternative groups of listeners—such as non-native speakers of English, children and older adults—have even more difficulty perceiving synthetic speech than young, healthy, college-aged listeners typically tested in perception studies. It has also been shown, however, that the ability to perceive synthetic speech improves rapidly with training and experience. Incorporating appropriate prosodic contours into synthetic speech algorithms—along with providing listeners with higher-level contextual information—can also aid the perception of synthetic speech. Listener difficulty in processing synthetic speech has been attributed to the impoverished acoustic-phonetic segmental cues—and inherent lack of natural variability and acoustic-phonetic redundancy—in synthetic speech produced by rule. The perceptual difficulties that listeners have in perceiving speech which lacks acoustic-phonetic variability has been cited as evidence for the importance of variability to the perception of natural speech. Future research on the perception of synthetic speech will need to investigate the sources of acoustic-phonetic variability and redundancy that improve the perception of synthetic speech, as well as determine the efficacy of synthetically produced audio-visual speech, and the extent to which the impoverished acoustic-phonetic structure of synthetic speech impacts higher-level comprehension processes. New behavioral methods of assessing the perception of speech by human listeners will need to be developed in order for our understanding of synthetic speech perception to keep pace with the rapid progress of speech synthesis technology.

## Introduction

Studying the perception of synthetic speech has proven to be useful in many different domains. When work on the perception of synthetic speech first began in the early 1970s, researchers were primarily interested in evaluating its segmental intelligibility in comparison to natural speech. These early studies were done with an eye toward improving the quality of synthetic speech for use in practical applications, such as reading machines for the blind or voice output communication devices. Early evaluation studies such as Nye and Gaitenby (1973) assessed listener's perception of synthetic and natural speech not only to quantify the intelligibility gap between the two types of speech but also to isolate and attempt to identify which synthetic speech segments were difficult for listeners to perceive correctly. Once such segments were identified, further research could be carried out to improve the intelligibility of these segments through the refinement of the text-to-speech algorithms which produced them. In this way, researchers hoped to be able to improve the overall segmental intelligibility of synthetic speech to the same level as natural speech.

The initial studies by Nye and Gaitenby (1973) have led to a long line of research which has developed continually more sophisticated metrics for evaluating the intelligibility of synthetic speech. Standards for assessing the quality of text-to-speech synthesis have been proposed by Pols (1989; 1992) and van Santen (1994). Pols (1989) grouped together the various assessment techniques into four broad categories: 1. *global* techniques, addressing acceptability, preference, naturalness and usefulness; 2. *diagnostic* techniques, addressing segmentals, intelligibility and prosody; 3. *objective* techniques, including metrics such as the Speech Transmission Index (STI) and the Articulation Index (AI); and 4.

*application-specific* techniques, addressing the use of synthetic speech in specific applied domains such as reading machines and televised weather briefings. The bulk of early research on the perception of synthetic speech focused primarily on the assessment of segmental intelligibility (Type 2 techniques) and some global measures of preference and acceptability (Type 1 techniques).

Another group of researchers realized that much could be learned about speech perception by simply trying to figure out why people perceived synthetic speech differently than natural speech. Instead of asking how the intelligibility of synthetic speech might be improved, they asked "*why* is it that synthetic speech is more difficult to understand than natural speech? Does synthetic speech lack certain fundamental characteristics of natural speech which might be helpful to perception?" And, can listeners adjust or attune their perceptual systems to overcome such shortcomings in the synthetic speech signal?

The answers to these kinds of questions potentially lie much deeper than the surface level of segmental intelligibility. In order to study these problems, therefore, Pisoni (1981) suggested that more research on synthetic speech perception ought to be done in the ten areas shown in Table 1:

1. Processing time experiments
2. Listening to synthetic speech in noise
3. Perception under differing attentional demands
4. Effects of short- and long-term practice
5. Comprehension of fluent synthetic speech
6. Interaction of segmental and prosodic cues
7. Comparisons of different rule systems and synthesizers
8. Effects of naturalness on intelligibility
9. Generalization to novel utterances
10. Effects of message set size

**Table 1.** Needed research on the perception of synthetic speech (adapted from Pisoni 1981)

All of these lines of research have been pursued—to varying extents—in the years following Pisoni (1981). The results of this work have shown consistent differences in perception between natural and synthetic speech at every level of analysis (Duffy & Pisoni, 1992; Pisoni, 1997; Pisoni, Nusbaum & Greene, 1985).

Researchers have attempted to account for the perceptual differences between natural and synthetic speech in terms of the acoustic-phonetic characteristics which differ between these two types of speech. Throughout most of the '80s and '90s, research on synthetic speech perception used text-to-speech (TTS) systems which produced synthetic speech by rule. Synthesis-by-rule operates just as its name implies—a synthesis algorithm takes an orthographic string of letters as input and automatically converts them into speech output by using a set of text-to-speech conversion rules. Using rules to produce speech in this way results in a synthetic speech signal which lacks much of the variability inherent in natural speech. Furthermore, synthetic speech produced by rule typically provides fewer redundant cues to particular segments or sound contrasts. Text-to-speech conversion rules also tend to use highly simplified coarticulation sequences between individual segments, and they often lack appropriate or natural-sounding sentential prosody contours. Researchers have typically focused on these acoustic aspects of synthetic speech produced by rule in attempting to account for why listeners have greater difficulty perceiving it. Accounting for the difficulties of synthetic speech perception in this way has led researchers to draw some important conclusions about the nature of normal human speech perception—it apparently relies on redundant cues to particular segments, and it also makes use of the natural acoustic-phonetic variability inherent in the speech signal. Redundancy and variability in the signal are fundamental properties of speech perception, and not just noise which needs to be filtered away through some kind of perceptual normalization process.

This chapter reviews evidence from the study of synthetic speech perception which has led researchers to draw these conclusions about normal speech perception. It details the various distinctions which have been found to exist between synthetic and normal speech perception, from the first segmental intelligibility studies to more global studies on comprehension, perceptual learning effects, alternative groups of listeners, and properties like naturalness and prosody. The final section concludes with some suggestions for potentially fruitful areas of future research in the context of the rapidly changing world of speech synthesis technology.

## Point 1: Synthetic Speech is Less Intelligible than Natural Speech

**The Modified Rhyme Test.** Investigations of the segmental intelligibility of synthetic speech have routinely shown that it is less intelligible than natural speech. For instance, one of the first studies on the perception of synthetic speech, Nye and Gaitenby (1973), assessed the intelligibility of natural speech and synthetic speech produced by the Haskins Parallel Formant Resonance Synthesizer (Mattingly, 1968). They measured segmental intelligibility with the Modified Rhyme Test (MRT). The MRT, which was originally developed by Fairbanks (1958) and House, Williams, Hecker and Kryter (1965), presents listeners with a spoken word and then requires them to select the word they heard from a set of six, real-word alternatives, all of which differ by only one phoneme. Examples of these response sets include the following:

| (1) | | (2) | |
|---|---|---|---|
| game | came | dull | dub |
| fame | name | duck | dun |
| tame | same | dug | dud |

Nye and Gaitenby (1973) adapted this task to the study of synthetic speech perception under the assumption that it would help them identify the particular segments for which the Haskins Pattern Playback produced poor, insufficient, or confusing cues.

In general, the results of Nye and Gaitenby's MRT study showed that synthetic speech was significantly less intelligible than natural speech. Listeners' overall error rate was 7.6% on the MRT for synthetic speech, but only 2.7% for natural speech. Nye and Gaitenby's results also revealed that synthetic obstruents were particularly unintelligible, as they induced error rates that were much higher than those for synthetic sonorants. However, Nye and Gaitenby pointed out that the MRT was of limited utility as an evaluative diagnostic for synthetic speech, since the closed response set provided a limited number of possible phonemic confusions and, moreover, it did not present all phonemes in equal proportions in the test stimuli.

Despite the shortcomings noted by Nye and Gaitenby (1973), the MRT has been used extensively in the ensuing years as virtually the standard way to assess the segmental intelligibility of various types of synthetic speech (Greene, Manous & Pisoni, 1984; Hustad, Kent & Beukelman, 1998; Koul & Allen, 1993; Logan, Greene & Pisoni, 1989; Mitchell & Atkins, 1988; Pisoni, 1987; Pisoni & Hunnicutt, 1980; Pisoni et al., 1985). Another problem with the original version of the MRT that Nye and Gaitenby used was that it was simply too easy. Ceiling effects are obtained in the MRT when natural or high-quality synthetic stimuli are presented under clear listening conditions. Hence, several studies have expanded upon the MRT paradigm by preserving the original MRT word list but eliminating the forced-choice response set. In this "open-response format" version of the MRT, listeners simply respond to each particular word by writing down whatever they think they heard.

Logan et al. (1989) showed that this version of the MRT can help reduce the ceiling effects in performance that often emerge from the relative ease of the original forced-choice task. These authors used both the "closed-set" and "open-set" response formats of the MRT in testing the segmental intelligibility of 10 different speech synthesizers along with natural speech. For each synthesizer and each response format, Logan et al. tested 72 different listeners. Figure 1 shows the percentages of errors that these groups of listeners made in attempting to identify words produced by all of these voices, in both the open and closed formats of the MRT. Listener error rates were higher for all versions of synthetic speech than for natural speech; they were also higher in the open-set form of the task than they were in the closed-set format. These results indicate, moreover, that the intelligibility of synthetic speech depends greatly on the type of synthesizer being used to produce the speech. The low error rate observed for DECTalk in the closed response format (≈5%), for instance, approximated that of natural speech, while the error rates for the worst synthesizers were consistently higher than 50% in the open response format.

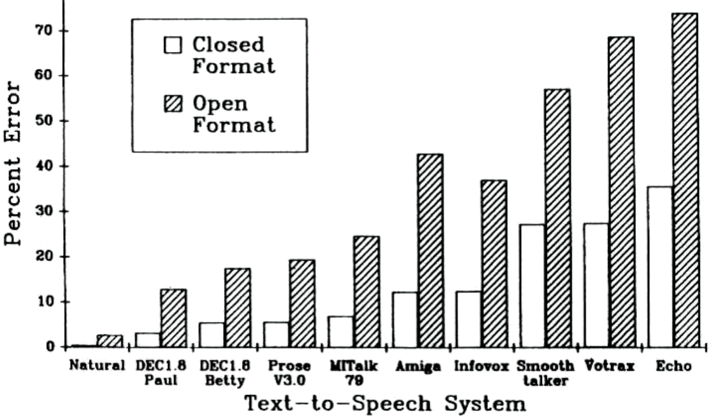


**Figure 1.** Error rates for word recognition in open and closed format MRT, by synthetic voice type (adapted from Logan et al. 1989).

**Intelligibility of Synthetic Speech in Noise.** The poor performance of most speech synthesizers in the open-response format in Logan et al. (1989) shows that the intelligibility gap between natural and synthetic speech increases significantly under more realistic testing situations. Offering listeners six possible response alternatives for each word they hear does not realistically reflect actual usage of speech synthesis applications, where listeners have to interpret each word they hear in terms of their entire lexicons. Likewise, testing the intelligibility of synthetic speech under ideal listening conditions in the laboratory does not reflect real-world usage either, where synthetic speech is often heard and produced in noisy environments. Several researchers have therefore tested the intelligibility of synthetic speech as it is played through noise and have found that such less-than-ideal listening conditions degrade the intelligibility of synthetic speech even more than they do the intelligibility of natural speech.

Pisoni and Koen (1981) were the first to report the results of testing synthetic speech in noise. They used the open- and closed-response formats of the MRT to test the intelligibility of natural and synthetic (MITalk) speech in various levels of white noise. Figure 2 shows the percentages of words correctly identified in the different listening conditions in Pisoni and Koen's study. These results replicate earlier findings that intelligibility is worse for synthetic than for natural speech—especially in the open-response format of the MRT. Pisoni and Koen's results also established that increasing the level of noise had a significantly greater detrimental effect on the intelligibility of synthetic speech than it had on the intelligibility of natural speech.

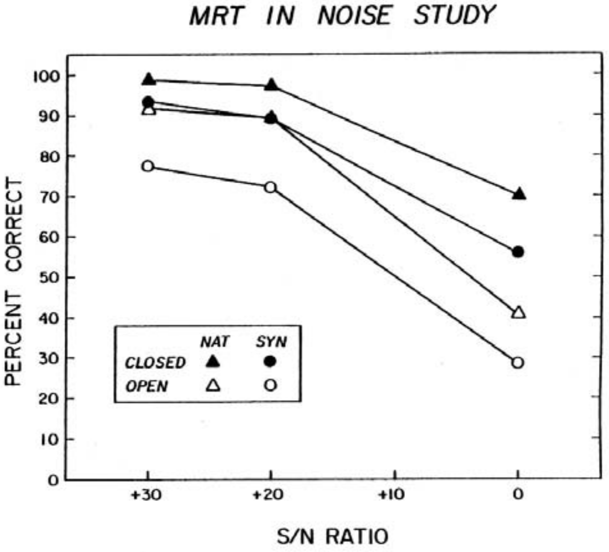


**Figure 2.** Percent correct word recognition in open and closed-format MRT, by voice type and signal-to-noise ratio (adapted from Pisoni & Koen, 1981).

In another study, Clark (1983) also tested the perception of synthetic speech, as produced by his own synthesis algorithms, in white noise. Instead of using the MRT, however, Clark measured the perception of vowels in /hVd/ sequences and consonants in /Ca:/ syllables, in an attempt to assess the susceptibility of particular segmental cues to degradation in noise. He found that the intelligibility of his synthetic vowels was surprisingly robust in noise, even to the point of being marginally more intelligible than natural vowels in the noisiest listening conditions (-6 db SNR). Clark did, however, find appreciable degradation of the intelligibility of synthetic consonants in noise, especially stops and fricatives. Despite the comparatively greater degradation of synthetic consonant intelligibility in noise, however, Clark found that the rank-ordering of individual consonant intelligibilities did not differ significantly between natural and synthetic speech in the noisy listening conditions. Clark therefore concluded that synthetic speech might be considered a "degraded" form of natural speech, with cues to individual segments that were simply not as salient as those found in natural speech.

Nusbaum, Dedina and Pisoni (1984) took issue with Clark's assessment and argued that the segmental cues in synthetic speech were not just "degraded" forms of the ones found in natural speech, but that they were "impoverished" and could also be genuinely misleading to the listener. In support of this claim, Nusbaum et al. presented results from a perception experiment in which they played CV syllables, produced by both a human voice and three different speech synthesizers, to listeners in various levels of noise. When Nusbaum et al. investigated the confusions that listeners made in trying to identify the consonants in this experiment, they found that listeners not only had more difficulty identifying synthetic consonants in noise, but that they also often misidentified synthetic consonants in ways that were never seen in the confusions of natural speech tokens. Even in listening to DECTalk speech, for instance—a form of synthetic speech that has been established as highly intelligible by studies like Logan et al. (1989)—listeners often misidentified /r/ as /b/—a perceptual confusion that never occurred for natural speech tokens of /r/. Nusbaum et al. therefore suggested that some synthetic speech cues could be misleading to the listener as well as being impoverished versions of their natural counterparts.

More recent investigations of the perception of synthetic speech in noise have incorporated multi-talker "babble" noise into the signal, rather than white noise. This methodology captures another level of ecological validity, since natural conversations—and the use of synthetic speech applications—often occur in public, multi-talker settings, where it is necessary to focus in on one particular voice through the "cocktail party effect" of other speakers (Bregman, 1990). Studies which have incorporated this type of noise in tests of synthetic speech intelligibility have found that it does not generally degrade the intelligibility of synthetic speech as much as white noise. Koul and Allen (1993), for instance, incorporated multi-talker noise into an open-response version of the MRT, using both natural and DECTalk voices. While they found that the natural voice was more intelligible than DECTalk—and that both voices were more intelligible at the higher signal-to-noise ratios (SNRs)—they did not find any interaction between voice type and the level of multi-talker babble noise. Both voices, that is, suffered similar losses in intelligibility at decreasing SNRs. For instance, at a +25 db SNR, listeners correctly identified 84% of natural words and 66% of synthetic words, while at a 0 db SNR, correct word identification scores decreased to 52% for natural words and 30% for the DECTalk items. Koul and Allen also found similar identification errors for both natural and synthetic segments. Different confusions emerged for /h/, however, which was often misidentified in DECTalk, but rarely confused in natural speech. Koul and Allen suggested that the "babble" noise might influence the perception of synthetic speech in this unique way because its acoustic energy is primarily focused in the lower frequencies, rather than evenly spread across the spectrum as in white noise. Aside from this particular difference, Koul and Allen's results indicated a high degree of similarity between the segmental cues used in DECTalk and those found in natural speech.

**Perception of formant vs. concatenative synthesis.** In recent years, speech synthesis technology has relied increasingly on "concatenative" synthesis as opposed to the "formant-based" synthesis techniques that were prevalent in the '70s and '80s (Atal & Hanauer, 1971; Dutoit & Leich, 1993; Moulines & Charpentier, 1990). Formant-based synthesis operated on the basis of an electronically implemented, source-filter model of the human articulatory system (Fant, 1960). These synthesizers produced speech by progressing through a series of source-filter targets, one segment at a time. This approach therefore focused on the acoustic quality of the synthetic speech within each individual segment, rather than on the acoustic transitions between the segments. Concatenative synthesis, on the other hand, uses an actual human voice as its source—rather than the output of an electronic model—and incorporates segmental boundaries within its basic unit of production. The size of the basic units used in concatenative synthesis may be as large as a sequence of words or as small as a portion of a phoneme, but a popular choice is a diphone-sized unit, which extends from the midpoint of one phoneme to the midpoint of the next phoneme. Each such diphone thereby contains the acoustic transition between a particular pair of phonemes. Concatenative synthesis algorithms operate by simply joining these basic diphone units together into the desired phonemic string.

Encoding the transitions between segments into the speech output—along with using a natural human voice as the speech source—is supposed to make concatenative synthesis sound more natural and aesthetically appealing to the listener than formant-based synthesis. While these aspects of concatenative synthesis may have motivated its increasing popularity, there has actually been little research demonstrating that it is either more intelligible or even more natural-sounding than formant-based synthesis. Studies testing early forms of diphone-based synthesis, such as RealVoice and SmoothTalker, indicated that they were consistently less intelligible than high-quality formant synthesis systems, such as DECTalk (Logan et al. 1989). Subsequent studies have shown, however, that more recent diphone-based synthesizers can match DECTalk in intelligibility. Rupprecht, Beukelman and Vrtiska (1995) tested the comparative intelligibility of DECTalk versus MacinTalk (an early form of diphone-based synthesis) and MacinTalk Pro (an improved version of MacinTalk). Rupprecht et al. played listeners sentences from the Speech Perception in Noise (SPIN) test (Kalikow, Stevens & Elliott, 1977) as produced by each of the synthetic voices, and asked listeners to identify the last word in each sentence. Rupprecht et al. found that correct identification rates for both DECTalk and MacinTalk Pro were significantly higher than the same rates for the MacinTalk synthesizer. There were, however, no significant differences between the intelligibility of the DECTalk and MacinTalk Pro voices. Hustad et al. (1998) also tested the intelligibility of DECTalk and MacinTalk Pro synthesis in an open-response format MRT. While the intelligibility of both voices was quite high in this task, Hustad et al. found a slight but significant advantage in intelligibility for the DECTalk voice. Hustad et al. suggested that Rupprecht et al.'s failure to find such a difference in their earlier study may have been the result of their listeners relying on higher-level contextual cues to identify words in the SPIN sentences.

In a more recent study, Venkatagiri (2003) also used words from the MRT to investigate the comparative intelligibility of four different synthesizers, each using a different combination of formant and concatenative synthesis in their speech production algorithms. These included AT&T's NextGen TTS, which uses half-phone based synthesis (a method of concatenating halves of individual phonemes together); Festival, which uses diphone-based synthesis; FlexVoice, which uses a combination of formant and diphone-based synthesis; and IBM ViaVoice, which uses formant-based synthesis only. Venkatagiri (2003) tested the intelligibility of each of these systems—along with a human voice—by playing productions of individual words from the MRT in neutral-sentence carrier phrases (e.g., "The word is ____.") in two different levels of multi-talker babble noise. Venkatagiri (2003) found that the listeners' ability to identify the natural speech tokens under these conditions was significantly greater than their ability to identify the same tokens as they were produced by all four of the synthetic voices. He also found that increasing the level of noise was significantly more detrimental to the intelligibility of all types of synthetic speech than it was for natural speech.

Among the four synthetic voices, the two voices that used concatenative synthesis techniques (NextGen and Festival) were significantly more intelligible than the one using formant-based synthesis (ViaVoice). In addition, FlexVoice, which used a combination of formant and concatenative algorithms was significantly less intelligible than the other three synthetic voices. Interestingly, formant-based synthesis outperformed the concatenative synthesizers on vowel intelligibility, even though it induced significantly more listener errors in the identification of consonant sounds. These results suggested that the ability of concatenative synthesizers to produce highly intelligible consonant sounds—presumably because they maintain the natural transitions between these transient segments and their surrounding phonemes—comes at the cost of being able to produce highly intelligible vowel sounds. Moreover, the relatively poor performance of the FlexVoice synthesizer—which combined the formant and concatenative approaches—indicated that low cost speech synthesis technology has not yet advanced to the point where it is possible to combine the best perceptual aspects of both production algorithms into one system. Venkatagiri (2003) also observed that the poor intelligibility of all synthetic voices in noisy conditions—no matter what their method of production—revealed that there are still serious limitations on their potential utility in real-world applications, where noisy listening conditions are the norm.

**Summary.** The available evidence suggests that the segmental intelligibility of synthetic speech is significantly worse than that of natural speech. Synthetically produced words have routinely been shown to be more difficult for listeners to identify in forced-choice tests of segmental intelligibility, such as the MRT. The gap between natural and synthetic speech intelligibility also increases substantially when noise is introduced into the speech signal, or if there are fewer constraints on the possible number of responses listeners can make in a particular testing paradigm. The poor segmental intelligibility of synthetic speech produced by rule appears to result from the use of segmental cues which are not only acoustically degraded relative to natural speech but also impoverished and therefore potentially misleading to the listener. The use of naturally produced speech segments in concatenative speech synthesis techniques may provide a method of overcoming such intelligibility limitations, but research on the perception of synthetic speech produced in this manner indicates that it is still not as intelligible as natural speech, especially in adverse listening conditions.

## Point 2: Perception of Synthetic Speech Requires More Cognitive Resources

**Lexical Decision.** One consequence of the poor segmental intelligibility of synthetic speech is that listeners may only be able to interpret it by applying more cognitive resources to the task of speech perception. Several studies have shown that listeners do, in fact, engage such compensatory mechanisms when they listen to synthetic speech. Pisoni (1981), for instance, had listeners perform a speeded lexical decision task, in which they heard both naturally produced and synthetically produced (MITalk) strings as test items. These test strings took the form of both words (e.g., "colored") and non-words (e.g., "coobered"). Pisoni found that listeners consistently took more time to determine whether or not the synthetic strings were words than if the natural strings were words, regardless of whether or not the test string was a real lexical item or a non-word. Since no significant interaction emerged between the voice type and the lexical status of the string, Pisoni concluded that listeners had to apply more cognitive resources to the task of interpreting the acoustic-phonetic surface structure of synthetic strings, prior to any higher-level lexical or semantic processing.

In a follow-up study, Slowiaczek and Pisoni (1982) suggested that the processing advantage for natural speech in the lexical decision task might be the result of greater listener familiarity with natural speech. They assessed whether the processing gap between natural and synthetic speech items in a speeded lexical decision task might be reduced as listeners became more familiar with the synthetic voice. They investigated this possibility by having listeners perform a speeded lexical decision task, as in Pisoni (1981), in which listeners heard both word and non-word strings as produced by both a natural voice and MITalk. Slowiaczek and Pisoni's listeners performed a speeded lexical decision task for five consecutive days, while listening to word and non-word strings produced by both a natural voice and MITalk. Figure 3 compares the response times for Slowiaczek and Pisoni's listeners, on the fifth day of performing this task, to the response times from Pisoni's (1981) listeners, on the only day on which they performed an identical task. Slowiaczek and Pisoni found that listener response times (RTs) decreased over the course of the five-day training process only for the strings produced by the synthetic voice; the RTs for the naturally produced items remained constant over time. The results in Figure 3 also show that, despite this improvement in performance, RTs for synthetic words never reached the same level as RTs for natural words (although RTs for both synthetic and natural non-words were quite close, after five days of testing). Slowiaczek and Pisoni concluded that the advantage for natural speech found in Pisoni (1981) reflected genuine differences in the processing of natural and synthetic speech, which could not be eliminated completely just by increasing listeners' familiarity with synthetic speech.

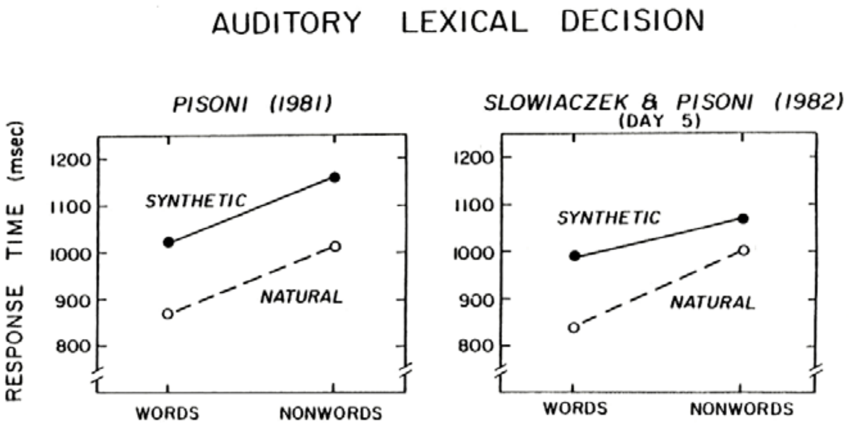


**Figure 3.** Lexical decision reaction times, for synthetic and natural words and non-words (adapted from Slowiaczek & Pisoni, 1982).

**Word Recognition.** In another study on spoken word recognition, Manous and Pisoni (1984) also found that synthetic speech puts greater demands on listeners in a word recognition task than natural speech. Using the gating paradigm developed by Grosjean (1980), Manous and Pisoni presented listeners with increasing amounts of individual words in a sentential context, as produced by both a human speaker and the DECTalk synthesizer. The amount of the word that each listener heard increased by 50 milliseconds on each successive trial; i.e., the listeners first heard 50 milliseconds of the word, then 100 milliseconds, and so on. Manous and Pisoni found that listeners needed to hear, on average, 361 milliseconds of naturally produced words before they could reliably identify them, whereas they needed to hear an average of 417 milliseconds of DECTalk-produced words before they could reach the same level of accuracy. Manous and Pisoni attributed this difference in performance to the acoustically impoverished nature of synthetic speech. Since DECTalk provides fewer redundant acoustic-phonetic cues for individual segments than natural speech does, listeners needed to hear and process more segmental information before they could reliably identify spoken lists of synthetic words.

**Word Recall.** The recall of synthetically produced words also requires more cognitive resources than the recall of naturally produced words. Luce, Feustel and Pisoni (1983) tested listeners' ability to recall unordered lists of words that were produced by both a human voice and MITalk. The authors found that listeners could recall more items from the natural word lists than they could from the synthetic word lists. They also found that there were more intrusions (words "recalled" by listeners that were not in the original lists) in the recall of the synthetic lists than in the recall of the natural lists. In a follow-up experiment, Luce et al. also measured listeners' ability to recall synthetic and natural word lists when they had a digit pre-load. Listeners first memorized a list of 0, 3 or 6 digits before they heard a list of 15 test words. The listeners were then asked to recall both lists—first the digits, in order, and then the words, in no particular order. Once again, the listeners recalled fewer words from the synthetic lists correctly than they did from the natural lists, and they also produced more "intrusions" in the recall of synthetic lists. Furthermore, the listeners' recall of the six-digit lists was worse when they were presented before lists of synthetic words. Luce et al. claimed that this interaction indicated that both digit and word recall shared the same, limited store of cognitive processing resources—and that the storage and maintenance of synthetic words absorbed more of these processing resources than the storage and maintenance of natural words.

Lastly, Luce et al. (1983) tested the serial recall of 10-word lists, produced by both natural and synthetic voices. In a serial recall task, listeners must recall a list of words in the correct order; this task thus requires the encoding of both item and order information. Luce et al. found both recency and primacy effects when listeners attempted to recall lists of synthetic and natural words in order; that is, they were best at recalling words that appeared both early and late in the various lists. However, these effects interacted with the type of voice used. While the recall of natural items was, in general, better overall than the recall of synthetic words, this advantage was significantly larger for early items than late items in the lists. Luce et al. hypothesized that this interaction might occur because the storage of synthetic words required more processing resources than the storage of natural words. Late items in the synthetic lists might therefore make use of memory resources that would otherwise be reserved for the storage of early items in the list. Thus, the recall of early synthetic items was disadvantaged not only by the fact that they required more memory resources than the early natural items, but also by the fact that subsequent items in the synthetic lists required more memory resources, as well.

Luce and Pisoni (1983) tested these hypotheses in another recall study, using mixed lists of synthetic and natural words. Following Luce et al.'s (1983) earlier logic, Luce and Pisoni hypothesized that synthetic words that appeared late in a list should adversely affect the recall of both natural and synthetic words that appeared earlier in the same list. Luce and Pisoni had listeners recall mixed lists in which five natural items were followed by five synthetic items (and vice versa). The results of this study did not support the prediction that the recall of either early synthetic or natural words would be hampered by late synthetic items. Luce and Pisoni's results also failed to replicate Luce et al.'s (1983) finding that the recall of early synthetic items was worse than the recall of early natural items. In order to determine whether or not these findings were the result of the poor intelligibility of the items in the synthetic word lists, Luce and Pisoni constructed new lists of both natural and synthetic words using only items that listeners identified correctly more than 98% of the time in an MRT task. Using these lists, Luce et al. found that the recall of natural words was better than the recall of synthetic words in positions 2, 3 and 5 of the 10 word-long lists; there were no significant differences in recall between natural and synthetic voices for the other positions in the word lists. Since differences in recall were found using lists of highly intelligible words whose acoustic-phonetic interpretation presumably required minimal amounts of extra cognitive effort, Luce and Pisoni concluded that the recall and higher-level processing of synthetic words genuinely did require more cognitive resources than the recall of natural words.

**Summary.** Evidence from lexical decision, word recognition and word recall studies suggest that the perception of synthetic speech requires more cognitive resources than the perception of natural speech. Listeners take longer to decide if synthetic strings are words in a lexical decision task, and they also need to hear more of a synthetic word before they can recognize it in a gated word recognition task. The existence of these processing deficits indicates that listeners must apply extra cognitive resources to the interpretation of the impoverished acoustic-phonetic cues of synthetic speech at the segmental level. Research showing that the storage and recall of synthetically produced words is more difficult than the storage and recall of naturally produced words also indicates that additional cognitive resources must be used in the encoding of synthetic speech items in memory.

## Point 3: Perception of Synthetic Speech Interacts with Higher-Level Linguistic Knowledge

**Perception of Words in Sentences and in Isolation.** Listeners also compensate for the poor segmental intelligibility of synthetic speech by relying on any available higher-level linguistic information to help them interpret a synthetic speech signal correctly. For this reason, the perception of synthetic words presented in sentential contexts has been found to be significantly better than the perception of synthetic words in isolation.

Hoover, Reichle, Van Tasell and Cole (1987) for example, demonstrated the influence of higher-level linguistic information on the perception of synthetic speech by comparing listeners' perception of synthetically produced words in isolation to the perception of synthetically produced "low probability" and "high probability" sentences. "Low" and "high" probability sentences were used as a means of testing the effects of semantic plausibility on the perception of words in a sentential context. Hoover et al. constructed sentences of these two types by asking participants, in a pre-test, to fill in the final word in a series of short, declarative sentences. The final words that the participants chose to fit these contexts more than 90% of the time was used in the high-probability sentences, whereas the final words that were chosen sparingly were used in the low-probability sentences. Hoover et al. recorded sentences of these two types, along with the final words for each sentence in isolation, as produced by a human speaker and both the Votrax and the Echo II synthesizers. They then presented these words and sentences to listeners, who were instructed to repeat what they had heard. Hoover et al. found that the listeners repeated sentences of both types more accurately than they repeated the individual words. Repetition accuracy for the synthetic items was still worse than the repetition accuracy for the natural words and sentences, however. Table 2 shows the correct identification rates for all voices and conditions in Hoover et al.'s study:

| | Single Words | Low-Prob. Sentences | High-Prob. Sentences |
|---|---|---|---|
| Votrax | 21.9 | 35.3 | 87.8 |
| Echo II | 19.5 | 23.8 | 77.3 |
| Natural | 99.9 | 100 | 100 |

**Table 2.** Percentage of words correctly repeated, by voice type and presentation context (adapted from Hoover et al., 1987).

Hoover et al. (1987) also found that Votrax speech was more intelligible than Echo II speech in the two different sentence contexts. In general, however, the intelligibility of both of these synthesizers was quite low. The correct identification rates for words in the high probability sentences produced by these synthesizers was less than 90%, even though the words in these contexts were chosen on the basis of their being selected more than 90% of the time by readers who were merely filling in the blanks at the ends of these sentences. Such results suggest that poor-quality synthetic speech may actually mislead a listener and thereby be less informative to a listener than no signal at all.

Mirenda and Beukelman (1987) undertook a similar investigation of the perception of synthetic words in isolation and in sentences, but they included tokens produced by the highly intelligible DECTalk synthesizer, along with the poorer quality Echo and Votrax synthesizers. These authors found that correct identification rates for DECTalk increased from 78% in isolated word contexts to 96.7% in sentences, for adult listeners. This approximated adult listener performance on natural speech versions of the same tokens, which reached ceiling levels of performance at 99.2% correct for individual words and 99.3% correct for words in sentences. Mirenda and Beukelman (1990) expanded the range of synthesizers used to produce the stimuli in a follow-up study, which used the same methodology, and found once again that the sentential contexts improved the intelligibility of all synthetic voices. The best synthesizer in this second study, SmoothTalker 3.0, was a diphone-based system, but it still failed to reach the intelligibility levels found with natural speech in both isolated words and sentential contexts.

**Semantically Anomalous Sentences.** Earlier research on the intelligibility of synthetic words in sentences suggests, however, that the perception of words in sentences actually becomes worse than the perception of words in isolation if the contents of the sentences lack semantic coherence and predictability. Nye and Gaitenby (1974), for instance, tested the intelligibility of the Haskins Parallel Formant Resonance Synthesizer with both the closed-set MRT and a set of syntactically normal but meaningless sentences (e.g., "The safe meat caught the shade.") Nye and Gaitenby found that correct identification rates were much lower for synthetic words in these meaningless sentences (78%) than they were for the words presented in isolation in the MRT (92%). This effect was also proportionally greater for synthetic speech than it was for natural speech, which scored a 97% correct identification rate in the MRT and a 95% correct identification rate in the meaningless sentence condition. Nye and Gaitenby attributed the detrimental effect of sentential contexts to the difficulty listeners had in parsing individual words out of a longer string of words. They also pointed out that higher-level sentential information might have biased listeners towards expecting to hear words which fit into the sentence's semantic context, rather than the words that did appear in the anomalous sentences that were presented to them.

Pisoni and Hunnicutt (1980) presented further evidence to support the hypothesis that semantically anomalous sentences produced detrimental effects on the perception of words. They tested listeners on the closed-set MRT, using both natural speech and the MITalk system. The listeners were also asked to transcribe both a set of semantically meaningless Haskins sentences and a set of meaningful Harvard sentences (Egan, 1948). Pisoni and Hunnicutt found that correct identification rates for individual words were comparable between the MRT (99.4% for natural speech, 93.1% for synthetic) and the Harvard sentences (99.2% for natural speech, 93.2% for MITalk), but were lower for the semantically meaningless Haskins sentences (97.3% for natural speech, 78.7% for synthetic speech). The loss of meaning in sentential contexts thus had a more detrimental effect on the intelligibility of synthetic speech than it did on the intelligibility of the natural voice.

**Gating in Sentences.** Duffy and Pisoni (1991) noted that the correct identification rates for both natural and DECTalk speech in Mirenda and Beukelman's (1987) tests of sentence transcription were close to ceiling levels of performance. They therefore developed a gating paradigm for words in sentential contexts in an attempt to tease apart the similar intelligibility levels of these two kinds of speech. Duffy and Pisoni's gating paradigm involved presenting sentences to listeners in which they heard increasing amounts of the final word on successive trials. On the first presentation, the listeners heard none of the final word; on the second presentation, they heard 50 ms of the word; on the third, they heard 100 ms, and so on. After each of these presentations, the listeners were instructed to guess what the final word in the sentence was. Duffy and Pisoni presented these words to the listeners in either a "congruent" or a "neutral" sentential context, using both DECTalk and human voices. The final words in the congruent sentences were semantically related to the words at the beginning of the sentences (e.g., "The soldiers flew in the helicopter.") while the final words in the neutral sentences had no clear semantic relation to the words at the beginning of those sentences (e.g., "The people were near the helicopter.")

Figure 4 shows the percentage of correct identifications of a word, in both natural and synthetic voices, for each presentation of that word at the various gating durations. On average, listeners needed to hear 68 ms more of the words if they were produced by DECTalk than if they were spoken in a natural voice before they could reliably identify them in these contexts. This effect of voice type interacted significantly with the type of sentence context; listeners had to hear 212 ms more of the synthetic words when they appeared in the "neutral" contexts than when they appeared in the "congruent" contexts before they could identify them reliably. These results indicated, once again, that the perception of synthetic speech not only requires more cognitive resources than the perception of natural speech, but also that listeners appear to draw much more heavily on higher-level syntactic and semantic information in order to compensate for the difficulties they incur in processing the impoverished acoustic-phonetic structure of synthetic speech.

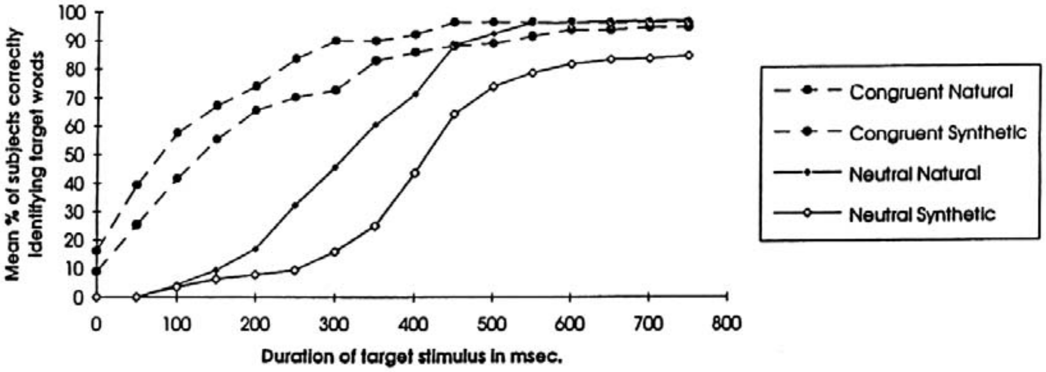


**Figure 4.** Percent of synthetic and natural words correctly identified, by duration of target stimulus in congruent and neutral sentential contexts (adapted from Duffy & Pisoni, 1991).

**Summary.** Research on the perception of synthetically produced sentences indicates that higher-level syntactic and semantic information can both facilitate and hinder the perception of synthetic speech. The transcription of synthetic words in meaningful sentences is typically better than the identification of those words in isolation. This finding suggests that listeners rely more extensively on the higher-level linguistic information in meaningful sentences to help them overcome difficulties they may have in interpreting the synthetic speech signal at the segmental level. However, listeners also have more difficulty identifying synthetic words in semantically anomalous sentences than they do identifying synthetic words in isolation, perhaps because the higher-level semantic information in these sentences encourages the listeners to develop misleading expectations about which words may follow. Semantic information has also been shown to interact with the perception of individual words at the ends of sentences in a gating paradigm, indicating that listeners may simultaneously apply both their knowledge of higher-level linguistic structure and additional cognitive resources to the challenging task of interpreting the impoverished acoustic-phonetic structure of synthetic speech.

## Point 4: Synthetic Speech is More Difficult to Comprehend than Natural Speech

Despite the wealth of evidence from many studies showing that synthetic speech is less intelligible than natural speech, early investigations of listeners' ability to comprehend synthetic speech often showed that it was no worse than their ability to comprehend natural speech. Many of these early studies, however, relied on post-perceptual measures of comprehension, using multiple-choice questions or free recall tasks. These studies assessed the "product" rather than the "process" of language comprehension. The results of these studies may have, therefore, reflected the influence of other, higher-level cognitive processes which extend far beyond the scope of perceiving and processing the synthetic speech signal itself.

**Post-perceptual Comprehension Tests.** Nye, Ingemann and Donald (1975) were the first researchers to study listeners' comprehension of synthetic speech. They played college-level reading passages to listeners and then asked them to answer a series of multiple-choice questions based on those passages. The listeners were instructed to take as much time as they needed to get as many of the questions right as they possibly could. They could even go back and re-play sections of the passages that they had difficulty understanding the first time. The dependent measure that Nye et al. looked at in this paradigm was the total amount of time it took listeners to answer the multiple-choice questions. Nye et al. found that this amount of time was significantly longer when the passages were played to the listeners using synthetic speech generated by the Haskins Parallel Resonant Synthesizer than when they were played using natural speech. The proportion of questions that listeners answered correctly did not vary according to the type of speech used, however.

Subsequent studies on the comprehension of continuous, connected passages of synthetic speech have yielded a similar pattern of results: it may take longer to process synthetic speech than natural speech, but the final levels of comprehension achieved for both types of speech are ultimately equivalent. In a replication of Nye et al.'s (1975) earlier study, Ingemann (1978) used the more advanced FOVE speech synthesizer and found no differences in performance between natural and synthetic voices in either the amount of time listeners took to complete the task or the number of multiple-choice questions they answered correctly. Pisoni and Hunnicutt (1980) assessed comprehension by having listeners answer multiple-choice questions based on passages that they had either read or heard spoken by a human voice or the MITalk text-to-speech system. They found that listeners answered more questions correctly if they had read the passages (rather than heard them), but that their level of performance did not differ between the natural voice and MITalk conditions. They also found that the percentage of questions that participants answered correctly improved most between the first and second halves of the experiment when the participants heard the MITalk versions of the passages, suggesting that some perceptual learning of this synthetic voice had occurred during the course of the experiment. Participants made similar, but smaller amounts of improvement between the two halves of the study in the natural voice and reading conditions.

**Online Measures: Sentence Verification.** Conclusive evidence showing that the comprehension of synthetic speech was more difficult for listeners than the comprehension of natural speech first began to emerge when researchers started applying more sensitive, online measures to the study of the comprehension process. Manous, Pisoni, Dedina and Nusbaum (1985) were the first researchers to use a sentence verification task (SVT) to study the comprehension of synthetic speech. They played listeners short, declarative sentences, which were either verified or falsified by the last word in the sentence and instructed the listeners to decide as quickly and as accurately as possible whether each sentence was true or false. The listeners recorded their responses by pressing one of two buttons on a response box and then writing down the sentence they heard. Manous et al. presented sentences to their listeners in this experiment using four different synthetic voices: DECTalk, Prose, Infovox and Votrax, as well as a human voice. They found that listener reaction time in the SVT was related to the intelligibility of the speech used, as determined by the number of transcription errors listeners made in writing down the sentences they heard. Manous et al. therefore concluded that the difficulty of encoding synthetic speech at the acoustic/phonetic level had a cascading effect on higher levels of the comprehension process.

In another study, Pisoni, Manous and Dedina (1987) followed up on Manous et al.'s (1985) earlier findings by investigating whether the increased latencies for synthetic speech in the SVT were due to the impoverished segmental cues in synthetic speech, or if they also resulted from independent difficulties the listeners incurred in processing synthetic speech beyond the level of acoustic-phonetic surface structure. Pisoni et al. assessed this hypothesis by testing listeners in a SVT using stimuli produced by human and synthetic voices that were matched in terms of segmental intelligibility. Pisoni et al. thus replicated the methodology from Manous et al. using a DECTalk voice that did not induce significantly more errors than a natural voice in a sentence transcription task. Figure 5 shows the average response latencies from this study, for true and false sentences, broken down by the type of voice used and the length of the sentence (three or six words). The reaction times for true sentences were significantly longer for the DECTalk voice than for the natural voice—despite the close match between the two voices in terms of segmental intelligibility. Pisoni et al. thus concluded that the comparatively impoverished acoustic-phonetic structure of DECTalk had detrimental effects on the comprehension process that did not emerge at the level of segmental encoding. In fact, this finding suggested that a level of processing devoted strictly to the segmental encoding of incoming speech might not even exist, since the ramifications of processing impoverished acoustic-phonetic cues seemed to persist beyond the segmental encoding stage. Pisoni et al.'s findings also suggested that the higher-level structures produced by the comprehension process were more impoverished—or more difficult to interpret—for the synthetic speech stimuli than they were for the natural speech stimuli, since the listeners' longer reaction times in the SVT did not correspond to any difficulties observed at the earlier segmental processing level.

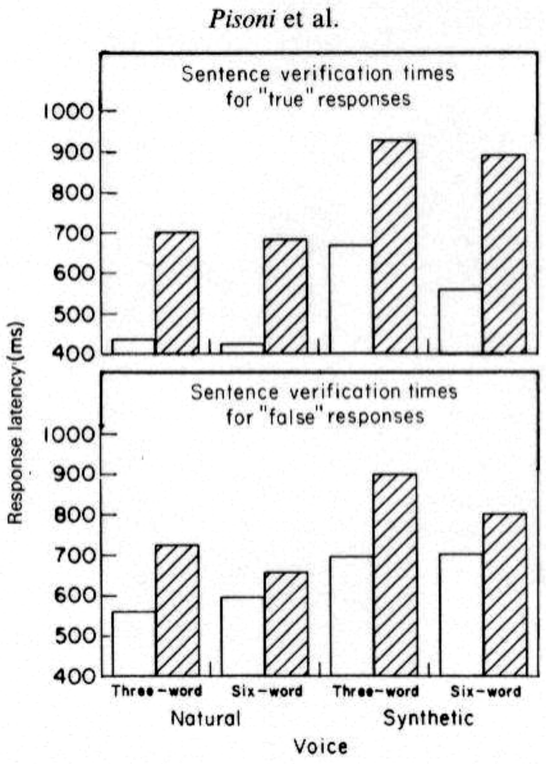


**Figure 5.** Response latencies for verification of false and true sentences, by voice type and length of sentence (adapted from Pisoni, Manous & Dedina, 1987).

**Comprehension and Recall.** Luce (1981), however, suggested that the difficulty of interpreting the acoustic-phonetic structure of synthetic speech may actually make the subsequent recall of synthetic speech easier than the recall of naturally produced speech. Luce drew upon this notion to account for a finding which showed that listener recall of individual words was better when they were produced by MITalk than when they were produced by a natural voice. Luce suggested that this followed from the extra cognitive effort listeners needed to make in order to encode synthetic speech successfully, which evidently resulted in a stronger memory trace for the particular acoustic details of the words the listener heard. Luce based this prediction on the results of a study in which listeners heard passages spoken in either MITalk or a natural voice. Luce then questioned the listeners about whether particular words, propositions or themes had appeared in those passage. Luce found that listeners were able to recall particular propositions and themes better when they heard the passage in a natural voice, rather than in synthetic speech. However, the listeners' memory for particular words was better when they had heard passages produced by the MITalk voice. Luce suggested that better recall for individual synthetic words followed from the fact that listeners had to expend greater amounts of cognitive effort and processing resources on encoding the acoustic-phonetic details of the synthetic speech into words. Allocating more cognitive effort to the process of word recognition, as it were, led to more robust lexical recall but produced poorer recall of more abstract propositional information.

Moody and Joost (1986) also measured the recall of words and propositions in synthetic and natural passages and found that recall interacted in unexpected ways with higher-level conceptual structures. They asked listeners multiple-choice questions from college and graduate entrance exams about passages which had been presented to them in DECTalk, LPC synthesis, and natural voices. Moody and Joost found that listeners answered more multiple-choice questions correctly when those questions dealt with topics in the naturally produced passages, rather than the synthetic ones. However, this difference in comprehension only held for the easier, or less complicated, propositions in the original passage; no differences in percent correct response were found for questions about more difficult passages. Their findings suggested once again that expending greater amounts of cognitive effort in interpreting the more difficult portions of a synthetically produced passage may help listeners recall those propositions just as well as their naturally produced counterparts later on. Their results also suggest that, as the comprehension and recall tasks become more difficult, post-perceptual measures of comprehension may be influenced more by alternate processing strategies and knowledge from beyond the immediate scope of the task.

Ralston, Pisoni, Lively, Greene and Mullennix (1991) further investigated the relationship between synthetic speech intelligibility and comprehension by having listeners perform a series of tests which used natural speech and Votrax synthetic speech stimuli. Their listeners first took an MRT, then had to monitor for a small set of words in a short passage, and then finally answered a series of true/false questions regarding the presence or absence of words and propositions in the passage they had just heard. In the word monitoring task, Ralston et al. found that accuracy was higher and response times lower for natural speech than for Votrax speech. Interestingly, word-monitoring latencies increased significantly for the more difficult (i.e., college-level) passages only for synthetic speech, indicating that the process of recognizing individual synthetic words shared processing resources with higher-level comprehension processes. However, in contrast to Luce (1981), Ralston et al. found that listeners' memory for both words and propositions was better when they had been presented in natural speech than when they had been produced by the Votrax synthesizer. Hence, despite the extra cognitive effort required to encode the phonetic details of these synthetic stimuli, they were still more difficult to recall from memory.

Ralston et al. (1991) also reported results from a novel sentence-by-sentence listening task which provided another on-line measure of the time course of speech comprehension. In this task, participants simply listen to a series of individual sentences and press a button when they are ready to move on from one sentence to the next. After listening to all of the sentences, the listeners are then tested on their memory of particular words or propositions in the sequence of sentences they just heard. Ralston et al. used this task to study the time it took listeners to comprehend sentences produced either by a human voice or by the Votrax synthesizer. Figure 6 shows the average amount of time it took listeners to move from one sentence to the next in this task, for both fourth grade- and college-level passages presented in synthetic and natural speech. Ralston et al. found that listeners in this experiment took significantly longer to complete this task when the sentences were produced by the Votrax system than when they were produced by a human voice. Since the listeners in this experiment had also taken the MRT, Ralston et al. looked for correlations between segmental intelligibility scores and the differences in the sentence-by-sentence listening times for the two voices. They found that the measures from the tasks were significantly correlated with one another, suggesting that the time course of comprehension depends on the segmental intelligibility of the speech. However, since the *r* values for the correlations between the listening times and the MRT scores varied between +.4 and +.6, this analysis showed that segmental intelligibility could not account completely for the corresponding differences observed in comprehension between synthetic and natural speech.

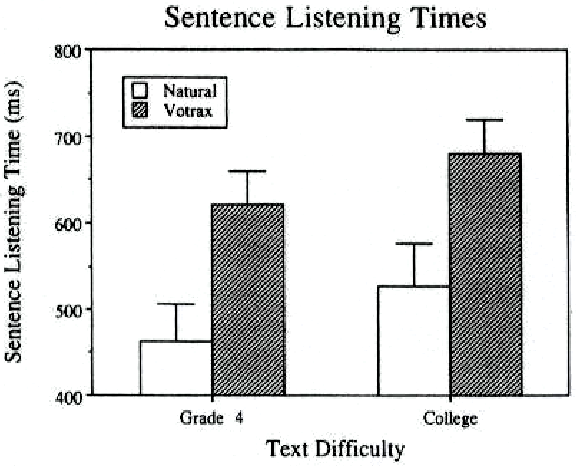


**Figure 6.** Average sentence-by-sentence listening times, by voice type and text difficulty. Error bars represent one standard error of the sample means (adapted from Ralston et al., 1991).

Paris, Gilson, Thomas and Silver (1995) followed up on Ralston et al.'s earlier study by investigating the comprehension of highly intelligible DECTalk speech, Votrax and natural speech. These investigators had listeners attend passively to passages produced by the three different voices, and then presented them with a series of true/false questions about the words and propositions which might have appeared in the various passages. In a separate condition, listeners were also asked to shadow (i.e., immediately repeat what they heard) two passages of synthetic or natural speech. Recall was better for both words and propositions produced by both DECTalk and natural voices than it was for items produced by Votrax. There were, however, no significant differences between either word or proposition recognition for DECTalk and natural speech. Nonetheless, shadowing accuracy was worse for DECTalk than it was for natural speech, and it was even worse for Votrax than it was for either of the other two voices. This on-line measure of perceptual processing indicated, once again, that there are perceptual difficulties in interpreting even the highest quality synthetic speech which disappear by the time the entire comprehension process has run its course.

**Summary.** Sensitive psycholinguistic tests of the time course of language comprehension have shown that it is more difficult to comprehend synthetic speech than natural speech. It takes longer, for instance, for listeners to verify whether or not synthetic sentences are true in a SVT than it does for them to verify naturally produced sentences. This effect holds even when natural and synthetic sentences are matched in terms of their segmental intelligibility, indicating that it is more difficult for listeners to process synthetic speech beyond the level of acoustic-phonetic interpretation than it is to process natural speech at this level. Processing deficits for synthetic speech also emerge in sentence-by-sentence listening tasks and shadowing tasks. Tests of the post-perceptual products of language comprehension have not always revealed greater listener difficulty in the comprehension of synthetic speech; however, listener recall of propositions from synthetically produced passages is often worse than recall of similar propositions from naturally produced passages. These findings suggest that listeners may be able to make up for the inherent difficulty of comprehending synthetic speech by implementing alternate processing strategies and drawing upon other sources of knowledge in order to complete a challenging comprehension task.

## Point 5: Perception of Synthetic Speech Improves with Experience

Despite the fact that listeners consistently have more difficulty perceiving synthetic speech than natural speech, their ability to perceive synthetic speech typically improves if they simply receive more exposure to it. Such perceptual learning of synthetic speech has been documented for a wide variety of intelligibility and comprehension tasks.

**Improvement on Broad Measures of Performance.** In their initial studies of the perception of synthetic speech produced by the Haskins parallel formant resonance synthesizer, Nye and Gaitenby (1973) found that listener performance on the MRT improved significantly over the course of their experiment. In the first of six testing sessions—each of which contained 150 different test items—listeners averaged 17.5% errors when they heard synthetic speech; by the sixth (and last) of these sessions, however, they averaged only 8% errors. Listener performance on natural speech tokens in the same task, on the other hand, maintained an average error rate of about 4 to 5% throughout the six different testing sessions, but it was close to ceiling.

Carlson, Granstrom and Larsson (1976) tested blind listeners' ability to repeat sentences of synthetic speech and found that their performance on this task also improved dramatically between the first and the last testing sessions in the experiment. Carlson et al.'s listeners took part in eight separate testing sessions, the final four of which took place one week after the first four. In each of these testing sessions, the listeners heard a unique list of 25 sentences—each of which they had to repeat—at two different speaking rates. They also listened to a 7-minute long short story. Carlson et al. tabulated the number of words and sentences the listeners repeated correctly and found that average percent correct scores increased from 52% words correct and 35% sentences correct in the first session to 90% words correct and 77% sentences correct in the final session. Furthermore, Carlson et al. found that the listeners maintained their improved performance over the week-long break between the fourth and the fifth sessions. In the fourth testing session, listeners averaged 85% words correct and 70% sentences correct; one week later, in the fifth testing session, listeners averaged 83% words correct and 65% sentences correct. This pattern of results showed that significant improvements in the perception of synthetic speech could not only be made with relatively little exposure to synthetic speech (i.e., 100 sentences), but also that such improvements could be maintained over comparatively long-term intervals.

Rounsefell, Zucker and Roberts (1993) reported an even more dramatic demonstration of the speed at which exposure to synthetic speech improves listeners' ability to perceive it. They tested the ability of high school-aged listeners to transcribe a pair of sentences as produced by either the DECTalk, VoicaLite or Echo II synthesizers. Half of the listeners received training on the particular synthetic voice they were to hear before the testing sessions began. This training consisted of three repetitions of three different sentences, as produced by the synthetic voice, each of which was followed by a live repetition of the same sentence, produced by the experimenter in a natural voice. Rounsefell et al. found that listeners who were trained in this way—essentially hearing nine tokens of synthetic sentences before testing began—were significantly better than untrained listeners at a subsequent sentence transcription task. The trained listeners correctly transcribed, on average, 9.11 syllables out of the 14 syllables in the two test sentences; the untrained listeners, on the other hand averaged only 3.75 syllables correct.

Venkatagiri (1994) observed that, even though listeners' ability to perceive synthetic speech may improve rapidly after a few initial exposures, this perceptual improvement may not necessarily continue indefinitely. Venkatagiri asked listeners to transcribe sentences produced by the Echo II synthesizer, and investigated how much their transcription accuracy improved over three consecutive days of testing. On each of these three days, the listeners transcribed a series of 20 different synthetic sentences. Venkatagiri found that listeners' transcription performance improved significantly between days one and two of testing (i.e., between the first 20 and the second 20 sentences), but that listener transcriptions were not significantly better on day three than they were on day two. Venkatagiri concluded that this failure to improve on day three may have been due to a ceiling effect, since the average percentage of correct transcriptions had already improved from 78.2% to 94.1% between days one and two, but were only able inch up to 96% on day three. Listeners' failure to post significantly higher scores on the third day of testing may therefore have been due to the fact that they had very little room left to improve beyond 94.1% correct. A task that does not set such easily reachable upper bounds on positive performance—as the sentence transcription task does—may yield continued perceptual improvements from exposure to synthetic speech over a longer interval of training.

**Improved Reaction Times.** As reviewed earlier in point 2, Slowiaczek and Pisoni (1982) found that listeners exhibit similar continued improvement in the perception of synthetic speech when tested on an auditory lexical decision task. They had listeners perform a lexical decision task, while listening to both natural and synthetic speech stimuli, over the course of five consecutive days. Figure 7 shows the average response time and the percentage of errors listeners made in this task, for each of the five consecutive days of training, for both synthetic and natural stimuli. This figure shows that listener response times for natural and synthetic non-word stimuli decreased significantly over the five days of training; however, the corresponding proportions of correct responses remained essentially constant for the duration of the study. This finding suggested that response time measurements might reflect continued improvements in perception even though performance on a coarser-grained measure such as response accuracy had already reached ceiling.

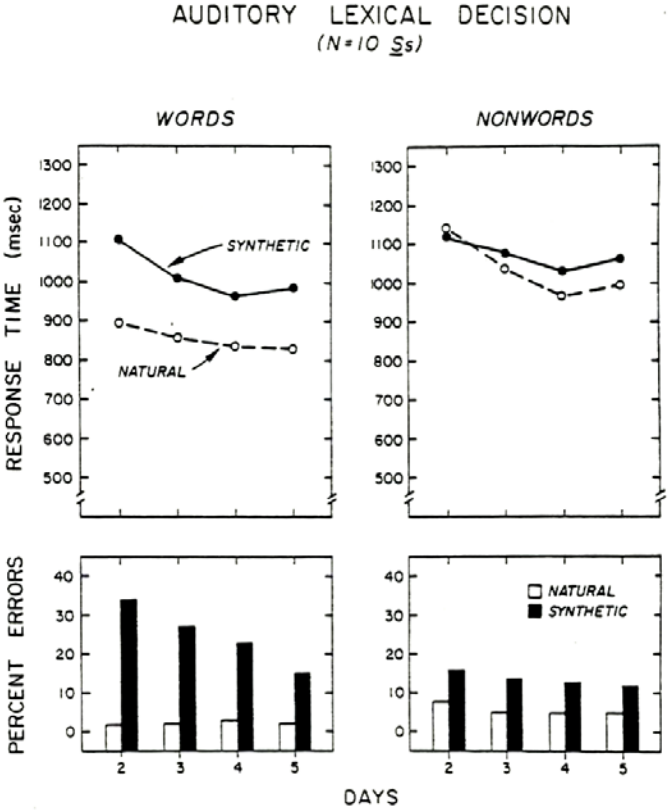


**Figure 7.** Lexical decision response times and errors, for words and non-words, by voice type and day of study (adapted from Slowiaczek & Pisoni, 1982).

Reynolds, Isaacs-Duvall, Sheward and Rotter (2000) also found that increased exposure to synthetic speech significantly decreased listeners' response times to synthetic speech sentences in a SVT. Reynolds et al. found that listeners exhibited such improvements even without explicit training on the SVT itself. One group of listeners in Reynolds et al. was trained for eight consecutive days—in between a sentence verification pre-test and post-test—on a sentence transcription task, in which they heard the same kinds of sentence stimuli that were used in the pre- and post-tests (as produced by DECTalk) without actually practicing the SVT itself. Another group of listeners in the study received no training or exposure to synthetic speech in between the sentence verification pre- and the post-tests. The authors found that trained listeners showed improvement between pre- and post-test in terms of the speed with which they verified synthetic <u>false</u> sentences; neither listening group's RTs decreased significantly for synthetic <u>true</u> sentences. Reynolds et al. suggested that this failure to improve may have resulted from a ceiling effect on performance with the true sentences, which had induced relatively short RTs in the initial pre-test. However, Reynolds et al.'s listeners also performed the SVT with naturally produced stimuli in both the pre- and the post-tests. Response times for these sentences were consistently shorter in both testing sessions than they were for the synthetically produced sentences. Moreover, Reynolds et al. found that the RTs for <u>both</u> groups of listeners decreased for these natural stimuli between the pre- and the post-test. Reynolds et al. suggested that this improvement may have resulted from increasing listener familiarity with the individual natural voices used to produce the stimuli.

Reynolds, Isaacs-Duvall and Haddox (2002) pursued the issue of listener improvement in the SVT further, questioning whether listeners' ability to comprehend synthetic speech could become as good as their ability to comprehend natural speech stimuli, after extended exposure to synthetic speech. Reynolds et al. thus had listeners participate in a five-day long testing sequence, during which they performed the SVT for 80 different sentences (40 produced by a natural voice, 40 produced by DECTalk) on each day. Reynolds et al. found that RTs were shorter for natural sentences than for synthetic sentences on all five days. Listener RTs also decreased significantly for both types of sentences between the first and the last days of the experiment. However, most of the improvement occurred between the first and second days of the experiment, and listener RTs had more or less bottomed out for both voices by the fifth day of testing. Nonetheless, Reynolds et al. charted regression lines to track the idealized future course of RT progress, beyond the fifth day of testing, for both the synthetic and the natural voices in this study. These regression lines indicated that the slopes of the natural and synthetic RTs were still diverging by the fifth day of testing. Even though listener performance improved for both voices in this study, that is, continued listener improvement on the <u>natural</u> voice was greater than continued improvement on the synthetic voice. This finding indicated that, no matter how much exposure listeners got to synthetic speech, their ability to comprehend it would still remain worse than their ability to comprehend natural speech, given equivalent amount of practice listening to both types of voices in a particular testing paradigm.

Pisoni and Hunnicutt (1980) also reported that listeners could rapidly improve their ability to comprehend synthetic speech over the course of an experiment. The listeners in this study answered multiple-choice questions based on passages they had either read, heard spoken by a natural voice, or heard spoken by the MITalk synthesizer. Table 3 shows the percentage of questions the listeners answered correctly in each half of the experiment, for all three presentation conditions. The percentage of questions that the listeners answered correctly increased between the first and second halves of the experiment for both types of auditory presentation, although the amount of improvement was greater when listeners heard the MITalk voice.

Although Pisoni and Hunnicutt did not run any tests to determine if the levels of improvement across both halves of the study were statistically significant, their results were nonetheless noteworthy because the comprehension of synthetic speech appeared to improve more than the comprehension of natural speech—even though comprehension for both types of spoken passages was roughly equivalent during the first half of the experiment.

| | 1st Half | 2nd Half |
|---|---|---|
| MITalk | 64.1 | 74.8 |
| Natural | 65.6 | 68.5 |
| Reading | 76.1 | 77.2 |

**Table 3.** Percentage of comprehension questions answered correctly, by voice type and experiment half (adapted from Pisoni & Hunnicutt, 1980).

**Perceptual Learning interacts with task type.** Schwab, Nusbaum and Pisoni (1985) reviewed the improvements reported in the earlier studies of Pisoni and Hunnicutt (1980) and Slowiaczek and Pisoni (1982) and questioned whether or not the findings might be the result of listeners simply becoming more proficient at carrying out the experimental tasks, rather than becoming better at perceiving synthetic speech. In order to test this hypothesis, Schwab et al. gave listeners extensive training on a variety of perceptual tasks and investigated whether the type of stimuli used in the training had any effect on the amount of improvement the listeners made in these tasks. The tasks included: the closed-set MRT, word recognition using phonetically balanced (PB) word lists, transcription of (meaningful) Harvard sentences, transcription of (meaningless) Haskins sentences, and yes/no comprehension questions based on a series of prose passages the listeners had heard. On the first day of testing, listeners participated in all of these tests, in which they were played stimuli produced by the Votrax synthesizer. On days two through nine of the study, the listeners were split into two groups—one which received training on all of these various tasks and another which received no training at all. The group of listeners that received training on the tasks was further split into two sub-groups: those who heard synthetic (Votrax) speech stimuli during the training tasks, and those who heard natural speech stimuli during the tasks. After eight days of receiving training—or no training—all groups of listeners repeated the same battery of tests as on the first day of the experiment, in which they once again heard all stimuli as produced by the Votrax synthesizer.

Schwab et al. (1985) found that those listeners who had been trained on the various perceptual tasks with synthetic speech stimuli showed significantly greater improvement on the word recognition and sentence transcription tasks between pre- and post-test than did either the untrained group or the listeners who had been trained with natural speech stimuli. Furthermore, the natural speech group did not show any more improvement than the untrained control group. Schwab et al. interpreted these results as evidence that listener improvement in perceiving synthetic speech was not merely due to a practice effect, since the group trained with natural speech stimuli had received the same amount of training and practice with the tasks as the synthetic speech group, and yet they showed less improvement in perceptual performance between the pre-test and the post-test. Schwab et al. argued that the observed perceptual improvement was due to exposure to synthetic speech *per se*—that is, listeners became better at extracting acoustic-phonetic information from the synthetic speech signal as they gained increasing amounts of exposure to it. This perceptual ability is evidently domain-specific and cannot be acquired by simply being exposed to equivalent amounts of natural speech.

Schwab et al. (1985) reported that their listeners made the largest amount of improvement in the PB word recognition and meaningless sentence transcription tasks. The authors pointed out that these tasks had less constrained response sets than did the closed-set MRT and the meaningful sentence transcription tasks. Since listeners could depend less on higher-level semantic and structural information to perform well in the PB and meaningless sentence tasks, their correspondingly greater improvement in these conditions must have been due primarily to developing their proficiency at interpreting the lower-level, acoustic-phonetic details in the synthetic speech signal. Corresponding improvements at interpreting such information in the MRT and the Harvard sentence tasks may have been masked by the listeners' ability to rely on higher-level linguistic information to perform these tasks well. Schwab et al. also ran a follow-up study, six months after the original training experiment, which showed that the group trained on synthetic speech stimuli still maintained better levels of performance on the various perceptual tasks than the group of listeners who were trained on natural speech. Schwab et al. therefore concluded that the effects of training had long-term benefits for the perception of synthetic speech.

**Learning Effects: Generalization.** Another way to interpret the results of Schwab et al. (1985) is that the natural speech group acquired little ability to generalize from their training with natural speech stimuli to the test conditions with synthetic speech stimuli. In another training study, Greenspan, Nusbaum and Pisoni (1988) questioned whether there were similar limits on the ability to make generalizations of particular synthetic speech training stimuli. Greenspan et al. thus adapted the training paradigm of Schwab et al. to determine how well listeners could generalize from particular sets of training stimuli to improve their performance on tests using novel synthetic speech stimuli. Greenspan et al. divided their listeners into five different training groups: one which received training on a new set of individual, novel words on each of the four successive days of training; a second group, which was trained on the same set of repeated words on each of the four training days; a third group, which listened to sets of novel sentences on all four days of training; a fourth group, which listened to the same set of repeated sentences on all four days of training; and a fifth group, which received no training on the four days between the pre and post-tests. The pre- and post-tests consisted of the closed-set MRT, the PB open-set word recognition task, and the transcription of Harvard and Haskins sentences. Each group of listeners—aside from the listeners who underwent no training at all—thus received training which was directly relevant to only two of these four testing conditions. Training on individual words was likely to benefit listeners on the MRT and PB word tasks, while training on sentences was likely to benefit listeners on the Haskins or Harvard sentence transcription tasks. Greenspan et al.'s training groups were also further divided between those who listened to novel stimuli everyday and those who listened to repeated sets of stimuli in order to determine whether it was easier for listeners to generalize from training sets which had greater amounts of variability in them.

While Greenspan et al. (1988) found no effects of novel versus repeated sets of training stimuli on listener performance in the testing conditions, they did find asymmetric effects of training between sentence and word recognition tasks. The groups that received training on the sentence transcription tasks showed improved performance on all four tasks in the study. The groups that received only training on individual novel or repeated words, however, only showed improved performance between pre- and post-test on the MRT and the PB word tasks; their performance levels on the two sentence transcription tasks remained the same. Greenspan et al. suggested that the failure of individual word training to transfer to the sentence transcription task may have been the result of the listeners' inability to parse the word boundaries in a stream of synthetic speech. The transfer of sentence training to the word recognition tasks, on the other hand, may have resulted from listeners simply being exposed to a large number of synthetic words in a wide range of contexts during the four days of training.

Greenspan et al. (1988) further investigated the perceptual consequences of variability in training stimuli in a follow-up experiment. In their second study, they trained two different listener groups on the PB word recognition task. After a 50-word pretest, one of these groups received training and feedback on the task with 200 different, novel words, while the other group received training on only 20 repetitions of 10 words which they had already heard in the pre-test. In a post-test, both groups of listeners were presented with a set of 10 words from the pre-test, along with 40 novel word stimuli. Both groups showed improvement in their ability to perceive these novel words; however, the group which had heard more diverse training stimuli showed more improvement than the other group. Greenspan et al. concluded that improvement in the perception of synthetic speech depends, to a large extent, on the amount of acoustic-phonetic variability in the synthetic speech samples that listeners are exposed to during training.

**Summary.** Listeners' ability to perceive synthetic speech improves as they receive more exposure to it. Such improvement has been shown in a wide variety of behavioral tasks, including word recognition, sentence transcription, lexical decision, and sentence verification. This improvement can occur quite rapidly, even manifesting itself after exposure to only a few sentences of synthetic speech. Such perceptual improvement may still persist for as long as six months—and maybe even longer—after initial exposure to synthetic speech. Research has also shown, however, that there are limitations on the amount of improvement that listeners can make in the perception of synthetic speech. Even after substantial amounts of training on synthetic speech stimuli in a sentence verification paradigm, for instance, listeners cannot verify synthetic sentences as quickly as they do natural sentences. Furthermore, improvement in the perception of synthetic speech depends to some extent on the type of training or exposure that listeners receive. Training on the transcription of synthetic sentences improves listeners' ability to identify individual synthetic words, for instance, but training on the identification of individual synthetic words does not improve listeners' ability to transcribe whole synthetic sentences. Other findings also indicate that the amount of improvement listeners make in the perception of synthetic speech depends on the amount of variability in the training stimuli they are exposed to. Increased familiarity with synthetic speech may thus alleviate, but not overcome, the fundamental limitations that poor acoustic-phonetic quality places on listeners' ability to perceive synthetic speech as well as they perceive natural speech.

## Point 6: Alternative Populations Process Synthetic Speech Differently

The studies reviewed above have found consistent perceptual deficits in the processing of synthetic speech by listeners who are typically college-aged, normal-hearing, native speakers of English. Studies which have investigated how other groups of listeners perceive synthetic speech have typically found that these alternative populations of listeners process synthetic speech differently than their college-aged, normal-hearing, native-speaker counterparts. In most cases, these alternative groups have even more difficulty perceiving and processing synthetic speech.

**Non-Native Listeners.** Greene (1986), for instance, found that non-native listeners of English have significantly more difficulty perceiving synthetic speech (in English) than native listeners do. Greene tested both native and non-native speakers of English on the MRT and a sentence transcription task, using stimuli produced by a natural voice and MITalk. Table 4 shows the percentage of synthetic and natural words that both groups of listeners correctly identified in the MRT. The non-native listeners in this study performed only slightly worse on this task for the natural voice than the native listeners did, with scores at or near ceiling. However, the performance gap between the two listening groups significantly increased when they listened to MITalk speech, with the native listeners remaining near ceiling but the non-native listeners decreasing substantially in accuracy.

| | Natural | MITalk |
|---|---|---|
| Natives | 99 | 93 |
| Non-Natives | 95 | 86 |

**Table 4.** Percent Correct on MRT by voice type and listener group (adapted from Greene, 1986).

However, Greene (1986) found no interaction between listening group and voice type in the transcription of either semantically anomalous or meaningful sentences. The non-natives' percent correct sentence transcription scores were consistently worse than the natives' scores by the same amount, for both natural and MITalk voices, as shown in Tables 5a and 5b below. This may be due to the fact that the native listeners performed at ceiling for the natural voice for both sentence types, and near ceiling for the MITalk voice for meaningful sentences.

| | Natural | MITalk |
|---|---|---|
| Natives | 98 | 79 |
| Non-Natives | 70 | 51 |

**Table 5a.** Words correctly transcribed from semantically anomalous sentences, by voice type and listener group (adapted from Greene, 1986).

| | Natural | MITalk |
|---|---|---|
| Natives | 99 | 93 |
| Non-Natives | 70 | 64 |

**Table 5b.** Words correctly transcribed from meaningful sentences, by voice type and listener group (adapted from Greene, 1986).

Despite the lack of an interaction in the sentence transcription scores, the gap between native and non-native listeners' scores was much greater in the sentence transcription task than it was in the MRT. This result indicated that native listeners were significantly better than non-natives in drawing upon higher-level linguistic knowledge to interpret whole sentences produced in synthetic speech—along with being more proficient at interpreting the low-level acoustic cues for individual synthetic speech segments in isolated words.

Greene (1986) also noted that non-native listeners not only scored worse on both the MRT and the sentence transcription tasks, but they also showed a wider range of variability in their percent correct scores than did their native-listener counterparts. Greene tested another group of non-native listeners on a sentence transcription task and found that their percent correct scores on this test correlated highly with their scores on the TOEFL (*r* = +.83) and on the English Proficiency Test (*r* = +.89). Greene thus concluded that the ability of non-native listeners to perceive synthetic speech depended greatly on their proficiency in the language being spoken.

More recently, Reynolds, Bond and Fucci (1996) also found that non-native listeners transcribed synthetic speech sentences less accurately than native English listeners. The participants in their study were asked to listen to pairs of thematically related sentences and then transcribe the second sentence in each pair. Reynolds et al. presented these sentence pairs to both native and non-native listeners of English in quiet and noisy (multi-talker babble at +10 dB SNR) conditions. The authors found that transcription accuracy was significantly higher for native listeners (96.3% correct) than non-native listeners (54.5% correct) in the quiet condition. The authors also found that introducing babble noise into the signal had a significantly greater detrimental effect on the non-natives' performance; their percent correct scores dropped 8.7% between quiet and noisy conditions, while the native listeners' scores dropped only 2.8%. Reynolds et al. suggested that this drop in performance probably occurred simply because of the non-native listeners' relative unfamiliarity with the English language and their corresponding inability to interpret unusual new forms of it. Furthermore, Reynolds et al. noted a wide variation in percent correct transcription scores in their study and suggested that, as in Greene (1986), non-natives' ability to perceive synthetic speech depended in large part on their proficiency in the English language.

**Children.** Children form another group of "alternative listeners." Greene and Pisoni (1988) studied the perception of synthetic speech by children and found that they, like their college-aged counterparts, had more difficulty processing synthetic speech than natural speech, on a variety of perceptual tasks. Greene and Pisoni developed a four-alternative, forced-choice, picture-word matching task, using items from the Peabody Picture Vocabulary Test, to investigate children's ability to understand synthetic speech. Kindergartners who participated in this test correctly identified 82% of words produced by the Prose 2000 speech synthesis system (Groner, Bernstein, Ingber, Pearlman & Toal, 1982). They correctly identified 94% of natural speech tokens, however. Second graders who took this test performed better, with 94% correct for synthetic speech tokens and 98% correct for natural speech tokens. Therefore, the performance of both groups of children on this task was analogous to the levels of performance obtained with adult listeners, in that they tended to perceive synthetic speech less well than natural speech.

Greene (1983) found that children show similar deficits in their ability to recall synthetic speech items. Greene tested the ability of fourth graders to recall in any order the items on lists of two to eight digits in length, as produced by either a natural or a synthetic voice. The children in this task correctly recalled 82.9% of the items in the naturally produced lists, but only 78.5% of the synthetically produced list items. Greene got similar results from a task involving the free recall of nouns by children. In this task, children listened to strings of nouns, spoken in either a natural or a synthetic voice, and were asked to repeat back the items in the string in any order. All of the strings were from two to 10 words in length. Children consistently recalled more words correctly from the naturally produced lists than they did from the synthetically produced lists. They also had more difficulty, for both types of voices, in recalling items from the longer lists. This pattern of results replicated earlier findings from studies using adult listeners such as Luce et al. (1983), which showed that the recall of synthetically produced items is more difficult than the recall of naturally produced items for college-aged listeners, as well, due to the comparative difficulty they have in interpreting and encoding the acoustic-phonetic structure of synthetic speech.

Mirenda and Beukelman (1987) directly compared the performance of adults and children on Greene's (1983) battery of perceptual tasks, using natural and synthetic stimuli. Table 6 shows the percentages of correct responses adults and children gave in transcribing individual words presented in four different voices (natural speech, DECTalk, Echo and Votrax synthesizers). These results showed that children consistently performed worse on this single-word transcription task than adults, for both natural and synthetic voices. Also, the younger children listeners—who were between 6 and 8 years old—performed worse than the older children, who were between 10 and 12 years old.

| | Natural | DECTalk | Echo | Votrax |
|---|---|---|---|---|
| Adults | 99.2 | 78.0 | 39.6 | 57.2 |
| 10-12 | 96.9 | 76.0 | 37.6 | 56.4 |
| 6-8 | 93.6 | 72.0 | 34.0 | 48.4 |

**Table 6.** Percent individual words correctly transcribed, by voice type and listener group (adapted from Mirenda & Beukelman, 1987).

Since children performed worse than adults in this task for all voices, the greater difficulties that children seemed to have in perceiving synthetic speech may, in fact, have been due to the greater difficulties they had in performing the perceptual task itself. Mirenda and Beukelman (1987) also administered a sentence transcription task to these child and adult listeners; Table 7 shows the percentage of words in these sentences that each group of listeners transcribed correctly, for each of the four voices. These results show an even greater discrepancy between the performance of adults and listeners in this task than there was in the individual word transcription task.

| | Natural | DECTalk | Echo | Votrax |
|---|---|---|---|---|
| Adults | 99.3 | 96.7 | 68.2 | 83.8 |
| 10-12 | 96.4 | 90.2 | 61.8 | 68.2 |
| 6-8 | 94.2 | 81.1 | 35.8 | 57.1 |

**Table 7.** Percent words correctly transcribed from meaningful sentences, by voice type and listener group (adapted from Mirenda & Beukelman, 1987).

All groups performed better on this task than they did on the individual word transcription task—especially for the synthetic voices. This result indicated that children, like adults, draw on higher-level linguistic knowledge to interpret synthetic speech. Since the adult transcription scores improved in the sentence transcription task more than the children's did, however, this pattern of results suggested that the adult listeners were better at accessing and using this higher-level linguistic information than the children were.

**Children's Comprehension of Synthetic Speech.** Reynolds and Fucci (1998) reported that children also have more difficulty comprehending synthetic speech than natural speech. They measured children's comprehension of naturally and synthetically (DECTalk) produced stimuli using a SVT. The children they tested, who were between the ages of 6 and 11, consistently responded faster to natural sentences than synthetic sentences. Reynolds and Jefferson (1999) expanded upon this finding by testing child listeners from two separate age groups, 6 to 7 years old and 9 to 11 years old, in an identical experimental paradigm. They found an equivalent processing advantage for natural versus synthetic sentences for all listeners, but also found that the 9 to 11 year-olds comprehended both kinds of sentences faster than the 6 to 7 year-olds. This pattern of results mirrored the findings of Mirenda and Beukelman (1987; 1990) for similar groups of listeners performing transcription tasks. It is thus possible that the differences in sentence verification times between 6-7 year-old and 9-11 year-old listeners may be due to the corresponding difficulties each group has in interpreting the acoustic-phonetic structure of synthetic speech, rather than higher-level comprehension difficulties in post-perceptual processing. However, the differences between the two groups may also reflect the extent to which older children are simply better at doing perceptual tasks of this kind.

Reynolds and Jefferson (1999) also tested their listeners again in a second session, either 3 or 7 days after the first day, on the same task. They found that the 6 to 7 year-olds had significantly faster response times in the second testing session than they did in the first, but that they 9 to 11 year-olds showed no significant differences in response times between sessions. The younger children's RTs improved primarily in their responses to synthetic sentences. Reynolds and Jefferson suggested that such improvement may have been due to the 6 to 7 year-old group's perceptual systems being more flexible than those of the older children, and thus more capable of quickly adjusting to novel stimuli. However, the 6 to 7 year-olds' RTs were quite long in the first session—especially for synthetic sentences—and thus left ample room for improvement, while the 9 to 11 year-olds' RTs may have initially been closer to ceiling performance.

**Older Adults.** Little is known about the extent to which older adults' perception of synthetic speech differs from that of younger adults. Sutton, King, Hux and Beukelman (1995) investigated perceptual differences between listener preferences for particular rates of synthetic speech. Two groups of listeners were used in this experiment: 20 older adults from 61 to 79 years of age and 20 younger adults from 21 to 28 years of age. All listeners heard sentences produced by DECTalk, which ranged in rate from 120 to 250 words per minute. After hearing each sentence, the listeners indicated to the experimenter what their subjective comfort level was in hearing the sentence, on a Likert scale from 1 to 7; a rating of 1 indicated "too slow" while a rating of 7 indicated "too fast." Sutton et al. found that younger adults had a broad "comfort range"—which was defined as sentences receiving scores between 3 and 5 on the Likert scale—for sentences spoken at rates between 150 to 220 words per minute. The older listeners, on the other hand, preferred slower rates of synthetic speech, from 130 to 210 words per minute. Sutton et al. suggested that a number of different factors might lead older listeners to prefer slower rates of synthetic speech--among them a decreased ability in temporal processing, hearing loss, and global changes in auditory perception.

**Hearing-Impaired Listeners.** Studies investigating older listeners' perception of synthetic speech have often been framed within the context of hearing impairment and its potential effects on the perception of synthetic speech. Kangas and Allen (1990), for instance, tested two groups of listeners between the ages of 49 to 64; the members of one group had normal hearing while the members of the other group had all acquired hearing losses. Both groups of listeners transcribed individual words that were produced by either DECTalk or a natural male voice. Kangas and Allen found that both groups of listeners transcribed naturally produced words correctly more often than synthetically produced words; they also found that hearing-impaired listeners performed worse on this task than normal-hearing listeners did. Importantly, there was no interaction between voice type and listener group; this suggested that synthetic speech did not exacerbate the difficulty these hearing-impaired listeners had in perceiving speech. Rather, the performance of hearing-impaired listeners in this task could best be understood by simply combining the deficits that synthetic speech and hearing impairment both impose individually on normal speech perception processes. Interestingly, Kangas and Allen also noted that there was much more individual variability in the hearing-impaired listeners' synthetic speech transcription scores than there was in the normal-hearing listeners' scores. Kangas and Allen concluded that most of this variability could be accounted for by the corresponding variability in the hearing-impaired listeners' performance on the natural speech transcription task.

In another study of hearing-impaired listeners, Humes, Nelson and Pisoni (1991) found that age also does not interact with hearing-impaired listeners' performance on the MRT. They tested a group of 66-73 year-old hearing-impaired listeners along with two groups of 19-22 year-old listeners on both the closed and open-response formats of the MRT. One of the groups of 19-22 year-old listeners heard the MRT items in the clear, while the other heard them through spectrally-shaped noise, which was designed to mimic the effects of hearing loss. Humes et al. included this condition in order to test the extent to which the decreased performance of the hearing-impaired listeners—in comparison to the younger listeners—was due to their hearing-impairment, rather than their increased age. All groups of listeners heard the MRT items as they were produced by a natural voice and by both the DECTalk and Votrax synthesizers. Figure 8 shows the percentage of word recognition errors each group of listeners made on the MRT while listening to the three different voices. All listeners in Humes et al. performed better on the open-response MRT when they heard the DECTalk and natural voices than when they heard the Votrax synthesizer. There were no significant differences between performance levels on DECTalk and the natural voice, except for the normal-hearing young listeners, who showed a small but significant advantage with the natural voice. These normal-hearing listeners were also better at the MRT task, for all three voices, than either the hearing-impaired or masked-noise listening groups. There were no significant differences between the hearing-impaired and masked-noise groups, indicating that the decreased level of performance on the MRT by the older, hearing-impaired listeners could be accounted for solely by their hearing impairment, regardless of their age. However, greater variability was observed within the group of hearing-impaired listeners than for either of the younger groups of listeners. Humes et al. pointed out that individual performance on the MRT by the hearing-impaired listeners corresponded closely to their level of hearing loss in pure-tone average (*r* = -.73, -.75, -.8 for the natural, DECTalk, and Votrax voices, respectively). Their performance in the synthetic speech condition was also strongly correlated with their performance on the natural speech condition (*r* = +.96 and +.90 for DECTalk and Votrax, respectively). As in the earlier study by Kangas and Allen (1990), therefore, hearing impairment did not interact with the degraded quality of the synthetic speech to produce further perceptual difficulties for the listener under the MRT testing conditions.

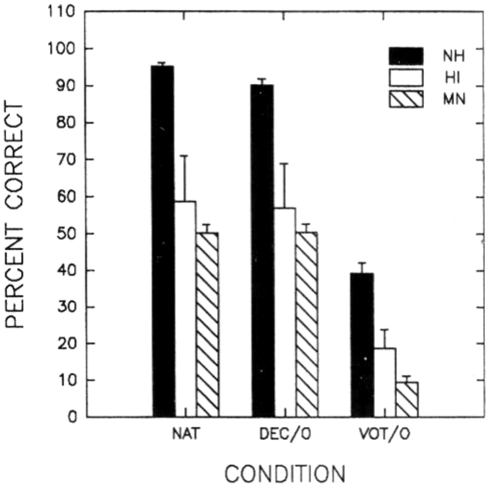


**Figure 8.** Percent of words correctly identified on the MRT, by voice type and listener group (NH = normal-hearing; HI = hearing-impaired; MN = masked-noise) (adapted from Humes et al., 1991).

Humes, Nelson, Pisoni and Lively (1993) tested older listeners' perception of synthetic speech using a serial recall task. The listeners in this study included both younger (21-24 years old) and older (61-76 years old) normal-hearing adults, who were presented with sequences of 10 words and were asked to repeat those words in the order that they heard them. The listeners heard these lists in both natural and synthetic voices. After performing the serial recall task, Humes et al. also had listeners transcribe all of the individual words that had been used in the lists (both synthetic and natural) in order to measure the intelligibility of the test words. The results of these two tests replicated earlier findings that synthetic speech was both less intelligible and more difficult to recall than natural speech. While the younger listeners had a slight advantage in the recall task, there were no significant differences between age groups on the transcription task, indicating that the processing of the acoustic-phonetic structure of synthetic speech does not significantly deteriorate with age in normal-hearing listeners.

**Expert Listeners.** One final group of alternative listeners—which the existing literature on synthetic speech perception has hardly studied—is expert listeners. With the increasing prevalence of speech synthesis products on the market today, this population of listeners is undoubtedly growing. Blind listeners, for example, may benefit from speech synthesis programs on their computers which can output text for them to hear in the form of speech. Such listeners have extensive practice listening to the specific form of synthetic speech that their computers produce. Their ability to extract information from such synthetic speech may be far more advanced than those that listeners in the short-term training experiments, such as Schwab et al. (1985), may be able to develop over a 10-day period.

While no studies that we know of have directly investigated the perceptual abilities of such "expert listeners," Hustad et al. (1998) did compare the perceptual abilities of a group of "speech synthesis experts" to those of a group of speech-language pathologists (who had never had more than incidental contact with synthetic speech) and another group of listeners who were completely unfamiliar with synthetic speech. Hustad et al.'s "speech synthesis experts" all had extensive experience listening to synthetic speech—they all had worked on a regular basis with both DECTalk and MacinTalk Pro (the two synthetic speech voices used in the study) for at least five hours a week for a year or more prior to the study. All listeners in this study took the MRT, listening to items produced by both DECTalk and MacinTalk Pro voices. Not surprisingly, the expert listeners outperformed both the speech-language pathologists and the inexperienced listeners on this task, in both voice conditions. Hustad et al. thus concluded that extensive experience listening to synthetic speech improved listeners' ability to extract information from the synthetic speech signal. Moreover, Hustad et al. reasoned that similar amounts of experience listening to natural speech in an analytic fashion could not improve the perception of synthetic speech, since the speech-language pathologists did not outperform the inexperienced listeners on the MRT. This finding is analogous to the results of Schwab et al. (1985), who found that only training on synthetic speech—and not natural speech—improved listener performance on recognizing words and transcribing sentences produced in synthetic speech. The extent to which expert listeners' performance on the same task might differ from performance by listeners who have only been trained on synthetic speech stimuli for a period of eight days, however, remains unknown.

**Summary.** The research on "alternative" groups of listeners indicates that these listeners display more variation in perceiving synthetic speech than is commonly revealed in experiments on the perception of synthetic speech by young, normal-hearing, native listeners. Non-native listeners and children, for instance, have more difficulty perceiving synthetic speech than native-speaking, college-aged listeners in a variety of behavioral tasks, including the MRT, sentence transcription, lexical recall and sentence verification. "Expert" listeners, on the other hand, appear to have developed an ability to extract information from synthetic speech signals which surpasses those which inexperienced listeners are able to develop in laboratory training experiments. However, the perceptual abilities of highly experienced listeners remain largely unstudied, along with those of older adults. Most research on older adults' perception of synthetic speech has focused on the effects of hearing impairment on the perception of synthetic speech; this research has shown that hearing impairment does not interact with the poor quality of synthetic speech to further degrade synthetic speech intelligibility. Accounting for this wide range of variability in listeners' perception of synthetic speech is an important consideration in the development of robust speech synthesis applications.

## Point 7: Prosodic Cues, Naturalness and Acceptability.

Research on the perception of prosodic or suprasegmental features in synthetic speech has often approached prosodic quality as a matter of subjective listener preference—a factor that could possibly make certain kinds of synthetic speech more acceptable, or natural-sounding, to listeners. Researchers have thus treated prosody primarily as a voice quality issue to be dealt with once the initial research goal of making synthetic speech as intelligible and comprehensible as natural speech had been met. The fact that synthetic speech may often sound monotonous or unnatural to listeners might at that point be investigated independently to improve highly intelligible speech synthesis for the sake of listener comfort and preference. This approach to researching prosody and naturalness in synthetic speech treats naturalness and prosody as independent of the intelligibility or comprehensibility of the synthetic speech signal (Nusbaum, Francis & Henly, 1995). It contrasts sharply with another line of research which has shown that the intelligibility of synthetic speech depends to some extent on the quality and appropriateness of the prosodic cues in the speech signal (Slowiaczek & Nusbaum, 1985; Paris, Thomas, Gilson & Kincaid, 2000; Sanderman & Collier, 1997).

**Prosodic Influences on Subjective Judgments of Naturalness, Preference, and Acceptability.** One early study on the subjective evaluation of synthetic speech was carried out by Nusbaum, Schwab and Pisoni (1984), who investigated subjective responses to synthetic and natural speech passages by simply asking listeners how they would rate the voices in the passages on a variety of subjective impressionistic scales (e.g., interesting vs. boring, grating vs. melodious, etc.). Nusbaum et al. also asked listeners to indicate how much they would trust each voice to provide them with particular kinds of information (e.g., tornado warnings, sports scores, etc.). Finally, Nusbaum et al. presented their listeners with a series of comprehension and recall questions based on the passages they had heard produced by the three different voices used in the study (natural speech, MITalk, and Votrax) and then asked the listeners how confident they were that they had comprehended the passages correctly.

The authors found that the listeners tended to give the natural voice higher ratings than both synthetic voices on descriptions that would be considered preferable (e.g., smooth, friendly, polished). A few adjectives (easy, clear, pleasant and fluent) also teased apart listener preference for the two different synthesizers in that MITalk (which was more intelligible than Votrax) was also judged more preferable by the listeners in these descriptive terms. Listeners also placed the most trust in the natural voice, followed by MITalk and then Votrax. Although listeners' comprehension of passages produced by all three voices was essentially equivalent, regardless of the voice used, their confidence in their ability to comprehend these passages was much lower for the two synthetic voices. This finding indicated that the post-perceptual comprehension task—which essentially asked listeners to recall particular facts and words, or to draw inferences based on the passages they had just heard—was not sensitive enough to capture subtler difficulties in the process of comprehending synthetic speech that the listeners were consciously aware of.

Terken and Lemeer (1988) assessed the interaction between naturalness and intelligibility in terms of their respective influences on the perceived "attractiveness" of individual Dutch sentences produced by speech synthesis. They recorded a passage of 21 sentences, read in a natural voice, and then re-analyzed the entire recorded passage using LPC synthesis. Terken and Lemeer did this twice, once using 30 LPC coefficients and another time using only 6; the re-analysis with 30 coefficients produced a version of the passage with relatively high segmental intelligibility, while the version with 6 LPC coefficients had low segmental intelligibility. Terken and Lemeer also created additional versions of these passages—for both LPC re-analyses—using either a copy of the natural prosodic contour or just a flat (monotone) pitch contour throughout. They then played all four versions of these passages—either whole or in a sequence of individual sentences—to a group of Dutch listeners. In each condition, the listeners were instructed to rate how "attractive" each stimulus sounded on a scale from 1 to 10. When listeners heard the entire passage of 21 sentences all at once, Terken and Lemeer found that both intelligibility and prosody affected the ratings of attractiveness: listeners rated highly intelligible passages as more attractive than the less intelligible passages, and they rated the versions with natural prosody as more attractive than those produced in monotone. Interestingly, the effect of segmental intelligibility on the attractiveness ratings was stronger than the effect of prosody.

The attractiveness judgments the listeners made when they heard the passages one sentence at a time yielded interesting interactions between the intelligibility and prosody factors. For the highly intelligible sentences, listeners judged the natural prosody versions to be more attractive than those produced in monotone; for the low intelligibility sentences, however, listeners rated both prosodic versions as equally attractive. Terken and Lemeer suggested that these results could be accounted for by assuming that the perception of prosody is strongly dependent on segmental intelligibility. Listeners evidently need time to adjust to speech with poor segmental quality before they can make judgments about its prosodic qualities. While listening to synthetic speech one sentence at a time, listeners apparently do not have enough time to perceive prosodic differences in poorly intelligible speech; however, with highly intelligible speech, they do, and it is under these conditions that the prosodic contributions to perceived "attractiveness" emerge.

In more recent studies of prosody, researchers have looked at listeners' subjective assessments as a means of evaluating the quality of particular prosodic features in synthetic speech. Terken (1993), for instance, evaluated systems of prosodic rules for synthetic speech by playing passages produced with those rules to "experienced phoneticians" and then asking them to rate the naturalness of each passage on a scale from 1 to 10. The listeners heard two versions of each passage—one which played the entire, 10 sentence-long passage in order, and another version which scrambled the order of the individual sentences within the passage. Both versions of these passages were produced by a diphone synthesizer using either the natural pitch contour or one of two different sets of synthetic prosody rules: an older set of prosodic rules calculated declination patterns over the entire utterance, while a newer set of rules established intonational boundaries within the utterance and reset the declination pattern after each one. Terken's listeners rated the passages produced with the newer set of synthetic prosody rules as more natural-sounding than the passages produced with the older set of rules, in both the scrambled and ordered-passage conditions. In the scrambled passages condition, the new set of rules not only improved the perceived naturalness of individual sentences over the old set of rules, but also made the sentences sound as natural to the "expert phoneticians" as the utterances produced with a natural intonation pattern. In the complete, ordered text condition, however, the passages with natural intonation sounded significantly more natural to the listeners than the passages with either set of synthetic prosody rules. Furthermore, the naturalness of only the passages with the natural intonation pattern was greater in the ordered condition than in the scrambled condition; the naturalness of the passages with the synthetic rule sets was slightly worse in the ordered condition. This pattern of results indicated that there are discourse-level prosodic patterns in the naturally produced speech which make it sound more natural than texts which have been produced with only sentence-level prosodic rules.

Sanderman and Collier (1996) further developed Terken's (1993) sets of prosodic rules and assessed how their improved rule sets affected listeners' acceptance of and preference for synthetic speech. Their new sets of prosodic rules essentially induced varying levels of boundary strength into synthetically produced pitch contours by independently adjusting features for phrase contour, boundary pause length, and declination reset. Sanderman and Collier played pairs of identical synthetic sentences—one of which had these prosodic phrasing rules and one of which did not—to a group of listeners and asked them which sentence of the pair they preferred. A listener preference for the prosodically phrased sentences emerged for sentences which were longer than about nine words; for sentences that were shorter than this, the listeners indicated no clear preference. In a subsequent experiment, Sanderman and Collier tested listener acceptability of sentences that had been produced with prosodic rule sets using different numbers of possible phrase boundary strengths. One set of sentences had two possible phrase boundary strengths, another had three, another had five, and yet another set of sentences was produced with natural prosody. Sanderman and Collier played sentences produced with each set of prosodic rules to untrained listeners and then asked them to rate how acceptable each sentence was on a scale from 1 to 10. In general, the listeners preferred the sentences produced with natural prosody to those produced with the artificial prosody rules. There was no significant difference in acceptability between the sentences with natural prosody and those with five different possible phrase boundary strengths, however. Sanderman and Collier's results therefore indicated that appropriate prosody not only becomes a more important contributor to the naturalness of synthetic speech as the synthetic speech segment becomes longer, but also that greater amounts of variability in the implementation of prosodic rules in synthetic speech significantly improves the perceived naturalness of that speech.

**Prosody and Naturalness Interact with Comprehension and Intelligibility.** Early studies on the perception of prosody in synthetic speech yielded only marginal evidence that prosodic cues contributed significantly to the overall intelligibility of synthetic speech. Slowiaczek and Nusbaum (1985) investigated the importance of prosody to synthetic speech intelligibility by presenting listeners with both meaningful (Harvard) and meaningless (Haskins) sentences, produced with either a “hat-pattern” (typical of declarative sentences) or a flat (monotone) pitch contour. Listeners also heard the sentences at two different speaking rates—150 and 250 words per minute—as produced by the Prose 2000 system; their task was to simply transcribe the words as they heard them. Slowiaczek and Nusbaum found that both speaking rate and semantic content influenced the accuracy of listeners’ transcriptions. Percent correct transcriptions were significantly higher for sentences produced at a slower rate of speech and for meaningful sentences. Slowiaczek and Nusbaum did not, however, find that the “hat pattern” increased transcription accuracy over the monotone pitch contour. Appropriate prosody, that is, did not have a consistent effect on individual word intelligibility.

In a follow-up experiment, Slowiaczek and Nusbaum (1985) explored the possibility that appropriate prosody might improve intelligibility if listeners heard a wider variety of syntactic constructions in the sentences they had to transcribe. They expanded upon the paradigm they used in their first experiment by including declarative sentences with active, passive and center-embedded syntactic constructions. These various sentences could also be either long or short. Slowiaczek and Nusbaum found that the listeners had less success transcribing the longer sentences and the center-embedded sentences in this study. They also found that sentences with the “hat pattern” prosody proved easier for the listeners to transcribe than sentences with monotone pitch contours. Their findings suggest that prosodic information in synthetic speech is useful to a listener when the syntactic structure of a sentence is not predictable.

Nusbaum et al. (1995) maintained that subjective judgments of preference or acceptability were highly dependent on the intelligibility of the synthetic speech signal. Nusbaum et al. therefore attempted to develop measures of naturalness that were independent of segment intelligibility. For example, in one experiment, Nusbaum et al. attempted to measure the “naturalness” of synthetic glottal pulse sources. The authors constructed one second-long vowels (/i/, /u/ and /a/) from sets of either one or five individual glottal pulses excised from both natural and synthetic (DECTalk, Votrax) speech. Nusbaum et al. played these vowels to listeners and asked them to identify—as quickly as possible—whether the vowels had been produced by a human or a computer. The “naturalness” of any particular vowel stimulus in this task was taken to be the probability that listeners would classify it as having been produced by a human. This classification experiment yielded an unexpected pattern of results: the “naturalness” of any given stimulus depended on both its vowel quality and the particular voice with which it was produced. The listeners consistently identified the /u/ vowels, for instance, as having been produced by a computer. /a/ vowels, on the other hand, exhibited an unexpected pattern of identification: DECTalk /a/ was more consistently identified as “human” than either of the natural /a/ productions (which, in turn, were more natural than the Votrax /a/). Only /i/ vowels were identified by listeners along the classification hierarchy that Nusbaum et al. expected: natural productions were consistently identified as “human” more often than DECTalk /i/, which was also more consistently classified as “human” than Votrax /i/. Nusbaum et al. suggested that this mixed pattern of results may have emerged because the higher frequency components of the glottal source waveform—which are present in the second formant of /i/ but not of /u/ or /a/—are important to identifying its naturalness.

In another experiment, Nusbaum et al. (1995) attempted to assess the naturalness of prosodic information independently of the influences of segmental intelligibility. To accomplish this, they low-pass filtered words produced by two human talkers and two synthetic speech voices (DECTalk and Votrax) at 200 Hz, thus removing segmental cues but preserving the prosodic information in the original speech sample. Nusbaum et al. then played these low-pass filtered stimuli to listeners and asked them to determine as quickly as possible whether the stimuli had been produced by a human or a computer. Listeners identified these stimuli just as they identified the /i/ stimuli in the previous experiment: the human voices were consistently more "natural" than the DECTalk voice, which, in turn, was more "natural" than the Votrax voice. Nusbaum et al. thus concluded that low-pass filtered words provided a reliable, intelligibility-free measure of the naturalness of any given voice. The results of this study also indicated that the lexical-level prosody of even a high-quality synthesizer such as DECTalk was still clearly inferior to the prosody of natural speech.

Sanderman and Collier (1997) showed that generating synthetic sentences with appropriate prosodic contours facilitates comprehension—and that, moreover, inappropriate prosodic contours may make comprehension more difficult. They demonstrated the importance of prosody to synthetic speech comprehension by investigating listener interpretations of a series of syntactically ambiguous sentences. For example, one of the sentences in Sanderman and Collier's study was (translated from the Dutch) "I reserved a room in the hotel with the fax." This sentence is ambiguous because "with the fax" may describe either how the room was reserved or a particular facility that the hotel has. Prosodic phrasing may help disambiguate these two interpretations. For instance, placing an intonation break between "the hotel" and "with the fax" discourages listeners from grouping these two phrases together within the same noun phrase; hence, this intonational phrasing supports the interpretation wherein "with the fax" describes how the reservation was made. Without an intonation break in "the hotel with the fax," listeners are more likely to interpret the fax as being one of the hotel's facilities. However, without any disambiguating information in sentences like these, listeners tend to be biased towards one interpretation of the sentence over another. To measure this bias, Sanderman and Collier presented written versions of sentences like the one above to untrained listeners of Dutch and asked them to circle a paraphrase of the most likely interpretation of the sentence. Using this method, they determined what the most likely interpretation of each ambiguous sentence was: the most likely interpretation of the example sentence, for instance, was the one in which "with the fax" described how the hotel reservation had been made.

Sanderman and Collier (1997) investigated how prosodic phrasing might interact with the inherent interpretive bias in these syntactically ambiguous sentences by having listeners answer questions which might further bias them towards one interpretation over another. Listeners read a question such as "How did I reserve the hotel room?" and then answered it based on information they heard spoken to them in a synthetically produced sentence. The prosodic rule set used to produce these sentences enabled five different levels of phrase boundary strength—this rule set being the one that produced the most natural-sounding stimuli in Sanderman and Collier (1996). Each particular sentence was produced with one of three phrasings: one which supported the most likely interpretation of the sentence, another which supported the less likely interpretation of the sentence, and another which had no phrasing boundaries. This "zero" prosody version was included in order to establish a response baseline with which to compare the effects of the other two prosodic phrasings.) Sanderman and Collier (1997) measured the amount of time it took listeners to respond to the written question after they had heard the inquired-after information in the synthetically produced answer (e.g., the <u>fax</u>). Sanderman and Collier (1997) found facilitory effects for matched questions and answers: a less likely answer combined with a less likely question reduced response times in comparison to the "zero" prosody answers, just as more likely answers in conjunction with more likely questions did. The mismatched conditions yielded a different set of effects: a more likely response matched to a less likely context question significantly increased response times over the baseline condition; however, less likely answers matched to more likely context questions did not. Thus, it appears that listeners have more difficulty undoing a bias towards a less likely interpretation of the sentence. The broader implication of Sanderman and Collier's (1997) work is that appropriate prosodic phrasing is necessary for optimal comprehension of synthetic speech, and inappropriate phrasing can actually make it more difficult under certain conditions for listeners to comprehend synthetic speech.

Paris et al. (2000) also found that prosodic information plays a role in enabling listeners to recall what they have heard in a sentence. These authors investigated the effects of prosody on lexical recall by constructing stimuli both with and without sentence-level prosodic information using both natural and synthetic (DECTalk, SoundBlaster) voices. Paris et al. also looked at the effects of meaningful, sentential contexts on listeners' ability to recall words, and created four different types of stimuli: 1. meaningful sentences with normal intonation; 2. meaningful sentences in monotone; 3. meaningless sentences with normal intonation; 4. meaningless strings of words with no sentence-level prosody. The first two sets of sentences were combinations of Harvard sentences, averaging 15 to 20 words in length; the third set consisted of similar sentences, except with the content words changed to make meaningless sentences (e.g., "Add house before you find the truck..."); and the fourth set consisted of strings of unrelated words (e.g., "In with she plate storm after of proof..."). Listeners were asked to listen to each of these stimuli once and then immediately recall as much of the stimulus as they could. The listeners recalled items from meaningful sentences more easily than items from meaningless strings, and naturally produced words were easier to recall than synthetically produced words. No significant differences were found in recall between DECTalk- and SoundBlaster-produced items. There were several interesting interactions between voice type and presentation context. Table 8 below shows the percentage of lexical items correctly recalled in each condition:

| | Condition 1 | Condition 2 | Condition 3 | Condition 4 |
|---|---|---|---|---|
| | Normal | No prosody | No meaning | Unstructured |
| Natural | 74 | 60 | 51 | 24 |
| DECTalk | 60 | 60 | 35 | 20 |
| SoundBlaster | 58 | 58 | 34 | 16 |

**Table 8.** Percentage of items correctly recalled, by voice type and presentation context (adapted from Paris et al., 2000).

In the two conditions using normal, sentence-level prosody—Conditions 1 and 3—natural speech items displayed a significant recall advantage over synthetic speech items. This advantage disappeared, however, in Condition 2, in which the listeners heard meaningful sentences with only lexical-level prosody. In Condition 4, there was only a significant difference in correct recall percentages between the natural items and the SoundBlaster items; there was not a significant difference between DECTalk and natural speech. These results indicated, once again, that broad, sentence-level prosodic cues can help listeners not only comprehend speech better, but also help them recall later what they have heard. The results also suggest that the comparative inability of listeners to recall synthetic speech items may not necessarily be due to the impoverished acoustic-phonetic structure of those items, but rather to the failure of the speech synthesis algorithms to incorporate those sentence-level prosodic cues in a natural way.

After the immediate recall task, Paris et al. (2000) played their listeners more samples of sentences produced by the three individual voices and asked them to make subjective judgments of how intelligible and natural-sounding each sentence sample was, on a scale from 1 to 10. The results of this second task are given in Tables 9 and 10.

| | Condition 1<br>Normal | Condition 2<br>No prosody | Condition 3<br>No meaning | Condition 4<br>Unstructured |
|---|---|---|---|---|
| Natural | 9.86 | 7.80 | 9.27 | 5.90 |
| DECTalk | 8.20 | 7.74 | 6.13 | 6.07 |
| SoundBlaster | 7.10 | 6.39 | 5.22 | 4.11 |

**Table 9.** Intelligibility ratings by voice type and presentation context (adapted from Paris et al., 2000).

| | Condition 1<br>Normal | Condition 2<br>No prosody | Condition 3<br>No meaning | Condition 4<br>Unstructured |
|---|---|---|---|---|
| Natural | 9.71 | 5.58 | 9.49 | 5.23 |
| DECTalk | 5.78 | 4.19 | 4.70 | 4.27 |
| SoundBlaster | 4.60 | 3.86 | 3.67 | 3.05 |

**Table 10.** Naturalness ratings by voice type and presentation context (adapted from Paris et al., 2000).

Sentence-level prosody had the same effect on intelligibility ratings as it did on free recall scores—with sentence-level prosody, listeners perceived natural speech tokens as much more intelligible than both types of synthetic speech tokens, but, without sentence-level prosody, there was no significant perceived intelligibility difference between the natural and DECTalk voices (and SoundBlaster was still rated as less intelligible than the other two voices). Sentence-level intonation played an even more significant role in the naturalness judgments, where the natural stimuli were rated considerably higher than the synthetic stimuli in Conditions 1 and 3, and still slightly higher than the synthetic stimuli in Conditions 1 and 2.

These results demonstrate that the perceived intelligibility of sentence stimuli corresponds closely to the listeners' ability to recall the content words from these sentences. Since sentence-level prosody influenced both intelligibility and naturalness ratings, it is likely that it is necessary to incorporate appropriate prosodic patterns into synthetic speech for it to match the intelligibility and retention of natural speech in all listening conditions. Paris et al.'s (2000) subjective naturalness ratings also indicated that sentence-level prosody dictates—more than any other factor—the potential level of perceived naturalness in synthetic speech stimuli. Since Paris et al.'s natural tokens were still judged to be more natural than synthetic tokens in conditions which lacked sentence-level prosody, however, some sub-prosodic features of human speech evidently contribute to perceived naturalness as well, such as glottal source characteristics, as noted by Nusbaum et al. (1995).

**Summary.** The available research has shown that the naturalness of synthetic speech depends on a variety of factors, including segmental intelligibility, the pragmatic appropriateness of particular pitch contours, and the amount of variability in the implementation of synthetic prosody. Assessing the naturalness of synthetic speech independently of these factors has proven difficult, and attempts to do so have primarily focused on evaluating the naturalness of isolated voice source information in the synthetic speech signal. Research has also shown that appropriate prosodic information may facilitate the comprehension and recall of sentences produced synthetically. Together, the findings of this body of research suggest that generating more naturalistic and appropriate sentence-level prosody will be an important research challenge in the effort to develop more natural-sounding, highly intelligible synthetic speech in the years to come.

## Conclusions and Directions for Future Research

Research on the perception of synthetic speech has always had both practical and theoretical motivations. Practically, research on the perception of synthetic speech can reveal what limitations there are on the comprehension of synthetic speech in applied settings, and what work needs to be done in order to overcome those limitations and improve the quality of speech synthesis. Theoretically, research on the perception of synthetic speech is important because it offers a unique window into how human listeners can deal with and extract information from unnatural and impoverished speech signals. By comparing the perception of such impoverished speech signals with natural speech, researchers can obtain new fundamental knowledge about which aspects of natural speech are important to the robust perceptual abilities of human listeners in a wide variety of listening environments.

The practical implications of research findings on the perception of synthetic speech over the past 30+ years are clear. Research on the perception of synthetic speech in noise, for instance, has revealed that, even though the segmental intelligibility of current speech synthesis systems may approximate natural speech under clear listening conditions, speech perception deteriorates rapidly and significantly in noisy listening conditions. This finding suggests that synthetic speech applications would be of limited utility in e.g., noisy work environments, in football stadiums, or even over a cell phone. Investigating ways to overcome the persistent limitations on synthetic speech intelligibility, however, provides several promising opportunities for future research. For instance, audio-visual speech synthesis systems provide a possible solution to the practical problem of improving the intelligibility of synthetic speech in noise (Cohen & Massaro, 1994; Ezzat & Poggio, 2000). The information in the speech signal that people can perceive in the visual domain is largely complementary to that which they can perceive in the auditory domain (Calvert, Spence & Stein, 2004; Massaro, 1997; Summerfield, 1987) and furthermore, provides robust cues to those features of speech which are most often misperceived in noisy listening conditions. Visual information about speech has been shown to provide substantial gains in intelligibility when audio-visual stimuli are presented in noise (Sumby & Pollack, 1954). Although viable systems for producing synthetic audio-visual speech have been around for some time, little is known about people's ability to perceive the visual information in the synthetic speech tokens that these systems produce. Is the visual-only perception of synthetic speech significantly different from the visual-only perception of natural speech? And, does synthetic visual speech boost the intelligibility of synthetic speech in noise as much as natural visual cues help improve the intelligibility of natural speech in noise? Answering basic research questions such as these may prove important not only to our future understanding of the utility of audio-visual speech synthesis, but also to our understanding of the fundamental characteristics that are important for multimodal perception of natural speech.

Research on the perception of synthetic speech also indicates that it requires more cognitive resources than the perception of natural speech. This finding suggests that synthetic speech applications might be of limited utility to listeners who are engaged in cognitively demanding tasks such as driving a car, flying an airplane, directing air traffic, etc. One way to limit such demands on listeners is to reduce the number of possible messages that a synthetic speech system can transmit. In diagnostic tests, such as the MRT, listeners can correctly identify synthetic speech tokens nearly as well as natural speech tokens when they only have to select responses from a small, limited set of response alternatives. In more cognitively demanding environments, therefore, synthetic speech may be better suited to transmitting only a small number of messages to listeners—such as, for example, warnings in an airplane cockpit (Simpson & Williams, 1980), or digits or letters of the alphabet. However, the production of a limited set of messages could be more easily handled by a recorded database of those messages, rather than a text-to-speech system which is designed to produce an unlimited number of messages (see Allen, Hunnicutt & Klatt, 1987).

The potential for success of any synthetic speech application also depends on the listener who is using the system. Research on the perception of synthetic speech has shown that alternative groups of listeners—e.g., children, non-native speakers, and listeners with hearing impairment—have more difficulty understanding synthetic speech than the average, normal-hearing, college-aged native speaker does. The ability of all groups of listeners, however, to extract information from synthetic speech tends to improve the more they are exposed to it. The limits of such improvement on the perception of synthetic speech are not precisely known. Improvement in the ability to perceive synthetic speech nonetheless has, by its very nature, the potential to create another alternative group of listeners. "Expert listeners," for instance, may emerge among people who listen to particular forms of synthetic speech for long periods of time on a daily basis. Such "expert listeners" may include, for instance, blind people who regularly use applications on their computers which output textual information to them in the form of synthetic speech. Such listeners would likely have far more experience listening to synthetic speech than any one listener might receive in a perceptual training experiment in the laboratory, and they may have thus developed correspondingly better abilities to extract acoustic-phonetic information from synthetic speech signals. Research on this potential group of listeners may reveal what upper limits (if any) exist on the fundamental human ability to extract information from synthetic speech, and may also determine if extensive amounts of listening experience can close the gap in intelligibility between synthetic and natural speech. It may also be instructive to investigate the extent to which expert knowledge of one form of synthetic speech may improve the perception of other forms of synthetic speech.

The finding that synthetic speech has been consistently shown to be less intelligible and less comprehensible than natural speech—across a wide variety of testing conditions and listener populations—has led researchers to draw a number of important theoretical conclusions about the aspects of natural speech which facilitate robust speech perception. These conclusions emerged from a consideration of how natural speech differs from synthetic speech that was produced by rule, since this was the form of synthetic speech that has been most commonly tested in synthetic speech perception research. Even very high quality synthetic speech produced by rule lacks both the rich variability and acoustic-phonetic cue redundancy characteristic of natural speech, and it is presumably the absence of these dynamic natural speech characteristics—along with the lack of appropriate prosodic information—which makes synthetic speech produced by rule difficult for listeners to perceive and comprehend.

However, not all synthetic speech is produced strictly by rule. "Concatenative" speech synthesis techniques, for instance, use natural human utterances as a voice source. Concatenative synthesis thus has a natural-sounding quality which has helped increase its popularity and use in recent years. Despite this increase in popularity, however, relatively little is known about the perception of speech produced by concatenative synthesis techniques and how it may differ from either synthetic speech produced by rule or natural speech. Using natural speech utterances as source material should improve the quality of synthetic speech in some putatively important respects—such as incorporating robust and redundant sets of cues to individual segments into the signal, for instance. However. using natural speech utterances as source material may also potentially damage the quality of synthetic speech in other ways—by introducing, for example, perceptible discontinuities between two adjoining units in the concatenated speech signal. It is perhaps not surprising, therefore, that existing research on the intelligibility of concatenative versus formant synthesis (e.g., Venkatagiri, 2003) indicates that concatenative synthesis produces highly intelligible consonants—for which it preserves the natural cues and formant transitions—but somewhat less intelligible vocalic segments, where the discontinuities between the concatenated source units typically exist.

The recent findings reported by Venkatagiri (2003) confirm that the acoustic-phonetic cue redundancy in natural speech is, indeed, important to perception. More research needs to be done, however, not only on the differences which may exist in perception between concatenative and formant synthesis—especially with respect to the vast body of research findings from the past 30+ years—but also on the specific role that acoustic-phonetic variability plays in speech perception. "Variability" has a wide variety of sources in speech—e.g., speaker, dialect, gender, age, emotional state, social identity, etc.—and little is known about what role (if any) these different sources of variability may play in facilitating the perception of natural speech. Incorporating different sources of variability into synthetic speech is a logical way to test their effects on speech intelligibility, and it may also provide a potential means of improving the overall intelligibility of synthetic speech. (Stevens, 1996)

Although the intelligibility of high quality speech synthesis systems currently approximates that of natural speech in clear listening conditions, improving its intelligibility in adverse listening conditions still remains an important research goal. One increasingly popular application of speech synthesis is providing spoken output of driving directions from computerized, on-board navigation systems in automobiles. The drivers who need to understand these spoken directions must do so in frequently noisy listening conditions and under a significant cognitive load. Research has shown that the perception of even high quality synthetic speech deteriorates significantly under such listening conditions. Determining how to improve the intelligibility of synthetic speech in noisy and cognitively demanding listening conditions should therefore help improve the viability and safety of such applications of speech synthesis.

Improving the quality of synthetic speech may also require a better understanding of what makes synthetic speech sound "unnatural" to most human listeners. However, even after three decades of research on the perception of synthetic speech, little is known about "naturalness" and its relationship to speech perception. Most research on the perception of synthetic speech has focused, instead, on studying the segmental intelligibility and comprehension of synthetic speech in comparison to natural speech. Now that the segmental intelligibility of synthetic speech has reached near-natural levels, however, determining what makes one kind of synthetic speech sound more natural or preferable to listeners should become an increasingly important research goal.

It is likely that incorporating appropriate prosodic contours into synthetic speech will increase its perceived "naturalness." Paris et al. (2000) found, for instance, that removing sentence-level prosody from natural speech not only diminished its perceived naturalness but also reduced its perceived intelligibility to a level comparable to that of synthetic speech. Listeners' ability to interpret synthetic speech therefore seems to depend to some extent on the presence of appropriate prosodic cues in synthetic speech. Testing the perceived "appropriateness" of particular prosodic contours, however, may prove more difficult because listeners cannot, in general, describe the prosodic information they perceive in linguistic terms. Thus, it may not be possible to investigate the perception of prosody in synthetic speech directly; instead, future research in this area may have to investigate the perception of prosody using indirect methods by focusing on the effects that prosodic structure has on, e.g., measures of memory, attention, processing load and processing speed, in quiet and noise, under a wide range of listening conditions, both with and without cognitive loads. Such research may require the development of entirely new assessment methods in order to better understand how prosodic information can facilitate or inhibit these elements of the perception of synthetic speech (Pisoni, 1997).

Developing new experimental methods which can target higher-level comprehension processes may also enable researchers to investigate and identify the scope of the detrimental effects of synthetic speech—as compared to natural speech—on the human language comprehension system. Duffy and Pisoni (1992) logically suggested that the phonological encoding of speech stimuli must occur before the interpretation of message-level information in comprehension, even though processing at both levels could go on concurrently. Synthetic speech appears to make processing more difficult at both the phonological and message levels; however, it is unclear whether this is due solely to the difficulties inherent in interpreting synthetic speech at the phonetic level—which may have cascading effects on subsequent stages in the comprehension process—or due to independent difficulties in the processing of synthetic speech that emerge at the stage of message-level semantic interpretation. Developing new behavioral tasks and better assessment methods which can target higher-level comprehension processes, independently of low-level phonemic encoding effects, could help clarify whether the difficulty in perceiving synthetic speech exists at the level of segmental interpretation alone, or whether it also causes specific, independent problems for semantic processing. Examining how these distinct levels of processing interact with one another in the perception of synthetic speech may shed light on how these levels operate together in the comprehension of natural speech as well.

Considerations such as these on the future directions of research on the perception of synthetic speech reflect the fact that speech scientists will likely have to develop more sophisticated methods as speech synthesis technology continues to improve. No matter what technological developments may come to pass, however, research on synthetic speech perception will continue to have direct practical benefits for speech synthesis technology, as well as provide a unique opportunity for speech scientists to investigate what makes synthetic speech hard for human listeners to understand and which aspects of natural speech help them perform the task of normal speech perception so well.

## References

Allen, J., Hunnicutt, M.S. & Klatt, D. (1987). *From Text to Speech: The MITalk System.* New York: Cambridge University Press.

Atal, B.S. & Hanauer, S.L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America, 50,* 637-655.

Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge: MIT Press.

Calvert, G., Spence, C. & Stein, B.E. (2004). *The Handbook of Multisensory Processing.* Cambridge, MA: MIT Press.

Carlson, R., Granstrom, B. & Larsson, K. (1976). Evaluation of a text-to-speech system as a reading machine for the blind. *Quarterly Progress and Status Report, STL-QPSR 2-3.* Stockholm: Royal Institute of Technology, Department of Speech Communications.

Clark, J.E. (1983). Intelligibility comparisons for two synthetic and one natural speech source. *Journal of Phonetics, 11,* 37-49.

Cohen, M.M. & Massaro, D.W. (1994). Synthesis of visible speech. *Behavior Research Methods, Instruments and Computers,* 22, 260-263.

Duffy, S.A. & Pisoni, D.B. (1991). Effects of sentence context on the signal duration required to identify natural and synthetic words. In *Research on Speech Perception Progress Report No. 17,* (Pp. 341-354). Bloomington, IN: Speech Research Laboratory, Indiana University.

Duffy, S.A. & Pisoni, D.B. (1992). Comprehension of synthetic speech produced by rule: a review and theoretical interpretation. *Language and Speech, 35(4),* 351-389.

Dutoit, T. & Leich, H. (1993). MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database. *Speech Communication, 13,* 435-440.

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

Ezzat, E. & Poggio, T. (2000). Visual Speech Synthesis by Morphing Visemes. *International Journal of Computer Vision, 38,* 45-57.

Fairbanks, G. (1958). Test of phonemic differentiation: the rhyme test. *Journal of the Acoustical Society of America, 30,* 596-600.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague, Netherlands: Mouton.

Greene, B.G. (1983). Perception of synthetic speech by children. In *Research on Speech Perception Progress Report No. 9,* (Pp. 335-348). Bloomington, IN: Speech Research Laboratory, Indiana University.

Greene, B.G. (1986). Perception of synthetic speech by nonnative speakers of English. In *Proceedings of the Human Factors Society,* 1340-1343. Santa Monica, CA.

Greene, B.G., Manous, L.M. & Pisoni, D.B. (1984). Perceptual evaluation of DECTalk: a final report on version 1.8. In *Research on Speech Perception Progress Report No. 10,* (Pp. 77-128). Bloomington, IN: Speech Research Laboratory, Indiana University.

Greene, B.G. & Pisoni, D.B. (1988). Perception of synthetic speech by adults and children: research on processing voice output from text-to-speech systems. In L.E. Bernstein (Ed.), *The Vocally Impaired: Clinical Practice and Research,* (pp. 206-248). Philadelphia: Grune & Stratton.

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.

Groner, G.F., Bernstein, J., Ingber, E., Pearlman, J. & Toal, T. (1982). A real-time text-to-speech converter. *Speech Technology, 1,* 73-76.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28,* 267-283.

Hoover, J., Reichle, J., Van Tasell, D. & Cole, D. (1987). The intelligibility of synthesized speech: Echo II versus Votrax. *Journal of Speech and Hearing Research, 30,* 425-431.

House, A.S., Williams, C.E., Hecker, M.H.L. & Kryter, K.D. (1965). Articulation-testing methods: consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America, 37,* 158-166.

Humes, L.E., Nelson, K.J. & Pisoni, D.B. (1991). Recognition of synthetic speech by hearing-impaired elderly listeners. *Journal of Speech and Hearing Research, 34,* 1180-1184.

Humes, L.E., Nelson, K.J., Pisoni, D.B. & Lively, S.E. (1993). Effects of age on serial recall of natural and synthetic speech. *Journal of Speech and Hearing Research, 34,* 1180-1184.

Hustad K.C., Kent R.D. & Beukelman D.R. (1998). DECTalk and MacinTalk speech synthesizers: intelligibility differences for three listener groups. *Journal of Speech Language and Hearing Research 41,* 744-752.

Ingemann, F. (1978). Speech synthesis by rule using the FOVE program. In *Haskins Laboratories Status Report on Speech Research, SR-54,* (pp. 165-173). New Haven, CT: Haskins Laboratories.

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(5),* 1337-1351.

Kangas, K.A. & Allen, G.D. (1990). Intelligibility of synthetic speech for normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Disorders, 55,* 751-755.

Koul, R., & Allen, G. (1993). Segmental intelligibility and speech interference thresholds of high-quality synthetic speech in presence of noise. *Journal of Speech and Hearing Research, 36,* 790-798.

Logan, J.S., Greene, B.G. & Pisoni, D.B. (1989). Segmental intelligibility of synthetic speech produced by rule. *Journal of the Acoustical Society of America, 86,* 566-581.

Luce, P.A. (1981). Comprehension of fluent synthetic speech produced by rule. In *Research on Speech Perception Progress Report No. 7,* (Pp. 229-242). Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P.A., Feustel, T.C. & Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors, 25,* 17-32.

Luce, P.A. & Pisoni, D.B. (1983). Capacity-demanding encoding of synthetic speech in serial-ordered recall. In *Research on Speech Perception Progress Report No. 9* (pp. 295-309). Bloomington, IN: Speech Research Laboratory, Indiana University.

Manous, L.M. & Pisoni, D.B. (1984). Effects of signal duration on the perception of natural and synthetic speech. In *Research on Speech Perception Progress Report No. 10* (pp. 311-321). Bloomington, IN: Speech Research Laboratory, Indiana University.

Manous, L.M., Pisoni, D.B., Dedina, M.J., & Nusbaum, H.C. (1985). Comprehension of natural and synthetic speech using a sentence verification task. In *Research on Speech Perception Progress Report No. 11* (pp. 33-57). Bloomington, IN: Speech Research Laboratory, Indiana University.

Massaro, D.W. (1997). *Perceiving Talking Faces: from Speech Perception to a Behavioral Principle.* Cambridge, MA: MIT Press.

Mattingly, I.G. (1968). *Synthesis by rule of General American English.* Ph.D. dissertation, Yale University. (Issued as supplement to Haskins Laboratories Status Report on Speech Research.)

Mirenda, P. & Beukelman, D.R. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. *Augmentative and Alternative Communication, 3,* 120-128.

Mirenda, P. & Beukelman, D. (1990). A comparison of intelligibility among natural speech and seven speech synthesizers with listeners from three age groups. *Augmentative and Alternative Communication, 6,* 61-68.

Mitchell, P. & Atkins, C. (1988). A comparison of the single word intelligibility of two voice output communication aids. *Augmentative and Alternative Communication, 4,* 84-88.

Moody, T. & Joost, M. (1986). Synthesized speech, digitized speech, and recorded speech: a comparison of listener comprehension rates. In *Proceedings of the Voice Input/Output Society.* Alexandria, VA.

Moulines, E. & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453-467.

Nusbaum, H.C., Dedina, J.J. & Pisoni, D.B. (1984). Perceptual confusions of consonants in natural and synthetic CV syllables. In *Speech Research Laboratory Technical Note, 84-02.* Bloomington, IN: Speech Research Laboratory, Indiana University.

Nusbaum, H., Francis, A., & Henly, A. (1995). Measuring the naturalness of synthetic speech. *International Journal of Speech Technology, 1,* 7-19.

Nusbaum, H.C. & Pisoni, D.B. (1984). Perceptual evaluation of synthetic speech generated by rule. In *Proceedings of the 4th Voice Data Entry Systems Applications Conference.* Sunnyvale, CA: Lockheed.

Nusbaum, H.C., Schwab, E.C. & Pisoni, D.B. (1984). Subjective evaluation of synthetic speech: measuring preference, naturalness and acceptability. In *Research on Speech Perception Progress Report No. 10* (pp. 391-408). Bloomington, IN: Speech Research Laboratory, Indiana University.

Nye, P.W. & Gaitenby, J.H. (1973). Consonant intelligibility in synthetic speech and in a natural speech control (modified rhyme test results). In *Haskins Laboratories Status Report on Speech Research, SR-33,* 77-91. New Haven, CT: Haskins Laboratories.

Nye, P.W. & Gaitenby, J.H. (1974). The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. In *Haskins Laboratories Status Report on Speech Research, SR-37/38,* 169-190. New Haven, CT: Haskins Laboratories.

Nye, P.W., Ingemann, F. & Donald, L. (1975). Synthetic speech comprehension: a comparison of listener performances with and preferences among different speech forms. In *Haskins Laboratories Status Report on Speech Research, SR-41,* 117-126. New Haven, CT: Haskins Laboratories.

Paris, C.R., Gilson, R.D., Thomas, M.H. & Silver, N.C. (1995). Effect of synthetic voice intelligibility upon speech comprehension. *Human Factors, 37,* 335-340.

Paris C.R., Thomas M.H., Gilson R.D., & Kincaid J.P. (2000). Linguistic cues and memory for synthetic and natural speech. *Human Factors 42,* 421-431.

Pisoni, D.B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. *Journal of the Acoustical Society of America, 70,* S98.

Pisoni, D.B. (1987). Some measures of intelligibility and comprehension. In J. Allen, M.S. Hunnicutt, & D.H. Klatt (Eds.), *From Text to Speech: the MITalk System,* Pp 151-171. Cambridge, UK: Cambridge University Press.

Pisoni, D.B. (1997). Perception of synthetic speech. In J.P.H. van Santen, R.W. Sproat, J.P. Olive & J. Hirschberg (Eds.), *Progress in Speech Synthesis,* Pp 541-560. New York: Springer-Verlag.

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.

Pisoni, D.B. & Hunnicutt, S. (1980). Perceptual evaluation of MITalk: The MIT unrestricted text-to-speech system. In *1980 IEEE International Conference on Acoustics, Speech and Signal Processing,* 572-575. New York: IEEE.

Pisoni, D.B. & Koen, E. (1981). Some comparisons of intelligibility of synthetic and natural speech at different speech-to-noise ratios. In *Research on Speech Perception Progress Report No. 7* (pp. 243-254). Bloomington, IN: Speech Research Laboratory, Indiana University.

Pisoni, D.B., Manous, L.M., & Dedina, M.J. (1987). Comprehension of natural and synthetic speech: effects of predictability on the verification of sentences controlled for intelligibility. *Computer Speech and Language, 2,* 303-320.

Pisoni, D.B., Nusbaum, H.C., & Greene, B.G. (1985). Perception of synthetic speech generated by rule. In *Proceedings of the Institute of Electrical and Electronics Engineers, 73 (11),* 1665-1675.

Pols, L.C.W. (1989). Assessment of text-to-speech synthesis systems. In A.J. Fourcin, G. Harland, W. Barry & V. Hazan (eds.), *Speech Input and Output Assessment*, (pp. 55-81). Chichester, England: Ellis Horwood.

Pols, L.C. W. (1992). Quality assessment of text-to-speech synthesis by rule. In S. Furui & M.M. Sondhi (eds)., *Advances in Speech Signal Processing*, (pp. 387-416). New York: Marcel Dekker.

Ralston, J.V., Pisoni, D.B., Lively, S.E., Greene, B.G. & Mullennix, J.W. (1991). Comprehension of synthetic speech produced by rule: word monitoring and sentence-by-sentence listening times. *Human Factors, 33,* 471-491.

Reynolds, M.E., Bond, Z.S. & Fucci, D. (1996). Synthetic speech intelligibility: comparison of native and non-native speakers of English. *Augmentative and Alternative Communication, 12,* 32-36.

Reynolds, M.E. & Fucci, D. (1998). Synthetic speech comprehension: a comparison of children with normal and impaired language skills. *Journal of Speech, Language and Hearing Research, 41,* 458-466.

Reynolds, M.E., Isaacs-Duvall C. & Haddox M.L. (2002). A comparison of learning curves in natural and synthesized speech comprehension. *Journal of Speech Language and Hearing Research 45,* 802-820.

Reynolds, M.E., Isaacs-Duvall, C., Sheward, B. & Rotter, M. (2000). Examination of the effects of listening practice on synthesized speech comprehension. *Augmentative and Alternative Communication, 16,* 250-259.

Reynolds, M.E. & Jefferson, L. (1999). Natural and synthetic speech comprehension: comparison of children from two age groups. *Augmentative and Alternative Communication, 15,* 174-182.

Rounsefell, S., Zucker, S.H. & Roberts, T.G. (1993). Effects of listener training on intelligibility of augmentative and alternative speech in the secondary classroom. *Education and Training in Mental Retardation, 28,* 296-308.

Rupprecht, S., Beukelman, D. & Vrtiska, H. (1995). Comparative Intelligibility of five synthesized voices. *Augmentative and Alternative Communication, 11,* 244-247.

Sanderman, A.A. & Collier, R. (1996). Prosodic rules for the implementation of phrase boundaries in synthetic speech. *Journal of the Acoustical Society of America 100,* 3390-3397.

Sanderman, A.A. & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech, 40,* 391-409.

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.

Simpson, C.A. & Williams, D.H. (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. *Human Factors, 22,* 319-330.

Slowiaczek, L.M. & Nusbaum, H.C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors, 27,* 701-712.

Slowiaczek, L.M. & Pisoni, D.B. (1982). Effects of practice on speeded classification of natural and synthetic speech. In *Research on Speech Perception Progress Report No. 7,* (pp. 255-262). Bloomington, IN: Speech Research Laboratory, Indiana University.

Stevens, K.N. (1996). Understanding variability in speech: a requisite for advances in speech synthesis and recognition. *Journal of the Acoustical Society of America, 100,* 2634.

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 Lipreading,* (pp. 3-51). Hillsdale, NJ: Lawrence Erlbaum & Associates.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212-215.

Sutton, B., King, J., Hux, K., & Beukelman, D. (1995). Younger and older adults' rate performance when listening to synthetic speech. *Augmentative and Alternative Communication, 11,* 147-153.

Terken, J. (1993). Synthesizing natural-sounding intonation for Dutch: rules and perceptual evaluation. *Computer Speech and Language, 7,* 27-48.

Terken, J. & Lemeer, G. (1988). Effects of segmental quality and intonation on quality judgments for texts and utterances. *Journal of Phonetics, 16,* 453-457.

van Santen, J.P.H. (1994). Assignment of segmental duration in text-to-speech synthesis. *Computer Speech and Language, 8,* 95-128.

Venkatagiri, H.S. (1994). Effect of sentence length and exposure on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10,* 96-104.

Venkatagiri, H.S. (2003). Segmental intelligibility of four currently used text-to-speech synthesis methods. *Journal of the Acoustical Society of America 113,* 2095-2104.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Some Effects of Feedback on the Perception of Point-Light and Full-Face Visual Displays of Speech: A Preliminary Report[1]

**Stephen J. Winters and David B. Pisoni**

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

[1] This research was supported by the NIH-NIDCD Research Grant DC-00111 and NIH-NIDCD Training Grant DC-00012 to Indiana University. We would like to thank Jeff Reynolds for his work on the point-light portion of this study and Luis Hernandez for conceptual inspiration and technical assistance.

# Some Effects of Feedback on the Perception of Point-Light and Full-Face Visual Displays of Speech: A Preliminary Report

**Abstract.** This study investigated the effects of feedback on the perception of words from point-light and fully-illuminated displays of speech. Participants attempted to identify words in these displays and were later given feedback about the identity of the words they had just seen. Feedback was presented to the participants in three forms: audio-visual, audio-only, and orthographic representations of the stimulus word. A control group of participants also received no feedback on the stimuli. It was expected that dynamic feedback—as found in the audio or audio-visual signals—would improve participant performance on the perceptual task more than static, orthographic feedback, or no feedback at all, due to the extra, event-based information about the original stimuli inherent in the dynamic representations. In general, we found that participants improved their ability to perform the perceptual task over the course of the experiment; however, their level of improvement did not depend on the type of feedback they received. This finding suggests that participants may not have improved their visual-only perception skills through attending to the feedback information but rather by simply practicing the experimental task and relying on what they already knew about the lawful acoustic consequences of articulatory gestures. Participants also had more success identifying the full-face stimuli than the point-light stimuli. However, they also exhibited different patterns of misidentification for some of the point-light and full-face stimuli, suggesting that the point-light stimuli were not merely impoverished representations of the full-face displays. Instead, the two representations of speech must be, to some extent, perceptually independent of one another.

## Introduction

This study investigated the extent to which observers can extract phonetic information from visual-only point-light and fully-illuminated displays of speech. Point-light displays (PLDs) are animated sequences of illuminated dot patterns. They are produced by attaching luminescent dots to various points on a person's body and then filming that person while they perform certain motions under a black light (Johansson, 1973). Figure 1 shows an example sequence of frames from a point-light movie; the person filmed in this particular point-light movie executed a placekick.

Point-light displays are valuable research tools because observers can perceive meaningful motion in them, even though they generally do not perceive anything more than random dot patterns when presented with any of the individual point-light frames in isolation. This suggests that perceivers are able to extract meaningful information from dynamic, time-varying properties in the sequences of the point-light patterns which do not exist in any of the individual, static point-light patterns per se. These findings are interesting because they might reflect aspects of event perception under normal conditions. For instance—does the visual perception of events ever rely on the recognition of individual static patterns? Or is event perception always based solely on dynamic, time-varying information in the visual array?

Several theorists have argued that the visual perception of speech involves the perception of meaningful articulatory events (e.g., Fowler & Rosenblum 1991). Lachs and Pisoni (2004) have shown that people can accurately perceive such events in visual-only, fully-illuminated displays of speech. Lachs and Pisoni (submitted) have also shown that people can perceive such articulatory events in PLDs of speech. Furthermore, Rosenblum, Johnson and Saldaña (1996) found that PLDs of speech provide an intelligibility gain to the perception of speech in noise, similar to the gain found for fully-illuminated visual displays in Sumby and Pollack (1954). Rosenblum and Saldaña (1996) also found that PLDs may induce a “McGurk Effect” on the perception of audio-visually mismatched stimuli, as was found for fully-illuminated visual displays in McGurk and McDonald (1976).

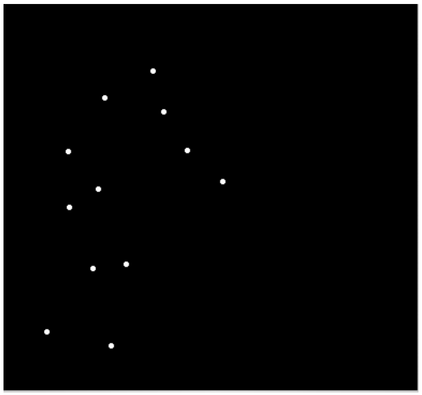


**Figure 1a.** Frame from a point-light display movie of a person executing a placekick.

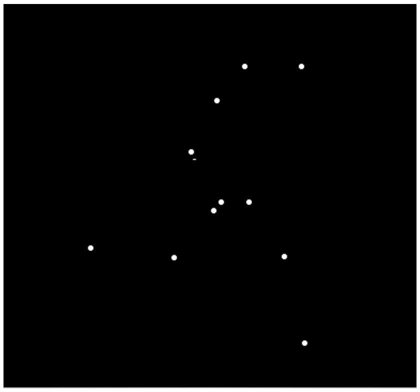


**Figure 1b.** Frame from a point-light display movie of a person executing a placekick.

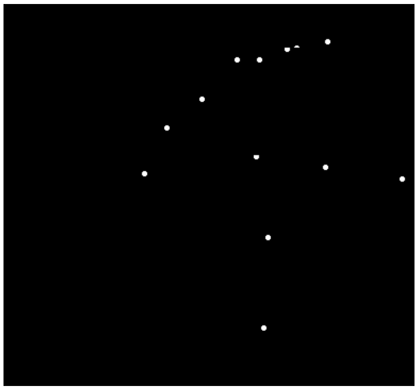


**Figure 1c.** Frame from a point-light display movie of a person executing a placekick (courtesy of http://astro.temple.edu/~tshipley/mocap/dotMovie.html).

Even though people can extract meaningful information from both visual-only point-light and fully-illuminated displays of speech, there are at least three reasons to believe that it might be more difficult for people to perceive visually presented words in PLDs than in fully-illuminated displays. Point-light displays are, first of all, visually impoverished versions of full-face displays; they contain only dynamic cues and no static visual information. Furthermore, PLDs do not necessarily contain all of the dynamic cues inherent in the full-face displays, since they only represent articulatory motions with a handful of illuminated dots attached to particular spots on a speaker's chin, lips, cheeks, teeth or tongue. The motions of such articulators in between these dots may provide meaningful information to perceivers which the PLDs cannot transmit. Finally, people might have difficulty interpreting PLDs of speech if only because they have little, if any, prior experience attempting to perceive speech in them.

Since PLDs present all of these perceptual challenges to human observers, it is remarkable that people can perceive speech in them at all. How do they perform this perceptual feat? One possible answer is that the PLDs preserve fundamental dynamic cues which are important to the visual perception of speech in full-face displays. Another possible answer is that they perceive meaningful events in the PLDs based on their knowledge of the physically lawful relationships which hold between articulation and acoustics. The perception of speech in PLDs might be direct, as it were, in the technical sense put forth by Fowler (1986). In this sense, speech perception does not need to be mediated by abstract knowledge of a symbolic system; instead, it is specified by visually apparent attributes in the speech signal itself. If this is the case, then it is unlikely that increased experience with the symbolic representations of speech events or gestures (as in, e.g., the written word) will improve a person's ability to perceive those events. By the same token, however, people may be able to perceive speech in representations with which they have had little if any experience if those representations happen to be based in physically lawful ways on the same articulatory events which constitute the original speech signal.

Fowler and Dekle (1991) made a compelling case for the "direct" nature of speech perception with an ingenious experiment that involved the perception of speech through the sense of touch. In this experiment, Fowler and Dekle briefly introduced their participants to the Tadoma method, which involves the placing of the fingers of one's hand on the lips, jaw and neck of a speaker in order to perceive what they are saying through (directly) feeling their articulators move (cf. Reed et al., 1989). Fowler and Dekle then presented their participants with McGurk-like stimuli which were mismatched between the auditory and tactile domains. The participants may have heard the syllable "ga," for instance, while they felt the syllable "ba." Fowler and Dekle also presented their participants (who were Ivy League undergraduates, with extensive reading experience) with stimuli that were similarly mismatched in audio and orthographic representations. They asked their participants to report what they had heard and what they had read/felt independently. Interestingly, Fowler and Dekle found that the participants' reports of what they had "heard" were not strictly independent of what they had felt in the mismatched-stimuli conditions. Participants who had heard "ga" but felt "ba," for instance, were more likely to report that they had heard "ba" than when they had heard the same "ga" stimulus in conjunction with a felt "ga" syllable. Mismatched audio and orthographic syllables did not induce similar cross-modal perceptual biases; in those cases, the participants' reports of what they had heard were not biased one way or another by what they had read.

Fowler and Dekle's (1991) participants experienced such McGurk-like effects in the combined auditory and tactile stimuli, even though they had never perceived speech through the Tadoma method before. The authors suggested that it was possible to induce this kind of McGurk effect for their participants in this way because of the physically lawful relationships which held between the movements the participants could feel and the acoustic sound patterns they could hear. The participants could extract information from both tactile and auditory modalities about the underlying speech events because those speech events structured the auditory and tactile modalities in a lawful way. Such lawful relationships between the two modalities are important to perception because they are not arbitrary, and are not learned by association. In contrast, the relationship between an auditory stimulus and its orthographic representation is arbitrary. Hence, Fowler and Dekle's participants experienced no McGurk-like effects for conflicting audio and orthographic stimuli, despite all of their experience associating the two kinds of representations with one another. Thus, Fowler and Dekle concluded that people perceive the events which produce the speech, through the lawfully structured consequences of those events in the acoustic, visual, or tactile domains. They do not perceive speech through reference to arbitrary relationships between sound and symbol which have been learned through experience.

In the present study we hypothesized that people might be able to perceive speech in PLDs—even without prior experience of them—because of the lawful relationships they have with the articulatory events that produced them. We investigated the effects that different kinds of feedback might have on participants' ability to improve their perception of speech in PLDs over the course of the perception experiment. Specifically, we reasoned that informing participants of which words they had seen being spoken under visual-only presentation might help improve their ability to perceive the words in subsequent visual-only stimuli. While we expected that providing such feedback information would help improve participants' perceptual performance in general, we also expected that some forms of feedback would be more helpful to participants if it came to them in the form of dynamic information that was lawfully based on the articulatory events they had seen in the visual-only stimuli. With feedback, participants might better learn how to perceive the visual-only stimuli if they actually heard what was being said in the stimuli tokens, rather than simply reading in orthographic form which words the speaker had said in the stimuli. We further hypothesized that auditory, event-based feedback would help the participants' perception of these events (whether they are represented in PLDs or fully-illuminated faces) because this kind of feedback would specify to the participants what the lawfully-based, acoustic consequences of those events are. Written words, on the other hand, would only inform participants of a static, abstract, idealized linguistic representation of those articulatory events, which—as Fowler and Dekle (1991) have shown—is unlikely to influence the direct perception of those events, no matter how much experience participants might have had with orthography.

The following experiment was designed to test this hypothesis about the nature of perceiving speech from PLDs by providing participants with different kinds of feedback on their performance in a visual-only speech perception experiment. The participants' primary task in this experiment was to watch a short video clip of a talker saying one English word at a time, in the form of a consonant-vowel-consonant (CVC) syllable. Half of the participants saw these videos in the form of PLDs, while the other half saw them as fully-illuminated displays of the face of the speaker; in neither case, however, did the participants hear what the speaker in the video was saying. Their task was to attempt to identify the word that the speaker had said, based only on what they could see in each video. After attempting to identify each word, the participants received feedback about their response. In some conditions of the experiment, they were then told what word had they had seen being spoken. Some participants received this information by seeing the video again, this time along with the original audio track; another group of participants just heard the audio track, while a third group of participants simply read an orthographic representation of the stimulus word on a computer screen. A fourth group in the control condition was not told which words they had seen, but simply moved on to the next video stimulus in the sequence after attempting to identify each stimulus word.

We expected that the participants who saw the full-face displays would be much better at identifying the stimulus words than the participants who saw the PLDs. We also expected that the performance of all participants on this task would improve over the course of the experiment, but that the extent of this improvement would vary depending on the type of feedback the participants received. The participants who received no feedback on their responses, for instance, were not expected to improve as much on the task as those who did receive feedback. Likewise, those groups who received static, symbolic feedback were not expected to improve as much as those who received dynamic, event-based feedback, in the form of either audio (A) or audio-visual (A+V) representations of the original visual stimulus. Such event-based feedback was expected to improve the participants' performance more than the symbolic feedback because it would not only provide the participants with information about the acoustic consequences of the dynamic events they were trying to perceive, but also reinstate the processing operations during the initial perception of the visual-only stimulus. Since audio-visual feedback would provide the participants with even more redundant information about the lawful relationships between articulation and acoustics than audio feedback alone, participants in the audio-visual feedback condition were also expected to show more improvement in the task than those in the audio feedback group. Finally, feedback of all kinds was expected to have more of an impact on the improvement of the participants' ability to identify words in the PLDs than in the full-face displays, since no participant had seen PLDs of speech prior to the experiment. Participants might thus depend more heavily on the feedback information they received during the experiment in order to interpret what they might be seeing in the PLDs. The participants' ability to identify words in the visual-only PLDs might also be closer to floor levels at the start of the experiment, which would leave considerably more room for improvement over the course of a short perception experiment.

## Methods and Procedures

**Participants.** All of the participants in this study were college undergraduates at Indiana University in Bloomington, Indiana, taking part for either course credit or a small fee. The experiment used a between-subjects design; participants took part in only one of the four feedback conditions, for either of the stimulus types. The exact number of participants in each of the feedback conditions is given below in Table 1.

| Full-face | N | Point-light | N |
|---|---|---|---|
| AVF | 21 | AVF | 21 |
| AF | 22 | AF | 20 |
| OF | 20 | OF | 24 |
| NF | 20 | NF | 16 |

**Table 1.** Number of participants in each experimental group, by feedback condition and stimulus type.

**Stimulus Materials.** The individual full-face video stimuli in this experiment were first produced for the Hoosier Audio-visual Multi-talker Database (Sheffert, Lachs, & Hernandez, 1996; Lachs & Hernandez, 1998). The point-light stimuli were originally produced for use in Lachs (submitted). Both point-light and full-face stimuli consisted of short (between one and two seconds long) videos of a female, native speaker of English saying a single monosyllabic, consonant-vowel-consonant (CVC) word. Different speakers were filmed in the point-light and full-face videos. Stimuli of both types consisted of close-ups of the speaker's face, from just above the shoulders up. In the point-light videos, luminescent dots were attached to the speaker's face according to the pattern seen in Figure 2, which shows an isolated frame from one of these videos. The visible dots in Figure 2 were attached to the cheeks, lips, nose and chin of the speaker. Two dots were also placed on both the upper and lower teeth of the speaker, as well as another dot on the blade of the speaker's tongue. Figure 3 shows a corresponding example frame from one of the fully-illuminated videos.

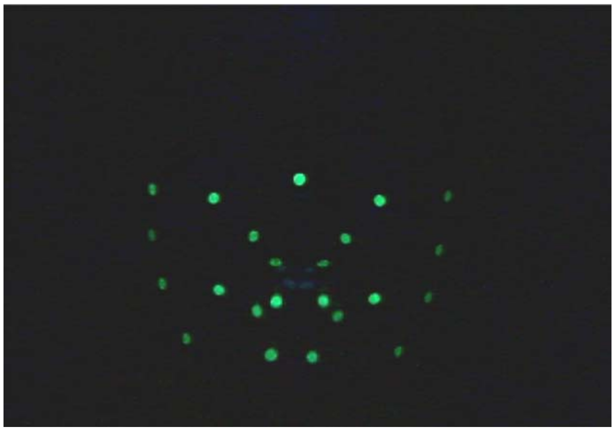


**Figure 2.** Example frame from a point-light display stimulus video.

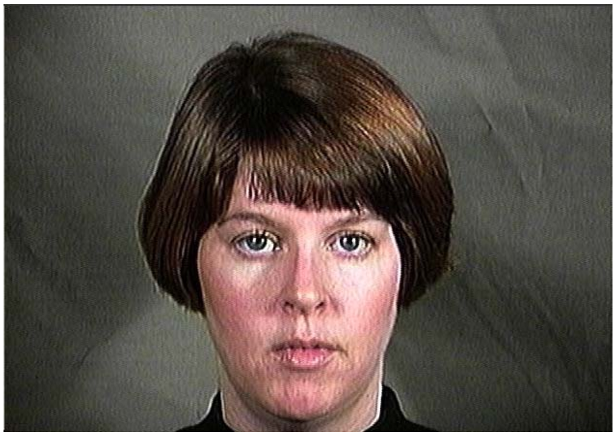


**Figure 3.** Example frame from a full-face display stimulus video.

Although originally recorded on videotape, all stimuli were digitized and transferred to Macintosh G3 computers for presentation in this experiment. All videos had a 640 x 480 aspect ratio and completely filled the entire monitor screen when they were presented to the participants. Each word in the stimulus set was a CVC, monosyllabic English word. There were 96 words in all; following Lachs (2002), half of these were "easy" words to identify in the sense that they were high-frequency items selected from low-density lexical neighborhoods, whereas the other half were "hard" in that they were low-frequency items selected from dense lexical neighborhoods. For example, easy words included "wife," "road," and "teeth," while hard words included "hag," "dame," and "toot."

**Design and Procedure.** Either a customized Psyscope routine (ver. 1.2.5.PPC) or SuperCard stack (ver. 4.1.1) were used to present the stimuli videos to the participants. On each individual trial, these programs presented visual stimuli—without sound—to the participants and then prompted them to type into the computer what word they thought the speaker in the video had said. The words were presented in random order to each participant. After the participants typed in their response to each stimulus, the customized programs then informed the participants in the feedback groups what word had actually been spoken in the video they had just seen. For the orthographic feedback group (OF), the programs presented the feedback to the listeners in the form of a written word, centered on the screen. For the audio feedback group (AF), the programs played the audio clip of the word to the listener over Beyer Dynamic DT-100 headphones, just as it was spoken in the original video stimulus. For the audio-visual feedback group (AVF), the programs played the original video stimulus to the participants again, together this time with the original audio track. The programs did not inform the participants in the control group—the no feedback group (NF)—which words had been spoken in the video stimuli; these participants just moved on to the next stimulus once they had finished typing in each of their responses.

All of the feedback conditions for the point-light stimuli, as well as the no feedback condition for the full-face stimuli, were run using a Psyscope program; it was not possible to use this program to run the orthographic feedback, audio feedback and audio-visual feedback conditions for the full-face stimuli because of computer memory limitations. Instead, these three conditions were run with the SuperCard stack. The Psyscope and SuperCard implementations of the experimental paradigm were essentially identical except for a few minor details, which were mostly aesthetic in nature. The most significant of these differences was that the SuperCard program presented the orthographic feedback to the participants for a full second (1000 milliseconds), whereas the Psyscope program only presented this information to the participants for 500 milliseconds. For both types of stimuli, audio and audio-visual feedback was only given to the participants for the inherent duration of the CVC word in the recording, which ranged between 1000 and 2000 milliseconds.

Prior to the experiment, the participants were informed that all of the words they would see being spoken were one-syllable, English words. They were also told that some words might be harder to identify than others; hence, they were encouraged to make guesses as to the identity of each word, even if they had no idea what the word was.

## Data Analysis

Since all stimuli were of the form consonant-vowel-consonant, the observers' responses could be scored not only in terms of whether or not they had identified the whole word correctly, but also in terms of whether or not they correctly identified each segmental portion of the stimulus: the onset (initial consonant), the nucleus (the vowel) and the coda (final consonant). In order to make such phoneme-by-phoneme evaluations, all stimuli and all responses were first converted into phonetic transcriptions by matching them up with entries in the Carnegie Mellon pronouncing dictionary (version 0.6; http://www.speech.cs.cmu.edu/cgi-bin/cmudict). Each entry in this dictionary consists of an English word listed in normal orthography along with a corresponding phonetic transcription encoded in an ASCII-based phonetic alphabet. All phonetic transcriptions in this dictionary offset each phoneme in the word with spaces and uniquely mark vowels with a number indicating their level of stress in the word. The phonetic transcription for the entry "hag," for instance, is /hh ae1 g/. For each stimulus word, the entire CMU dictionary was searched for a corresponding orthographic entry. A perl script was written that searched the CMU dictionary for orthographic entries corresponding to each of the stimulus items. Once such a match had been found, its corresponding phonetic transcription was segmented into an "onset," a "nucleus" and a "coda." Since all the words were of the form CVC, the vowel of each word was considered to be the "nucleus," the initial consonant the "onset," and the final consonant the "coda." In the case of "hag," for example, the /hh/ formed the onset, the /ae1/ vowel formed the nucleus, and the /g/ formed the coda.

Even though all participants were informed, prior to the task, that they would only see monosyllabic words, many of their responses had two or more syllables in them. For all responses—no matter how many syllables they contained--the "nucleus" was considered to be the vowel with the highest stress level in the response. All segments—including any consonants or vowels—which preceded this response nucleus were then taken to be the "onset" of the response, and all segments which succeeded it were taken to be the response's "coda." For example, one participant gave the response "camera" to the point-light stimulus "thumb." The phonetic transcription for "camera" in the CMU pronouncing dictionary is /k ae1 m ax0 r ax0/. Since the /ae1/ vowel has the highest stress level in the word, it was taken to be the "nucleus" of the response. Thus, the /k/ which preceded it formed the response "onset," while the final /m ax0 r ax0/ sequence formed the "coda."

Response onsets, nuclei, or codas—as determined in this fashion--were only considered to be correct identifications of their counterparts in the original stimuli if the two matched perfectly. Thus, response onsets or codas which contained more than one segment were considered to be incorrect even if one of those segments formed the original stimulus onset or coda. Thus, the /m ax0 r ax0/ coda of "camera" did not count as a correct identification of the /m/ coda in the "thumb" stimulus, even though an /m/ formed part of the response coda. Many of the participants' responses could not be matched to any entry in the CMU pronouncing dictionary. Those that were obvious misspellings (e.g., "cheif") were simply corrected in the original data file and then matched with the corresponding dictionary entry, while those responses that were not obviously English words (e.g., "rith") were given onset-nucleus-coda transcriptions by hand and then scored accordingly.

Participant responses were also scored in terms of whether or not the participants had correctly identified the place of articulation and viseme category of the stimulus coda and onset. Each coda and onset consonant in the stimulus was thus classified in terms of both its place of articulation (alveolar, bilabial, interdental, glottal, labio-dental, labio-velar, palato-alveolar, velar), and its viseme type (bilabial, interdental, dorso-lingual, glottal, lateral, labio-dental, labio-velar, palato-alveolar, retroflex, /s/). "Visemes" is a concept that was first introduced by Walden et al. (1977) in order to account for the broad categories of consonantal phonemes that can be consistently identified in visual-only speech perception. Most viseme categories correspond primarily to a particular place of articulation, but they also include a few categories which are determined by manner of articulation (e.g., lateral /l/, retroflex /r/, and alveolar fricative /s/. Similar classifications were also made for the place of articulation and viseme category of the consonants in the response onsets and codas. Those response onsets and codas which contained more than one segment were considered to have "mixed" places of articulation or viseme categories--unless all of the segments in those onsets and codas happened to agree in either viseme type or place of articulation. In this case the common viseme category or place of articulation was then taken to be the appropriate classification for that portion of the response.

The place of articulation or viseme type of the onsets, nuclei and codas of the responses were only counted as correct identifications if they exactly matched the corresponding sub-phonemic features of the stimulus. One participant, for instance, gave the response "damp" to the "dame" stimulus. In "damp," the coda /mp/ was considered to have the bilabial place of articulation, since both /m/ and /p/ are bilabial consonants. This was scored as a correct identification of the place of the stimulus coda consonant, since the coda /m/ in "dame" also has a bilabial place of articulation. Another participant, however, identified the same "dame" stimulus as "table." Since the coda of "table" includes both /b/ and /l/ segments, which have bilabial and alveolar places of articulation, respectively, it was categorized as having a "mixed" place of articulation. This response was therefore scored as an incorrect identification of the bilabial place of articulation in the stimulus coda /m/.

## Results and Discussion

Prior to the experiment, we expected to find two general trends in the response data. First, we expected the percentage of correct responses to be much higher for the full-face stimuli than for the point-light stimuli. Second, we expected the participants' perceptual performance to improve more with event-based feedback than with either symbolic, orthographic feedback or no feedback at all. In order to quantify the amount of improvement participants made in the perceptual task over the course of the experiment—and thereby test this second prediction—percent correct scores at all levels of analysis (whole words, phonemes, and visemes/places of articulation) were tallied independently for the responses in the first and the second halves of the experiment. Comparing the differences between the percent correct scores across the two halves of the experiment thus provided a rough but straightforward way to gauge the amount of improvement participants made in identifying the visual-only stimuli over the course of the experiment.

**Words Correct.** Table 2 lists the mean percentages of whole words correctly identified in both halves of the experiment. This table lists these percentages separately by feedback condition and stimulus type; each set of scores thus reflects the performance of a different group of participants. The amount of "improvement" participants made over the course of the experiment can be assessed by subtracting their mean percentage correct scores in the first half of the experiment from their mean percentage correct scores in the second half of the experiment. This difference is listed for each experimental condition under the "Improvement" column in Table 2. The final column—labeled "% Improvement"—normalizes this improvement score by dividing it by the difference between 100% and the mean percentage correct score in the first half of the experiment for that condition—i.e., the potential amount the participants' mean percentage correct score could improve after their first-half performance.

| **Point-Light** | **N** | **First Half** | **Second Half** | **Average** | **Improvement** | **% Improvement** |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 2.3% | 3.3% | 2.8% | 1.0% | 1.0% |
| audio | 20 | 2.1% | 2.8% | 2.4% | 0.7% | 0.7% |
| orthographic | 24 | 2.2% | 1.7% | 2.0% | -0.4% | -0.4% |
| none | 16 | 1.7% | 1.8% | 1.8% | 0.1% | 0.1% |

| **Full-Face** | **N** | **First Half** | **Second Half** | **Average** | **Improvement** | **% Improvement** |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 22.2% | 23.9% | 23.1% | 1.7% | 2.2% |
| audio | 22 | 21.9% | 23.3% | 22.6% | 1.4% | 1.8% |
| orthographic | 20 | 19.9% | 22.6% | 21.3% | 2.7% | 3.4% |
| none | 20 | 20.2% | 22.4% | 21.3% | 2.2% | 2.7% |

**Table 2.** Mean percentages of whole words correctly identified by listeners in all four feedback conditions, for each stimulus type.

The data summarized in Table 2 confirm the first of the general predictions—the mean percentage of whole words correctly identified was much higher for the full-face stimuli (around 20%) than it was in the point-light stimuli (from 2% to 3%). However, the data does not provide correspondingly convincing evidence to confirm the second set of predictions—that participants' performance would improve more in the audio-visual and audio feedback conditions than in either of the other two feedback conditions. Independent, two-way Analyses of Variance (ANOVA) were run for the percentage correct data from each stimulus condition in order to test the effects that feedback type (audio-visual, audio, orthographic, none)—a between-subjects factor—and experiment half (first, second)—a within-subjects factor—had on the raw percentages of whole words correctly identified. Neither of these ANOVAs revealed any significant effects of either feedback type or experiment half. Only the experiment half factor in the ANOVA for the full-face stimuli came marginally close to significance ($F$ = 3.848*;* $df$ = 1,79*;* $p$ = .053).

It is perhaps not surprising that participants showed no significant improvement in their ability to correctly identify whole words in this experimental paradigm. This experiment provided participants with feedback on individual word stimuli they had just seen, but it never presented any of those stimuli to the participants again during the rest of the experiment. Feedback on an incorrect response to a particular stimulus would only be likely to help improve a participant's ability to identify that stimulus if the participant saw that same stimulus again, later on in the experiment (cf. Pashler, Cepeda, Wixted & Rohrer, in press). In the present experiment, participants never saw the same word stimulus twice. They did, however, see different tokens of the same phoneme more than once. The phoneme /p/, for instance, appeared in the onset position of five different stimulus words: pool, peace, pet, push, and page. Any participant who might have misidentified the /p/ in "pet," then, would have been informed of the correct identity of this consonant (in the feedback conditions) before getting another chance at identifying the same phoneme, in the same syllabic position, in a different word—e.g., "push." Learning the correct identities of previously misidentified phonemes in this way should have helped improve the participants' ability to identify those phonemes in subsequent stimuli. The same holds true for the sub-phonemic features of viseme type and place of articulation. The anticipated effects of feedback on the improvement of visual-only perception may be more likely to emerge in a more detailed analysis of the percentages of phonemes and features correctly identified.

**Phonemes Correct.** The results of the analyses of the number of phonemes correctly identified in the responses did indeed support this prediction. Tables 3 and 4 break down the mean percentages of phonemes correctly identified across the various experimental conditions and halves in the same way that Table 2 did for whole words. Table 3 lists the correct percentage scores for onset phonemes while Table 4 lists the same scores for phonemes in coda position.

| Point-Light | N | First Half | Second Half | Average | Improvement | % Improvement |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 19.9% | 22.0% | 21.0% | 2.1% | 2.6% |
| audio | 20 | 17.6% | 23.5% | 20.6% | 5.9% | 7.2% |
| orthographic | 24 | 17.0% | 23.2% | 20.1% | 6.2% | 7.4% |
| none | 16 | 19.1% | 18.1% | 18.6% | -1.0% | -1.3% |

| Full-Face | N | First Half | Second Half | Average | Improvement | % Improvement |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 46.8% | 51.7% | 49.3% | 4.9% | 9.1% |
| audio | 22 | 44.5% | 47.2% | 45.8% | 2.7% | 4.8% |
| orthographic | 20 | 44.3% | 47.6% | 45.9% | 3.3% | 6.0% |
| none | 20 | 44.4% | 48.1% | 46.3% | 3.8% | 6.7% |

**Table 3.** Mean percentages of onset phonemes correctly identified by listeners in all four feedback conditions, for each stimulus type.

| Point-Light | N | First Half | Second Half | Average | Improvement | % Improvement |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 9.8% | 11.6% | 10.7% | 1.8% | 2.0% |
| audio | 20 | 9.4% | 11.4% | 10.4% | 2.0% | 2.2% |
| orthographic | 24 | 9.1% | 8.9% | 9.0% | -0.2% | -0.2% |
| none | 16 | 8.3% | 10.0% | 9.2% | 1.7% | 1.8% |

| Full-Face | N | First Half | Second Half | Average | Improvement | % Improvement |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 34.7% | 36.7% | 35.7% | 2.0% | 3.0% |
| audio | 22 | 36.4% | 37.3% | 36.8% | 0.9% | 1.5% |
| orthographic | 20 | 35.1% | 39.1% | 37.1% | 4.0% | 6.1% |
| none | 20 | 33.9% | 37.1% | 35.5% | 3.2% | 4.9% |

**Table 4.** Mean percentages of coda phonemes correctly identified by listeners in all four feedback conditions, for each stimulus type.

These numbers reflect the same general pattern seen in the whole-word data: the scores for the full-face stimuli were much higher than those for the PLDs. Also, the scores tended to be much higher, in general, for onset phonemes than they are for coda phonemes. Independent two-way ANOVAs were run for the data from each stimulus type and syllabic position in order to test the effects of feedback type and experiment half on the percentages of phonemes correctly identified. All four of these ANOVAs revealed significant effects of experiment half on the phoneme percentage correct scores. For the ANOVA on onset phoneme identification in full-face stimuli, experiment half was significant at *p* < .006 (*F* = 7.916*; df* = 1,79). Likewise, experiment half was significant for the coda phoneme data in the full-face stimuli (*F* = 5.889*; df* = 1, 79*; p* = .018), as well as for the point-light onset phonemes (*F* = 15.7*; df* = 1,77*; p* < .001) and point-light coda phonemes (*F* = 4.234*; df* = 1,77*; p* = .043).

The ANOVA for the point-light onset phoneme identification scores also revealed a significant feedback by half interaction (*F* = 4.034*; df* = 3,77*; p* = .010). This is the only significant feedback by half interaction that emerged in the analysis of the data. Figure 4 presents this interaction graphically; it shows that the improvements made by the audio and orthographic feedback groups were significantly greater than those made by the audio-visual and no feedback groups.

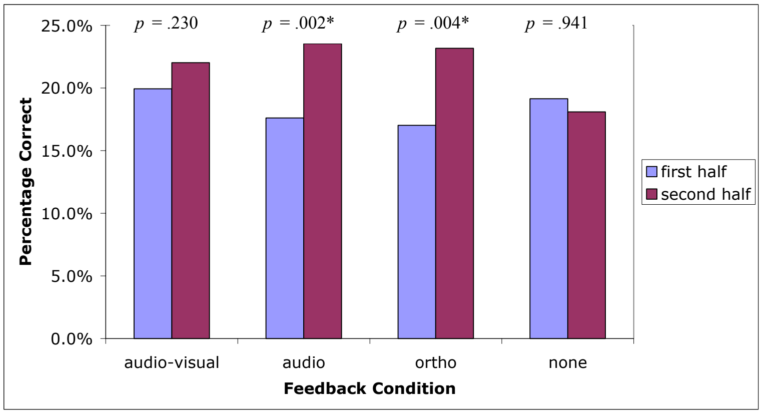


**Figure 4.** Percentage of onset phonemes correctly identified, point-light stimuli (p values are from t-test of significance for percentage correct scores between first and second halves of the experiment in that particular feedback condition).

It is not immediately clear why audio-visual feedback did not help participants' visual-only perception at least as much as audio feedback. One possible explanation is that the onset of the participants' perceptual improvement in the audio-visual feedback condition occurred earlier than it did in the audio or orthographic groups; the percentage of correct onset phoneme identifications is, in fact, higher in the first half of the experiment for the AVF group (19.6%) than it is for either the AF (17.6% or OF (17.0%) groups. The decision to compare correct percentage scores between the first and second halves of the experiment is arbitrary, and grouping the response data together in halves may have thus glossed over improvements the participants made during the first half of the experiment alone. However, this explanation may be confounded by the fact that different groups of participants took part in each feedback condition, and any one group of participants may have entered the task with a greater or lesser ability to perceive speech from visual-only PLDs than any of the other groups. For instance, the no feedback group performed relatively well in the first half of this condition—scoring 19.1% onset phonemes correctly identified, which was nearly as good as the percentage posted by the AVF group. However, the no feedback group's performance then declined to 18.1% in the second half of the experiment. Their relatively high percentage correct score in the first half of the experiment is probably, therefore, not due to significant improvement during the first half alone. On the other hand, the AVF group may, on the other hand, not have been able to improve as much as the AF or OF groups because their performance started out closer to ceiling level in this task, which may be near 23% or 24% correct. Since the OF and AF groups started with a lower baseline performance, their ability to attain near-ceiling performance in the second half of the experiment produced a significant effect of experiment half in the Analysis of Variance, even though similarly high levels of performance by the AVF group in the second half were not significantly different from their higher performance baseline in the first.

It is also important to emphasize that the three groups who received feedback also showed improvement between the first and second halves of the experiment that were in the expected (positive) direction. For the group that did not receive feedback, however, participant performance on onset phoneme identification dropped between the first and the second halves of the experiment. This pattern of results suggests that feedback does have a positive effect on the participants' ability to perform this visual-only speech perception task. The fact that the feedback by experimental half interaction only emerged among the groups who saw point-light stimuli is also suggestive, since they were predicted to be more susceptible to the positive effects of feedback, due to their unfamiliarity with that type of stimulus. The broad trend for the full-face stimuli, on the other hand, was improvement between first and second halves for all feedback groups, including the group that received no feedback at all. This suggests that the participants in the full-face conditions could simply get better at the experimental task through practice, rather than having to depend on the information they received through feedback in order to interpret the visual-only stimuli.

The fact that the participants did not show significant amounts of improvement—in all experimental conditions—in their ability to correctly identify the vowel phonemes in the stimuli is also suggestive. Table 5 shows the mean percentages of vowels correctly identified for all feedback groups and both stimulus types, broken down by experiment half.

Independent, two-way (feedback type x experimental half) ANOVAs were run on the percentages of vowels correctly identified in both the full-face and point-light stimuli, considering feedback type and experimental half as factors. Neither of these ANOVAs revealed any significant factors or interactions in the vowel identification scores. Nonetheless, the percentages of vowels correctly identified in Table 5 reveal some interesting trends. In general, the percentages of vowels the participants correctly identified were much higher for the full-face stimuli (about 50%) than it was for the point-light stimuli (about 20%). The participants' ability to identify vowels correctly for the full-face stimuli was also marginally better than their ability to identify consonant phonemes in the same stimuli in either onset or coda position. For the full-face stimuli, participants identified between 50% and 55% of the vowels correctly, while their percentages for the onset phonemes in the same stimuli were between 45% and 50%. While this difference may seem significant at first glance, it may simply reflect the fact that participants had to identify vowels from a smaller set of responses (15) than they had for consonants (22 alternatives in onset position).

| Point-Light | N | First Half | Second Half | Average | Improvement | % Improvement |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 19.0% | 22.2% | 20.6% | 3.2% | 3.9% |
| audio | 20 | 19.3% | 23.1% | 21.2% | 3.9% | 4.8% |
| orthographic | 24 | 20.3% | 19.8% | 20.1% | -0.5% | -0.7% |
| none | 16 | 19.7% | 17.8% | 18.8% | -1.8% | -2.3% |

| Full-Face | N | First Half | Second Half | Average | Improvement | % Improvement |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 56.3% | 56.3% | 56.3% | 0.0% | 0.0% |
| audio | 22 | 53.5% | 53.5% | 53.5% | 0.0% | 0.0% |
| orthographic | 20 | 49.2% | 51.9% | 50.5% | 2.7% | 5.3% |
| none | 20 | 54.0% | 54.6% | 54.3% | 0.6% | 1.4% |

**Table 5.** Mean percentages of vowel phonemes correctly identified by listeners in all four feedback conditions, for each stimulus type.

Seen in this light, the fact that nearly equivalent percentages of vowels and onset phonemes were identified correctly in the point-light stimuli (≈20%) reflects poorly on the participants' vowel identification skills. However, the patterns of improvement for vowel identification in the point-light conditions revealed an interesting trend: participants in the AVF and AF groups actually got better (improvements of 3.2% and 3.9% correct, respectively), while the OF and NF groups got marginally worse (decreases of .5% and 1.8% correct, respectively). Even though these levels of improvement are not statistically significant, this pattern roughly confirms the prediction that dynamic, event-based feedback could help improve participant performance more than either static, symbolic feedback or no feedback at all. However, the full-face vowel identification data exhibits the exact opposite pattern—participants showed no improvement at all between halves in the AF and AVF conditions (0% for both, exactly), while they showed marginal improvement in the OF and NF conditions (2.7% and 0.6%, respectively). Future work might be able to better clarify the importance of these patterns of improvement by increasing the number of stimuli in the experiment. This could enable some of these suggestive—but conflicting—trends in the improvement of vowel identification to reach statistical significance.

**Place of articulation and visemes correct.** Participants' failure to significantly improve their vowel identification scores may have, however, reflected a lack of clearly defined place of articulation cues in vowels. In contrast to their performance on vowel identification, participants in all the experimental conditions improved their ability to correctly identify both place of articulation and viseme type between the two halves of the experiment. Tables 6 and 7 show the mean percentage scores for visemes correctly identified in both halves of the experiment, broken down by the eight participant groups in the study.

The scores here mirror those in the phoneme correct identification scores, except that the overall percentage means were considerably higher. For the full-face stimuli, for instance, viseme scores were around 70% in onset position and 65% in coda position, while they were only between 45 to 50% for onset phonemes and around 35% correct for coda phonemes. Independent, two-way ANOVAs—using experiment half and feedback type as factors—were carried out on the raw percentages of visemes correctly identified in both syllabic positions, for both stimulus types. These ANOVAs yielded similar significant effects to those found in the ANOVAs for the phoneme identification data. Experiment half was a significant factor for onset visemes in the full-face stimuli (*F* = 15.98; *df* = 1,79; *p* < .001), as well as for the coda visemes in the same condition (*F* = 8.287; *df* = 1,79; *p* = .005). The same factor was also significant for both onset visemes (*F* = 21.26; *df* = 1, 77; *p* < .001) and coda visemes (*F* = 4.487; *df* = 1,77; *p* = .037) in the point-light stimuli. All of these factors were also significant in the corresponding ANOVAs for the phoneme identification data; the only difference between the Analyses of Variance at the two different levels of linguistic structure was that the feedback by half interaction did not reach significance for the identification of point-light onset visemes, even though it did for the identification of point-light onset phonemes.

| **Point-Light** | **N** | **First Half** | **Second Half** | **Average** | **Improvement** | **% Improvement** |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 37.5% | 44.6% | 41.1% | 7.1% | 11.4% |
| audio | 20 | 37.4% | 43.2% | 40.3% | 5.8% | 9.3% |
| orthographic | 24 | 34.3% | 41.2% | 37.8% | 6.9% | 10.6% |
| none | 16 | 35.5% | 37.4% | 36.5% | 1.8% | 2.8% |

| **Full-Face** | **N** | **First Half** | **Second Half** | **Average** | **Improvement** | **% Improvement** |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 69.5% | 73.5% | 71.5% | 4.0% | 13.0% |
| audio | 22 | 67.5% | 69.7% | 68.6% | 2.2% | 6.7% |
| orthographic | 20 | 67.5% | 72.4% | 69.9% | 4.9% | 15.1% |
| none | 20 | 67.4% | 71.0% | 69.2% | 3.6% | 11.2% |

**Table 6.** Mean percentages of onset visemes correctly identified by listeners in all four feedback conditions, for each stimulus type.

| **Point-Light** | **N** | **First Half** | **Second Half** | **Average** | **Improvement** | **% Improvement** |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 27.5% | 28.8% | 28.1% | 1.3% | 1.8% |
| audio | 20 | 26.7% | 33.2% | 29.9% | 6.6% | 8.9% |
| orthographic | 24 | 28.4% | 27.4% | 27.9% | -1.0% | -1.3% |
| none | 16 | 26.3% | 27.9% | 27.1% | 1.6% | 2.1% |

| **Full-Face** | **N** | **First Half** | **Second Half** | **Average** | **Improvement** | **% Improvement** |
|---|---|---|---|---|---|---|
| audio-visual | 21 | 62.1% | 64.9% | 63.5% | 2.8% | 7.3% |
| audio | 22 | 64.4% | 66.1% | 65.2% | 1.7% | 4.8% |
| orthographic | 20 | 65.4% | 67.7% | 66.6% | 2.3% | 6.6% |
| none | 20 | 61.1% | 65.3% | 63.2% | 4.2% | 10.7% |

**Table 7.** Mean percentages of coda visemes correctly identified by listeners in all four feedback conditions, for each stimulus type.

The close parallels between the phoneme and viseme data seem to indicate that the overall phoneme scores depended to a large extent on the viewers' ability to identify each phoneme's viseme type. The 20% discrepancy between viseme scores and phoneme scores in most conditions may thus be attributed largely to the difficulty the participants had in identifying the sub-phonemic features of the onset and coda which are not integrated into the categorization of visemes—i.e., voicing and, to a lesser extent, manner of articulation. The difficulty the participants had in identifying manner and voice in the point-light stimuli may also account for the aforementioned differences in the results at the two different levels of linguistic structure. The significant feedback by half interaction which emerged in the point-light onset phoneme correct scores indicated that scores improved in the conditions where participants received feedback but got worse in the condition where participants received no feedback at all. A similar feedback by half interaction failed to emerge for the viseme correct data because these scores improved—at least marginally--in all four feedback conditions. This pattern of results suggests that the participants could improve their ability to perceive particular viseme categories even without feedback but that their ability to pick up manner and voice information in the point-light stimuli could only improve if they received feedback about those stimuli.

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Alveolars** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 44.9% | 52.6% | 40.9% | 46.1% |
| audio | 36.3% | 43.3% | 36.1% | 46.3% |
| orthographic | 32.2% | 38.2% | 40.9% | 50.8% |
| none | 34.2% | 23.7% | 43.0% | 46.5% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Bilabials** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 65.6% | 67.3% | 90.9% | 92.3% |
| audio | 72.4% | 71.9% | 87.0% | 90.6% |
| orthographic | 70.9% | 70.4% | 89.6% | 94.7% |
| none | 61.6% | 48.3% | 89.9% | 92.9% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Glottals** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 15.9% | 42.5% | 33.3% | 42.2% |
| audio | 12.0% | 16.0% | 38.2% | 42.4% |
| orthographic | 32.6% | 37.7% | 61.0% | 61.5% |
| none | 15.6% | 3.1% | 47.1% | 56.5% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Interdentals** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 0.0% | 2.9% | 92.3% | 91.9% |
| audio | 0.0% | 0.0% | 90.2% | 92.0% |
| orthographic | 2.9% | 0.0% | 96.6% | 80.6% |
| none | 0.0% | 0.0% | 82.1% | 87.5% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Labio-dentals** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 42.2% | 60.6% | 86.0% | 90.3% |
| audio | 33.3% | 49.2% | 76.5% | 82.8% |
| orthographic | 24.4% | 39.8% | 83.0% | 88.3% |
| none | 21.2% | 29.0% | 88.3% | 84.9% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Labio-velars** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 57.1% | 72.9% | 95.9% | 91.9% |
| audio | 73.4% | 80.2% | 98.7% | 93.6% |
| orthographic | 73.7% | 80.4% | 98.5% | 94.7% |
| none | 35.1% | 5.5% | 95.7% | 95.7% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Laterals** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 18.4% | 19.6% | 71.4% | 89.3% |
| audio | 10.4% | 15.5% | 86.8% | 84.2% |
| orthographic | 1.6% | 12.1% | 84.3% | 85.7% |
| none | 13.2% | 9.5% | 94.1% | 93.9% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Palato-alveolars** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 15.3% | 30.4% | 93.5% | 97.7% |
| audio | 14.9% | 22.0% | 94.2% | 96.6% |
| orthographic | 14.1% | 26.8% | 90.0% | 94.0% |
| none | 10.3% | 7.3% | 95.0% | 96.7% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Retroflex** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 1.9% | 7.3% | 59.5% | 70.0% |
| audio | 0.7% | 2.9% | 48.7% | 55.3% |
| orthographic | 3.9% | 5.2% | 45.3% | 54.4% |
| none | 2.4% | 3.9% | 40.6% | 51.3% |

| | Point Light | | Full Face | |
|---|---|---|---|---|
| **Velars** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 7.6% | 8.8% | 37.2% | 29.7% |
| audio | 7.6% | 18.5% | 32.0% | 31.7% |
| orthographic | 9.9% | 17.3% | 29.2% | 30.0% |
| none | 10.0% | 5.3% | 28.9% | 23.2% |

**Table 8.** Percent hits, by place of articulation, across feedback condition, experiment half, and stimulus type for phonemes in onset position.

Interestingly, the discrepancy between the percentage of visemes correctly identified and the percentage of phonemes correctly identified increased to about 30% for the coda segments in the full-face stimuli. This shift in correct identification scores may have been the result of a lack of phonetic balance for the various places of articulation in the stimuli codas combined with the fact that the participants' ability to identify place of articulation was not uniform across all place categories.

Tables 8 and 9 show the percentages of correct identifications participants made for each individual place of articulation, for both point-light and full-face stimuli, across both halves of the experiment. Table 8 shows these percentages for the various places of articulation in onset position, while Table 9 shows the corresponding percentages for the coda places of articulation.

| | **Point Light** | | **Full Face** | | | **Point Light** | | **Full Face** | |
|---|---|---|---|---|---|---|---|---|---|
| **Alveolars** | **1st** | **2nd** | **1st** | **2nd** | **Laterals** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 31.1% | 35.5% | 58.4% | 63.1% | audio-visual | 12.3% | 14.7% | 34.1% | 40.0% |
| audio | 30.9% | 34.5% | 59.6% | 62.3% | audio | 4.5% | 8.9% | 39.8% | 45.2% |
| orthographic | 33.5% | 30.0% | 63.4% | 65.5% | orthographic | 7.3% | 6.3% | 43.5% | 44.0% |
| none | 33.8% | 34.4% | 53.5% | 58.4% | none | 6.3% | 7.7% | 38.3% | 35.4% |

| | **Point Light** | | **Full Face** | | | **Point Light** | | **Full Face** | |
|---|---|---|---|---|---|---|---|---|---|
| **Bilabials** | **1st** | **2nd** | **1st** | **2nd** | **Palato-Alveolars** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 46.2% | 34.4% | 59.3% | 65.0% | audio-visual | 11.9% | 25.4% | 92.2% | 88.7% |
| audio | 37.4% | 40.9% | 57.4% | 65.6% | audio | 23.6% | 26.9% | 79.7% | 88.9% |
| orthographic | 40.0% | 34.9% | 60.0% | 64.2% | orthographic | 31.0% | 31.5% | 81.8% | 83.3% |
| none | 30.9% | 17.1% | 56.5% | 80.0% | none | 7.1% | 4.4% | 87.3% | 87.7% |

| | **Point Light** | | **Full Face** | | | **Point Light** | | **Full Face** | |
|---|---|---|---|---|---|---|---|---|---|
| **Interdentals** | **1st** | **2nd** | **1st** | **2nd** | **Retroflex** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 1.9% | 9.6% | 70.7% | 80.9% | audio-visual | 8.0% | 9.1% | 41.5% | 40.4% |
| audio | 12.5% | 4.9% | 70.9% | 80.0% | audio | 8.8% | 10.3% | 59.1% | 36.4% |
| orthographic | 4.4% | 4.0% | 70.6% | 81.6% | orthographic | 12.3% | 12.7% | 30.4% | 38.9% |
| none | --- | 7.5% | 63.3% | 66.7% | none | 0.0% | 4.5% | 44.9% | 49.0% |

| | **Point Light** | | **Full Face** | | | **Point Light** | | **Full Face** | |
|---|---|---|---|---|---|---|---|---|---|
| **Labio-dentals** | **1st** | **2nd** | **1st** | **2nd** | **Velars** | **1st** | **2nd** | **1st** | **2nd** |
| audio-visual | 16.7% | 27.0% | 73.7% | 77.5% | audio-visual | 9.9% | 6.3% | 26.2% | 20.1% |
| audio | 17.8% | 17.6% | 75.0% | 79.7% | audio | 8.5% | 14.0% | 17.2% | 17.9% |
| orthographic | 5.7% | 23.8% | 71.6% | 78.1% | orthographic | 9.2% | 14.8% | 22.8% | 19.7% |
| none | 1.3% | 2.8% | 81.4% | 80.0% | none | 4.3% | 2.8% | 21.6% | 16.4% |

**Table 9.** Percent hits, by place of articulation, across feedback condition, experiment half, and stimulus type for phonemes in coda position.

The data in these tables indicates that, in general, those places of articulation nearer the front of the mouth (bilabial, labio-dental, labio-velar) are easier for participants to identify than those which are further back (e.g., velar, glottal). The alveolar consonants—which have a place of articulation in between the labials and the velars—show a moderately high rate of correct identification in onset position. For the full-face stimuli, for instance, participants correctly identified alveolars in onset position between 40% and 50% of the time. This falls in between the percentage of bilabials they correctly identified (90%) and the percentage of velars they correctly identified (30%). In coda position, however (see Table 9), the percentage of correct alveolar identifications (about 60%) was much closer to that of bilabials (60% to 65%) than it was to that of velars (about 20%). This increase in the participants' success at identifying alveolars in coda position reflects both an inherent response bias in the participants, as well as an imbalance in the places of articulation in the stimuli codas. Table 10 shows the distribution of the various places of articulation in both the onsets and the codas of the experimental stimuli.

| **Onset** | **N** | **%** | **Coda** | **N** | **%** |
|---|---|---|---|---|---|
| Bilabials | 21 | 21.9% | Bilabials | 9 | 9.4% |
| Labio-dentals | 10 | 10.4% | Labio-Dentals | 7 | 7.3% |
| Labio-velars | 7 | 7.3% | | | |
| Interdentals | 3 | 3.1% | Interdentals | 5 | 5.2% |
| Alveolars | 19 | 19.8% | Alveolars | 42 | 43.8% |
| Laterals | 5 | 5.2% | Laterals | 8 | 8.3% |
| Retroflex | 11 | 11.5% | Retroflex | 5 | 5.2% |
| Palato-Alveolars | 5 | 5.2% | Palato-Alveolars | 6 | 6.3% |
| Velars | 11 | 11.5% | Velars | 14 | 14.6% |
| Glottals | 4 | 4.2% | | | |
| **Total** | **96** | | **Total** | **96** | |

**Table 10.** Distribution of places of articulation for onset and coda consonants in experimental stimuli.

Table 10 clearly shows that nearly half of the stimuli codas were alveolars. This predominance of alveolars in coda position reflects a general trend in the English language. 48.2% (61,234 out of all 127,006 items in the CMU pronouncing dictionary end in alveolar consonants). This is, in part, due to the predominance of word-final alveolar inflections in English (e.g., plural /-s/, past-tense /-d/, etc.), but it holds true of monomorphemic items, as well—32.4% (194 out of 598) of the monomorphemic CVC words in the CMU pronouncing dictionary also end in alveolars. Thus, the ability of the participants in this study to identify alveolar coda consonants well probably reflects their realization that they should expect this place of articulation to appear often in the coda position of English words. It may also reflect a tendency on the part of the participants to simply choose words from the English lexicon—at random, even—which happened to end in alveolar consonants. Such a bias towards responding with coda consonants which end in alveolar consonants may have converged with the greater likelihood of this place of articulation in the stimuli codas to increase the overall hit rate for coda alveolars across all experimental conditions. To the extent to which this increased hit rate was due to response bias, it does not reflect an increased sensitivity to the visual cues for the alveolar place of articulation in coda position. Nonetheless, this increase in the percentage of hits for alveolars in coda position probably accounts in large part for the increased difference between the percentage of correct viseme identifications and the percentage of correct phoneme identifications for the coda consonants in the full-face conditions.

Breaking down correct identification scores by individual places of articulation also reveals a number of interesting differences between the perception of full-face and point-light stimuli. In general, the identification of place in the full-face stimuli was much better than it was in the point-light stimuli. For example, the average percent correct for bilabial place of articulation in onset position, across all four feedback conditions, hovered around 90% for the full-face stimuli, but was only a modest 70% in most of the point-light conditions. For other places of articulation, however, the point-light groups did far worse than might be expected. In particular, the participants had almost no ability to identify the interdental place of articulation in the onset of the point-light stimuli, scoring little better than 0% correct for interdental consonants in all feedback conditions. The corresponding percentages correct for the full-face groups were, on the other hand, close to 90%. The complete inability of the point-light groups to identify a place of articulation which was relatively easy for the full-face groups to perceive probably reflects significant gaps in the particular pattern of fluorescent dots that were placed on the speaker's faces, lips, teeth and tongues during the production of the point-light stimuli. While one dot was placed on the blade of each speaker's tongue, along with two dots each on both rows of teeth, there were no dots placed on the edge or tip of the speaker's tongue. Such dots may not have provided salient cues to the interdental place of articulation. The lack of such cues in the point-light stimuli meant the participants could perceive nothing in them that indicated that the speaker had produced an interdental consonant. This seems to have been especially true in onset position; for the interdentals in coda position, on the other hand, the participants' performance improved modestly to between 5% and 10% correct identifications. This improvement was probably due in large part to the transitional cues afforded the viewers by the visual offset of the vocalic portion of the CVC stimuli, rather than any particular configuration or dynamic transformation of the fluorescent dots on the interdental articulators.

The point-light groups also had particular difficulty identifying the retroflex consonant /r/ in both onset and coda positions. They correctly identified less than 5% of the /r/ tokens in onset position, while the full-face groups correctly identified between 40% and 60% of the /r/ tokens in the onset of their stimuli. For the /r/s in coda position, the point-light groups' percentages correct increased modestly to between 5% and 10%, while the full-face groups' performance decreased slightly to around 40% correct, in general. The reason behind the participants' inability to identify /r/ in the point-light stimuli may, in essence, be the same as that proposed for the difficulty they had in identifying interdentals—i.e., a lack of fluorescent dots at the appropriate places on the retroflex articulators. However, the two groups' perception of /r/ differed in more than just their ability to identify this segment correctly; the point-light groups consistently misidentified /r/ tokens in a different way than the full-face groups misidentified them. Tables 11 and 12 show confusion matrices for the retroflex stimuli, broken down by stimulus type, feedback condition and experiment half. Table 11 shows this data for the /r/s in onset position while Table 12 provides the same data for /r/s in coda position. Table 12 reveals that participants most often misidentified /r/ in the onset of the full-face stimuli as labio-velar /w/. For example, one participant misidentified "rang" as "ways." For the point-light stimuli, however, the participants primarily misidentified onset /r/ as a bilabial (e.g., /b/, /p/ and /m/). One participant, for example, misperceived "rang" as "bounce." That /r/ might be misperceived as /w/ in the full-face stimuli is not surprising, since both consonants have similar lip-rounding gestures (Johnson, 2002). Children often substitute /w/ for /r/, in fact, before they have learned to produce the (invisible) tongue-curling and pharynx-constricting gestures necessary to make an adult-like /r/ sound in English. Why the /r/ tokens in the point-light stimuli were so often misidentified as bilabials and not labio-velars, however, is not clear. This may reflect a bias in the participants to identify any sound with a labial gesture as bilabial, since this is the place of articulation they are most likely to identify correctly and therefore receive positive feedback on.

Whatever the reason behind the point-light groups' misidentification of /r/ may be, however, it is interesting to note that they were able to identify labio-velars in onset position quite well—especially those groups who received feedback. Those three groups correctly identified between 60% and 80% of labio-velars in onset position, whereas the no feedback group scored 35% to 5% for the same stimuli across both halves of the experiment. The scores for the point-light groups were, in fact, nearly as high as those for the full-face groups, which topped out near ceiling between 95% and 100%. These results suggest that the participants could readily identify the cues to labio-velar place of articulation that were preserved in the point-light stimuli, and that feedback (of any kind) was also particularly helpful in attuning their perceptual systems to these cues. However, this pattern of results suggests that whatever aspects of articulation appear to be visually similar between retroflex /r/ and labio-velar /w/ in the full-face stimuli were not preserved in the PLDs used in this experiment, since the point-light groups did not have difficulty identifying these cues in the labio-velar stimuli themselves. In a broader sense, this pattern of misidentifications also indicates that the point-light stimuli do not just transmit a simplified representation of the dynamic cues in the full-face displays. Instead, there is some degree of dissociation between the two visual representations of the same set of articulatory gestures.

Point-Light

| AVF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 67 | 5 | 2 | 1 | 8 | 5 | **2** | 4 | 1 | 0 | 5 | 8 | 108 |
| 2nd | 75 | 3 | 11 | 0 | 4 | 4 | **9** | 3 | 3 | 0 | 2 | 9 | 123 |
| | | | | | | | | | | | | | |
| AF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
| 1st | 96 | 4 | 2 | 0 | 10 | 0 | **1** | 4 | 4 | 2 | 8 | 8 | 139 |
| 2nd | 83 | 5 | 9 | 0 | 13 | 0 | **4** | 5 | 2 | 2 | 2 | 11 | 136 |
| | | | | | | | | | | | | | |
| OF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
| 1st | 78 | 4 | 5 | 1 | 8 | 3 | **5** | 3 | 2 | 5 | 5 | 10 | 129 |
| 2nd | 85 | 5 | 8 | 0 | 7 | 1 | **7** | 1 | 3 | 4 | 7 | 7 | 135 |
| | | | | | | | | | | | | | |
| NF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
| 1st | 65 | 5 | 4 | 0 | 13 | 2 | **3** | 1 | 3 | 3 | 9 | 17 | 125 |
| 2nd | 33 | 0 | 4 | 0 | 3 | 0 | **2** | 0 | 1 | 1 | 3 | 4 | 51 |

Full-Face

| AVF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 2 | 0 | 38 | 0 | 0 | 0 | **66** | 0 | 0 | 0 | 2 | 3 | 111 |
| 2nd | 2 | 0 | 32 | 0 | 1 | 1 | **84** | 0 | 0 | 0 | 0 | 0 | 120 |
| | | | | | | | | | | | | | |
| AF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
| 1st | 3 | 0 | 57 | 0 | 0 | 1 | **58** | 0 | 0 | 0 | 0 | 0 | 119 |
| 2nd | 3 | 0 | 48 | 0 | 3 | 0 | **68** | 0 | 0 | 1 | 0 | 0 | 123 |
| | | | | | | | | | | | | | |
| OF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
| 1st | 5 | 2 | 47 | 0 | 0 | 0 | **48** | 0 | 0 | 1 | 1 | 2 | 106 |
| 2nd | 2 | 1 | 46 | 0 | 1 | 0 | **62** | 0 | 1 | 0 | 0 | 1 | 114 |
| | | | | | | | | | | | | | |
| NF | bi | ld | lv | id | al | la | **re** | pa | ve | gl | None | Other | total |
| 1st | 4 | 1 | 48 | 0 | 0 | 0 | **41** | 0 | 1 | 1 | 2 | 3 | 101 |
| 2nd | 2 | 0 | 54 | 0 | 1 | 0 | **61** | 1 | 0 | 0 | 0 | 0 | 119 |

**Table 11.** Response totals for /r/ phonemes in the onset of the stimulus. (Response place key: bi = bilabial, ld = labio-dental, lv= labio-velar, id = interdental, al = alveolar, la = lateral, re = retroflex, pa = palato-alveolar, ve = velar, gl = glottal).

The two groups of participants also showed a different set of response biases when they misidentified retroflex /r/ in coda position. Table 12 provides more details about these biases; note that the point-light group often misidentified coda /r/s as either alveolars or bilabials. Over the course of the experiment, however, this bias appears to shift away from the bilabials and towards more anterior places of articulation, such as lateral and velar. For the full-face stimuli, on the other hand, there are very few alveolar or bilabial misidentifications in either half of the experiment. There are, however, many misidentifications of retroflex /r/ as lateral /l/, which seem to increase in frequency—across most feedback conditions—between the first and the second halves of the experiment. One participant in the

Point-light

| AVF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 7 | 0 | 2 | 13 | 5 | **4** | 1 | 7 | 7 | 4 | 50 |
| 2nd | 3 | 1 | 0 | 16 | 7 | **5** | 1 | 8 | 3 | 11 | 55 |

| AF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 12 | 1 | 2 | 12 | 1 | **5** | 3 | 4 | 9 | 8 | 57 |
| 2nd | 5 | 1 | 1 | 23 | 9 | **7** | 0 | 11 | 5 | 6 | 68 |

| OF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 11 | 0 | 1 | 13 | 7 | **7** | 0 | 2 | 7 | 9 | 57 |
| 2nd | 6 | 4 | 0 | 7 | 9 | **8** | 1 | 7 | 7 | 14 | 63 |

| NF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 10 | 0 | 2 | 12 | 0 | **0** | 0 | 1 | 5 | 6 | 36 |
| 2nd | 8 | 0 | 3 | 12 | 1 | **2** | 1 | 3 | 6 | 8 | 44 |

Full-face

| AVF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 0 | 1 | 0 | 5 | 7 | **22** | 0 | 1 | 5 | 12 | 53 |
| 2nd | 1 | 1 | 0 | 0 | 14 | **21** | 0 | 1 | 5 | 9 | 52 |

| AF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 1 | 0 | 0 | 1 | 2 | **26** | 0 | 1 | 0 | 13 | 44 |
| 2nd | 0 | 1 | 0 | 5 | 11 | **24** | 0 | 0 | 5 | 20 | 66 |

| OF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 0 | 2 | 0 | 5 | 8 | **14** | 0 | 2 | 6 | 9 | 46 |
| 2nd | 1 | 0 | 0 | 6 | 6 | **21** | 1 | 3 | 5 | 11 | 54 |

| NF | bi | ld | id | al | la | **re** | pa | ve | None | Other | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 3 | 1 | 0 | 3 | 9 | **22** | 1 | 1 | 3 | 6 | 49 |
| 2nd | 0 | 1 | 0 | 0 | 12 | **25** | 0 | 1 | 4 | 8 | 51 |

**Table 12.** Response totals for /r/ phonemes in the coda of the stimulus. (Response place key: bi = bilabial, ld = labio-dental, lv= labio-velar, id = interdental, al = alveolar, la = lateral, re = retroflex, pa = palato-alveolar, ve = velar, gl = glottal).

full-face condition, for example, misidentified "beer" as "bowl" in the full-face condition, while a different participant misidentified the corresponding point-light stimulus as "put." It is not clear why the participants tended to misidentify coda /r/s as /l/s in the full-face stimuli; it is possible that they may have picked up on a shared element of semi-syllabicity in the codas ending in these two liquids. The tendency for the participants in the point-light groups to misidentify coda /r/s as alveolars, on the other hand, may simply reflect the participants' broader tendency to respond with words which ended in alveolars, especially when there was insufficient evidence in the stimulus to specify an alternative place of articulation in the coda consonant.

## Conclusions

This study was designed to investigate the potential differences in perception between the visual-only PLDs and full-face displays of speech. We explored the effects that different kinds of feedback might have on the participants' ability to perceive speech in either kind of visual-only display. The results of this investigation confirmed the prediction that it should be easier to perceive speech in full-face displays than in PLDs. The participants in the full-face conditions correctly identified the words, phonemes and visemes at consistently higher levels than the participants in the point-light conditions. These results are likely due to the fact that full-face displays provide more visual information about articulatory gestures to perceivers than PLDs do. The full-face displays not only show all of a speaker's face but also potentially provide complementary static cues to perceivers instead of just the dynamic, time-varying cues found in PLDs. The participants in this study also had no experience perceiving speech in PLDs prior to this experiment, and may therefore, have found them perceptually difficult simply because they were unfamiliar with them. The results of this study do confirm, however, that--despite the impoverished visual signals in the point-light stimuli and the fact that none of the participants had seen them before—observers can, to a limited extent, accurately perceive speech in visual-only PLDs (cf. Rosenblum et al., 1996, Rosenblum & Saldaña, 1996; Lachs & Pisoni, submitted). This finding indicates that observers can perceive speech using only dynamic cues—i.e., moving points of light—without necessarily being able to recognize the pattern of the talker's face or their individual articulators. It is presumably possible for observers to do this because the pattern of point-light movements is highly constrained and lawfully determined by the articulatory events in speech that the observers perceive as meaningful.

Investigating the differences in the participants' ability to perceive particular places of articulation in the two different types of stimuli revealed that the PLDs are not just simplified or impoverished versions of the full-face displays. Certain places of articulation—e.g., interdental and retroflex—were nearly impossible for participants to perceive in the PLDs despite their relative ease of perceptibility in the full-face stimuli. Furthermore, certain systematic misidentifications emerged in the perception of point-light stimuli (e.g., bilabials for retroflexes) which differed from more common patterns of misidentification found in full-face displays (e.g., labio-velars for retroflexes). These results indicate that the particular patterns of dot placement used in the production of the point-light stimuli in this experiment were probably less than ideal; they failed to capture certain aspects of the dynamics of articulation which are clearly perceptible in the full-face displays and also gave viewers unusually misleading information about certain places of articulation. For this reason, the point-light stimuli developed a peculiar form of perceptual independence from their full-face counterparts; the participants perceived articulatory events in them which they would not have perceived in full-face displays of the same events. This finding serves as an important caution against concluding too much about the visual perception of speech under normal viewing conditions based only on the results of studies which have shown that observers can successfully perceive speech in PLDs (e.g., Rosenblum & Saldaña, 1996; Lachs & Pisoni, submitted); the perception of speech in one type of display does not necessarily reflect the perception of speech in the other. It is also not known at this time whether different patterns of dot placement might eliminate this perceptual independence between the two visual representations of speech. However, determining a pattern of dots to use in point-light speech stimuli which can more faithfully represent the dynamics of articulation in fully-illuminated displays may provide a fruitful line of future research.

The results of this preliminary study on the effects of feedback on participants' improvement in the perceptual task were unfortunately somewhat inconclusive. No main effects of feedback were found in any of the experimental conditions, and the one significant feedback by experimental half interaction which did emerge yielded a pattern of improvement across the four feedback conditions which was not consistent with any of the effects that feedback was predicted to have on the participants' perceptual improvement. Thus, while it may seem rational to suggest that dynamic, event-based feedback—in audio or audio-visual form—may improve viewer performance in a visual-only speech perception task more than either static, symbolic feedback or no feedback at all would, the results of this experiment do not provide any solid empirical support for that hypothesis. This null result leaves open the question of just how important dynamic, articulatory event-based feedback really is to viewers in improving their skills in the visual-only perception of speech. It also remains unclear, for that matter, whether feedback is of any particular use to participants in a visual-only perception task, when they can evidently improve just as much without feedback as they can when they receive either orthographic or audio-visual feedback on the stimuli they have just seen. The lack of statistically significant differences in perceptual improvement between the various feedback groups suggests, in fact, that participants may have been able to improve by simply becoming more familiar with the experimental task. Practice with the visual-only perception task may have helped the participants fine-tune their perceptual systems into what they already knew about the acoustic consequences of particular articulations. For example, one participant remarked after the experiment that she had often tried to articulate the response she had in mind in order to determine if it matched the articulatory gestures she had seen the speaker make on the computer screen. Tapping into such tacit knowledge of articulation may thus provide a better guide to correctly identifying visual-only stimuli than being informed—after the fact—what a speaker in a silent video has just said.

However, the results of this preliminary study do not necessarily close the theoretical door on the possible efficacy of feedback on improvement in a visual-only speech perception experiment. There are a number of ways in which the paradigms used in this experiment might be modified in order to provide more optimal conditions for the emergence of the anticipated effects of feedback on perceptual improvement. For example, even though the various amounts of improvement made by the participants in the different feedback groups were not significantly different from one another, some groups did show a trend at improving more than other groups in the identification of certain aspects of the stimulus words. The audio and audio-visual feedback groups, for instance, improved more on vowel identification than either the orthographic or no feedback groups. Such a trend might become statistically significant if the experiment contained more than just 96 visual-only stimuli, thereby giving the participants more time (and practice) to improve their perceptual skills.

Feedback may also become more efficacious if participants get second chances to identify stimuli they have already seen before and received feedback on. In this experiment, participants only received feedback on individual words, all of which they saw only once. In this vein, it may also be more informative to analyze the efficacy of feedback in terms of how the ability of the participants to identify a particular stimulus (or part thereof) improves with each successive presentation of that stimulus in the experiment. The analyses carried out in this report looked at the effects of feedback over the comparatively broad scopes of the first and second halves of the experiment. Small-scale improvements within each half may have thus been lost in the analysis. Analyzing feedback effects in terms of the number of times a particular stimulus has been seen might also be made easier by balancing the proportions of places of articulation in coda and onset consonants across all stimuli. The coda consonants in this experiment's stimuli included a large proportion of alveolars, while the onset consonants had—to a lesser extent--disproportionate amounts of bilabials and alveolars. It was therefore not possible to make equitable assessments of the participants' improvement in identifying particular places of articulation after a certain number of presentations of that place of articulation in the experiment—the participants saw too many tokens of some places of articulation and too little of others. Furthermore, the predominance of alveolars—which do not have strong visual cues to their place of articulation--in the coda position of the experiment's stimuli also facilitated deceptively positive effects of response bias on the percentage of correct coda identifications. This disproportionate representation of alveolars in coda position may also have inhibited the improvement of participants' abilities to identify more visually salient—but less well-represented—places of articulation in the coda consonants. Balancing the stimuli for place of articulation in future studies may help eliminate some of the confounding influences that the variability of ease of place identifiability may have on the participants' ability to improve in a visual-only speech perception task.

Although such possibilities for improving on the experimental paradigm in this study remain, its results did nonetheless demonstrate that participants could accurately identify individual words and phonemes from PLDs, and that their ability to perform this perceptual task improved over the course of a one-hour experiment. That human observers can perceive speech from only the impoverished, dynamic cues represented in these displays is quite remarkable; studying further just how much this ability may improve—through either feedback or practice—may shed further light on the importance of dynamic cues in visual-only speech perception, as well as the role that event-based information—in any sensory modality—might play in helping people perceive the significant articulatory events which produce the dynamic features of speech.

## References

CMU Pronouncing Dictionary, version 0.6. Available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

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 and Performance, 17,* 816-828.

Fowler, C.A. & Rosenblum, L.D. (1991). Perception of the phonetic gesture. In I.G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory*. Pp. 33-59. Hillsdale, NJ: Lawrence Earlbaum.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201-211.

Johnson, K. (2002). *Acoustic and Auditory Phonetics.* 2nd ed. Oxford: Blackwell.

Lachs, L. (2002). *Vocal tract kinematics and crossmodal speech information.* (Research on Speech Perception Technical Report No. 10). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., & Pisoni, D.B. (2004). Crossmodal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 30,* 378-396.

Lachs, L., & Pisoni, D.B. (submitted). Specification of crossmodal source information in isolated kinematic displays of speech. *Journal of the Acoustical Society of America*.

Lachs, L. & Hernandez, L.R. (1998). Update: the Hoosier Audiovisual Multitalker Database. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 377-388). Bloomington, IN: Speech Research Laboratory, Indiana University.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Pashler, H., Cepeda, N., Wixted, J., & Rohrer, D. (In press). When does feedback facilitate learning of words and facts? *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Reed C.M., Durlach N.I., Braida L.D., & Schultz M.C. (1989) Analytic study of the Tadoma method: effects of hand position on segmental speech perception. *Journal of Speech and Hearing Research, 32*, 921-929.

Rosenblum, L.D., Johnson, J.A. & Saldaña, H.M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research, 39,* 1159-1170.

Rosenblum, L.D. & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *22,* 318-331.

Sheffert, S.M., Lachs, L. & Hernandez, L.R. (1996). The Hoosier Audiovisual Multitalker Database. In *Research on Spoken Language Processing Progress Report No. 21* (pp. 578-583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212-215.

Walden, B.E., Prosek, R.A., Montgomery, A.A., Scher, C.K. & Jones, C.J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20,* 130-145.

# RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 26 (2003-2004)
*Indiana University*

# Nonword Repetition and Reading in Deaf Children with Cochlear Implants[1]

**Caitlin M. Dillon and David B. Pisoni**

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

[1] This work was supported by NIH-NIDCD Research Grant R01 DC00111 and T32 Training Grant DC00012 to Indiana University. We are grateful to Ann Geers, Chris Brenner, and the research staff at Central Institute for the Deaf in St. Louis, MO in 1999 and 2000 for testing the cochlear implant users and making the literacy data available to us for this study. We would also like to thank Rose Burkholder, Allyson Carter, and Miranda Cleary for their valuable insights and help with this investigation.

# Nonword Repetition and Reading in Deaf Children with Cochlear Implants

**Abstract.** In the present study, we report analyses of nonword repetition responses from 76 experienced pediatric cochlear implant (CI) users. Immediate repetition of a spoken nonword stimulus requires a child to correctly perceive a novel sound pattern, maintain a representation and rehearse it in phonological working memory, and then reproduce a phonological pattern as an output response. Nonword repetition performance by normal-hearing participants has been shown to be affected by the structure of the language. Systematic effects of the structure of the ambient language on nonword repetition performance reflect the participants' reliance on their phonological knowledge, which developed based on their experience with the language. The acquisition of reading and literacy skills in normal-hearing children has also been found to be dependent on the development of phonological knowledge and the use of phonological processing skills. In the present study conducted with deaf children who use CIs, we report findings showing that nonword repetition performance is strongly correlated with several measures of phonological awareness and reading comprehension, and is related to lexical diversity. These new findings provide additional converging evidence for the proposal that the development of phonology is an important prerequisite for the acquisition of reading and literacy skills in this clinical population. The children's ability to provide nonword repetition responses, as well as correlations obtained between their performance on this task and several measures of reading indicate that they relied on phonological knowledge to complete both the nonword repetition and reading tasks. The ability to encode, decompose, and reassemble sound patterns of speech appears to be a fundamental prerequisite for both speech perception and reading.

## Introduction

Studies of children in pre-reading and early reading stages often discuss "reading readiness" skills (Adams, 1990). Reading readiness skills reflect phonological awareness, the extent to which the child is consciously aware that individual words have an internal structure that is composed of sequences of speech sounds (such as phonemes), which can be represented orthographically with graphemes (Rayner & Pollatsek, 1995). Phonological awareness is typically measured by the child's performance on behavioral tasks in which he/she is required to demonstrate implicit or explicit awareness of the existence of phonological structure. For example, in some procedures the child is asked to recognize whether words rhyme, or whether they start or end with the same sound or are minimal pairs. A child's conscious awareness of the existence of phonological structure (e.g., phonemes), indexed by these phonological awareness or reading readiness tasks, is a necessary prerequisite for him/her to develop the ability to map orthographic representations of speech (graphemes) onto phonemes. The ability to easily and rapidly complete grapheme-to-phoneme conversion is related to reading ability in young normal hearing children (Adams, 1990; Marschark, 2003; Rayner & Pollatsek, 1995). When learning to read, children who have developed phonological representations of the sounds of their ambient language can take greater advantage of such processes as "inner speech" (Conrad, 1979) and verbal rehearsal processes in working memory (Baddeley & Gathercole, 1992).

Phonological awareness is one set of cognitive operations involved in phonological processing of speech, along with retrieval of phonologically coded information from the lexicon, and encoding of sound patterns in phonological working memory (Troia, 2004; Wagner & Torgeson, 1987). Phonological processing abilities and reading and literacy development have been found to be interdependent, with development in each causing further development in the other (see Brady, 1997; Troia, 2004). For example, Bradley and Bryant (1983) found that training children using phoneme awareness tasks led to better phonemic awareness and improved reading skills several years later in comparison to children in control groups. On the other hand, development of reading and spelling skills can also lead to increased phonological awareness (see, e.g., Cassar & Treiman, 2004).

The relationship between hearing status and reading skills has been a topic of interest for centuries (e.g., Dalgarno, 1680). In the more recent past, research on speech and reading has been guided at least in part by the development of theories and the completion of empirical studies in areas such as reading, visual word recognition, speech perception, spoken word recognition, and phonological working memory. Studies of the reading skills of deaf children and adults in the past 50 years have consistently shown that deaf children's reading readiness and reading skills are significantly delayed relative to their normal-hearing peers, and often do not exceed a 4th-grade level (Paul, 2003). Phonological knowledge and phonological processing skills in visual word recognition tasks and reading have also been shown to be utilized to some extent by deaf readers (Hanson, 1991).

The reading and literacy skills of deaf children with cochlear implants have been studied recently by several researchers. Spencer, Tomblin, and Gantz (1997) reported that a group of 2- to 13-year-old children with cochlear implants completed a reading comprehension task with greater accuracy than deaf children without cochlear implants. Although over one-fourth of the children with cochlear implants achieved reading levels that were 30 or more months below their grade levels, almost one-fourth of the children with cochlear implants achieved reading levels at or above their grade levels. Spencer et al. concluded that the auditory information about speech provided by a cochlear implant may facilitate a deaf child's ability to decode or recode orthographic representations of speech into a "speech code" (see Conrad, 1979).

Further support for Spencer et al.'s (1997) conclusions was provided by Geers (2003), who reported the results of a study that included all of the children who participated in the Central Institute for the Deaf (CID) Education of the Deaf Child program (N=181, including the 76 children described in the present study). The children in this study were all 8 and 9 years old and had received their cochlear implant before the age of 5. Geers found that the children averaged mid to high 2nd grade reading levels on the PIAT Recognition and PIAT Comprehension measures. They used the children's scores on these two measures to calculate total reading scores for each child, for which standard scores are available. The standard scores were based on the expected grade level of the children based on their chronological age. The total reading standard scores revealed that 52% of the children scored within the average range of children their age, and 48% were below average. On the rhyming task, children performed relatively well. The results suggested that the children were using both phonological and visual cues to complete this task. Incorrect responses were provided most often for word pairs that rhymed but were orthographically dissimilar.

In the present study on nonword repetition, we were interested in the extent to which the children's performance on traditional reading readiness, single-word reading, and reading comprehension measures was related to their performance on an auditory-only task that measures their sublexical phonological abilities to perceive, rapidly encode, rehearse, and then reassemble a novel sound pattern for speech production. Nonword repetition has been shown to be strongly correlated with the development of reading skills in normal-hearing children with and without phonological disorders (see Brady, 1997). Although simple on the surface, nonword repetition is a complex information processing task that loads heavily on phonological processing skills and verbal rehearsal processes (Pisoni, this volume). In order to accurately reproduce a nonword auditory pattern, it is necessary for a child to accurately complete the following subprocesses:

- Perceive and encode a novel sound pattern in an auditory-only mode without the aid of speech-reading or other context or content
- Store and verbally rehearse the novel sound pattern in immediate memory
- Reassemble and translate the perceived novel sound pattern into a sensory-motor articulatory program to produce speech output

We report analyses of a subset of the children in Geers' (2003) study, who also completed a nonword repetition task. If nonword repetition skills are highly correlated with reading skills as the literature on normal-hearing children would suggest, we would predict that better phonological processing skills would be related to better reading skills in deaf children with CIs.

## Method

### Participants

Eighty-eight children who participated in the CID Education and the Deaf Child program in 1999 or 2000 (see Geers & Brenner, 2003) participated in the both the nonword repetition task and the reading tasks described below. Twelve children were excluded from the analysis because they provided responses to less than 75% of the target nonwords. The remaining 76 children were included in the present study. Thirty-six were male and 40 were female. Seventy-four children used a Nucleus 22 CI and the SPEAK coding strategy. One child used a Nucleus 24 CI and one child used a Clarion CI.

Table 1 provides a summary of the demographic characteristics of the children. Their mean chronological age at the time of testing was 8.9 years (range 7.8-9.9, *SD* = 0.6), as shown in Table 1. Sixty-four of the children were congenitally deaf, six became deaf before the age of one year, and the remaining six became deaf by the age of 3 years. The children's mean duration of deafness was 37.2 months (range 7-65, *SD* = 13.1). The children's mean age at time of implantation was 3.3 years (range 1.9-5.4, *SD* = 1.0). The children had used their implant for a mean of 5.6 years at the time of testing (range 3.8-7.5, *SD* = 0.8). The children's mean communication mode scores were based on a parent questionnaire. Children with Communication Mode scores of 15 or higher were considered Oral Communication (OC) users (i.e., their educational programs emphasized oral communication methods). Children with communication mode scores below 15 were considered Total Communication (TC) users (i.e., both manual and oral communication methods were used in their educational environment; see Geers & Brenner, 2003).

| Demographic Variable | Mean (SD) | Range |
|---|---|---|
| Age at Onset of Deafness (months) | 2.3 (6.4) | 0-36 |
| Duration of Deafness (months) | 37.2 (13.1) | 7-65 |
| Age at Implantation (years) | 3.3 (1.0) | 1.9-5.4 |
| Duration of Implant Use (years) | 5.6 (0.8) | 3.8-7.5 |
| Chronological Age (years) | 8.9 (0.6) | 7.8-9.9 |
| Number of Active Electrodes | 18.4 (2.3) | 8-22 |
| Communication Mode Score | 19.8 (7.7) | 6-30 |

**Table 1.** Summary of the demographic make-up of the 76 children.

## Nonword Repetition Task

**Stimulus Materials and Procedure.** The 20 target nonwords used in the present study were a subset of the nonwords in the Children's Test of Nonword Repetition (Gathercole, Willis, Baddeley, & Emslie, 1994; see also Carlson, Cleary, & Pisoni, 1998). The nonwords, shown in Table 2, were balanced in terms of syllable number and included 112 target consonants and 68 target vowels. Each child was asked to listen to the novel nonwords, presented one at a time, and attempt to repeat the nonword aloud. The children heard digital recordings of a female native speaker of American English played over a loudspeaker at approximately 70 dB SPL. The stimuli and responses were recorded onto digital audio tape for later analysis.

| Number of Syllables | | | |
|---|---|---|---|
| 2 | 3 | 4 | 5 |
| ballop | bannifer | comisitate | altupatory |
| prindle | berrizen | contramponist | detratapillic |
| rubid | doppolate | emplifervent | pristeractional |
| sladding | glistering | fennerizer | versatrationist |
| tafflist | skiticult | penneriful | voltularity |

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

**Nonword Transcriptions.** All of the nonword repetition responses were independently transcribed by two phonetically trained listeners. Disagreements were resolved by consensus (93% agreement). A third listener resolved the remaining 7% disagreements. These phonemic transcriptions were then used to calculate two "suprasegmental scores" for a subset of the children (N = 24), who provided responses to all 20 target nonwords: (1) percent of imitations with the correct number of syllables, and (2) percent of imitations with correct primary stress placement. The phonemic transcriptions were also used to calculate "segmental scores" for all 76 children: (1) percent consonants correct, based on the number of consonants reproduced with correct place (labial, coronal, dorsal), manner (stop, fricative, liquid, nasal), and voicing (voiced or voiceless), both out of the total number of target consonants (N = 112), and out of the total number of target consonants in the nonwords for which the child provided a response; and (2) percent vowels correct, based on the number of vowels reproduced with correct height (high, mid, or low) and backness (front, central, or back), both out of the total number of target vowels in the target nonwords (N = 68), and out of the total number of target vowels in the nonwords for which the child provided a response. In addition to these scores, the children's nonword responses were played back to naïve listeners to obtain perceptual goodness ratings.

**Scores Based on Perceptual Accuracy Ratings.** The target nonword patterns and the child's attempted nonword repetitions were played back to groups of normal-hearing college-age adult listeners who were asked to make similarity judgments. On each trial, the listener heard the target nonword followed by a child's attempt to repeat that target nonword. Listeners were asked to provide goodness ratings of the child's response on a scale of 1 (poor) to 7 (perfectly accurate), which were used to calculate a mean rating score per child.

## Reading Outcome Measures

**Stimulus Materials and Procedures.** Three measures of reading were also obtained from these children. The Word Attack subtest of the Woodcock Reading Mastery Tests - Revised (WRMT; Woodcock, 1987) was administered to all of the children. The Word Attack subtest is a nonword reading task that includes 45 nonwords or extremely rare real words. Each child was asked to read aloud the nonwords one at a time. The child cannot complete this task by relying on visual recognition or reading skills because the stimuli are unfamiliar nonwords. Instead, the Word Attack subtest measures the child's "ability to apply phonic and structural analysis skills to pronouncing words that are not recognizable by sight" (Woodcock, 1987: 6).

The children also completed the two subtests of the Peabody Individual Achievement Test-Revised (PIAT; Dunn & Markwardt, 1989). The Reading Recognition subtest of the PIAT includes 100 items. The first 16 items consist of four-alternative forced choice questions requiring a pointing response. This measure was designed to test "reading readiness" skills, which are assumed to be essential pre-requisites for a child learning to read (Markwardt, 1998). Several types of items are included in the reading readiness part of the PIAT Reading Recognition subtest. For example, the child is shown a letter or word such as "B," "GO," or "to" and is asked to point to one like it from among four choices; or, the child is asked to name the object shown in four pictures and then choose the picture of an object whose name does not start with the same sound as the other three objects, such as "ball" from among pictures of a ball, pencil, pan and pie. Several other items require the child to choose an item that begins with the same sound as a stimulus picture, from among four pictures or four written words. Items 17-100 all involve single real-word reading. The questions are ordered in terms of increasing difficulty, ranging from kindergarten level to 12th-grade level. In the Reading Recognition subtest, the child earns one point for every correct answer to items 1 through 16, and for every correct pronunciation of items 17-100, with each pronunciation counted as either correct or incorrect after one attempted pronunciation.

The Reading Comprehension subtest of the PIAT was also given to all of the children. This reading measure includes 82 four-alternative forced-choice items that require a pointing response. The test items are meaningful narrative sentences designed to test literal reading comprehension (as opposed to interpretation of information or recognition of inferences; Markwardt, 1998). For each item, the child is shown a sentence and is told to read it to him/herself only once. Then the child is shown a page with four pictures and is asked to point to the picture that best represents the meaning of the sentence. As in the Reading Recognition subtest, the items in the Reading Comprehension subtest are ordered in terms of increasing difficulty over a wide range, e.g., *There is the sun.*, *The eagle floats on its wings as it travels in search of a feast.*, and *The residence has been essentially reduced to rubble, the remainder being only the foundation.* The child is given one point for each correct response.

Finally, the children also participated in a Rhyming Task (Geers, 2003) in which, on each trial, they were presented with two words and asked to state whether or not the two words rhymed. The two words in each pair either rhymed or did not rhyme, and were either orthographically similar or dissimilar. The word pairs were counterbalanced in terms of these two characteristics. All of the reading tasks described above are referred to as "reading outcome measures" in the present report.

**Scores.** Grade Equivalent Scores were determined for the Word Attack (Woodcock, 1998), the Reading Recognition, and Reading Comprehension tasks (Markwardt, 1998). A Total Reading standard score was also calculated for each child. The child's raw scores on the two PIAT reading subtests were summed and converted to a standard score using the child's expected grade levels based on his/her age, because grade levels were not available for all children (Markwardt, 1998). Forty-one children were considered 3rd-graders and 35 children were considered 4th-graders. The Rhyming task was scored for "rhyme errors," the percentage of word pairs for which the child responded incorrectly (Geers, 2003).

### Lexical Diversity

Recent findings have revealed that nonword repetition performance is related to vocabulary size in normal-hearing adults, typically-developing children, and children with phonological disorders (Edwards, Beckman, & Munson, 2004; Munson, Edwards, & Beckman, in press). A direct measure of vocabulary size was not available for the deaf children with cochlear implants in the present study. However, the children had participated in a conversational oral interview as part of the larger CID study (Geers, Nicholas, & Sedey, 2003). The number of different words used by each child during the interview was calculated, and considered to be a measure of "lexical diversity," which is likely to reflect overall vocabulary knowledge. In the present study, we used this measure to investigate the relationship between lexical diversity, nonword repetition performance, and reading skills.

## Results

### Nonword Repetition Task

All of the children described in the present study provided a response to at least 15 of the 20 original nonword stimuli. More detailed summaries of the nonword repetition task results are reported in earlier studies by Carter, Dillon, and Pisoni (2002), Dillon, Cleary, Pisoni, and Carter (2004), Dillon, Pisoni, Cleary, and Carter (2004), and Dillon, Burkholder, Cleary, and Pisoni (in press). The children's nonword responses varied in terms of suprasegmental, consonant, vowel, and overall perceptual accuracy. A summary is provided in Table 3.

| Nonword Repetition Score | Mean (SD) | Range |
|---|---|---|
| % Correct # of syllables (N=24) | 65% (18%) | 35 - 95% |
| % Correct primary stress placement (N=24) | 62% (13%) | 30 - 85% |
| % Correct Cs out of Cs in all 20 target NWs (N=76) | 30% (17%) | 1 - 76% |
| % Correct Cs out of target Cs in responses (N=76) | 33% (17%) | 1 - 76% |
| % Correct Vs out of Vs in all 20 target NWs (N=76) | 44% (17%) | 9 - 75% |
| % Correct Vs out of target Vs in responses (N=76) | 48% (17%) | 13 - 78% |
| Mean  Perceptual Accuracy Ratings (N=76) | 3.1 (1.1) | 1.1 - 5.7 |

**Table 3.** Summary of means, standard deviations (SD), and ranges for the nonword repetition scores.

The 24 children for whom suprasegmental accuracy scores (i.e., number of syllables and placement of primary stress) were calculated had all produced responses to the complete set of 20 target nonwords. Overall, they produced a mean of 65% (range = 35-95%, *SD* = 18%) of their responses with the correct number of syllables, and 62% (range = 30-85%, *SD* = 13%) of their responses with the correct placement of primary stress. Percent consonants (Cs) and vowels (Vs) correct scores were calculated first out of the total number of target Cs in all 20 target nonwords (N = 112) and target Vs in all 20 nonwords (N = 68), respectively. Because some of the 76 children did not produce a repetition response to all 20 target nonwords, we also calculated individual percent Cs and Vs correct scores out of the total number of target consonants in only the target nonwords for which the child provided a response. The mean percent consonants correct scores, calculated in both ways described above, were 30% (range = 1 - 76%, *SD* = 17%), and 33% (range = 1-76%, *SD* = 17%), respectively. The mean percent vowels correct scores were slightly higher, 44% (range = 9-75%, *SD* = 17%) and 48% (range = 13-78%, *SD* = 17%), respectively.

The children’s nonword responses received mean accuracy ratings that ranged from 1.1 to 5.7 out of 7 (*M* = 3.1, *SD* = 1.1). As shown in Table 4, the different methods of scoring the nonword repetition task yielded scores that were strongly intercorrelated with each other.

| **Nonword Repetition Score** | **1. Syls** | **2. Str** | **3. Cs 1** | **4. Cs 2** | **5. Vs 1** | **6. Vs 2** | **7. Ratings** |
|---|---|---|---|---|---|---|---|
| 1. % Correct # of Syllables (N=24) | 1 | +.40 | +.69*** | +.69*** | +.70*** | +.70*** | +.67*** |
| 2. % Correct Primary Stress Placement (N=24) | | 1 | +.63** | +.63** | +.51* | +.51* | +.69*** |
| 3. % Correct Cs out of Cs in all 20 target NWs (N=76) | | | 1 | +.99*** | +.87*** | +.87*** | +.92*** |
| 4. % Correct Cs out of target Cs in responses (N=76) | | | | 1 | +.88*** | +.88*** | +.92*** |
| 5. % Correct Vs out of Vs in all 20 target NWs (N=76) | | | | | 1 | +.98*** | +.88*** |
| 6. % Correct Vs out of target Vs in responses (N=76) | | | | | | 1 | +.87*** |
| 7. Mean Accuracy Ratings (N=76) | | | | | | | 1 |

**p* < .05, ***p* < .01, ****p* < .001

**Table 4.** Intercorrelations among the nonword repetition scores.

Correlations between the measures of nonword repetition accuracy and age at implantation, duration of CI use, chronological age at the time of testing, and number of active electrodes did not reach significance. Correlations between nonword repetition accuracy and the demographic factors that did reach significance are shown in Table 5.

| **Nonword Repetition Score** | **Age at Onset** | **Comm. Mode** | **PIQ** |
|---|---|---|---|
| % Correct # of syllables (N=24) | +.38 | -.18 | +.01 |
| % Correct primary stress placement (N=24) | +.52** | +.36 | -.01 |
| % Correct Cs out of Cs in all 20 target NWs (N=76) | +.31** | +.54*** | +.22 |
| % Correct Cs out of target Cs in responses (N=76) | +.28* | +.54*** | +.22 |
| % Correct Vs out of Vs in all 20 target NWs (N=76) | +.24* | +.47*** | +.25* |
| % Correct Vs out of target Vs in responses (N=76) | +.20 | +.45*** | +.27* |
| Mean accuracy ratings (N=76) | +.32** | +.51*** | +.26* |

**Table 5.** Significant correlations between the children’s demographic characteristics and their nonword repetition scores.

## Reading Outcome Measures

A summary of the children's scores on the reading outcome measures is shown in Table 6. We report grade equivalent scores for the WRMT Word Attack subtest and the PIAT Recognition and Comprehension subtests. PIAT Total Reading standard scores were calculated using estimated grade levels based on the children's ages because grade levels were not available for all children (see also Geers, 2003). Forty-one children were considered 3rd-graders and 35 children were considered 4th-graders. Fifty-three children (70%) obtained Total Reading standard scores within the normal range for children their age. The remaining 23 children (30%) had Total Reading standard scores that were below the normal range for children their age (based on norms from Markwardt, 1998; Geers, 2003 results are based on norms from Dunn & Markwardt, 1987).

| Reading Outcome Measure | Mean (SD) | Range |
|---|---|---|
| WRMT Word Attack (Grade Equivalent Scores) | 3.3 (2.8) | 0.0 - 12.6 |
| PIAT Recognition (Grade Equivalent Scores) | 2.9 (1.0) | 0.4 - 6.1 |
| PIAT Comprehension (Grade Equivalent Scores) | 2.9 (1.5) | 0.0 - 8.7 |
| PIAT Total Reading (Standard Scores) | 87.6 (6.5) | 72 - 106 |
| Rhyme Errors (Percent) | 12.4 (7.2) | 0 - 37 |

**Table 6.** Summary of means, standard deviations (SD), and ranges for the reading outcome measures (N=76).

The number and corresponding percentage of children that performed at several grade levels on the WRMT Word Attack subtest and the PIAT Recognition and Comprehension subtests are shown in Table 7. Seventeen children (22%) scored above the 4th-grade level on the Word Attack, two children (3%) on Reading Recognition, and eight children (11%) on Reading Comprehension. Ten children (13%) scored below the first-grade level on at least one of the Word Attack, Reading Recognition, and Reading Comprehension tasks. As shown in Table 8, scores on the reading outcome measures were all highly intercorrelated; the percentage of rhyme errors was moderately correlated with the other reading outcome measures.

| Reading Outcome Measure | Below 1st grade level | 1st grade level | 2nd grade level | 3rd grade level | 4th grade level | Above 4th grade level |
|---|---|---|---|---|---|---|
| WRMT Word Attack (GE Scores) | 7 (9%) | 23 (30%) | 16 (21%) | 7 (9%) | 6 (8%) | 17 (22%) |
| PIAT Recognition (GE Scores) | 2 (3%) | 10 (13%) | 29 (38%) | 25 (33%) | 8 (11%) | 2 (3%) |
| PIAT Comprehension (GE Scores) | 6 (8%) | 6 (8%) | 37 (49%) | 16 (21%) | 3 (4%) | 8 (11%) |

**Table 7.** The number and corresponding percentage of children with grade equivalent scores below 1st grade level to above 4th grade level on the three reading outcome measures for which grade equivalency scores were available (N=76).

| Reading Outcome Measure | WRMT Word Attack | PIAT Reading Recog. | PIAT Reading Comp. | PIAT Total Reading | Rhyme Errors |
|---|---|---|---|---|---|
| WRMT Word Attack | 1 | +.83*** | +.68*** | +.82*** | -.37** |
| PIAT Reading Recognition | | 1 | +.78*** | +.88*** | -.40*** |
| PIAT Reading Comprehension | | | 1 | +.89*** | -.41*** |
| PIAT Total Reading | | | | 1 | -.42*** |
| Rhyme Errors | | | | | 1 |

***p* < .01, ****p* < .001

**Table 8.** Intercorrelations among the reading outcome measures (N=76).

We did not find any significant correlations between the reading measures and age at onset of deafness, duration of deafness, age at implantation, duration of CI use, or chronological age at the time of testing (all *p*'s > .11) or number of electrodes (after one outlier was removed, all *p*'s > .08). A t-test revealed no differences in performance by gender (*p*'s = .69, 71, .96, .81, .09 for the WRMT Word Attack, PIAT Recognition, PIAT Comprehension, PIAT Total Reading, and Rhyme Errors tasks, respectively). Correlations between the reading measures and both communication mode and performance IQ (Wechsler, 1991) reached significance (see Table 9).

| Reading Outcome Measure | Comm. Mode | PIQ |
|---|---|---|
| WRMT Word Attack | +.41*** | +.34** |
| PIAT Raw Recognition Scores | +.26* | +.35** |
| PIAT Raw Comprehension Scores | +.15 | +.39*** |
| PIAT Reading Standard Scores | +.25* | +.45*** |
| Rhyme Errors | -.14 | -.21 |

**p* < .05, ***p* < .01, ****p* < .001

**Table 9.** Significant correlations between the children's demographic characteristics and their scores on the reading outcome measures (N=76).

## Correlational Analysis

Because the different methods of scoring the nonword repetition task were all highly correlated with each other, we report only the correlations between the nonword repetition accuracy ratings and the reading outcome measures. We found that age at onset of deafness, communication mode, and performance IQ (measured using the WISC III, Wechsler, 1991; see also Geers, 2003 and Dillon et al., in press) were all significantly correlated with nonword repetition performance. We also computed partial correlations between nonword repetition accuracy ratings and the reading measures to control for these demographic variables. After these potentially confounding demographic factors were partialled out, we found that the children's performance on nonword repetition, a phonological processing task, and their performance on the measures of reading readiness and reading still remained significantly correlated. Finally, we computed partial correlations in which lexical diversity was also controlled, in addition to the potentially confounding demographic characteristics (age at onset of deafness, communication mode, and performance IQ). When lexical diversity was controlled, several of the correlations between children's nonword repetition performance and their reading scores no longer reached significance. The reading recognition scores and total reading scores remained significantly correlated with nonword repetition performance, but were substantially decreased.

| Reading Outcome Measure | Bivariate correlation | Partial correlation 1 | Partial correlation 2 |
|---|---|---|---|
| WRMT Word Attack | +.61*** | +.49*** | +.22 |
| PIAT Reading Recog. Scores | +.57*** | +.50*** | +.26* |
| PIAT Reading Comp. Scores | +.43*** | +.41*** | +.15 |
| PIAT Total Reading Scores | +.59*** | +.55*** | +.32** |
| Rhyme Errors | -.37** | -.29* | -.12 |

*$p < .05$, **$p < .01$, ***$p < .001$

**Table 10.** Correlations between nonword repetition accuracy ratings and reading outcome measures (N=76): Simple bivariate correlations, partial correlations 1 (controlling for age at onset of deafness, communication mode, and performance IQ), and partial correlations 2 (controlling for age at onset of deafness, communication mode, performance IQ, and lexical diversity).

## Discussion and Conclusions

Nonword repetition is a difficult information processing task that requires immediate and rapid phonological processing. Most of the deaf children with cochlear implants in the present study were able to complete the nonword repetition task with some measurable level of accuracy, although their performance was substantially worse than normal-hearing children. Several methods of scoring their performance were all highly intercorrelated. The deaf children with cochlear implants in this study demonstrated higher level reading skills than have traditionally been reported in deaf children (but see also Geers, 2003; Spencer et al., 1997). Many of the children's reading scores fell within the range of their normal-hearing age-mates. We also found that nonword repetition performance was strongly correlated with measures of reading readiness (such as letter-sound correspondences and rhyme recognition), single-word reading, nonword reading and read-sentence comprehension.

The strong correlation between the children's nonword repetition performance and their performance on a nonword reading task, the Word Attack, suggests that the children used the same phonological processing skills to read nonwords out loud as they did to repeat spoken nonword stimuli. Because the nonword reading task requires a spoken response, differences in speech production and speech intelligibility could be potentially confounding factors. However, we also found that the children's nonword repetition performance was strongly correlated with their performance on the PIAT Reading Comprehension subtest, a sentence comprehension task that does not involve processing speech or spoken language input signals. Children who were better able to "decompose" and "reassemble" spoken nonwords were also better at reading and comprehending meaningful written sentences.

This correlation is revealing and is theoretically significant. Because the stimuli used in the nonword repetition task are nonword phonological patterns, the children could not rely directly on the retrieval and access of previously developed lexical representations of the stimuli. However, we also found that the correlations between nonword repetition and reading scores decreased when the children's lexical diversity was statistically controlled. This finding suggests that when a child who uses a greater number of real words in conversation performs nonword repetition and reading tasks, he/she relies on phonological representations that are more robust and stable due to the existence of a larger number of lexical items in the child's mental lexicon.[2] If any similarity between the nonword and real words in English affected the children's performance, the effect was dependent on the children's ability to abstract and generalize phonological regularities and structure of the nonwords to real words in their lexicon. These linguistic skills rely on the use of phonology and the construction of phonological representations (both new representations of the nonwords and pre-existing representations of real words), not access to semantic representations. The nonword repetition task thus relies on phonological processing skills: the ability to perceive, encode, decompose, rehearse, and then reassemble for speech production a novel phonological pattern.

In contrast, the PIAT Reading Comprehension measure requires the child to match a written sentence with a picture that represents what is described by the sentence. In this sentence comprehension task, the child must be able to recognize and read written words, and access lexical representations (at least semantic if not phonological) and syntactic representations. This task also does not require the child to produce any spoken responses. Theoretically, it does not require the use of phonological representations at all. A child could recognize the words in the sentence directly using pre-existing visual representations of the letters and/or words, and process the sentence for meaning without constructing or accessing phonological representations of spoken words from his/her lexicon. On the other hand, there is now a great deal of evidence in the literature that suggests that young children convert the visual (graphemic) representations of print into phonological representations (phonemes or lexical phonological representations), as a part of the process of reading comprehension (Liberman, Shankweiler, & Liberman, 1989).

In the present study, we found that the children's nonword repetition performance and their reading comprehension scores were also correlated ($r$ = +.43, $p$ < .001). This indicates that nonword repetition, which involves speech perception, encoding and verbal rehearsal in phonological working memory, and speech production, is related to reading comprehension, which on the surface does not involve speech perception or speech production. The correlation between these two measures suggests that these deaf children were utilizing phonological processing skills in order to complete the reading comprehension task, and were relying on the same phonological processing skills for reading comprehension as they did for nonword repetition. Previous research on reading in young children has shown that nonword repetition relies heavily on preexisting phonological processing skills (see Brady, 1997). The present findings indicate that the strategies used by deaf children with cochlear implants to complete a reading comprehension task also rely on phonological knowledge and phonological processing skills. Like normal-hearing children, in order to carry out the reading comprehension task, the children in this study had to convert graphemes to phonemes (without excluding the possibility that they used more direct visual word recognition as well), and create a phonological representation of the sentence (or parts of it) and maintain that in phonological working memory to successfully complete this information processing task.

This use of phonological processing skills is a prerequisite to processing the sentence for meaning. The use of phonological processing in reading comprehension is predicated upon the previous existence of phonological representations. That is, in order to benefit from the use of phonological processing to complete a reading comprehension task, the child must be able to use abstract representations of the contrasting sounds of his/her ambient language; he/she must be able to map the

[2] In a preliminary investigation into the effect of the wordlikeness and phonotactic probability of the target nonwords on the children's nonword repetition performance, we found that overall performance on the 20 target nonwords (averaged across all 76 children) was not significantly correlated with the nonwords' wordlikeness (see Carlson et al., 1998) or phonotactic probability, when calculated either according to individual phoneme frequency by word position or according to biphone frequency (based on Vitevitch & Luce, submitted).

visual graphemes onto abstract phonological units. The more robust their phonological representations were, the more reliably these children could decompose and reassemble a spoken nonword. In addition, the present study reveals that the children who are better able to decompose and reassemble an auditorily-presented nonword (i.e., whose nonword repetition scores are higher), also tended to be the children who were better able to comprehend meaningful written sentences. We hypothesize that this correlation is based on the fact that these better performing children used phonological processing in completing the reading comprehension task. Thus, performance on both tasks relies at least to some extent on the construction of phonological representations and the use and development of phonological processing skills. The more accurate and robust a child's phonological representations are, the more useful they will be in phonological processing, and the better the child will perform on a wide range of information processing tasks that involve phonological processing such as nonword repetition, reading comprehension, and rhyme detection.

The findings obtained in the present study are consistent with other findings reported recently in the literature for children with phonological disorders. Using a nonword repetition task in which the nonwords were systematically varied in terms of biphone frequency, Munson et al. (in press) found that children with phonological disorders (PD) and typically-developing (TD) children repeated nonwords with low frequency sequences less accurately than nonwords with high frequency. Although the children with PD repeated the nonwords less accurately overall than the TD children, the children with PD were no more affected by frequency differences than typically developing children. Munson et al. also found that across both groups, children with larger vocabularies repeated the nonwords with greater accuracy than children with smaller vocabularies. Furthermore, they found that nonword repetition performance was not dependent on the speech perception or articulatory ability of the children in their study. Based on their findings, Munson et al. concluded that poorer overall performance by the children with PD in comparison to TD children in the nonword repetition task was not related to difficulties with speech perception, articulation, or even the ability to form abstract representations, but rather to having abstract representations that were not as robust or well specified as those of the TD children. According to this view, nonword repetition tasks index the robustness of the participant's abstract phonological representations, which is related to vocabulary size and the building of lexical representations. Similarly, Edwards et al. (2004), in a study of adults and TD children, found that nonword repetition performance (on the same task used in Munson et al.) was related to vocabulary size, and that performance by children who have larger vocabularies was less affected by biphone frequency differences than performance by children with smaller vocabularies.

The findings of Edwards et al. (2004) and Munson et al. (in press) provide support for a proposal of Studdert-Kennedy (2002: 11), who stated that "If segmentation of words into their phonological components is an emergent consequence of lexical growth, as several authors have proposed… we may hypothesize that a smaller than usual lexicon will result in defective ('fuzzy'/'weak') phonological representations, and so defective phoneme awareness." Insomuch as nonword repetition involves segmentation of a novel sound pattern into units that are encoded as abstract phonological segments, Munson et al.'s interpretation of their finding is consistent with Studdert-Kennedy's hypothesis that a smaller lexicon will lead to deficits in phonological representations. Studdert-Kennedy's position is that deficits in phonemic awareness are related to the fuzziness or weakness of phonological representations, which stems from poor speech-specific perception rather than a general auditory processing deficit (as proposed by Tallal and colleagues, e.g., Tallal, Miller, & Fitch, 1993). Munson et al. reject the idea that weak phonological representations stem from poor speech perception. Thus, Studdert-Kennedy and Munson et al. have disparate views on the *source* of 'weak' or non-robust phonological representations, but their views appear to be similar in that they see deficits in phonological awareness and phonological disorders (respectively) as directly related to poorly specified or incomplete phonological representations.

Taken together, studies such as those summarized in Studdert-Kennedy (2002; see also Mody, Studdert-Kennedy, & Brady, 1997), Edwards et al. (2004), Munson et al. (in press), and the present results lend converging support to the notion that robust phonological representations are behind better performance on tasks such as nonword repetition and phonemic awareness tasks. Further research is warranted regarding the *sources* of the development of weak abstract phonological representations, and the extent to which the sources vary across populations (e.g., NH children with phonological disorders versus deaf children with cochlear implants in OC or TC environments). Insights into the sources of the development of phonological representations should come from further behavioral research, and may also come from genetic studies. Several recent studies have demonstrated a connection between nonword repetition performance and three specific chromosomes (SLI Consortium, 2002; Watkins, Dronkers, & Vargha-Khadem, 2002, and Stein et al., 2004; see Kent, 2004). Such research could also provide insight into methods of identifying and treating children who fall into multiple clinical populations, e.g., deaf children who have phonological deficits independent of those caused by auditory deprivation.

The correlation between nonword repetition performance and reading comprehension suggests that the deaf children with cochlear implants in the present study were utilizing phonological processing skills in order to complete the reading comprehension task. A child's ability to use abstract phonological representations of the linguistically significant sound contrasts in his/her ambient language contributes to reading readiness and reading skills. This finding is not informative regarding the causality or directionality of this relationship. However, it provides important motivation for investigation into whether pediatric cochlear implant users' participation in tasks specifically aimed at strengthening phonological representations and processing skills would also lead to and contribute to increased reading and, ultimately, literacy skills. Support for such investigation is also warranted by the fact that numerous studies have found that normal-hearing children's reading skills can benefit significantly from training in phonemic awareness and grapheme-phoneme mapping skills (phonics) (see National Institute of Child Health and Human Development, 2000; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001).

Further understanding of the development of phonology and phonological processing skills may inform treatment of phonological disorders and habilitation of children with cochlear implants, and conversely, studies of the effects of treatment and habilitation may also provide insight into the sources and characteristics of robust categories in a phonological system. In a recent summary of treatment studies of children with phonological disorders, Gierut (2004) reported that phonological learning is greatest when linguistically complex or difficult sounds or sound pairs are used as the targets in treatment. Perhaps the development of more robust abstract phonological representations in deaf children with CIs could be aided by a focus on linguistically complex sounds in treatment for phonological disorders.

In summary, our results are consistent with the notion that phonological processing and knowledge are important for both nonword repetition and reading performance in deaf children with cochlear implants. The findings from this study suggest that the children's use of abstract representations (robust or not) of phonological structure is reflected in their performance on these tasks. The correlations obtained between spoken nonword repetition and phonological awareness and reading comprehension tasks suggest that early intervention, including explicitly training children in ways that may help them to develop robust, stable phonological representations, may be crucial in the habilitation of deaf children if they are ultimately to achieve maximal literacy levels. Methods developed in clinical phonology to help children with phonological disorders improve their speech intelligibility may prove to be useful in helping the clinical population of deaf children with cochlear implants to development phonological knowledge and skills.

## References

Adams, M.J. (1990). *Beginning to Read*. Cambridge, MA: MIT Press.

Baddeley, A., & Gathercole, S. (1992). Learning to read: The role of the phonological loop. In J. Alegria, D. Holender, J. de Morais, & M. Radeau (Eds.), *Analytic Approaches to Human Cognition* (pp. 153-167). New York: Elsevier Science Publishers.

Brady, S.A. (1997). Ability to encode phonological representations: An underlying difficulty of poor readers. In B. Blachman (Ed.), *Foundations of Reading Acquisition and Dyslexia* (pp. 21-47). Mahwah, NJ: Lawrence Erlbaum Associates.

Bradley, L., & Bryant, P.E. (1983). Categorizing sounds and learning to read - a causal connection. *Nature, 301*, 419-421.

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.

Carter, A.K., Dillon, C.M., & Pisoni, D.B. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: Suprasegmental analyses. *Clinical Linguistics & Phonetics, 16*, 619-638.

Cassar, M., & Treiman, R. (2004). Developmental variations in spelling: Comparing typical and poor spellers. In C.A. Stone, E.R. Silliman, B.J. Ehren, & K. Apel (Eds.), *Handbook of Language and Literacy: Development and Disorders* (pp. 627-643). New York: The Guilford Press.

Conrad, R. (1979). Deafness and reading. In R. Conrad (Ed.), *The Deaf Schoolchild* (pp. 140-175). London: Harper and Row.

Dalgarno, G. (1680; 1971). Didascalocophus or The deaf and dumb man's tutor. In R.C. Alston (Ed.), *English Linguistics 1500-1800*. Menston, England: The Scholar Press Limited.

Dillon, C.M., Burkholder, R.A., Cleary, M., & Pisoni, D.B. (in press). Perceptual ratings of nonword repetition responses by children who are deaf after cochlear implantation. *Journal of Speech, Language, and Hearing Research*.

Dillon, C.M., Cleary, M., Pisoni, D.B., & Carter, A.K. (2004). Imitation of nonwords by hearing-impaired children with cochlear implants: Segmental analyses. *Clinical Linguistics & Phonetics, 18*, 39-55.

Dillon, C.M., Pisoni, D.B., Cleary, M., & Carter, A.K. (2004). Nonword imitation by children with cochlear implants: Consonant analyses. *Archives of Otolaryngology - Head and Neck Surgery, 130*, 587-591.

Dunn, L.M., & Markwardt, F.C. (1989). *Peabody Individual Achievement Test-Revised*. Circle Pines, MN: American Guidance Service, Inc.

Edwards, J., Beckman, M.E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*, 421-436.

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. (2003). Predictors of reading skill development in children with early cochlear implantation. *Ear and Hearing, 24*, 59S-68S.

Geers, A., & Brenner, C. (2003). Background and educational characteristics of prelingually deaf children implanted by five years of age. *Ear and Hearing, 24*, 2S-14S.

Geers, A., Nicholas, J.G., & Sedey, A.L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24*, 46S-58S.

Gierut, J.A. (2004). Enhancement of learning for children with phonological disorders. Paper presented at From Sound to Sense, June 11-13, 2004, MIT, Cambridge, MA.

Hanson, V.L. (1991). Phonological processing without sound. In S.A. Brady & D.P. Shankweiler (Eds.), *Phonological Processes in Literacy* (pp. 153-161). Hillsdale NJ: Lawrence Erlbaum Associates.

Kent, R.D. (2004). Development, pathology and remediation of speech. Paper presented at From Sound to Sense, June 11-13, 2004, MIT, Cambridge, MA.

Liberman, I.Y., Shankweiler, D., & Liberman, A.M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I.Y. Liberman (Eds.), *Phonology and Reading Disability: Solving the Reading Puzzle* (pp. 1-33). Ann Arbor MI: University of Michigan Press.

Markwardt, F.C. (1998). *Peabody Individual Achievement Test-Revised*. Circle Pines, MN: American Guidance Service, Inc.

Marschark, M. (2003). Interactions of language and cognition in deaf learners: From research to practice. *International Journal of Audiology, 42*, S41-S48.Mody, M., Studdert-Kennedy, M., & Brady, S.A. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 64*, 199-231.

Munson, B., Edwards, J., & Beckman, M.E. (in press). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Paul, P.V. (2003). Processes and components of reading. In M. Marschark & P.E. Spencer (Eds.), *Deaf Studies, Language, and Education* (pp. 97-109). New York: Oxford University Press.

Pisoni, D.B. (this volume). Speech perception skills of deaf children with cochlear implants. In *Research on Spoken Language Processing Progress Report No. 26.* Bloomington IN: Speech Research Laboratory, Indiana University.

Rayner, K., Foorman, B.R., Perfetti, C.A., Pesetsky, D., & Seidenberg, M.S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31-54.

Rayner, K., & Pollatsek, A. (1995). *The Psychology of Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

SLI Consortium (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *American Journal of Human Genetics, 70*, 384-398.

Spencer, L., Tomblin, B., & Gantz, B. (1997). Reading skills in children with multi-channel cochlear implant experience. *Volta Review, 99*, 193-202.

Stein, C.M., Schick, J.H., Taylor, H.G., Shriberg, L.D., Millard, C., Kundtz-Kluge, A., Russo, K., Minich, N., Hansen, A., Freebairn, I.A., Elston, R.C., Lewis, B.A., & Iyengar, S.K. (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *American Journal of Genetics, 74*, 283-297.

Studdert-Kennedy, M. (2002). Deficits in phoneme awareness do not arise from failures in rapid auditory processing. *Reading and Writing: An Interdisciplinary Journal, 15*, 5-14.

Tallal, P., Miller, S., & Fitch, R.H. (1993). Neurobiological basis of speech: A case for the preeminenc of temporal processing. In P. Tallal, A.M. Galaburda, R.R. Llinas, & C von Euler (Eds.), *Temporal Information Processing in the Nervous System. Annals of the New York Acaedmy of Science*, Vol. 82 (pp. 27-47). New York: New York Academy of Sciences. Paper reprinted in *Irish Journal of Psychology* (1995), 16: 194-219.

Troia, G.A. (2004). Phonological processing and its influence on literacy learning. In C.A. Stone, E.R. Silliman, B.J. Ehren, & K. Apel (Eds.), *Handbook of Language and Literacy: Development and Disorders* (pp. 271-301). New York: The Guilford Press.

Vitevitch, M.S., & Luce, P.A. (submitted). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*.

Wagner, R.K., & Torgeson, J.K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Watkins, K.E., Dronkers, N.F., & Vargha-Khadem, F. (2002). Behavioral analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain, 125*, 452-464.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–III*. San Antonio, TX: The Psychological

Corporation.

Woodcock, R.W. (1987) *Woodcock Reading Mastery Tests-Revised*. Allen, TX: DLM Teaching Resources.

Woodcock, R.W. (1998). *Woodcock Reading Mastery Tests-Revised*. Allen, TX: DLM Teaching Resources.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Development of Visual Attention Skills in Prelingually Deaf Children Who Use Cochlear Implants[1]

**David L. Horn,[2] Rebecca A.O. Davis[2] 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 T32 DC00012, NIH-NIDCD Research Grants R01 DC00064 and DC00423 to Indiana University. We would like to thank Karen Iler Kirk, Derek Houston, and Rose Burkholder for insightful comments on earlier versions of this paper. This is a draft of a paper to be submitted for publication.

[2] Devault Otologic Research Laboratory, Department of Otolaryngology-Head & Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana 46220

# Development of Visual Attention Skills in Prelingually Deaf Children Who Use Cochlear Implants

**Abstract.** The goal of this paper is to determine the effect of length of cochlear implant (CI) use and other factors on the development of sustained visual attention in prelingually deaf children, and to investigate the relations between sustained visual attention and audiological outcomes with an implant. It includes retrospective analysis of data collected before cochlear implantation and over several years after implantation. Two groups of prelingually deaf children, one >6 years old (n=41) and one <6 years old (n=47) at testing, were given an age-appropriate Continuous Performance Task (CPT). In both groups, children monitored visually-presented numbers for several minutes and responded whenever a target number appeared. Hit rate, false alarm rate, and signal detection parameters were dependent measures of sustained visual attention. A number of independent variables were tested for effects on CPT performance. Several audiological outcome measures were found to be correlated with CPT scores. Mean CPT performance was low compared to published norms for normal-hearing children. CPT scores improved as a function of length of CI use over at least two years of use. In the younger group, a higher number of active electrodes predicted better CPT performance. In the older group, CPT performance after two years of CI use was correlated with vocabulary knowledge. These findings suggest that cochlear implantation in prelingually deaf children leads to improved sustained visual processing of numbers over two or more years of CI use.

## Introduction

Prelingually deaf children with cochlear implants (CIs) are a unique clinical population to study because they have experienced a period of auditory deprivation prior to the development of significant speech and language skills. The success of many deaf children in acquiring oral language skills from the electrical stimulation provided by a CI is impressive. Although clinical outcomes have been found to be quite variable in this population (Blamey, Sarant, Paatsch, Barry, Bow, Wales, Wright, Psarros, Rattigan, & Tooher, 2001; Pisoni, Cleary, Geers, & Tobey, 2000; Sarant, Blamey, Dowell, Clark, & Gibson, 2001), most children's scores on a range of speech perception, language, and speech intelligibility tests improve with CI use (Miyamoto, Svirsky, Kirk, Robbins, Todd, & Riley, 1997; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000; Tyler, Fryauf-Bertschy, Kelsay, Gantz, Woodworth, & Parkinson, 1997).

Physical device limits, such as reduced spectral resolution and compressed dynamic range, have been shown to play a role in limiting the speech perception abilities developed by prelingually deaf children who use CIs. Nevertheless, some children are able to use these degraded signals to develop age appropriate speech perception and language skills with a CI. In contrast, other children obtain only awareness of sound with a CI and their speech and language skills remain significantly delayed compared to their normal-hearing peers. The sources of this outcome variability have yet to be fully revealed. Studies that have looked into pre-implant predictors of CI performance have reported effects of early implantation, communication mode, device type, and dynamic range (Geers, Brenner, & Davidson, 2003; Geers, Nicholas, & Sedey, 2003; Kirk, 2000; Tobey, Geers, Brenner, Altuna, & Gabbert, 2003). However, a large portion of outcome variance remains unexplained by these demographic and medical factors (Pisoni et al., 2000; Sarant et al., 2001).

A major paradigm shift in the study of pediatric CI outcomes has been emerging from recent investigations of the neurocognitive abilities of prelingually deaf children (Pisoni et al., 2000). The premise of this approach is that individual differences in speech and language outcomes in prelingually deaf children may reflect underlying differences in central auditory processing and related neurocognitive functions such as working memory, attention, and executive functions. These central processes are not specific to audition or language, yet they may play important roles in perceiving speech, acquiring language, and developing the sensory-motor abilities crucial to producing highly intelligible speech.

Two major research objectives can be identified in the study of neurocognitive processes of prelingually deaf children with CIs. The first is to describe how individual cognitive processes may be altered in children who have experienced a period of auditory deprivation. The second is to assess relationships between individual cognitive functions and speech/language outcomes. Both objectives may reveal important predictors of benefit and new insights into sources of individual differences in clinical outcomes with a CI. Furthermore, research on central auditory and cognitive factors may provide the theoretical basis for therapeutic cognitive and behavioral interventions for deaf children who are struggling with their implants.

Sustained visual attention is one cognitive process that has been studied in prelingually deaf children with CIs. Sustained attention is a form of attention that is responsible for the “continuous allocation of processing resources for the detection of rare events,” (Parasuraman, 1998). This ability is crucial for the efficient exploration of the environment and the support of other perceptual and cognitive processes. Experimentally, sustained visual attention capacity can be measured reliably by continuous performance tasks (CPTs) in which subjects are asked to respond to low frequency target stimuli. Experimental manipulations of these tasks in normal hearing subjects have demonstrated the fragile nature of sustained visual attention. Performance has been shown to decline with increased task length, lower signal salience, faster stimulus rate, and higher memory load (Parasuraman, 1998).

The most widely used task for measuring sustained visual attention is the CPT. Originally developed by Rosvold, Mirskey, Sarason, Bransome, and Beck (1956), several newer CPTs have since been designed and normed for clinical use (Conners, 2000; Gordon, McClure, & Aylward, 1996; Greenberg, 1991). These CPTs possess moderate clinical validity for detecting attentional disorders such as ADHD in children at risk for these conditions (Barkley, 1991; Forbes, 1998). In a CPT, children are asked to monitor a stream of visually presented stimuli and respond by pressing a key each time they detect a target stimulus. Detection rate depends on two independent factors: perceptual sensitivity (ability to distinguish targets from non-targets), and expectations of target frequency (Parasuraman, 1998). Because CPTs are visual information processing tasks that have no auditory demands, they have been used to assess the visual attention skills of both deaf children and adults.

In the first study of sustained visual attention skills of prelingually deaf children with CIs, Quittner, Smith, Osberger, Mitchell, and Katz (1994) administered a CPT to three groups of children: prelingually deaf children with CIs, prelingually deaf children who used hearing aids, and normal-hearing children. Children were assigned to one of two age groups: younger (6- to 8-year olds) and older (9- to 13-year olds). Overall, Quittner et al. found that children in the older group performed better than children in the younger group demonstrating an effect of chronological age on sustained visual attention. In addition, Quittner et al. reported a main effect of hearing. Across age groups, normal-hearing children performed better than both children with CIs and children with hearing aids (HAs). However, the authors also found an interaction between age and group. In the early age range, both groups of prelingually deaf children had lower mean CPT scores than the normal-hearing children. In the older age range, however, the children with CIs performed as well as normal-hearing children while the mean performance of children with HAs was below that of both the children with CIs and normal-hearing children.

In a second experiment, Quittner et al. reported findings from a longitudinal study of CPT performance in prelingually deaf children with CIs and a group of deaf children who used HAs. Each child was tested twice, about eight months apart. The children with CIs were tested only after implantation (about 10 and 18 months post-implantation). The results from this experiment were consistent with the results of the first experiment: children with CIs showed more improvement in mean CPT performance than children with HAs over the 8 month interval. From the results of these two experiments, Quittner et al. concluded that children who experienced a period of profound prelingual deafness prior to cochlear implantation are delayed in their development of visual attention skills. Moreover, they interpreted the findings from their two experiments as evidence that access to sound with a CI produced more typical development of visual attention skills than a HA in prelingually deaf children.

Although no attempt was made to control for length of CI use, Quittner et al. suggested that the effects of a CI on visual attention were relatively rapid as indicated by their findings in Experiment 2. They also suggested that visual attention skills were unlikely to be related to speech and language outcomes in the children with CIs' because gains in these skills typically emerge more slowly, after several years of CI use (Fryauf-Bertschy, Tyler, Kelsay, & Gantz, 1992; Miyamoto, Osberger, Robbins, Myres, & Kessler, 1993). However, Quittner et al. did not report any speech and language outcome data, nor did they assess the relations between CPT performance and speech/language outcomes in either of their two experiments.

In a follow-up study with a larger sample of children, Smith, Quittner, Osberger, and Miyamoto (1998) attempted to replicate the earlier findings reported by Quittner et al. Once again, CPT scores were obtained from prelingually deaf children with CIs, prelingually deaf children with HAs, and normal-hearing children. Each of these groups were subdivided into 6 age groups: 6-, 7-, 8-, 9-, 11-, and 13-years-old. Like Quittner et al., Smith et al. found an interaction between chronological age and group. At 6, 7, and 8 years of age, both groups of deaf children had lower mean CPT scores than normal-hearing children. In contrast, at 9, 11, and 13 years of age, deaf children with CIs performed equally to normal-hearing children on the CPT. Both of these groups had higher mean CPT performance than the children who used HAs.

To explain the interaction between cochlear implantation and chronological age, Smith et al. hypothesized that access to sound (either with a normal cochlea or a CI) before or during a critical developmental period was important for children to acquire typical sustained visual attention skills. The implantation of deaf children with a CI before or during the critical period lead to faster maturation of their visual attention skills, compared to children with HAs, until they reached age appropriate levels compared to children with normal hearing. The authors reasoned that, prior to the onset of the critical period for visual attention development, deaf children with CIs would show no developmental advantage over deaf children with HAs in acquiring these skills. Smith et al. attributed the lack of an effect of a CI in the 6-8 year old children as evidence that this critical period began at around age 7-8 years old.

However, in a second experiment, Smith et al. tested a younger group of prelingually deaf children (4-7 years old) who used either CIs or HAs on an easier version of the CPT. In the original CPT, children were asked to detect a two-number sequence. In the easier CPT, children were only asked to detect single target stimuli. This new version had a lower working memory load than the original CPT, and is recommended and normed for children aged 3-5 years (Gordon et al., 1996). Children with CIs displayed higher mean CPT scores than children with HAs, indicating that CI use affected the development of sustained visual attention even in younger children than those who were tested previously. Therefore, the lack of an observed effect of a CI in 6-8 year-old children on the original CPT may have been due to the difficulty of the task as Smith et al.'s second experiment demonstrated that auditory experience had an effect on visual attention at even younger ages than previously tested. Although a critical period might exist for the development of the visual attention, the age of onset remained undefined by the results reported by Smith et al.

Any interpretation of the interaction between chronological age and cochlear implantation reported by Quittner et al. and Smith et al. is limited by the fact that neither study controlled for length of CI use, which varied between 6 months and 6 years. This confounding variable may have been responsible for the observed interaction because children in the younger CI group may have had fewer years of experience with their CIs than children in the older group. Although Smith et al. reported that length of CI use and CPT performance were not correlated in their study, these results do not rule out an effect of length of CI use over the first few years of CI use. Indeed, Quittner et al. had concluded that the effect of a CI on visual attention skills was relatively rapid. However, because chronological age and length of CI use were not controlled, the source of the interaction between chronological age and cochlear implantation cannot be determined from either Quittner et al.'s or Smith et al.'s data.

Although Smith et al. and Quittner et al. attributed the CI effect to deaf children having access to sound, they also did not systematically examine the role of demographic, medical, or audiological factors on CPT performance. They did assess the effect of mode of communication training and found no differences in CPT performance between children in oral communication (OC) programs (deaf children who learn to communicate primarily through listening, lip-reading, and oral skills) and total communication (TC) programs (deaf children who incorporate a manual correlate of their spoken language along with oral skills). However, they did not assess other variables such as IQ, etiology of deafness, CI type, gender, or other factors which may have been at least partially responsible for the reported CI effect.

Why should access to sound provided by a CI, and not a HA, provide benefits for visual attention development? The aided auditory thresholds of the two groups of deaf children in Quittner et al. were quite similar, indicating that both devices were equal in terms of their effect on sound detection. Perhaps, something about the nature of the auditory information provided by a CI lead to the improved sustained visual attention skills in this population. HAs function by amplifying ambient noise to stimulate the functional hair cells remaining in a cochlea damaged by sensori-neuronal hearing loss (SNHL). Processing of this amplified signal by the profoundly deaf cochlea leads to a significant amount of spectral distortion, limiting the effectiveness of HAs for improving speech perception (Niparko, Kirk, Mellon, McConkey-Robbins, Tucci, & Wilson, 2000). In contrast, CIs extract specific speech cues encoded in the time-varying spectrum of speech and use this information to directly stimulate the auditory nerve (Niparko et al., 2000). Despite carrying impoverished acoustic-phonetic information about speech, a CI may be better suited than an HA for speech perception in profoundly deaf individuals who have great difficulty understanding speech (Niparko et al., 2000).

Possibly, the benefits of a CI for speech and language development in profoundly deaf individuals are related to the neural mechanisms responsible for the improvements in sustained visual attention skills with a CI. For instance, improved auditory perceptual acuity, rather than simple access to sound, might lead to improvements in sustained visual attention skills. Perhaps gains in language skills and other related cognitive processes enable deaf children to perform better on the CPT. One way to begin to answer these hypotheses would be to calculate correlations between speech and language test scores and CPT scores. These relations were not examined by either Quittner et al. or Smith et al. or by any other study of sustained visual attention in deaf children.

In a recent report, Tharpe, Ashmead, and Rothpletz (2002) have raised several criticisms about the conclusions of Quittner et al. and Smith et al. and suggested that their results may have reflected differences in IQ between the test populations. To deal with this concern, Tharpe et al. conducted a study of CPT performance in prelingually deaf children with CIs, prelingually deaf children with HAs, and normal-hearing children in which non-verbal IQ of each subject was measured and used as a covariate in all analyses. Tharpe et al. failed to find any differences in mean CPT performance between children with CIs and deaf controls. They were unable to replicate the earlier differences in CPT scores between normal-hearing children and either group of deaf children reported by Quittner et al. and Smith et al. Tharpe et al. interpreted these results as evidence that the benefit of a CI on sustained visual attention reported by Quittner et al. and Smith et al. was due to differences in IQ rather than differences in visual attention.

While Tharpe et al.'s conclusions might be correct, their findings must be interpreted cautiously for several reasons. First, the sample size used in their study was extremely small (only 8 subjects in each subject group) compared to the larger samples used by Quittner et al. and Smith et al. Second, the study design used by Tharpe et al. was fundamentally different from the designs used by Quittner et al. and Smith et al. in which chronological age was an independent variable. Although Tharpe et al. have raised several important issues and suggested further areas of research on visual attention, their study was not a comprehensive attempt to replicate either of the earlier studies.

Finally, the CPT scores reported by Tharpe et al. were close to ceiling performance. They used a different CPT than the one used by Quittner et al. and Smith et al., and the mean d's were 3.5 and above. d' is a parametric measure of perceptual sensitivity computed using the principles of signal detection theory and d's above 3 suggest ceiling performance. It is possible that Tharpe et al. failed to detect any differences in CPT performance between deaf children with CIs and deaf children with HAs because their CPT was simply too easy for these children. Indeed, Quittner et al. and Smith et al. reported a much wider range of means for d' in their studies (1.4 – 4.5 in Quittner et al.).

To summarize, three studies of sustained visual attention development in prelingually deaf children with CIs have been reported in the literature to date. In the recent study by Tharpe et al., no differences between deaf children with CIs, deaf children with HAs, or normal-hearing children were observed when non-verbal IQ was controlled. However, several methodological factors make Tharpe et al.'s results difficult to compare with the results from the previous studies by Quittner et al. and Smith et al. which both found a beneficial effect of a CI on the development of sustained visual attention in prelingually deaf children. More work is needed in order to understand the effect of length of CI use and chronological age as these variables were confounded in the previous studies. Furthermore, we do not know what other demographic or medical variables might influence sustained visual attention skills of deaf children. Finally, the relations between oral speech and language skills and sustained visual attention skills have not yet been systematically assessed. The present study was designed to explore these unanswered questions.

Our first aim was to investigate the effect of CI use on sustained visual attention skills using a design which controlled for the independent effects of length of CI use and chronological age. We hypothesized that visual attention skills would be affected by both length of CI use and chronological age. We were also interested in determining the time-course of the effects of CI use: Do visual attention skills increase rapidly during the first year of use as suggested by Quittner et al. and Smith et al., or do they change more gradually over time after implantation? Finally, we were interested in determining whether there was an interaction between length of CI use and chronological age as predicted by Smith et al.'s critical period hypothesis: the effect of a CI use would be seen in children during a critical age range, but not before or after this age range.

Our second aim of this study was to determine if several medical, audiological, or demographic characteristics of prelingually deaf children would affect the development of sustained visual attention skills after cochlear implantation. Although some of these factors are known to play a role in the development of auditory/oral speech and language skills with a CI, the effects of these variables on other cognitive skills in prelingually deaf children such as sustained visual attention are largely unknown.

Finally, our third aim was to assess the relations between sustained visual attention skills and several traditional audiological outcome measures of speech perception, language acquisition, and speech intelligibility. As mentioned previously, no attempt was made in any of the previous studies to examine the relations between sustained visual attention skills and clinical outcome measures of speech and language. If gains in sustained visual attention in deaf children are correlated with gains in speech and language skills, this would have important implications for predicting individual clinical outcomes with a CI, and for understanding the underlying cognitive mechanisms of audiological benefit with a CI.

We report two experiments with similar longitudinal designs, measures, and independent variables. The major difference between the two experiments was the ages of the children. The first experiment included children older than 6 years of age and the second experiment included a younger group of children. Two different age-appropriate CPTs were used to assess sustained visual attention skills in each group of children.

## Experiment I

### Method

**Participants.** We conducted a retrospective analysis of longitudinal clinical data gathered at the Indiana University School of Medicine Cochlear Implant Program. The subjects were part of a larger comprehensive, prospective clinical study of speech and language outcomes in deaf children with CIs. Table 1 provides a summary of the medical, audiological, and demographic characteristics of our sample. The group included 41 prelingually deaf children (profoundly deaf by 2.5 years old) who received CIs by 9 years of age. All children were implanted with Nucleus 22 devices. Mean age of implantation was 6.2 years old. Fifteen children used OC and 26 used TC at the time of testing. Over 68% had a congenital hearing loss and the most common acquired etiology was meningitis. A subset of children (n=19) had pre-implant non-verbal IQ scores (WISC or WIPSI) in their charts and the mean standardized score was 101.3.

Children were tested once every 6-12 months from before implantation to three years post-implantation. Interval data were collapsed into one of four intervals: pre-implant, one year post-implant, two years post-implant, and three years post-implant. Not all children were tested at each interval, as is common in clinical populations, creating missing data cells. Missing data occurred for several reasons. Some children moved away from the Indianapolis area after implantation and were unable to continue the study. Also, because our clinical participants are given a large number of tests during each visit, they are often too tired or not cooperative enough to complete all of the tests.

| | | |
|---|---|---|
| **Medical** | **Etiology** | **1 genetic, 1 CMV, 11 meningitis, 28 unknown** |
| | **Ear of Implantation (n=37)** | **17 right, 20 left** |
| **Audiological** | **Mean Pure Tone Average (n=26) (dB HL)** | **111.8 (6.478)*** |
| | **Mean Number of active electrodes (n=34)** | **20.1 (3.59)*** |
| | **Mean Age of implantation (yrs.)** | **6.20 (1.60)*** |
| **Demographic** | **Gender** | **19 female, 22 male** |
| | **Communication mode** | **15 OC, 26 TC** |
| | **Mean non-verbal IQ (n=19)** | **101.3 (16.53)*** |

**Table 1.** Medical, audiological, and demographic characteristics of participants in Experiment I (standard deviations in parentheses)

**Procedures.** The CPT used in the present study was the “School-age CPT” (Gordon et al., 1996), the same CPT used in the studies by Quittner et al. and Smith et al. This test format is recommended for use in normal-hearing children from 6-16 years old. The experimental apparatus consists of a free-standing button box with one blue key below an LCD display. While this apparatus is capable of running a number of different tests, the school-age CPT is a 9 minute CPT in which target and non-target numbers appear at random in the center of a viewing screen at 1 second intervals. Children were instructed to press a key every time a “9” appeared following a “1.” They were instructed to refrain from pressing the key when non target numbers appeared or when “9” appeared following any number other than “1.” No feedback was given during the task other than general encouragement. Instructions were given in the child’s preferred mode of communication. Children were tested with an experimenter, but not their caregiver, present in the room.

Each time a correct response (key press) was made when the target number appeared, the computer internally scored a “hit.” Each time a key press was made when a non-target number appeared, the computer internally scored a “false alarm.” The Gordon CPT apparatus automatically computes the total number of hits, misses, and false alarms from which the hit rates and false alarm rates were computed manually.

We used these raw scores to compute two additional measures based on methods used in signal detection theory (SDT; Green & Swets, 1966). Perceptual sensitivity, or d', was computed as the difference between the z scores for hit rate and false alarm rates. This measure is used to assess the subject’s ability to discriminate visual targets from non-targets. A second measure, β, was computed as the ordinate for hit rate divided by the ordinate for false alarm rate. β reflects the response bias of the subject. More conservative (i.e. less impulsive) responders will have higher β scores than less conservative, or more impulsive responders. Although there is some controversy over the use of parametric measures in cases where the assumptions of SDT are not necessarily true (Pastore, Crawley, Berens, & Skelly, 2003), we included these measures in our analyses along with hit rate and false alarm rate because earlier studies of CPT performance in deaf children and adults have used d' and β as measures of sustained visual attention. Thus, we report results in terms of d' and β for the sake of comparison to previous studies, however, we also report results in terms of hit rate and false alarm rate where appropriate.

Because we did not collect data from a control group of age-matched normal-hearing children, we used the published percentile norms as a benchmark for comparative purposes. The Gordon CPT manual publishes norms for children from 6-16 years old (Gordon et al., 1996). We did not perform any statistical analyses using these standardized data, and used them only for descriptive purposes.

At our center, a large battery of tests are used to assess oral speech and language skills of children with CIs. From this battery, we selected five tests as representative outcome scores for our sample of children. Open-set speech perception was measured using the Phonetically Balanced Kindergarten (PBK) test (Haskins, 1949). This test is administered using live voice presentation and is scored by word and phoneme. Children hear a spoken word and are asked to repeat the word aloud to the examiner. The words used in this test are phonetically balanced, English monosyllables.

Sentence comprehension was measured with the Common Phrases test (Osberger, Miyamoto, Zimmerman-Philips, Kemink, Stroer, Firszt, & Novak, 1991) in which percent correct scores reflect the child's ability to repeat a sentence or follow a command. This test is typically administered in auditory (CPA), visual (CPV), and auditory plus visual (CPAV) modalities. We used scores from all these presentation formats.

Vocabulary knowledge was assessed with the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997). At our center, this test is administered in the child's preferred mode of communication. Each vocabulary item is presented either orally or manually. The child then chooses from four pictures, one of which correctly corresponds to the meaning of the word.

The Reynell Developmental Language Scales 3rd edition (RDLs; Reynell & Huntley, 1985) was administered to assess both receptive (RR) and expressive (RE) language skills. The receptive scales measure 10 different skills including word recognition, sentence comprehension, and verbal comprehension of ideational content. The expressive language scales assess skills such as spontaneous expression of speech and picture description. Like the PPVT, the RDL was administered in the child's preferred mode of communication.

Speech intelligibility was assessed using the Beginner's Intelligibility Test (Osberger, Robbins, Todd, & Riley, 1994). Audio recordings were made of children repeating a list of 10 sentences given to them by a clinician. These recordings were then presented to three naïve adult listeners who were asked to transcribe what they thought the children were saying. Intelligibility scores are based on the number of words correctly transcribed by the adult listeners. Each of these tests, including the CPT, was administered in an Otolaryngology clinical setting by licensed health professionals who had received special training in working with children with CIs.

## Results

As noted earlier, because our participants were drawn from a larger clinical population enrolled in a long-term longitudinal study of CI outcomes, we could not always obtain CPT scores for each child at each test interval. A traditional repeated measures ANOVA would, therefore, eliminate data from any child who was not tested at each interval. However, such an analysis can often lead to skewed results as well as underestimates of variability (Schafer & Graham, 2002). We therefore employed several statistical techniques using the SAS Mixed Procedure (Wolfinger & Chang, 1995) to analyze our data. The SAS Mixed procedure utilizes a maximum-likelihood estimation method to create a model without eliminating any participants (Schafer & Graham, 2002). In this manner, systematic selection biases can be avoided by using data from all children, even those who were not tested at each interval, in the test design. Table 2 lists the number of participants who were tested on the CPT at each of the test intervals.

| | Preimplant | Year 1 | Year 2 | Year 3 |
|---|---|---|---|---|
| **Experiment 1 (schoolage CPT)** | 6 | 17 | 30 | 15 |
| **Experiment 2 (preschool CPT)** | 19 | 30 | 23 | 4 |

**Table 2.** Number of participants tested on the CPT tasks at each testing interval

**CPT performance of prelingually deaf children with CIs compared to normative sample.** As described in the previous section, we calculated normative percentile scores from the CPT raw scores. Table 3 shows the percentile means for hit rates and false alarm rates at each interval along with the number of participants at each interval. Examination of these scores reveals that the prelingually deaf children performed poorly on the school-age CPT compared to the normative sample in term of both hits and false alarms. The highest mean normative scores were found in children who had used their implants for two years. However, these means were only at the 36th and 16th percentiles for hit rate and false alarm rate respectively.

| Gordon Test | Test Interval | Percentile Mean (SD) |
|---|---|---|
| **Hit Rate** | 0 (n=3) | 23.0 (27.0) |
| | 1 (n=16) | 27.5 (34.8) |
| | 2 (n=26) | 36.6 (32.9) |
| | 3 (n=15) | 17.9 (25.8) |
| **False Alarms** | 0 (n=2) | 3.0 (0.0) |
| | 1 (n=16) | 12.5 (11.2) |
| | 2 (n=25) | 16.3 (18.9) |
| | 3 (n=14) | 14.9 (22.0) |

**Table 3.** Percentile CPT Scores at each interval of testing from Experiment 1 (school-age CPT)

**Effects of chronological age.** The results obtained from the SAS mixed model revealed significant effects of chronological age on hit rate ($F(1, 34.7) = 15.93$, $p < 0.01$), and d' ($F(1, 39.5) = 13.15$, $p < 0.01$). No significant effects were found for β or false alarm rate. Thus, our sample of prelingually deaf children with CIs demonstrated an increase in hit rate and increase in perceptual sensitivity as a function of age. This effect is illustrated in Figures 1a and 1b as a function of chronological age. In all of our subsequent analyses with the mixed model, we controlled for the effect of chronological age.

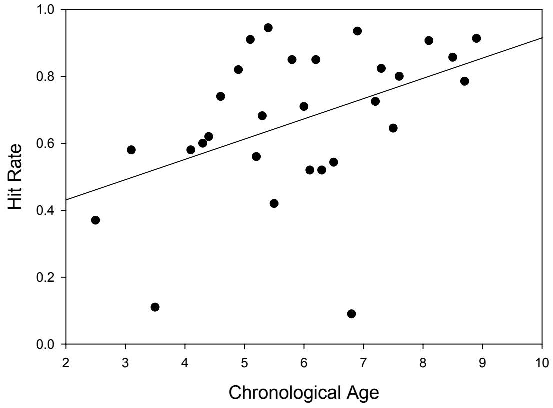


**Figure 1a.** Hit rate on the school-age CPT as a function of chronological age. Some data points overlap; all individual data points are plotted with a line of best

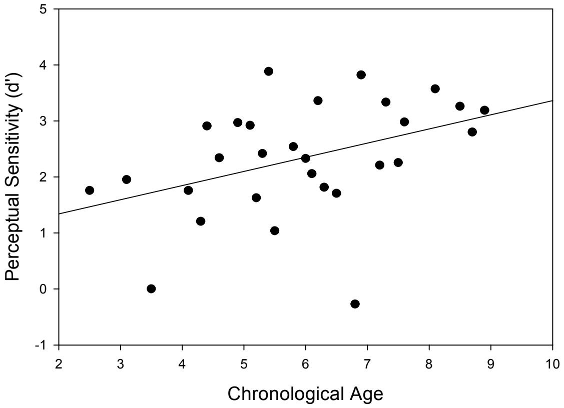


**Figure 1b.** Perceptual sensitivity (d') on the school-age CPT as a function of chronological age. Some data points overlap; all individual data points are plotted with a line of best fit.

**Effects of length of CI use.** The SAS mixed model produced estimated means for hit rate, false alarm rate, d' and β as a function of length of CI use. The estimated mean hit rate increased from 0.60 (*SE*=0.11) before cochlear implantation to 0.76 (*SE*=0.06) after three years of CI use, a significant effect (*F*(3,27.9) = 3.08, *p* <0.05). The estimated mean for d' also increased with length of CI use from 2.08 (*SE*=0.57) to 2.86 (*SE*=0.25) pre implant and after 3 years of use respectively (*F*(3,23.8) = 4.46, *p* <0.05). In contrast, length of CI use did not significantly affect β or false alarm rate. Figures 2a and 2b illustrate the significant effects of length of CI use.

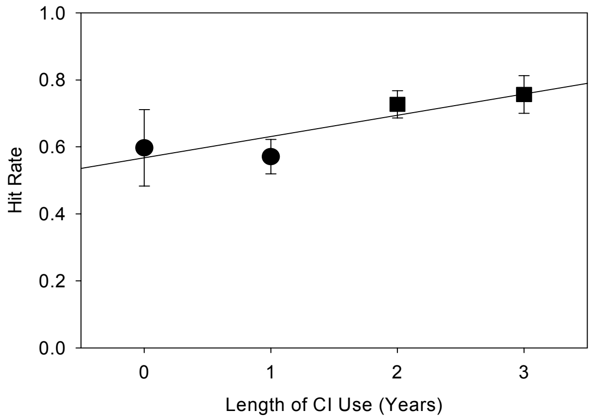


**Figure 2a.** Hit rate on the school-age CPT as a function of length of CI use in years. Different shaped data points are used to represent significantly different means at different lengths of CI use.

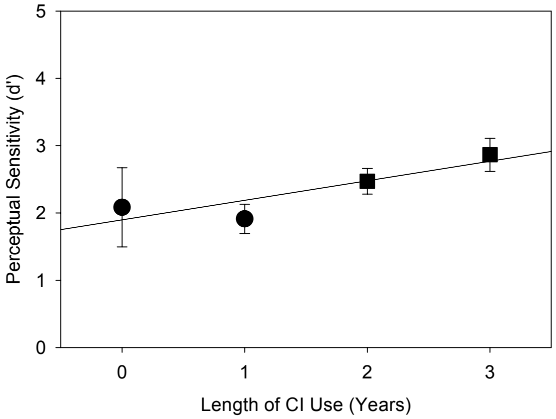


**Figure 2b.** Perceptual sensitivity (d') on the school-age CPT as a function of length of CI use. Different shaped data points are used to represent significantly different performance means at different lengths of CI use.

Tukey's post hoc comparisons were carried out to assess differences between each post-implant interval. In Figure 2a and 2b, means which differed significantly from each other are represented by data points of different shapes. Figure 2a shows that hit rate increased significantly, as a function of length of CI use, between post-implant years 1 and 2 (Tukey's $p < 0.05$), and years 1 and 3 (Tukey's $p < 0.05$), but no significant increase was found between years 2 and 3 (Tukey's $p = 0.97$). Figure 2b shows that perceptual sensitivity (d') followed a similar pattern, increasing between years 1 and 2 (Tukey's $p < 0.01$) and years 1 and 3 (Tukey's $p < 0.01$), but not between years 2 and 3 (Tukey's $p = 0.39$).

**Effects of medical, audiological, and demographic variables.** The effects of medical, audiological and demographic variables on CPT scores were examined in separate analyses using SAS mixed models which controlled for chronological age and length of CI use. We found no significant effects on school-age CPT scores for any of the variables described in Table 1.

**Correlations between visual sustained attention and speech and language outcome measures.** We computed simple bivariate correlations between the four CPT measures and each of the speech and language outcome measures using scores obtained at two years post-implantation. As shown in Table 4, significant correlations (*p* <0.05) were only found between vocabulary (PPVT) and both hits and false alarms on the CPT (*r*'s 0.45 and –0.47 respectively). No other significant correlations were observed between CPT scores and speech/language outcome measures. Because many of the speech and language outcome measures are often correlated with chronological age, we conducted univariate regression analyses with PPVT score as the dependent measure, hit rate or false alarm rate as the independent measures, and chronological age as a covariate. Using this model, the relations observed between false alarms and PPVT scores remained significant (*F*(1,19)=6.1, *p* <0.05). However, the relations between hits and PPVT scores were not significant when chronological age was controlled (*F*(1,21)=2.277, *p* >0.05). Thus, a lower rate of false alarms on the CPT was associated with greater vocabulary knowledge scores obtained after two years of implant use.

| Outcome Measure | Hit Rate | False Alarms | d' | Beta |
|---|---|---|---|---|
| **PBK** | ns | ns | ns | ns |
| **CP V** | ns | ns | ns | ns |
| **CP AV** | ns | ns | ns | ns |
| **PPVT** | 0. 45* | - 0.53* | ns | ns |
| **RE** | ns | ns | ns | ns |
| **RR** | ns | ns | ns | ns |
| **BIT** | ns | ns | ns | ns |

**Table 4.** Bivariate correlations between school-age CPT scores and outcome measures
* *p*< 0.05 level (two tailed)

## Discussion

The low mean percentile CPT scores demonstrated by the children in our study are consistent with the earlier findings reported by Quittner et al. and Smith et al. that, across ages, prelingually deaf children show atypical sustained visual attention skills compared to normal-hearing children. Depending on the length of CI use, approximately one-third to one-half of children fell into the "abnormal" clinical range as defined in the Gordon CPT manual (Gordon et al., 1996). However, similar proportions of children performed in the "normal" range suggesting that not all deaf children displayed atypical sustained visual attention skills. From the means of the percentile scores, it appeared that children showed more atypical performance for false alarms than for number of hits.

The SAS mixed model used here allowed us to measure the effects of chronological age and length of CI use separately as continuous independent variables. Similarly to Quittner et al. and Smith et al., we found a significant effect of chronological age on CPT performance. We also found that CPT performance increased with CI use. Over three years of CI use, prelingually deaf children showed an increase in hit rate and perceptual sensitivity when the effect of chronological age was partialled out.

These findings suggest that sustained visual attention abilities of prelingually deaf children may begin to improve during the first year after implantation and, on some measures, continue to improve over at least 3 years of CI use. Because we did not have sufficient data to analyze children following greater lengths of CI use, we cannot say whether this improvement levels off or continues to increase with greater CI use. Given the previous findings of Quittner et al. and Smith et al., we would expect continued improvement in CPT performance as a function of CI use until the deaf children are, on average, performing equally to normal-hearing children. We did not find a significant interaction between chronological age and length of CI use to suggest a critical period for development of sustained visual attention skills. We did not find any significant effects for the demographic, medical, or audiological variables on CPT performance.

Our correlational analysis revealed a significant relation between false alarm rate on the school-age CPT and vocabulary knowledge as assessed by the PPVT in prelingually deaf children who had used a CI for two years. No other relations were uncovered involving the speech perception, language, or speech intelligibility measures. These findings suggest that individual differences in sustained visual attention abilities may explain some portion of the variability of vocabulary acquisition, but not other aspects of speech and language development, in prelingually deaf children with CIs.

## Experiment II

### Methods

**Participants.** Experiment II was also a retrospective analysis of longitudinal clinical data gathered at the Indiana University School of Medicine Cochlear Implant Program. Participants included a younger group of 47 prelingually deaf children who received CIs by 9 years of age. The medical, audiological, and demographic characteristics of the sample are summarized in Table 5.

| | | |
|---|---|---|
| **Medical** | **Etiology** | **3 genetic, 13 meningitis,<br>31 unknown** |
| | **Ear of Implantation** | **24 right, 19 left** |
| **Audiological** | **Mean Pure Tone Average (n=29) (dB HL)** | **110.1 (6.162)** |
| | **Mean Number of active electrodes** | **19.8 (3.42)** |
| | **Mean Age of implantation (yrs.)** | **4.80 (1.34)** |
| **Demographic** | **Gender** | **21 female, 26 male** |
| | **Communication mode** | **22 OC, 25 TC** |
| | **Mean non-verbal IQ (n=18)** | **104.1 (15.28)** |

**Table 5.** Medical, audiological, and demographic characteristics of participants in Experiment I (standard deviations in parentheses)

Testing protocols in this experiment were identical to the previous study. Children were tested once every 6-12 months until two years post-implantation. Interval data were collapsed into one of three intervals: pre-implant, one year post-implant, and two years post-implant. As in the previous experiment, not all children were tested on all measures at each interval. The numbers of children who were tested at each interval are shown in Table 2.

**Procedures.** The preschool CPT is recommended for use in normal-hearing children from 3-5 years of age (Gordon et al.). This 6-minute CPT task uses the same testing apparatus and visual stimuli as the school-age CPT described previously. In this new task, children were also required to monitor a stream of visually presented numbers presented at 1-second intervals. In contrast to the school-age CPT, however, the preschool CPT required children to only respond whenever a "1" appeared on the screen. This CPT version is easier than the school-age task because children are not required to remember the previously presented number when the target appears. Therefore, the preschool CPT has a lower working memory load than the school-age CPT. The results were scored in the same manner and dependent variables for sustained visual attention were computed as described previously for Experiment I.

The same five traditional speech and language outcome measures used in Experiment I were also used as outcome measures in this second experiment. However, in this younger population of children, we did not have a sufficient sample size with 2 years of CI use to carry out correlations with scores obtained at this interval. Therefore, in Experiment II, we only examined relations between measures obtained after 1 year of CI use.

## Results

To test for independent effects of chronological age and length of implant use, we constructed a model using the SAS Mixed Procedure described earlier.

**Preschool CPT performance of prelingually deaf children with CIs compared to normative sample.** Although the preschool CPT is recommended and normed for children from 3-5 years of age, a number of children were older than 5 at the time of testing. The most stringent criteria for calculating percentile scores would be to exclude those children who were older than 5 years old. However, this would leave us with a very small group of children to compare to the normative samples. Therefore, we made the judgment to report normative scores for children who were 6 years old or younger at the time of testing, although these data should be interpreted with caution and are included only for descriptive purposes. Table 6 shows the percentile means for hit rate and false alarm rate at each interval along with the number of participants who were normed at each interval. These data suggest that, like the older deaf children in Experiment I, deaf children with CIs performed poorly on the preschool CPT compared to the normative sample. The highest mean normative scores, 38th and 20th percentile for hits and false alarms respectively, were found in children who had used their CI for two years.

| Gordon Test | CI Use (Years) | Percentile mean (SD) |
|---|---|---|
| **Hit Rate** | 0 (n=19) | 20.6 (25.6) |
| | 1 (n=25) | 20.2 (23.3) |
| | 2 (n=19) | 38.5 (36.1) |
| **False Alarms** | 0 (n=19) | 13.8 (14.8) |
| | 1 (n=25) | 22.7 (23.0) |
| | 2 (n=19) | 25.4 (20.6) |

**Table 6.** Percentile CPT scores at each interval of testing from Experiment II (preschool CPT).

**Effects of chronological age and length of CI use.** The results obtained from the SAS mixed model revealed significant effects of chronological age on both hit rate (*F*(1, 47.7) = 64.69, *p*=<0.01), and d' (*F*(1, 50.9) = 55.2, *p*=<0.01). Thus, this sample of deaf children with CIs demonstrated an increase in hit rate and increase in perceptual sensitivity as a function of chronological age. No effect of chronological age was found on β or false alarm rate, although the latter just missed significance (*F*(1,51.6) = 3.18, *p*=0.08). Figures 3a-b illustrate the mean hit rate and d' of individual participants plotted as a function of chronological age. In all subsequent analyses with the SAS mixed model, we controlled for the effect of chronological age.

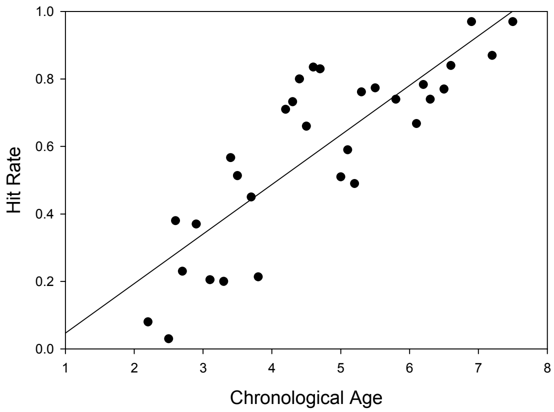


**Figure 3a.** Hit rate on the preschool CPT as a function of chronological age. Some data points overlap; all individual data points are plotted with a line of best fit.

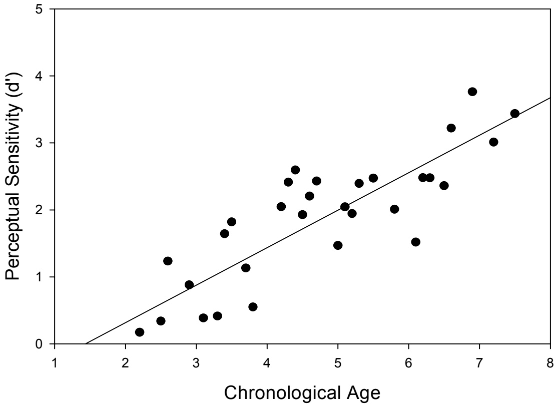


**Figure 3b.** Perceptual sensitivity (d') on the preschool CPT as a function of chronological age. Some data points overlap; all individual data points are plotted with a line of best fit.

Estimated mean hit rate increased from 0.45 (*SE*=0.05) before implantation to 0.74 (*SE*=0.04) after two years of CI use, a significant effect (*F*(2,47.8) = 11.38, *p* <0.01). A significant effect of length of CI use was also found for mean false alarm rate which decreased from 0.14 (*SE*=0.02) before cochlear implantation to 0.07 (*SE*=0.02) after two years of CI use (*F*(251.3) = 5.55, *p* <0.01). Finally, d' increased from 1.04 (*SE*=0.19) prior to implantation to 2.49 (*SE*=0.17) after two years of CI use, also a significant effect (*F*(2,50.2) = 16.25, *p* <0.01). In contrast, length of CI use did not significantly effect β. Figures 4a-c illustrate the significant effects of length of CI use on preschool CPT performance.

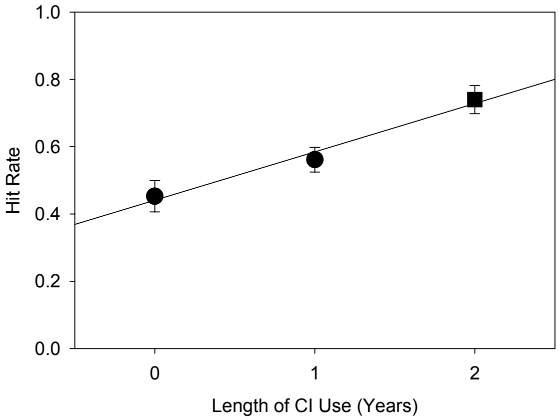


**Figure 4a.** Hit rate on the preschool CPT as a function of length of CI use. Different shaped data points are used to represent significantly different performance means between different lengths of CI use.

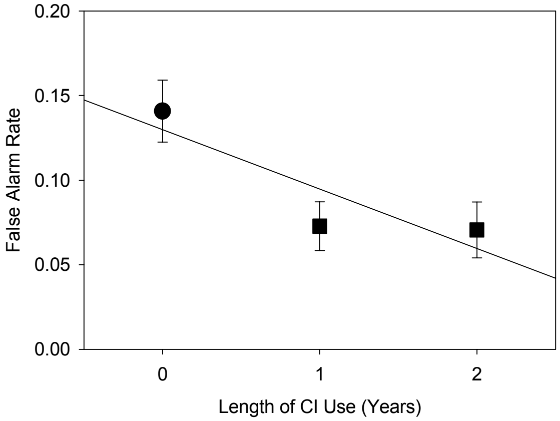


**Figure 4b.** False alarm rate on the preschool CPT as a function of length of CI use. Different shaped data points are used to represent significantly different performance means between different lengths of CI use.

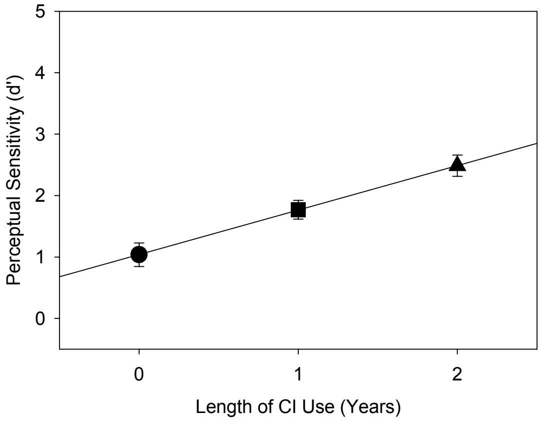


**Figure 4c.** Perceptual sensitivity (d') on the preschool CPT as a function of length of CI use. Different shaped data points are used to represent significantly different performance means between different lengths of CI use.

Tukey's post hoc analyses were carried out to assess performance differences on the preschool CPT between each post-implant interval. Figure 4a shows that preschool CPT hit rate increased significantly, as a function of length of CI use, between post-implantation years 1 and 2 (Tukey's $p$<0.01), but not between pre-implant and post-implant year 1 although the latter was a trend (Tukey's $p$=0.06). Figure 2b shows that false alarm rate on this task decreased significantly between the pre-implant interval and post-implantation year 1 (Tukey's $p$ <0.01), but not between post-implantation years 1 and 2 (Tukey's $p$ =0.92). Figure 2c shows that perceptual sensitivity (d') increased significantly from pre-implantation to post-implantation year 1 (Tukey's $p$ <0.01) and then again from post-implantation year 1 to post-implantation year 2 (Tukey's $p$ <0.01).

**Effects of medical, audiological, and demographic variables on preschool CPT performance.** No significant effects on preschool CPT scores were found for any of the medical, audiological, or demographic variables with the exception of the number of active electrodes in the implant array. To explore this effect further, we divided the participants into one of two groups, using a median split, at an electrode number of 21. Thus, the two groups consisted of children with a full array of active electrodes and children with less than a full array of active electrodes. Children in the full array group had higher d' scores on the preschool CPT ($F$(1,48)=6.79, $p$ <0.05) and a higher hit rate ($F$(1,43.9)=4.03, $p$ =0.05) than children who had less than 22 electrodes. Not only does this effect remain significant when we split at lower electrode numbers, but the significance p values become even lower suggesting that the effect is likely carried by those children with electrode numbers far lower than 21. However, because we found electrode number to be skewed toward higher numbers, we only present the findings using a median split.

**Correlations between preschool CPT scores and speech and language outcome measures.** We computed bivariate correlations between each preschool CPT measure and each of the speech and language outcome measures using scores obtained at one year post-implantation. As shown in Table 7, significant correlations ($p$ <0.05) were found between expressive and receptive language (RDLS) and hit rate ($r$'s =0.48 and 0.52, respectively). No other significant correlations were found between preschool CPT scores and speech/language outcome measures. We conducted two univariate regression analyses with RDLS receptive and expressive scores as dependent variables, CPT hit rate as the independent variable, and chronological age as a covariate. In this model, the relations between hit rate and RDL expressive and receptive language were no longer significant. Thus, when we controlled for the effects of chronological age, we found no significant relations between the preschool CPT scores and speech and language outcomes in deaf children who had used their CIs for 1 year.

| **Outcome Measure** | **Hit Rate** | **False Alarms** | **d'** | **Beta** |
|---|---|---|---|---|
| **PBK** | ns | ns | ns | ns |
| **CP V** | ns | ns | ns | ns |
| **CP AV** | ns | ns | ns | ns |
| **PPVT** | ns | ns | ns | ns |
| **RE** | 0.517* | ns | ns | ns |
| **RR** | 0.483* | ns | ns | ns |
| **BIT** | ns | ns | ns | ns |

**Table 4.** Bivariate correlations between school-age CPT scores and outcome measures
* *p*= 0.05 level (two tailed)

## Discussion

The children in Experiment II, who were all younger than 6 years of age, performed poorly on the preschool CPT compared to the normative sample tested by Gordon et al. Although no study to date has compared the preschool CPT scores of prelingually deaf children to normal-hearing children, our results suggest that atypical development of sustained visual attention in deaf children is detectible by ages 3-6 years old. However, future comparisons between deaf and normal-hearing children on the preschool CPT will help to confirm this interpretation.

As we showed in Experiment I for the older children, scores on the preschool CPT improved as a function of length of CI use. Over two years of CI use, prelingually deaf children showed an increase in hit rate, decrease in false alarm, and an increase in perceptual sensitivity when the effect of chronological age was partialled out. While the decrease in false alarm rate on the preschool CPT leveled off after one year of CI use, the hit rate did not increase until after 1 year of use. Perceptual sensitivity, d', increased significantly at each interval of CI use. Given the earlier findings of Quittner et al. and Smith et al., we would expect continued improvement in CPT performance as a function of CI use until the deaf children are, on average, performing at levels that are comparable to normal-hearing children.

We did not observe any effects of the medical, audiological, or demographic variables on preschool CPT scores except for the number of active electrodes. This is the first report of an effect for the number of active electrodes on the development of visual attention skills in children with CIs. Our sample size and skewed distribution of electrode number do not allow us to determine whether this finding was carried by a small number of children who have fewer active electrodes than most of the children. Other studies have found that only a relatively small number of electrodes (less than 12) is required to achieve maximal speech perception scores with CIs or CI simulations (Dorman, Loizou, Kemp, & Kirk, 2000; Friesen, Shannon, Baskent, & Wang, 2001).

There are at least two reasons why some children with CIs may not have a full array of active electrodes. First, some of these children may have had partial insertions due to anatomic cochlear abnormalities such as a Mondini malformation. Secondly, during programming it is sometimes necessary to deactivate electrodes which are functioning inappropriately (stimulating the facial nerve, requiring a disproportionate large amount of current, not working at all). Thus, it is hard to determine precisely whether the effect we found is due to the actual number of stimulation points in the cochlea, or to some other confounding variable associated with a lower number of electrodes.

Our correlational analysis of the data in Experiment II revealed no significant relations between the preschool CPT scores and the speech perception, language, vocabulary, or speech intelligibility measures when we controlled for chronological age. These findings demonstrate that individual differences in sustained visual attention abilities and speech/language outcomes do not share a significant amount of variance. However, our correlations were obtained from scores at 1 year post-implantation. This is admittedly quite early in terms of assessing speech/language benefits with a cochlear implant. It is possible that relations between CPT performance and speech/language skills would emerge if we were to test children with 4 or more years of implant experience. Indeed, this is a research area for future exploration.

## General Discussion

Experiments I and II investigated the development of sustained visual attention skills of prelingually deaf children who used cochlear implants. Overall, our results are consistent with the earlier findings of Quittner et al. and Smith et al. which showed that the sustained visual attention skills of prelingually deaf children are atypical compared to normal-hearing children. The effect of CI use on CPT performance, independent from chronological age, suggests that auditory experience leads to a gradual maturation of sustained attention skills over a period of years. Although we did not find direct evidence for a critical period for sustained visual attention development as suggested by Quittner et al.'s and Smith et al.'s data, the lack of an interaction between length of CI use and chronological age in our study does not disprove their critical period hypothesis. It is possible that the age ranges of children in the present study fell within a critical period during which CI experience influenced sustained visual attention development. Further research on the sustained visual attention skills of our youngest and oldest children with CIs may reveal evidence of a critical period.

We found little evidence to suggest that demographic, medical, and audiological variables influenced CPT performance, and did not find any interactions of these variables with the effect of length of CI use. Therefore, the impact of early auditory experience appears to shape the development of cognitive processing in the visual modality. Moreover, sustained visual attention abilities display some degree of plasticity and ability to reorganize in the presence of the cross-modal auditory input provided by a CI. However, the underlying neural and cognitive processes responsible for deaf children's atypical CPT performance, and for the effect of CI use, have yet to be sufficiently defined.

The hypothesis proposed by Quittner et al. and Smith et al. was that early auditory deprivation leads to remodeling of visual attention processes. Reasoning that deaf children are required to utilize vision to monitor their environment, Quittner et al. and Smith et al. argued that visual attention processes in these children would reorganize and adapt to maintain a wide spatial focus rather than a narrow, task-specific focus such as a CPT. In describing their "division of labor" hypothesis, Smith et al. suggested that normal-hearing children learn to sustain focused visual attention with a remarkable degree of acuity in part because of their ability to utilize auditory signals to detect environmental events. Because many environmental events may occur outside the field of visual attention, audition frees the visual system from utilizing capacity-demanding resources to detect these events.

However, the CPT used in the present and earlier studies does not explicitly assess selective allocation of visual attention. In selective attention experiments, distracting stimuli are used to test for effects on visual target processing time or accuracy (Parasnis, Samar, & Berent, 2003). Although selective attention may be useful in any task requiring concentration (there are almost always small sounds/events occurring in the environment), we do not know to what degree these abilities play a role in performance on sustained attention tasks. To test the predictions of the division of labor hypothesis, therefore, the appropriate type of task is a selective attention task.

The Gordon Diagnostic System includes a task called the distractibility CPT (Gordon et al., 1996). This task is similar to the previous CPTs (which we will call vigilance CPTs for distinction) except for the addition of two additional streams of numbers which appear to the right and left of the original number sequence. Participants are instructed to ignore these two additional number sequences and focus only on the numbers which appear in the center of the display. By comparing performance on the distractibility CPT with the vigilance CPT, the ability of a subject to selectively attend to the center display can be measured.

Mitchell and Quittner (1996) administered both the distractibility and vigilance CPTs to a population of normal-hearing children and prelingually deaf children with CIs. They found that prelingually deaf children scored lower on average than the normal-hearing children on both CPTs. The authors computed a distraction decrement score by subtracting each subject's score on the distractibility CPT from their score on the vigilance CPT. Mitchell and Quittner reported that prelingually deaf children with CIs showed greater mean distraction decrements than the normal-hearing children. Thus, prelingually deaf children with CIs were less able to ignore the distracting numbers which flanked the stimulus stream of interest. This finding is consistent with Smith et al.'s prediction that selective visual attention skills of prelingually deaf children develop atypically compared to those of normal-hearing children.

A number of visual information processing studies conducted with prelingually deaf adults have revealed compelling evidence that auditory deprivation leads to reorganization of visual attention processes (Bavelier, Brozinsky, Tomann, Mitchell, Neville, & Liu, 2001; Bavelier, Tomann, Hutton, Mitchell, Corina, Liu, & Neville, 2000; Neville & Lawson, 1987a, 1987b, 1987c; Proksch & Bavelier, 2002; Rothpletz, Ashmead, & Tharpe, 2003). In general, the findings of these electrophysiological, functional magnetic resonance imaging (fMRI), and behavioral studies have demonstrated that deaf adults, compared to normal hearing adults show evidence for increased processing of peripheral visual stimuli and wider spatial distribution of selective attention. The deaf adults in these studies appeared to have developed a visual system geared toward processing a wider spatial area than typically observed in normal hearing participants (Proksch & Bavelier, 2002).

The deaf adult participants in these experiments typically use American Sign Language (a manual language which is linguistically distinct from any spoken language). Therefore, several studies have tested normal hearing adults who learned ASL at an early age (due to having deaf parents) to control for effects of acquiring a manual language. The effect of ASL fluency itself has not turned out to be significant, suggesting that the reorganization of peripheral visual processing and selective attention results from auditory deprivation (Bavelier et al., 2001; Neville et al., 1987c; Proksch & Bavelier, 2002).

These studies of visual processing in deaf adults do not tell us at what age this visual reorganization occurs. However, the results reported by Mitchell and Quittner with the distractibility CPT showing that deaf children were more distractible than normal-hearing children do suggest that some reorganization occurs in deaf children by the school-age years. Furthermore, if deaf children are processing a wider part of their visual world than normal-hearing children, this might lead to atypical performance on the vigilance CPT as well (although this task is not a selective attention task). Given the capacity limitations of young children, these additional processing demands might be responsible for the poor performance of prelingually deaf children on the tasks in this study and the earlier studies of Quittner et al. and Smith et al.

Presumably, the effect of auditory experience on the development of sustained and selective visual attention reflects some degree of cortical reorganization. Studies of cross-modal plasticity have found evidence to suggest that cortical areas normally responsible for auditory processing can be recruited and reorganized for visual processing. In an early study, Rebillard, Carlier, Rebillard, and Pujol (1977) reported increased visual evoked responses in the primary auditory cortex of congenitally deafened cats compared to normal-hearing cats. Their findings were taken as evidence for recruitment of unused auditory cortex by visual cortical functions. Although a more recent study of congenitally deaf cats failed to demonstrate cross-modal recruitment of primary auditory cortex (Kral, Schroder, Klinke, & Engel, 2003), other evidence has been found for cross-modal plasticity in secondary auditory areas.

Finney and colleagues (Finney, Clementz, Hickok, & Dobkins, 2003; Finney, Fine, & Dobkins, 2001) have employed neuro-imaging techniques to demonstrate visual evoked activity in the auditory cortex of prelingually deaf adults. In particular, increased activity Brodmann's areas 41, 42, and 43 in deaf adults compared to normal-hearing adults when participants processed visual motion stimuli (Finney et al., 2001). This visually-evoked activity in the auditory cortex of deaf adults occurs rapidly after stimulus presentation, suggesting that it results from cross-modal recruitment of auditory cortex by visual thalamic afferents (Finney et al., 2003).

Lee and colleagues (Lee, Lee, Oh, Kim, Kim, Chung, Lee, & Kim, 2001; Oh, Kim, Kang, Lee, Lee, Chang, Ahn, Hwang, Park, & Koo, 2003) have found PET evidence that cross-modal recruitment of auditory cortex may have important consequences for speech/language outcomes in prelingually deaf children with CIs. These authors found a negative predictive relationship between the degree of pre-implant auditory cortex metabolism and post-implant speech perception scores. Children with less auditory metabolism (an indication of cross-modal recruitment) demonstrated better outcomes with their CI. The results reported by Lee et al. and Oh et al. suggest that cross-modal reorganization is one cortical mechanism which may be responsible for shaping the speech perceptual processes of deaf children with CIs. However, it is not clear whether cross modal recruitment of auditory cortex plays a role in the observed atypical visual processes of deaf children and adults. Relations between cross-modal recruitment of auditory cortex and visual attention processes have not yet been reported and may prove to be a promising area for future research.

The effect of CI use on sustained visual attention skills of deaf children, and their atypical skills relative to normal-hearing children, might be due to additional factors other than cross-modal reorganization. An alternative hypothesis is that these findings may reflect underlying deficiencies in processing of verbally-encoded visual stimuli in deaf children. Because visually presented numbers were likely to be verbally-encoded by the deaf children in our study, we would expect that verbal fluency would play a role in CPT performance. Several recent studies of working memory capacity in prelingually deaf children with CIs have reported evidence that verbal encoding and rehearsal skills are also atypical in these children.

Prelingually deaf children with CIs have been shown to have impaired immediate recall for sequences of stimuli regardless of modality of presentation (Cleary, Pisoni, & Geers, 2001; Dawson, Busby, McKay, & Clark, 2002; Pisoni & Cleary, 2003). Compared to normal-hearing children, prelingually deaf children with CIs display reduced overall span (number of items they can recall) for auditory presentation of numbers (Pisoni & Cleary, 2003), visual presentation of colored lights (Cleary et al., 2001), auditory presentation of color names (Cleary et al., 2001), and for auditory color names and lights presented together (Cleary et al., 2001). Dawson et al. tested deaf children with CIs on a number of immediate recall tasks using either auditory or visually-presented stimuli. They found impairments in deaf children on the auditory tasks as well as on some of the visual tasks. Dawson et al. found that, on the visual tasks which employed stimuli which could be verbally encoded or named, deaf children showed shorter spans than normal-hearing children. However, they did not show impairments in recall of visual stimuli which could not be easily verbally encoded. Taken together these studies suggest that verbal encoding of stimuli is impaired in young deaf children regardless of whether these stimuli are heard, seen, or both.

The CPTs used in the present study utilized visually-presented numbers, stimuli which can be verbally encoded by our participants. Therefore, deficits in verbal encoding of these stimuli may have played a limiting role in the performance of deaf children on the CPT. In other words, the deaf children might have normal visual attention mechanisms, yet have difficulty encoding and processing stimuli into phonological representations. Indeed, phonological storage and sub-vocal rehearsal of stimulus representations via the phonological loop have been shown to protect stimulus representations in working memory from time-related decay (Baddeley, Gathercole, & Papagno, 1998). Studies which manipulate use of the phonological loop by participants have shown that deaf children do not take advantage of this mechanism to the same degree as normal-hearing children (Chincotta & Chincotta, 1996; Pisoni & Cleary, 2003). Therefore, the CPT used in this and other studies may be biased against prelingually deaf children with atypical verbal encoding skills.

To date, only two studies of sustained visual attention in normal hearing adults and prelingually deaf adult ASL users have utilized CPTs in which stimuli were unlikely to be verbally encoded. Dittmar, Berch, and Warm (1982) used a CPT which required 45 minutes of visual monitoring of a light-bar which executed a string of paired movements. The detection task was to respond when the degree of movement differed between the two paired movements. Dittmar et al. reported that deaf adults showed greater performance on this task than normal hearing adults, a result which initially appears to contradict the present results. However, the visual processing required by this task is clearly different from the current CPT because visual detection of movement differences does not involve or require verbal encoding. Thus, differences in CPT performance between deaf and normal hearing individuals may depend on the type of visual stimulus used, and whether or not this stimulus is verbally encoded.

Recently, Parasnis et al. (2003) reported CPT results obtained from prelingually deaf adult ASL users and normal hearing adults. Their CPT used visually-presented geometrical shapes as stimuli and the task was to detect a certain type of shape based on whether the object had a hole toward its top or bottom. In contrast to Dittmar et al.'s findings, Parasnis et al. reported deficits in performance of deaf adults compared to normal-hearing participants. It is unclear to what degree the geometric stimuli were verbally encoded (for instance as, "top" or "bottom"). If we assume that verbal encoding was utilized by participants in Parasnis et al.'s task, then their results appear consistent with the current and earlier studies by Quittner et al. and Smith et al. which found deficits in deaf children's' abilities on a CPT which used numbers as stimuli. Future work on sustained visual attention skills of deaf children with CIs should include tasks which are not likely to induce or encourage verbal encoding.

In summary, we have proposed two general mechanisms to explain the effect of a CI on the sustained visual attention skills of deaf children. One possible explanation is that visual attention processing is altered by cross-modal auditory experiences obtained during a period of auditory deprivation and is reorganized after implantation. A second possible explanation is that atypical phonological encoding and verbal rehearsal skills, which result from a period of early auditory deprivation begin to improve after cochlear implantation. Both processes may interact to affect CPT performance in deaf children, and certainly other explanations for the current and previous findings may exist. Future research into sustained visual attention skills of deaf children with CIs should utilize tasks which do not contain stimuli which can be easily named or encoded phonologically. Testing of these children on tasks of selective attention may help to define the nature of reorganization of visual attention resulting from auditory experience. In parallel to these studies, future work on the longitudinal development of the phonological loop in deaf children may provide valuable information as to how the processing and storage of visual stimuli may be altered by early auditory experience. Finally, the testing of children from a wide range of ages may reveal critical periods during which the development of these cognitive processes may be altered by auditory deprivation and cochlear implantation.

We did not find evidence to suggest that the skills assessed by the CPT were closely related to speech and language acquisition with CIs, although we did find one relationship between the school-age CPT and vocabulary knowledge. However, before we make the conclusion that sustained visual attention skills do not partially explain individual differences in audiological outcomes of these children, it would be important to examine relations between CPT and speech/language scores after longer intervals of CI use. Relations may emerge after many years of implantation which we were not able to detect in this study. Therefore, more research on CPTs and other laboratory measures of sustained visual attention are needed to fully assess the clinical usefulness of these measures for deaf children with cochlear implants.

## References

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

Barkley, R. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology, 19*(2), 149-178.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., & Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *Journal of Neuroscience, 21*, 8931-8942.

Bavelier, D., Tomann, A., Hutton, C., Mitchell, T., Corina, D., Liu, G., et al. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. *Journal of Neuroscience, 20*, RC93.

Blamey, P., Sarant, J., Paatsch, L., Barry, J., Bow, C., Wales, R., et al. (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.

Chincotta, M., & Chincotta, D. (1996). Digit span, articulatory suppression, and the deaf: a study of the Hong Kong Chinese. *American Annals of the Deaf, 141*, 252-257.

Cleary, M., Pisoni, D., & Geers, A. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing, 22*, 395-411.

Conners, C. (2000). *Conners' CPT II for Windows*. North Tonawanda, NY: Multi-Health Systems.

Dawson, P., Busby, P., McKay, C., & Clark, G. (2002). Short-term auditory memory in children using cochlear implants and its relevance to receptive language. *Journal of Speech Language and Hearing Research, 45*, 789-801.

Dittmar, M., Berch, D., & Warm, J. (1982). Sustained visual attention in deaf and hearing adults. *Bulletin of the Psychonomic Society, 19*, 339-342.

Dorman, M., Loizou, P., Kemp, L., & Kirk, K. (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.

Dunn, L., & LM, D. (1997). *Peabody picture vocabulary test, 3rd edition*. Circle Pines, MN: American Guidance Service.

Finney, E., Clementz, B., Hickok, G., & Dobkins, K. (2003). Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG. *Neuroreport, 14*, 1425-1427.

Finney, E., Fine, I., & Dobkins, K. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience, 4*, 1171-1173.

Forbes, G. (1998). Clinical utility of the test of variables of attention (TOVA) in the diagnosis of attention-deficit/hyperactivity disorder. *Journal of Clinical Psychology, 54*, 461-476.

Friesen, L., Shannon, R., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110*, 1150-1163.

Fryauf-Bertschy, H., Tyler, R., Kelsay, D., & Gantz, B. (1992). Performance over time of congenitally deaf and postlingually deafened children using a multichannel cochlear implant. *Journal of Speech Language and Hearing Research, 35*, 913-920.

Geers, A., Brenner, C., & Davidson, L. (2003). Factors associated with development of speech perception skills in children implanted by age five. *Ear and Hearing, 24*, 24S-35S.

Geers, A., Nicholas, J., & Sedey, A. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24*, 46S-58S.

Gordon, M., McClure, F., & Aylward, G. (1996). *The Gordon diagnostic system (GDS) instruction manual & interpretive guide* (3 ed.). Dewitt, NY: Michael Gordon, PhD.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Greenberg, L. (1991). *T.O.V.A. interpretation manual*. Minneapolis, MN.

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

Kirk, K. (2000). Cochlear implants: new developments and results. *Opinion in Otolaryngology Head and Neck Surgery, 8*, 415-420.

Kral, A., Schroder, J., Klinke, R., & Engel, A. (2003). Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Experimental Brain Research, 153*, 605-613.

Lee, D., Lee, J., Oh, S., Kim, S., Kim, J., Chung, J., et al. (2001). Cross-modal plasticity and cochlear implants. *Nature, 409*, 149-150.

Mitchell, T., & Quittner, A. (1996). Multimethod study of attention and behavior problems in hearing-impaired children. *Journal of Clinical Child Psychology, 25*, 83-96.

Miyamoto, R., Osberger, M., Robbins, A., Myres, W., & Kessler, K. (1993). Prelingually deafened children's performance with the nucleus multichannel cochlear implant. *Amercian Journal of Otology, 14*, 437-445.

Miyamoto, R., Svirsky, M., Kirk, K., Robbins, A., Todd, S., & Riley, A. (1997). Speech intelligibility of children with multichannel cochlear implants. *Annals of Otology Rhinology Laryngology Supplement, 168*, 35-36.

Neville, H., & Lawson, D. (1987a). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. I. Normal hearing adults. *Brain Research, 405*, 253-267.

Neville, H., & Lawson, D. (1987b). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Research, 405*, 268-283.

Neville, H., & Lawson, D. (1987c). Attention to central and peripheral visual space in a movement detection task. III. Separate effects of auditory deprivation and acquisition of a visual language. *Brain Research, 405*, 284-294.

Niparko, J., Kirk, K., Mellon, N., McConkey-Robbins, A., Tucci, D., & Wilson, B. (2000). *Cochlear Implants: Principles & Practices*. Philadelphia, PA: Lippincott Williams & Wilkins.

Oh, S., Kim, C., Kang, E., Lee, D., Lee, H., Chang, S., et al. (2003). Speech perception after cochlear implantation over a 4-year time period. *Acta Oto-Laryngologica, 123*(2), 148-153.

Osberger, M., Miyamoto, R., Zimmerman-Philips, S., Kemink, J., Stroer, B., Firszt, J., et al. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus-22 channel cochlear implant system. *Ear and Hearing, 12*, S66-80.

Osberger, M., Robbins, A., Todd, S., & Riley, A. (1994). Speech intelligibility of children with cochlear implants. *Volta Review, 96*, 169-180.

Parasnis, I., Samar, V., & Berent, G. (2003). Deaf adults without attention deficit hyperactivity disorder display reduced perceptual sensitivity and elevated impulsivity on the Test of Variables of Attention (T.O.V.A.). *Journal of Speech Language and Hearing Research, 46*, 1166-1183.

Parasuraman, R. (1998). *The Attentive Brain*. Cambridge, MA: MIT Press.

Pastore, R., Crawley, E., Berens, M., & Skelly, M. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10*, 556-569.

Pisoni, D., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*, 106S-120S.

Pisoni, D., 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.

Proksch, J., & Bavelier, D. (2002). Changes in the spatial distribution of visual attention after early deafness. *Journal of Cognitive Neuroscience, 14*, 687-701.

Quittner, A., Smith, L., Osberger, M., Mitchell, T., & Katz, D. (1994). The impact of audition on the development of visual attention. *Psychological Science, 5*, 347-353.

Rebillard, G., Carlier, E., Rebillard, M., & Pujol, R. (1977). Enhancement of visual responses on the primary auditory cortex of the cat after an early destruction of cochlear receptors. *Brain Research, 129*, 162-164.

Reynell, J. K., & Huntley, M. (1985). *Reynell Developmental Language Scales* (2nd ed.). Windsor, UK: NFER-Nelson.

Rosvold, H., Mirskey, A., Sarason, I., Bransome, E., & Beck, L. (1956). A continuous performance test of brain damage. *Journal of Consulting and Clinical Psychology, 20*, 343-350.

Rothpletz, A., Ashmead, D., & Tharpe, A. (2003). Responses to targets in the visual periphery in deaf and normal hearing adults. *Journal of Speech Language and Hearing Research, 46*, 1378-1386.

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

Schafer, J., & Graham, J. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*, 147-177.

Smith, L., Quittner, A., Osberger, M., & Miyamoto, R. (1998). Audition and visual attention: the developmental trajectory in deaf and hearing populations. *Developmental Psychology, 34*, 840-850.

Svirsky, M., Robbins, A., Kirk, K., Pisoni, D., & Miyamoto, R. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science, 11*, 153-158.

Tharpe, A., Ashmead, D., & Rothpletz, A. (2002). Visual attention in children with normal hearing, children with hearing aids, and children with cochlear implants. *Journal of Speech, Language, & Hearing Research, 45*, 403-413.

Tobey, E., Geers, A., Brenner, C., Altuna, D., & Gabbert, G. (2003). Factors associated with development of speech production skills in children implanted by age five. *Ear and Hearing, 24*, 36S-45S.

Tyler, R., Fryauf-Bertschy, H., Kelsay, D., Gantz, B., Woodworth, G., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology Head and Neck Surgery, 117*, 180-187.

Wolfinger, R., & Chang, M. (1995). *Comparing the SAS GLM and MIXED procedures for repeated measures.* Paper presented at the Proceedings of the Twentieth Annual SAS Users Group International Conference, Orlando, Florida.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Mothers' Speech to Hearing-Impaired Infants with Cochlear Implants: Some Preliminary Findings[1]

**Tonya R. Bergeson[2] and Kasi McCune[2]**

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

[1] This work was supported by NIH-NIDCD Training Grant T32DC00012 to Indiana University and by NIH-NIDCD Research Grant R01DC00064 to the Indiana University School of Medicine. We would like to thank Carrie Hansel and Becky Davis for their help with acoustic analyses.

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

# Mothers' Speech to Hearing-Impaired Infants with Cochlear Implants: Some Preliminary Findings

**Abstract.** Caregivers typically speak to their normal-hearing (NH) infants using a distinct style known as infant-directed (ID) speech. With the recent expansion of cochlear implant (CI) criteria to include infants, it is critical to assess the changes in mothers' speech to infants as they acquire auditory skills via CIs, compared to mothers' speech to NH infants. To establish mothers' vocal style when speaking to their NH and implanted infants, we digitally recorded mothers speaking to their 4- to 10-month-old NH infants and mothers speaking to their 17- to 37-month-old hearing-impaired infants who use a CI. For the implanted infants, the average duration of implant use (i.e., "hearing age") was 11 months. We also recorded mothers' adult-directed (AD) speech. We analyzed the recordings in terms of the acoustic features known to characterize ID speech. The results were similar to those of previous studies. These preliminary findings suggest that mothers speak to NH and implanted infants in similar styles despite the chronological age difference in the two infant populations. That is, mothers' speech to infants is influenced by the child's "hearing age" rather than chronological age.

## Introduction

Caregivers across cultures speak to their infants and children in a distinct style compared to how they speak to adults. Infant-directed (ID) speech is characterized by higher pitch, increased pitch range, shorter utterances, and longer pauses compared to adult-directed (AD) speech (Bergeson & Trehub, 2002; Fernald, 1991, 1992; Fernald & Simon, 1984; Papoušek, Papoušek, & Bornstein, 1985). Both mothers and fathers change their speaking style when talking to infants (Fernald, Taeschner, Dunn, Papoušek, de Boysson-Bardies & Fukui, 1989). Moreover, ID speech styles are similar across many languages such as French, Italian, German, Japanese, Mandarin, Swedish, and Russian (Fernald & Simon, 1984; Grieser & Kuhl, 1988; Jacobson, Boersma, Fields, & Olson, 1983; Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova,, Sundberg, & Lacerda., 1997).

Infants also show attentional and affective preferences for this style of speech. Very young infants will look longer at a visual stimulus in response to ID speech compared to AD speech (Cooper & Aslin, 1990, 1994; Fernald, 1985; Pegg, Werker, & McLeod, 1992). Infants also produce appropriate affective signals such as smiling in response to female ID speech rather than AD speech (Fernald, 1993; Werker & McLeod, 1989).

In turn, maternal speech is sensitive or responsive to several infant factors, including age (Kitamura & Burnham, 2003; Stern, Spieker, Barnett, & MacKain, 1983), linguistic skill (Burnham, Kitamura, & Vollmer-Conna, 2002; Fernald & Mazzie, 1991; Kuhl et al., 1997; Ratner, 1984), and affective or social context (Burnham, Kitamura, Vollmer-Conna, 2002; Trainor, Austin, & Desjardins, 2000). Caregivers' simulations of ID speech in the absence of their infants are easily differentiated from infant-present speech (Jacobson et al., 1983). Thus, the affective and cognitive characteristics of infants themselves contribute greatly to caregivers' communication style, even though the caregivers may not be consciously aware of fine-tuning their performances.

If we assume that infants' response to ID speech encourages caregivers' continued use of this vocal register, it would be important to determine if caregivers with hearing-impaired infants decrease their use of ID speech when they discover their infants are not responding to the auditory information.

Previous researchers have shown that when NH mothers first learn of their child's hearing loss they tend to increase their use of vocal exaggeration, but over time such vocal exaggeration decreases (Wedell-Monnig & Lumley, 1980). NH mothers who have hearing-impaired children tend to be more controlling and less responsive than NH mothers who have NH children (Cheskin, 1981; Goss, 1970; Henggeler & Cooper, 1983), and tend to repeat utterances rather than expand on them when speaking to hearing-impaired children compared to when speaking to NH children (Cross, Nienhuys, & Kirkman, 1985; Nienhuys, Cross, & Horsborough, 1984). Interestingly, mothers' speech to hearing-impaired children was also more similar to speech directed to NH children of the same linguistic age than to NH children of the same chronological age (Cross et al., 1985; Nienhuys et al., 1984). Finally, NH mothers tend to produce fewer and less complex verbal utterances but more nonverbal attention-getting behaviors in interactions with hearing-impaired infants and children compared to interactions with NH infants and children (Goldin-Meadow & Saltzman, 2000; Koester, Brooks, & Karkowski, 1998; Koester, Karkowski, & Traci, 1998).

It also appears that hearing-impaired infants and children behave differently than NH infants and children when interacting with their NH mothers. Koester (1995) found that 9-month-old infants with hearing loss did not actively elicit their mothers' attention by means of smiling, greeting, or reaching, in contrast to NH 9-month-olds. Instead, the hearing-impaired infants displayed more self-comforting and repetitious motor behaviors than NH infants. Other studies have found that hearing-impaired children are more passive and less responsive than NH children when interacting with their NH mothers (Henggeler & Cooper, 1983; Wedell-Monnig & Lumley, 1980). These studies suggest that hearing loss does have an effect on infants' and children's interactions with their caregivers.

The findings from these studies are both clinically and theoretically important because maternal responsiveness and sensitivity to their infants and children, particularly those with hearing loss, has been linked to the development of cognitive and linguistic skills (Hart & Risley, 1995; Kaplan, Bachorowski, Smoski, & Hudenko, 2002; Liu, Kuhl, & Tsao, 2003; Meadow-Orlans & Spencer, 1996; Pressman, Pipp-Siegel, Yoshinaga-Itano, & Deas, 1999; Spencer & Meadow-Orlans, 1996). For example, NH children who heard more parental utterances between the ages of 11 and 18 months were more likely to have much better language skills at ages 3 and 8 years than those children who heard fewer parental utterances during infancy (Hart & Risley, 1995). Moreover, Liu and colleagues (2003) found that maternal vowel clarity was positively correlated with 6- to 8-month old and 10- to 12-month old NH infants' speech discrimination abilities. NH 4-month-old infants display better learning skills in response to ID speech of nondepressed mothers compared to the much less exaggerated ID speech of depressed mothers (Kaplan et al., 2002). Finally, maternal sensitivity and responsiveness predicts language gain and representational play in hearing-impaired children (Pressman et al., 1999; Spencer & Meadow-Orlans, 1996).

With the recent FDA expansion of CI criteria to include profoundly deaf infants, it is critical to assess the changes in mothers' speech and singing to infants as they acquire auditory skills via CIs, compared to mothers' speech and singing to NH infants. Despite the growing support for effects of early experience on the speech perception skills in hearing-impaired infants and children with CIs (e.g., Houston, Ying, Pisoni, & Kirk, 2003; Miyamoto, Kirk, Robbins, Todd, Riley, & Pisoni, 1997; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000), there has been very little research on the nature of the input these infants and children receive on a daily basis from their caregivers. Detailed investigation of caregivers' communicative interactions with their hearing-impaired infants who use CIs is needed to determine the optimal input for the development of language and other complex cognitive abilities.

To our knowledge, there has been only one published study on the early linguistic experience of hearing-impaired children with CIs. In a recent study from our laboratory, Stallings, Kirk, Chin, and Gao

(2000) found that parents' familiarity with uncommon words was positively correlated with the vocabulary and language skills of their hearing-impaired children with CIs. However, no research has assessed the acoustic characteristics, such as pitch level, in caregivers' speech and singing to hearing-impaired infants and children with CIs. The purpose of the current study was to determine whether NH mothers of NH infants and mothers of hearing-impaired infants with CIs use similar ID vocal styles.

## Method

### Participant Characteristics

NH mothers of implanted infants (N=6) were recruited from the clinical population at the Indiana University School of Medicine, Department of Otolaryngology – Head and Neck Surgery, and NH mothers of NH infants (N=6) were recruited from the local community. Table 1 shows demographic data across individual infants. The mean age of the CI infants was 26.9 months, the mean age at stimulation was 15.6 months, and the mean duration of CI use (i.e., "hearing age") was 11.4 months. Mean age of the NH infants was 7.6 months. All mothers were reimbursed $10 per visit.

| Hearing Status | Gender | Chronological Age (months) | "Hearing Age" (months) |
|---|---|---|---|
| NH116 | F | 10.30 | 10.30 |
| NH141 | F | 8.45 | 8.45 |
| NH147 | F | 8.49 | 8.49 |
| NH151 | F | 9.90 | 9.90 |
| NH170 | M | 4.05 | 4.05 |
| NH171 | F | 4.51 | 4.51 |
| CI03 | F | 23.85 | 11.64 |
| CI08 | F | 37.07 | 17.99 |
| CI10 | M | 23.00 | 12.63 |
| CI12 | M | 23.16 | 6.18 |
| CI16 | M | 17.5 | 3.51 |
| CI21 | F | 37.04 | 16.12 |

**Table 1.** Infant demographic information. Note that chronological age of normal-hearing infants is the same as their "hearing age."

### Procedure

We digitally recorded mothers speaking to their infants or an experimenter in a double-walled copper-shielded sound booth (IAC). In the ID speech condition, we asked mothers to sit with their child on a blanket or a chair, whichever option was most comfortable for them. We also provided the same group of quiet toys for all mother-child dyads. Mothers were instructed to speak to their child as they normally would at home. In the AD speech condition, an experimenter conducted a short interview with each mother. The order of ID and AD performances was counterbalanced across mothers. Mothers' speech was recorded by a hypercardioid microphone (Audio-Technica ES933/H), powered by a phantom power source. The microphone was linked to an amplifier (DSC 240) and a digital/audio tape recorder (Sony DTC-690). We also videotaped the recording sessions using a digital camera (Sony DCR-TRV 120/TRV 320).

Acoustic features known to characterize maternal ID speech were analyzed using Praat speech analysis software (Boersma & Weenink, 1996): average fundamental frequency (Hz), minimum fundamental frequency (Hz), maximum fundamental frequency (Hz), utterance duration (s), and duration of pauses between utterances (s). All features were measured for each utterance in a two-minute speech sample in both ID and AD conditions and then averaged across utterances. An utterance was defined in this study as a complete sentence or a complete thought.

## Results

Figure 1 shows the average pitch results in NH mothers' speech to NH infants, hearing-impaired infants with CIs, and a NH adult. In terms of average pitch level, a 2 (speech type: ID vs. AD) x 2 (NH infant vs. hearing-impaired infant with CI) repeated measures ANOVA revealed a significant main effect of speech type (*F*(1, 10) = 65.19, *p* < .0001). There was no effect of hearing status, and no interaction between speech type and hearing status. Pitch was higher in mothers' speech to infants than to an adult experimenter, regardless of their infant's hearing status.

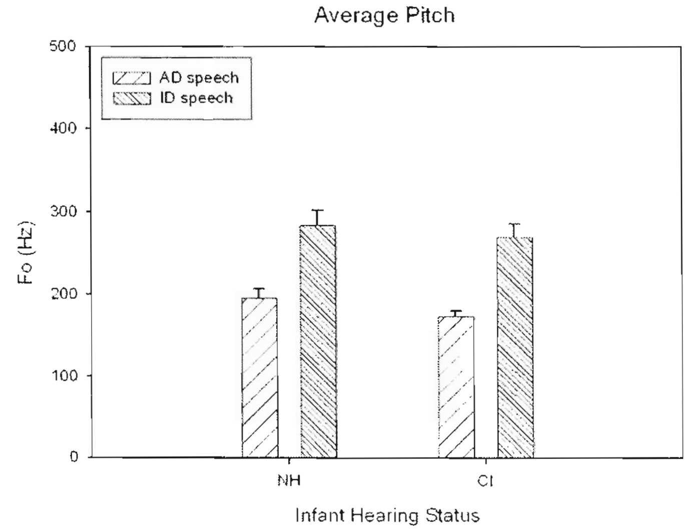


**Figure 1.** Average pitch in normal-hearing mothers' speech to normal-hearing infants, hearing-impaired infants with cochlear implants, and a normal-hearing adult. Error bars represent standard error.

Figures 2 and 3 show the minimum and maximum pitch results in NH mothers' speech to NH infants, hearing-impaired infants with CIs, and a NH adult. We found a significant main effect of speech type (minimum pitch: *F* (1, 10) = 67.12, *p* < .0001; maximum pitch: *F*(1, 10) = 37.67, *p* < .0001). Again we found no effect of hearing status and no interaction between speech type and hearing status. Both minimum and maximum pitch levels were higher in mothers' ID speech compared to mothers' AD speech, regardless of infant hearing status.

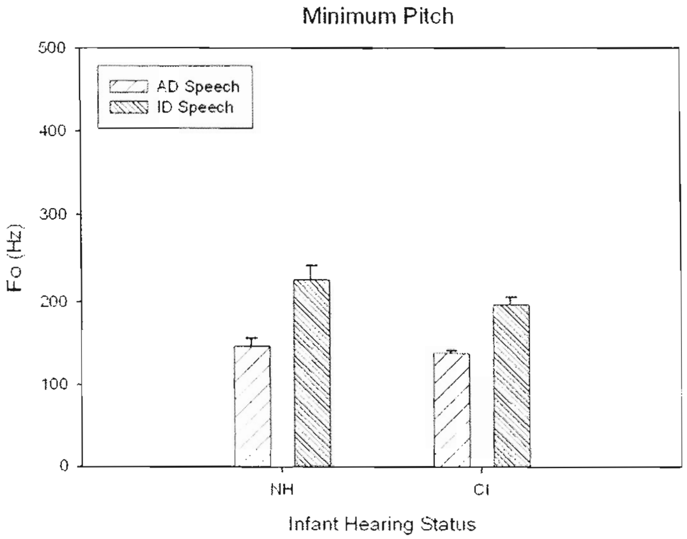


**Figure 2.** Minimum pitch averaged across normal-hearing mothers' speech to normal-hearing infants, hearing-impaired infants with cochlear implants, and a normal-hearing adult. Error bars represent standard error.

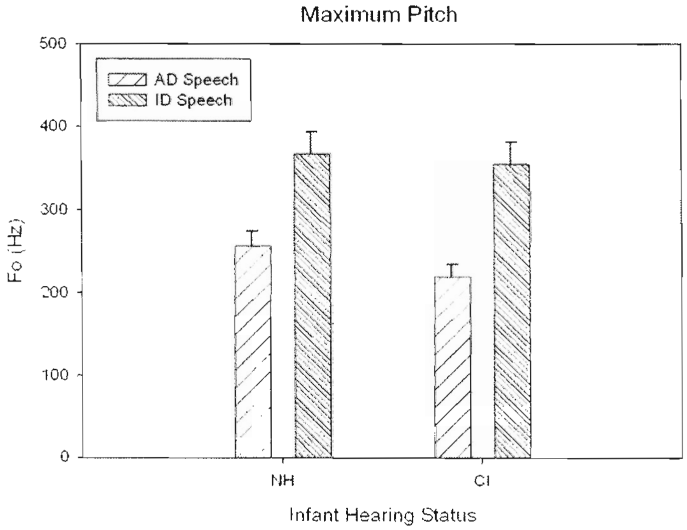


**Figure 3.** Maximum pitch averaged across normal-hearing mothers' speech to normal-hearing infants, hearing-impaired infants with cochlear implants, and a normal-hearing adult. Error bars represent standard error.

Figures 4 and 5 show the average utterance durations and the average duration of pauses between utterances in NH mothers' speech to NH infants, hearing-impaired infants with CIs, and a NH adult. We performed 2 (speech type: ID vs. AD) x 2 (NH infant vs. hearing-impaired infant with CI) repeated measures ANOVAs on the measures of utterance duration and pause duration. One NH mother/NH infant dyad was excluded from this analysis due to experimenter error. We found a significant main effect of speech type for both utterance duration ($F$ (1, 10) = 39.90, $p$ < .0001) and pause duration ($F$ (1, 9) = 9.76, *p* = .01). There was no main effect of hearing status for either duration measure. Although there was no significant interaction between speech type and hearing status for utterance duration, this interaction approached significance for the measure of pause duration (*F* (1, 9) = 4.20, *p* = .07). Mothers' utterances were shorter in duration when speaking to their infant than to an adult experimenter, regardless of infant hearing status. On the other hand, pauses between utterances were longer when directed to infants than adults, and this difference was much more pronounced in mothers of NH infants compared to mothers of hearing-impaired infants with CIs.

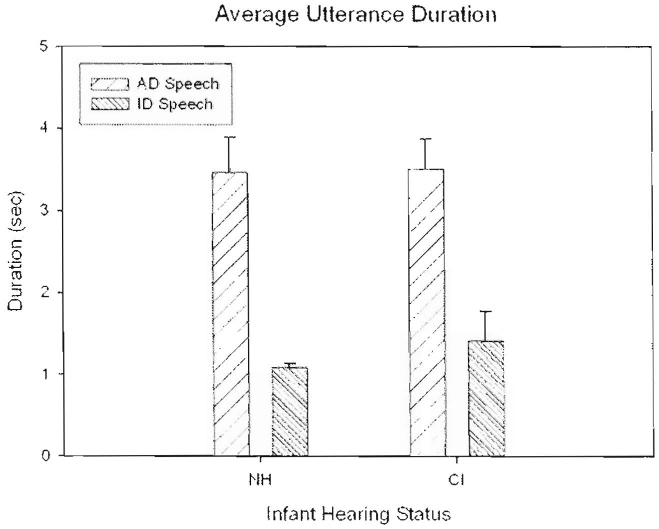


**Figure 4.** Utterance duration averaged across normal-hearing mothers' speech to normal-hearing infants, hearing-impaired infants with cochlear implants, and a normal-hearing adult. Error bars represent standard error.

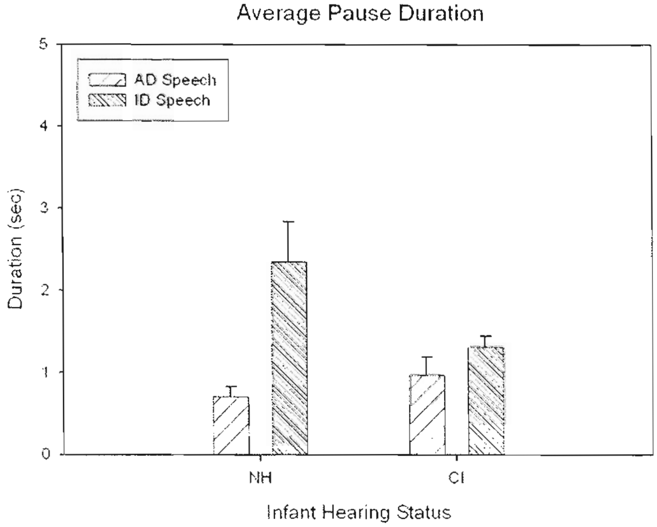


**Figure 5.** Pause duration averaged across normal-hearing mothers' speech to normal-hearing infants, hearing-impaired infants with cochlear implants, and a normal-hearing adult. Error bars represent standard error.

## Discussion

As expected from previous literature, the results of the present study revealed that the average, minimum, and maximum pitch levels were higher in ID speech than AD speech, and utterances were shorter in duration in ID speech than AD speech, regardless of infant hearing status. Average pause duration was longer in ID speech compared to AD speech when directed to NH infants, but not when directed to CI infants. These preliminary findings suggest that mothers speak to NH and CI infants in similar styles despite the chronological age difference in the two infant populations. That is, mothers' speech to infants is influenced by "hearing age" rather than chronological age.

These results are both clinically and theoretically significant. Not only has ID speech quality been linked to infants' development of language and other cognitive skills (Kaplan et al., 2002; Liu et al., 2003; Pressman et al., 1999; Spencer & Meadow-Orlans, 1996), but recent studies have also shown that very early auditory and audiovisual experiences and activities have significant effects on hearing-impaired children's development of speech perception and language skills (e.g., Bergeson, Pisoni, & Davis, 2003; Yoshinaga-Itano, Sedey, Coulter, & Mehl, 1998). These studies simply group hearing-impaired children with CIs into broad categories such as Oral Communication (i.e., auditory/verbal communication) versus Total Communication (i.e., simultaneously signed and spoken English) and Early-implanted versus Late-implanted. Although differences in speech and language performance across these groups are informative and clinically useful, there is still a great deal of variability within these broad categories and it is unclear exactly what types of specific experiences and activities children in each group are receiving. Thus, it is extremely important to investigate other factors that may underlie this variability. Further studies of mothers' vocal communication styles while interacting with their hearing-impaired infants and children with CIs should contribute greatly to understanding the large variability in speech and language outcome measures. Our findings could also be used to develop the optimal speech therapy tools to increase speech perception and language performance in hearing-impaired infants and children who receive CIs.

## References

Bergeson, T.R., Pisoni, D.B., & Davis, R.A.O. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review, 103,* 347-370.

Bergeson, T.R., & Trehub, S.E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science, 13,* 72-75.

Boersma, P. & Weenink, D. (1996). *PRAAT, a system for doing phonetics by computer* (Rep. No. 132). Amsterdam, The Netherlands: University of Amsterdam, Institute of Phonetic Sciences.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296,* 1435.

Cheskin, A. (1981). The verbal environment provided by hearing mothers for their young deaf children. *Journal of Communication Disorders, 14,* 485-496.

Cooper, R.P., & Aslin, R.N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584-1595.

Cooper, R.P., & Aslin, R.N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development, 65,* 1663-1667.

Cross, T.G., Nienhuys, T.G., Kirkman, M. (1985). Parent-child interaction with receptively disabled children: Some determinants of maternal speech style. In K. E. Nelson (Ed.), *Child Language: Volume 5* (pp. 247-290). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181-195.

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of Child Development, 8,* 43-80.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262-282). Cambridge: Cambridge University Press.

Fernald, A. (1993). Approval and disapproval: Infants' responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64,* 657-674.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477-501.

Goldin-Meadow, S., & Saltzman, J. (2000). The cultural bounds of maternal accommodation: How Chinese and American mothers communicate with deaf and hearing children. *Psychological Science, 11,* 307-314.

Goss, R. (1970). Language used by mothers of deaf children and mothers of hearing children. *American Annals of the Deaf, 115,* 93-96.

Grieser, D.L., & Kuhl, P.K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24,* 14-20.

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

Henggeler, S.W., & Cooper, P.F. (1983). Deaf child-hearing mother interaction: Extensiveness and reciprocity. *Journal of Pediatric Psychology, 8,* 83-95.

Houston, D.M., Ying, E.A., Pisoni, D.B., & Kirk, K.I. (2003). Development of pre-word-learning skills in infants with cochlear implants. *The Volta Review, 103,* 303-326.

Jacobson, J.L., Boersma, D.C., Fields, R.B., & Olson, K.L. (1983). Paralinguistic features of adult speech to infants and small children. *Child Development, 54,* 436-442.

Kaplan, P.S., Bachorowski, J.-A., Smoski, M.J., & Hudenko, W.J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Science, 13,* 268-271.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy, 4,* 85-110.

Koester, L.S. (1995). Face-to-face interactions between hearing mothers and their deaf or hearing infants. *Infant Behavior and Development, 18,* 145-153.

Koester, L.S., Brooks, L.R., & Karkowski, A.M. (1998). A comparison of the vocal patterns of deaf and hearing mother-infant dyads during face-to-face interactions. *Journal of Deaf Studies and Deaf Education, 3,* 290-301.

Koester, L.S., Karkowski, A.M., & Traci, M.A. (1998). How do deaf and hearing mothers regain eye contact when their infants look away? *American Annals of the Deaf, 143,* 5-13.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277,* 684-686.

Liu, H.-M., Kuhl, P.K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6,* F1-F10.

Meadow-Orlans, K.P., & Spencer, P.E. (1996). Maternal sensitivity and the visual attentiveness of children who are deaf. *Early Development and Parenting, 5,* 213-223.

Miyamoto, R.T., Kirk, K.I., Robbins, A.M., Todd, S., Riley, A., & Pisoni, D.B. (1997). Speech perception and speech intelligibility in children with multichannel cochlear implants. In I. Honjo & H. Takahashi (Eds.), *Cochlear implant and related sciences update. Advances in otorhinolaryngology* (pp. 198-203). Basel, Switzerland: Karger.

Nienhuys, T.G., Cross, T.G., & Horsborough, K.M. (1984). Child variables influencing maternal speech style: Deaf and hearing children. *Journal of Communication Disorders, 17,* 189-207.

Papoušek, M., Papoušek, H., & Bornstein, M.H. (1985). The naturalistic vocal environment of young infants: On the significance of homogeneity and variability in parental speech. In T.M. Field & N.A. Fox (Eds.), *Social perception in infants* (pp. 269-297). Norwood, NJ: Ablex.

Pegg, J.E., Werker, J.F., & McLeod, P.J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15,* 325-345.

Pressman, L.J., Pipp-Siegel, S., Yoshinaga-Itano, C., & Deas, A. (1999). Maternal sensitivity predicts language gain in preschool children who are deaf and hard of hearing. *Journal of Deaf Studies and Deaf Education, 4,* 294-304.

Ratner, N. B. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language, 11,* 557-578.

Spencer, P.E., & Meadow-Orlans, K.P. (1996). Play, language, and maternal responsiveness: A longitudinal study of deaf and hearing-impaired infants. *Child Development, 67,* 3176-3191.

Stallings, L.M., Kirk, K.I., Chin, S.B., Gao, S. (2000). Parent word familiarity and the language development of pediatric cochlear implant users. *The Volta Review, 102,* 237-258.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language, 10,* 1-15.

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.

Trainor, L.J., Austin, C., & Desjardins, R. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11,* 188-195.

Wedell-Monnig, J., & Lumley, J.M. (1980). Child deafness and mother-child interaction. *Child Development, 51,* 766-774.

Werker, J.F., & McLeod, P.J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43,* 230-246.

Yoshinaga-Itano, C., Sedey, A.L., Coulter, K.D., & Mehl, A.L. (1998). Language of early- and later-identified children with hearing loss. *Pediatrics, 102,* 1161-1171.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Effects of a Cochlear Implant Simulation on Immediate Memory Span in Normal-Hearing Adults[1]

**Rose A. Burkholder, David B. Pisoni, and Mario A. Svirsky[2]**

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

[1] This work was supported by NIH-NIDCD R01 Research Grants DC00111 and DC03937 and Training Grant T32 DC00012 as well as the American Hearing Research Foundation. We are grateful for the time and support from past and present Speech Research Laboratory members Brianna Conrey, Caitlin Dillon, Kuai Hinojosa, and Jeff Reynolds who participated in the piloting of this experiment. We also thank Shivank Sihna and Luis Hernandez for their technical assistance and expertise on this project.

[2] Indiana University School of Medicine and Purdue University Departments of Biomedical and Electrical Engineering.

# Effects of a Cochlear Implant Simulation on Immediate Memory Span in Normal-Hearing Adults

**Abstract.** Measures of immediate memory span were obtained from 25 normal-hearing adults who listened to an 8-channel, frequency shifted acoustic simulation of a cochlear implant. A short period of digit identification training was conducted before forward and backward digit spans were obtained under both processed and unprocessed conditions. As expected, forward and backward digit spans were significantly shorter when the stimuli were processed than when they were presented in unprocessed form. Participants' digit spans in unprocessed conditions and their accuracy in identifying digits in isolation were used to calculate predicted digit span scores in the processed speech condition. The observed digit spans in the transformed speech conditions did not differ significantly from the predicted digit spans. This result suggests that the decrease in immediate memory span is related to misidentification or misencoding of digits, due to the nature of the signal degradation, rather than to inefficient subvocal verbal rehearsal or serial scanning processes of the phonological representations in short-term memory under these unusual auditory conditions.

## Introduction

Several previous studies have found that recall performance on auditory memory span tasks is adversely affected in normal-hearing listeners when the auditory signals are degraded. For instance, decreased signal-to-noise ratios have been found to produce poorer memory performance in both working memory and short-term memory tasks (Dallett, 1964; Rabbitt, 1966, 1968; Pichora-Fuller, Schneider, & Daneman, 1995). In addition, Luce, Feustel, and Pisoni (1983) found that immediate memory capacity could also be reduced by using synthetic speech stimuli. Although they differed in methods, these studies lead to similar conclusions and interpretations by the authors. The main conclusion from these studies is that the perceptual difficulties in the tasks caused cognitive load to increase when items were encoded in memory. Increased cognitive load leads to the depletion or reallocation of resources normally used in other memory processes such as subvocal verbal rehearsal and serial scanning. Thus, as a result of reduced access to limited memory resources, processing capacity decreases under noisy or degraded auditory conditions.

Recently, working memory capacity and rehearsal processes have also been examined in individuals for whom auditory stimuli are always degraded because they use electrical stimulation provided by a cochlear implant to perceive speech and other sounds. In particular, several studies have reported that deaf children with cochlear implants have shorter forward and backward auditory digit spans compared to their normal-hearing peers (Pisoni & Geers, 2000). Deaf children with cochlear implants may have shorter memory spans simply because they have difficulty in correctly perceiving and encoding some of the test stimuli prior to recall. For instance deaf children may "recall" items that were not even presented in the list. These perceptually-driven errors are defined as item errors (Conrad, 1965).

However, deaf children using cochlear implants have also been found to perform poorly on immediate memory tasks even when the stimuli are not presented auditorily and recall does not require spoken responses. In a recent study, Cleary and colleagues (2001) demonstrated that deaf children using cochlear implants had shorter memory spans than their normal-hearing peers in a memory task that required the child to simply reproduce sequences of colored lights by manually pressing colored and illuminated response buttons. Their findings suggest that problems with memory processes other than the early encoding of auditory input may also contribute to the shorter digit spans of deaf children who use cochlear implants. Cleary and colleagues proposed that deaf children using a cochlear implant performed poorly even on a task of visual memory span because they were inefficient at coding visual sequences verbally and were slower at verbally rehearsing phonological representations of color names in working memory.

Measures of overt speaking rates suggest that deaf children with cochlear implants perform subvocal verbal rehearsal more slowly than age-matched, normal-hearing children. Speaking rate is widely accepted as an estimate of subvocal verbal rehearsal speed based on a series of studies in normal-hearing adults (Baddeley, Thompson, & Buchanan, 1975; Schweickert, Guentert, & Hersberger, 1990) and children (Cowan, et al., 1998; Hulme & Tordoff, 1989; Kail & Park, 1994) that demonstrated that overt speaking rate is linearly related to memory capacity. People who speak faster have longer digit spans. The explanation of this result is that as the rate of overt speech and subvocal verbal rehearsal speed increases, items can be refreshed more rapidly in the short-term memory store. The rapid cycling of verbally encoded items through the short-term memory store increases and facilitates recall of items in immediate memory (Baddeley et al., 1975).

Recently, Pisoni and Cleary (2003) reported that the speaking rates of deaf children using cochlear implants were strongly correlated with their digit spans. Children with the fastest speaking rates had the longest memory spans. Deaf children with cochlear implants speak more slowly than their normal-hearing peers, which may further explain why their digit spans are shorter than the digit spans of normal-hearing children (Burkholder & Pisoni, 2003; Pisoni & Cleary, 2003).

In addition to measuring subvocal verbal rehearsal processes in deaf children with cochlear implants, serial scanning speed has also been studied in this population. According to the recent work of Cowan (1999), interword pauses in immediate serial recall tasks reflect serial scanning, which is the process in which digits in the list are retrieved and scanned in serial order until the next item to be recalled is located (Sternberg, 1966). In a recent study, Burkholder and Pisoni (2003) measured interword pause durations during the digit span recall tasks in deaf children using cochlear implants and age-matched, normal-hearing children. They found that the interword pauses of the deaf children using cochlear implants were nearly twice as long as the interword pauses of the normal-hearing children.

The findings of Burkholder and Pisoni (2003) suggest that, in addition to having slower subvocal verbal rehearsal processes, deaf children with cochlear implants are also slower at serially scanning and retrieving verbal items (digits) in short-term memory. Taken together, the slowed subvocal verbal rehearsal speeds and slower serial scanning processes are both likely to contribute to the shorter memory spans of deaf children using cochlear implants. Thus, in addition to the initial encoding problems that lead to item errors by deaf children using cochlear implants, memory processing problems also contribute to their shorter memory spans. In contrast to initial auditory encoding, verbal rehearsal and serial scanning processes play a critical role in correctly maintaining the serial order of items in memory. Thus, problems associated with subvocal verbal rehearsal and serial scanning are likely to lead to order errors in immediate serial recall or the inability to maintain and recall the correct sequence of items in memory (Conrad, 1965; Gupta, 2003). It seems evident then that a primary problem facing researchers examining memory capacity in deaf children with cochlear implants or any other population of hearing-impaired listeners is delineating item errors from order errors in auditory memory span tasks.

However, unlike earlier studies measuring memory span in normal-hearing adults and children, it is impossible to measure the memory spans of deaf children using cochlear implants both before and after being exposed to the degraded auditory input that they receive through a cochlear implant. It is also difficult to determine exactly how much of an impact the degraded auditory input or item errors have on memory capacity. Although there is evidence for subvocal verbal rehearsal and serial scanning problems in deaf children using cochlear implants, the magnitude of these problems cannot be reliably measured unless order errors are first separable from pure encoding errors.

However, it is possible to observe directly how memory span is affected in normal-hearing listeners who are exposed to auditory stimuli modeled after a cochlear implant's unique auditory input. To measure how memory capacity in normal-hearing listeners is influenced by listening to stimuli similar to a cochlear implant, Eisenberg and her colleagues (2000) used an acoustic simulation of a cochlear implant. Normal-hearing adults and children completed a digit span recall task in clear auditory conditions and while listening to stimuli filtered into eight different frequency bands designed to simulate output from a cochlear implant. As expected, both adults and children performed significantly worse on digit span recall when the digits were processed by the simulator.

Although the authors found weak correlations between digit spans, word and sentence recognition, and speech feature discrimination under the degraded auditory conditions, they did not attempt to determine whether item errors were entirely responsible for this decrease in digit span or whether order errors may have also been induced while listening to the degraded stimuli. In other words, although standard clinical tests of spoken word recognition and speech perception abilities were administered to these normal-hearing listeners to assess their general ability to understand speech through the cochlear implant simulation, insufficient data were collected on how accurate the listeners were at identifying test stimuli in isolation (Eisenberg et al., 2000). Measuring stimulus identification in isolation is necessary in order to estimate perceptual accuracy of each stimulus item in the absence of the additional cognitive load associated with the immediate serial recall task. Thus, a pretest evaluating the ability to recognize digits in their processed form, before being embedded in a memory task, is needed to determine more precisely the magnitude of item errors in digit span recall of normal-hearing listeners exposed to an acoustic simulation of a cochlear implant. Although the authors did conduct a pretest to ensure that all participants identified the degraded digits in isolation, they only used one presentation of each degraded digit. Using such a limited number of stimuli to test identification in isolation may have resulted in an overestimation of how accurate participants really were at the task.

In the present study, we utilized a stimulus pretest to predict and determine the contributions of perceptual errors to normal-hearing adults' digit spans while listening to an acoustic simulation of a cochlear implant. The acoustic simulation used in this study was similar to Eisenberg et al.'s (2000) with the exception that a basalward frequency shift was also included to make the task even more difficult by increasing the frequency or pitch of the stimuli. We expected that, similar to previous studies testing memory in degraded or noisy auditory conditions, memory capacity would decrease substantially (Luce et al., 1983; Rabbitt, 1966, 1968). In addition, we predicted that the nature of this decrease will primarily be accounted for by item errors due to the degraded nature of the auditory stimuli. However, we also expected that the perceptual difficulty of this task would lead to an increase in serial order errors in normal-hearing adults despite their intact memory processing abilities.

## Method

### Participants

Twenty-five undergraduate students enrolled at Indiana University participated in this study. They received partial course credit for the introductory Psychology class. The group of participants included 18 females and 7 males. A brief hearing screening was administered by the first author to determine whether the participants' hearing was within normal limits. Using a standard, portable pure-tone audiometer (Maico Hearing Instruments, MA27) and headphones (TDH-39P), each participant was tested at 250, 500, 1000, 2000, and 4000 Hz at 20 dB first in the right ear and then in the left ear. None of the participants showed any evidence of a hearing loss. All participants also reported that they were monolingual native speakers of American English and had no prior history of speech, language, hearing, or attentional disorders at the time of testing.

### Stimuli and Materials

**Simulation Strategy.** All auditory stimuli were processed offline using a personal computer equipped with DirectX 8.0 and a Sound Blaster Audigy Platinum sound card. The signal processing procedure used for the cochlear implant simulation was adapted from real-time signal processing methods designed by Kaiser and Svirsky (2000). The signal processing strategy used bandpass filtering with a cutoff frequency of 1200 Hz. Eight filters were then used to simulate the speech processing capabilities of an 8-channel cochlear implant. The output of each filter modulated noise bands of a higher frequency range than the initial analysis filters. This mismatch was designed to model the natural frequency mismatch that occurs when the electrodes of a cochlear implant are shifted more basalward in the cochlea. The basalward shift used in this model was equivalent to a 0.5 mm shift within the cochlea.

**Stimuli.** Several familiar nursery rhymes (i.e. *Twinkle, Twinkle Little Star*; *Jack and Jill*) were used to familiarize the listeners with the processed speech. The nursery rhymes included in the familiarization phase are included in the Appendix. While listening to these passages, the participants were provided with the written text of the nursery rhymes so they could read along with them. The stimuli used for pretest digit identification included isolated utterances of the digits 1 through 9. The digit span lists were taken from the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991) and Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997). All stimuli used for familiarization and the digit span training task were recorded digitally in a sound attenuated booth by the first author using an individualized version of a speech acquisition program (Dedina, 1987; Hernandez, 1995). The stimuli were sampled and digitized at 22,050 Hz with 16-bit resolution and then equated for amplitude using the Level16 software program (Tice & Carrell, 1998).

### General Procedures

**Familiarization Task.** Prior to testing, a sound level meter (Triplett Model 370) was used to adjust the amplitude of the stimuli to 70 dB SPL. In order to familiarize the participants with the processed speech, the five nursery rhymes were played through a high-quality tabletop loudspeaker (Cyber Acoustics MMS-1) while the participants read along with the written text. Nursery rhymes were chosen for familiarization, because they are well known to most listeners and they have a distinctive prosody and rhythm that may assist in recognizing the degraded stimuli as real speech. In addition, these stimulus materials were chosen because the experimental procedure was specifically designed for future use with normal-hearing children. Thus, the utilization of these tasks in adults served as a pilot study for a companion project that will be carried out with normal-hearing children.

**Stimulus Pretest.** Prior to obtaining any digit span measures, the identification of processed digits in isolation was measured and feedback was provided. The digits 1 through 9 were each played five times in random order which resulted in the presentation of 40 digits. Participants indicated verbally what digit they thought they heard on each trial. If the response was correct, the experimenter simply affirmed that the correct answer had been given. However, if the response was incorrect, the experimenter played the correct response.

**Digit Spans.** Following digit identification training, participants completed several digit span tasks under both normal and processed auditory conditions. The order of presentation of the processed and unprocessed conditions was counterbalanced over participants. However, following the traditional digit span administration procedures in the WISC manual, backward digit span was always administered after forward digit span.

## Scoring Procedures

**Observed Digit Span Scores.** The digit span tasks were not scored according to traditional methods that quantify digit span in terms of how many lists of digits are correctly repeated. Rather, the results from both the forward and backward digit span tasks, conducted in unprocessed conditions, were used to obtain two different scores. The first score derived from the digit span tasks was a measure of the average length of the two longest lists of digits that each participant could repeat correctly in both the unprocessed or processed conditions. This score reflected a participant's memory capacity.

**Predicted Digit Spans.** A second score was also calculated using an algorithm that combined memory capacity in unprocessed conditions and the accuracy of identifying digits in isolation when they were processed. This score provided an estimate of the predicted digit span in the transformed speech condition. For example, if a participant had a digit span of 3 in the unprocessed digit condition and in the pretest only identified digits correctly 60% of the time, for each digit, there would be a 40% chance that it will be misheard and recalled incorrectly. Thus, there is a 40% chance of missing the first digit of the first list and having a digit span of 0. This "predicted" digit span score is meant to reflect the degree of digit span recall problems strictly related to the inability to correctly identify digits. These errors will have a negative effect on the digit span measured in processed speech conditions. The size of the effect of perceptual errors on digit span can be estimated mathematically. Equation 1 illustrates the basic steps taken to calculate the predicted digit span score given the example that the digit span in unprocessed conditions is 3 and digits were incorrectly identified in the pretest 40% of the time. The calculation of predicted digit span through this method was achieved using a MatLab script.

a) the probability of missing the first digit of a list and having a digit span of 0 is expressed as: (1)

   i. *.40*

b) the probability of correctly recalling the first digit but misidentifying the second digit is expressed as:

   i. *(.60)(.40)*

c) the probability of repeating the first two digits correctly and incorrectly recalling the third is:

   i. *(.60)(.60)(.40)*

d) and the probability of correctly recalling all digits is:

   i. *(.60)(.60)(.60)*

e) therefore, taking into account all the probabilities of having a digit span of 0, 1, 2, 3, etc., the predicted digit span is:

   i. $.4(0) + (.6)(.4)(1) + ((.6)^2(.4)(2) + (.6)^3(3)) = 1.176$

Thus, in this example, the listener has an observed digit span of 3 in unprocessed conditions but has a predicted digit span of only 1.176 in the processed condition due to the effect of misidentification of digits at the time of initial encoding.

**Digit Span Error Scoring.** In order to determine whether item or encoding errors were more numerous in digit span recall conducted in the processed speech condition, the participants' digit span errors in all incorrectly recalled lists were classified according to error type. In addition to classifying the item and order errors, omission and combination errors were also recorded. An item error was recorded if a digit(s) that did not appear in the original list was recalled in the place of an intended digit(s) (Example: *6, 1, 5, 8* repeated as "6, 9, 4, 8"). Order errors included responses in which all the correct digits of a list were repeated but in an incorrect order or in a combination of incorrect orders (Example: *6, 1, 5, 8* repeated as "5, 6, 1, 8"). An error of omission was scored when one or more numbers were omitted from the list. Errors in digit span recall that consisted of several different types of errors were considered to be combination errors.

A 2 x 2 x 4 factorial design was used to determine in which processing of the conditions the four types of errors were most likely to be committed by the participants (unprocessed or processed and recall forward or backward). The error rate of each type of error committed by the participants was the dependent variable. The error rate was calculated by dividing the raw number of errors made by the total number of possible errors for each recall condition. The total number of errors possible was determined by the number of lists completed and the number of digits administered in each list. According to this method, it is assumed that each possible digit could be a source of error.

Error rates were expressed as proportions rather than using raw scores in order to equate for the different number of possible errors in the two conditions. For instance, when using the standard digit span administration procedures, a greater number of errors are possible in forward digit span recall, making it necessary to equate the conditions according to how many lists were administered and how many possible errors could have been made. More lists were administered in the forward digit span condition, and typically participants progress through more lists, resulting in more opportunities for the participants to commit errors. In addition, participants were generally able to complete more lists in the unprocessed condition, resulting in a greater number of errors possible during recall under these conditions.

## Results

As expected, both forward and backward digit span recall were significantly worse under the processed speech conditions. Figure 1 illustrates the digit spans obtained in the processed and unprocessed speech conditions. The mean forward digit span obtained under the processed speech conditions (*M* = 5.78, *SD* = 1.13) was nearly one digit shorter (*t*(24) = 2.62, *p* = .015) than the mean forward digit span observed in the unprocessed conditions (*M* = 6.36, *SD* = .97). Similarly, backward digit spans were nearly one digit shorter (*t*(24) = 3.41, *p* = .002) under processed conditions (*M* = 4.22, *SD* = 1.49) than in unprocessed conditions (*M* = 5.02, *SD* = 1.42).

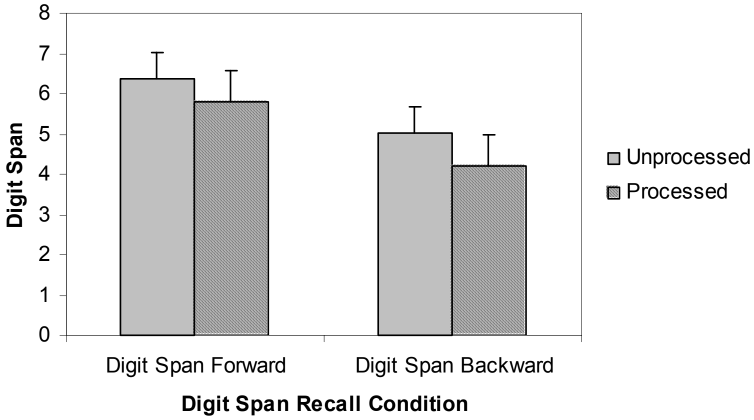


**Figure 1.** Mean digit span scores in the processed and unprocessed speech condition. Error bars represent the standard error of the mean.

Prior to calculating the predicted digit spans, the data obtained from the digit pretest were analyzed. Scores on the isolated digit pretest ranged from 76%-100% with nearly half of the participants capable of identifying the digits with 100% accuracy. Figure 2 displays a frequency distribution of the digit identification scores obtained when digits were played in isolation under processed speech conditions. Using each participant's processed digit identification accuracy and the unprocessed digit spans of each participant, predicted processed digit spans were calculated using the procedures summarized above.

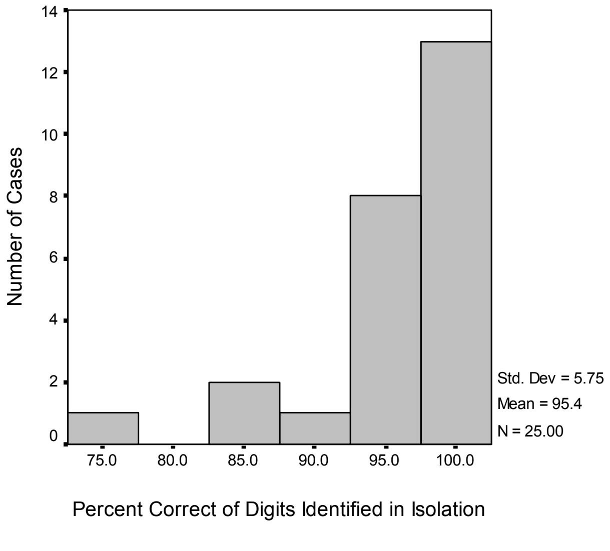


**Figure 2**. Frequency distribution of the digit identification scores.

A paired-samples t-test revealed no significant difference between the forward and backward predicted digit spans and the observed digit spans in processed speech conditions. Figure 3 illustrates the mean predicted and processed digit spans. Predicted forward digit spans (*M* = 5.50, *SD* = 1.28) were lower than the observed digit spans (*M* = 5.78, *SD* = 1.28), however this difference did not reach significance (*t*(24) = .703, *p* = .489). Predicted backward digit spans (*M* = 4.38, *SD* = 1.42) were slightly higher than the observed backward digit spans (*M* = 4.22, *SD* = 1.49) but this difference also did not reach significance either (*t*(24)= 1.19, *p* = .246).

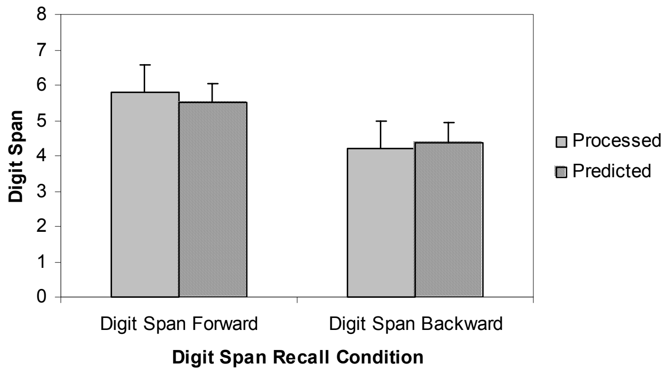


**Figure 3.** Mean predicted digit span scores and observed digit spans in the processed speech condition. Error bars represent the standard error of the mean.

Figure 4 shows a graph of the proportion of errors and type of digit span errors committed by the participants in both the forward and backward recall conditions. A univariate ANOVA revealed a main effect of error type (*F*(3, 416) = 8.76, *p* = .000) and processing condition (*F*(1, 416) = 10.16, *p* = .002). In addition, there was an interaction effect between the processing condition and error type (*F*(3, 416) = 7.17, *p* = .000). The main effect of recall direction approached significance (*F*(1, 416) = 3.29, *p* = .071) and no interaction effects involving recall condition were found despite the finding that processing errors did increase more in the backward recall condition when the auditory stimuli were processed.

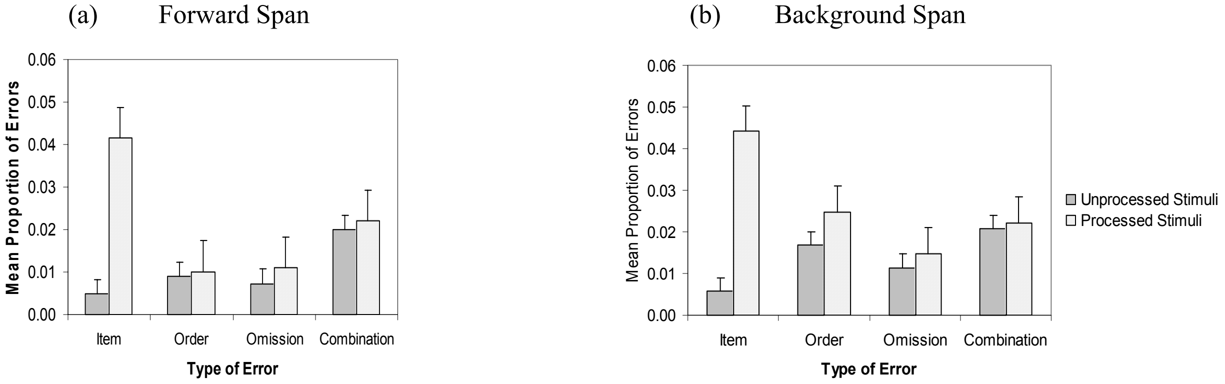


**Figure 4.** Mean proportion of errors committed by participants in the digit span task under processed speech condition in (a) forward and (b) backward recall conditions. Error bars represent the standard error of the mean.

## Discussion

As expected, the results demonstrate that immediate memory span for digits declined when the auditory stimuli were processed by a simulator that was designed to model the signal transmitted to cochlear implant users. The decline in digit span in these normal-hearing participants appears to be primarily due to item errors at the time of encoding, because the observed digit spans under processed speech conditions were no worse than would be expected after accounting for accuracy in identifying digits in isolation. In addition, using a classification system to identify the digit span recall errors, we found that item errors increased the most under the simulator. Although the proportion and number of order errors increased in the backward digit span task while remaining unaffected in forward digit span recall, this difference was not significant. Previous research suggesting that memory resources may be recruited to assist in auditory processing during memory tasks presented in noise and contribute to decreased memory span performance was not supported by the results obtained in this study (Luce et al., 1983; Pichora-Fuller et al., 1995).

However, the memory task used in the present study differed in several ways from the previous studies. One difference that may have contributed to the lack of a primary influence of order errors in this task was that it involved the recall of a small set of digits rather than recall of words. In addition, the words to be recalled in previous studies were embedded within sentences in which the listeners were tested for comprehension, making it a more traditional working memory task rather than a serial recall task (Daneman & Carpenter, 1980). The additional processing demands required in these tasks and the less limited range from which auditory stimuli could be selected makes them more challenging than the serial digit recall task used in the present study. However, backward digit span recall is considered to be a more complex cognitive task requiring additional executive planning, controlled attention, verbal rehearsal, and recall strategies than forward span (Li & Lewandowsky, 1995; Thomas, Milner, & Haberlandt, 2003). Therefore, increasing difficulty of the task through a cochlear implant simulation utilizing a frequency shift was expected to cause more processing difficulties as well. Order errors occurred more often in the backward recall condition, but these differences were not significant.

Any evidence of more order errors due to the degradation of the auditory stimuli is noteworthy, because the normal-hearing adults used in the present study undoubtedly had typical working memory processing skills. However, research suggests that deaf children using cochlear implants may also have poorer memory processing skills because of their slower subvocal verbal rehearsal speeds and serial scanning rates (Burkholder & Pisoni, 2003; Cleary et al., 2001). It seems then that testing deaf children with cochlear implants in an auditory digit span task would certainly contribute further to the incidence of order errors and not only elicit item errors that reflect their hearing impairment and speech and word recognition skills. However, because of the substantial and dominant role that item errors played in the decreased digit span performance of these normal-hearing adults, it should still be assumed that pure perceptual errors may also underlie some of the difficulties that deaf children with cochlear implants have in digit span recall.

Unfortunately, however, we cannot determine what the memory spans of deaf children with cochlear implants would be had they not had a hearing impairment or can we derive what their predicted memory span would be when first exposed to the degraded auditory input. However, the ability to make these predictions about memory span in normal-hearing listeners, exposed to a cochlear implant simulation, could be a useful benchmark for further interpretations that can be made about memory processes in deaf individuals using cochlear implants. In addition, with further analyses of the verbal digit span responses of the deaf children with cochlear implants, an error categorization process similar to the one used in this study could be used to determine the proportion of item and order errors that contribute to their shorter memory spans (Burkholder & Pisoni, 2004).

Although the present study on memory span with normal-hearing adults was motivated by earlier research on immediate memory span in deaf children using cochlear implants, it is necessary to acknowledge several limitations that may be associated with these kinds of comparisons. For example, although the acoustic simulation of a cochlear implant used in this study does accurately represent how speech is processed by a cochlear implant, it is currently unknown what differences may occur in the way the auditory stimuli are processed and perceived at peripheral and cortical levels in normal-hearing and deaf populations. Neuroplasticity associated with a period of prolonged deafness followed by cochlear implantation has been documented extensively in both animals and humans (e.g., Leake, Hradek, & Snyder, 1999; Moore, Vollmer, Leake, Snyder, & Rebscher, 2002; Ponton, 2001) and may be a substantial source of the limitations involved in studies utilizing acoustic simulations of cochlear implants.

Additionally, it may not be appropriate to examine the effects of a simulation on memory by making comparisons between adults and children. Developmental differences related to the speed of subvocal verbal rehearsal and serial scanning processes could potentially confound the interpretation of results obtained by making comparisons between adults and children (Cowan, 1999; Cowan et al., 1998). However, other memory mechanisms and processes and their relationship to other cognitive tasks have been found to be similar between adults and children, making an examination of these relationships interesting in normal-hearing adults and deaf children using cochlear implants (Gupta, 2003). A final drawback in making comparisons between normal-hearing individuals' performance on tasks using an acoustic simulation of a cochlear implant and the performances of actual cochlear implant users is related to the large differences in the amount of exposure that each group of listeners has had with the auditory input. However, the lack of experience with such a degraded auditory input by normal-hearing listeners provides an ideal situation in which to further examine the time course of perceptual learning, adaptation, and cognitive performance while listening to the unusual stimuli for only a short period of time.

## References

Baddeley, A.D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Behavior, 14*, 575-589.

Burkholder, R.A., & Pisoni, D.B. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85*, 63-88.

Burkholder, R.A., & Pisoni, D.B. (2004). *Analysis of Digit Span Recall Error in Paediatric Cochlear Implant Users and Normal-Hearing Children.* Poster session presented at the European Symposium of Paediatric Cochlear Implantation, Geneva, Switzerland.

Cleary, M., Pisoni, D.B., & Geers, A. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing, 22*, 395-411.

Cleary, M., Dillon, C., & Pisoni, D. (2002). Imitation of nonwords by deaf children following cochlear implantation. *Annals of Otology, Rhinology, and Laryngology, 111*, 91-96.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning & Verbal Behavior, 4*, 161-169.

Cowan, N. (1999). The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology, 72*, 193-209.

Cowan, N., Wood, N., Wood, P., Keller, T., Nugent, L., & Keller, C. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology, 127*, 141-160.

Dallett, K.M. (1964). Intelligibility and short-term memory in the repetition of digit strings. *Journal of Speech and Hearing Research, 7,* 362-368.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Dedina, M.J. (1987). SAP: A speech acquisition program for the SRL-VAX. In *Research on Speech Perception Progress Report No. 13* (pp. 331-337). Bloomington, IN: Speech Research Laboratory, Indiana University.

Eisenberg, L., Shannon, R., Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America, 107*, 2704-2709.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology*, 56A, 1213-1236.

Hernandez, L.R. (1995). Current computer facilities in the Speech Research Laboratory. In *Research on Speech Perception Progress Report No. 13* (pp. 389-393). Bloomington, IN: Speech Research Laboratory, Indiana University.

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology, 47*, 72-87.

Kail, R., & Park, Y. (1994). Processing time, articulation time, and memory span. *Journal of Experimental Child Psychology, 57*, 281-291.

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 Proceedings of the IXth IEEE-DSP Workshop., Hunt, TX.

Leake, P.A., Hradek, G.T., & Snyder, R.L. (1999). Chronic electrical stimulation by a cochlear implant promotes survival of spiral ganglion neurons after neonatal deafness. *Journal of Comparative Neurology, 412*, 543-562.

Li, S-C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition 21*, 837-847

Luce, P.A., Feustel, T.C., & Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors, 25*, 17-32.

Moore, C., Vollmer, M., Leake, P., Snyder, R., & Rebscher, S. (2002). The effects of chronic intracochlear electrical stimulation on inferior colliculus spatial representation in adult deafened cats. *Hearing Research, 164*, 82-96.

Pichora-Fuller, M.K., Schneider, B.A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*, 593-607.

Pisoni, D.B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*, 106S-120S.

Pisoni, D., & 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, and Laryngology, 185*, 92-93.

Ponton, C.E.J. (2001). Of kittens and kids: Altered cortical maturation following profound deafness and cochlear implantation. *Audiology & Neuro Otology, 6*, 363-380.

Rabbitt, P. (1966). Recognition memory for words correctly heard in noise. *Psychonomic Science, 6,* 383-384.

Rabbitt, P. (1968). Channel-capacity, intelligibility, and immediate memory. *Quarterly Journal of Experimental Psychology, 20,* 241-248.

Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science, 1*, 74-77.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652-654.

Thomas, J., Milner, H., & Haberlandt, K. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science, 14*, 169-174.

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

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

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – III*. San Antonio, TX: The Psychological Corporation.

## Appendix

Stimulus materials used for familiarization with the acoustic model of a cochlear implant.

*Hickory, Dickory, Dock*
*Hickory, dickory, dock,*
*The mouse ran up the clock.*
*The clock struck one,*
*The mouse ran down.*
*Hickory, dickory, dock!*

*Jack and Jill*
*Jack and Jill went up the hill*
*To fetch a pail of water*
*Jack fell down and broke his crown,*
*And Jill came tumbling after*

*One, Two, Buckle My Shoe*
*One, two, buckle my shoe,*
*Three, four, knock at the door.*
*Five, six, pick up sticks,*
*Seven, eight, lay them straight.*
*Nine, ten, big fat hen.*

*Star Light*
*Star light, star bright*
*First star I see tonight*
*I wish I may, I wish I might*
*Have the wish I wish tonight*

*Twinkle, Twinkle*
*Twinkle, twinkle, little star,*
*How I wonder what you are!*
*Up above the world so high,*
*Like a diamond in the sky.*
*Twinkle, twinkle, little star,*
*How I wonder what you are!*

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Visual and Visual-Spatial Memory Measures in Children with Cochlear Implants[1]

**Miranda Cleary[2] and David B. Pisoni**

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

[1] Research supported by NIH-NIDCD Training Grant T32DC00012 and NIH-NIDCD Research Grant R01DC00111 to Indiana University. Work also supported by NIH-NIDCD Research Grant DC03100 to Central Institute for the Deaf. We would like to thank Dr. Ann Geers and her staff at Central Institute for the Deaf for making this project possible and also Caitlin Dillon and Rose Burkholder for their help in data collection.

[2] Miranda Cleary is currently an NIH post-doctoral research fellow in the Speech and Hearing Sciences Program at the Graduate Center of the City University of New York (mcleary@gc.cuny.edu).

# Visual and Visual-Spatial Memory Measures in Children with Cochlear Implants

**Abstract.** Recent findings suggest that children developing spoken language while using a cochlear implant (CI) perform more poorly than normal-hearing (NH) children on short-term visual/visual-spatial sequence memory tasks, particularly when verbal recoding and verbal rehearsal strategies for the visual/visual-spatial information are possible. The present study examined the performance of children with CIs using two measures from the Children's Memory Scale (CMS): memory for static patterns of dots and recognition memory for unfamiliar faces. The study included 31 eight- and nine-year old deaf children implanted before age five and 31 age- and gender-matched normal-hearing children. Published norms for normally-developing children were also available for comparison. On the dot locations task, the implanted children performed comparably to the published norms, but not as well as the normal-hearing comparison group. On the face recognition task, the CI group did more poorly than the published norms. Their scores were also lower than the normal-hearing comparison group. Although small, these differences reached statistical significance. Interestingly, face memory scores in the implant group showed a modest but significant positive correlation with short-term memory for sequences of auditory stimuli but not with short-term memory for sequences of visual stimuli. These results suggest that the pediatric CI users may have attempted some form of verbal recoding for these complex static (non-sequenced) visual/visual-spatial face stimuli. The degree to which children with CIs have developed age-typical verbal coding and rehearsal skills may influence how they process certain complex visual stimuli such as unfamiliar faces, sequences of colored light or other visual patterns that can be rapidly encoded using phonological representations.

## Introduction

Hearing-impairments in children have long been associated with delays in spoken language acquisition (Conrad, 1979; Furth, 1966; Mayberry, 1992). Of particular concern has been the case of the child who has experienced a severe to profound hearing loss since birth, making auditory sensory input unavailable during infancy and early childhood. In the study of these children, an understanding of their language-learning difficulties is an area of primary concern. In addition, however, our present understanding of human information processing suggests that it may be important to examine the performance of hearing-impaired children in certain tasks formerly believed to involve more "amodal," general cognitive processing. Many abilities related to learning and memory, for example, are now believed to be interconnected and closely coupled with linguistic knowledge and language skills (Adams & Gathercole, 2000; Montgomery, 2000). Also of interest is whether certain cognitive processing skills show *enhancement* attributable to hearing-impairment, perhaps due to greater reliance on sensory cues from the non-auditory sensory modalities. Several models of neural development suggest that compensatory strategies and neural reorganization arise when organisms encounter sensory deprivation. These models predict that enhanced skills may be observed for hearing-impaired individuals in domains such as visual processing (e.g., Bavelier, Tomann, Hutton, Mitchell, Corina, Liu, & Neville, 2000; Bellugi, O'Grady, Lillo-Martin, O'Grady-Hynes, Van-Hoek, & Corina, 1990).

The present study examined visual memory in children who were born with or acquired a profound hearing-impairment very early in life and who experienced a period of auditory deprivation before receiving a cochlear implant. Cochlear implants are a relatively new form of treatment for sensorineural hearing loss, available to children in the U.S. only since the early 1990s. Because of their relatively recent introduction and because most implant research has focused rather narrowly on basic audiological and language measures (Pisoni, 2000), there is relatively little data available on the long-term effects of auditory deprivation followed by cochlear implant use on the development of memory skills in this clinical population.

There does exist, however, sizable literature on memory and cognitive processing in non-implanted children with hearing-impairments. Although cochlear implants provide an input signal unlike that provided by conventional hearing aids, past findings on memory function in non-implanted hearing-impaired children can provide some insights regarding memory skills in children with CIs.

Not surprisingly, many studies have found that children with hearing-impairments tend to score more poorly than age-matched peers on auditory list memory tasks, even when care is taken to insure that the stimuli are familiar and audible to the hearing-impaired children (e.g., Conrad, 1972; 1979; Pisoni & Cleary, 2003). More interestingly, it has also been repeatedly observed that children with hearing-impairments tend to score more poorly than age-matched peers when they are asked to recall lists of stimuli presented via the visual modality and these stimuli readily lend themselves to spoken language labeling. This finding has been reported for stimuli such as printed letter names, printed numerals, orthographically presented words, drawings of common objects, and pictures of shapes (Conrad, 1979; Furth, 1966; Kelly & Tomlinson, 1976; Pintner & Paterson, 1917; Waters & Doehring, 1990).

Lower average levels of spoken language fluency in children with hearing impairments appear to contribute to these findings. Although manually-signed equivalents to verbal labeling and verbal rehearsal can be used by signing individuals to aid memory, effective use of such skills appears to be predicated on the stimuli being suitable for encoding using signs, and on the individual being a fluent user of a sign system (Bebko, 1984; Mayberry, 1992). Because most children who receive cochlear implants come from homes in which spoken language is used (most deaf children are born to hearing parents), fluency in a manual communication system is unlikely to develop spontaneously for these children.

Thus, because lower levels of language fluency are associated with poorer memory performance for stimuli that can be linguistically labeled, it is reasonable to expect that children with cochlear implants may, on average, display reduced performance on memory tasks involving readily namable visual stimuli, relative to children with more advanced language skills. Several recent studies have provided data supporting this hypothesis.

Cleary, Pisoni, and Geers (2001) examined list memory in eight- and nine-year old children with cochlear implants. All of the children had a congenital or early-acquired hearing-impairment, had undergone cochlear implantation before the age of 5 years, and had used an implant for at least four years. Normal-hearing children matched for age and gender were also studied for comparison. Cleary et al. compared performance across three different presentation conditions of a list memory task. In one condition, sequences of colored-lights were presented using a response box consisting of four backlit colored buttons. In a second condition, sequences of colored lights were presented in synchrony with auditorily presented color-names of the lights. Finally, in a third condition, only the auditory color-names were presented. In all conditions, the children were asked to reproduce the patterns by pressing the colored response buttons on the response box in the correct order. The stimulus lists differed from trial to trial and the child's ability to reproduce successively longer lists of items was assessed and compared across conditions.

Cleary et al. found that normal-hearing children scored significantly higher on average, than the children with cochlear implants in all three presentation conditions. Even when the stimulus patterns were visual-spatial "lights-only" sequences, the children with cochlear implants had more difficulty with the task than did the normal-hearing children. This result contrasted with data previously reported for congenitally deaf users of sign language who performed at least as well as normal-hearing students on a closely analogous lights-only list memory task (Tomlinson-Keasey & Smith-Winberry, 1990). Although their correlational analyses did not entirely support such a hypothesis, several aspects of Cleary et al's results suggested that the children with cochlear implants were attempting to use a verbal-labeling strategy to help them remember the sequences of visually-presented colored lights. Their decreased facility with such a verbal-labeling strategy may have contributed to the finding of lower scores in this lights-only condition for the CI group as compared to the NH group.

More recently, Dawson, Busby, McKay, and Clark (2002) reported data from pediatric cochlear implant users consistent with the findings of Cleary et al. (2001). Dawson et al. studied the list memory skills of 24 early-deafened children, ranging in age from 5 to 11 years, implanted in their preschool-age years, and having, on average, at least 4.5 years of cochlear implant experience. Twenty-four age- and gender-matched normal-hearing children were also tested as a comparison group. Dawson et al. compared performance across five different list memory conditions. One set of conditions used two auditory stimuli—the recorded words, "fish" and "dog." Children heard sequences of these two auditory words, and, in one condition, were asked to reproduce the sequence in their preferred communication mode (speech or speech and sign). In another condition, the children used associated manual button presses to reproduce the auditory sequence. In a third condition, picture sequences of "fish" and "dog" images were presented and the child again was asked to use associated button presses to reproduce the sequence. Two additional conditions included a "hand movement" imitation task in which an examiner presented a sequence of fist vs. open-palm gestures that the child was asked to reproduce, and a tone imitation task in which the child used associated button press responses to reproduce each target sequence.

Dawson et al. found that the CI group scored more poorly than the NH group, on average, in all five of the conditions tested, with only differences in two of the conditions failing to reach statistical significance (hand movements and tones).[3] Significant differences were found even when the picture stimuli of "dog" and "fish" were used. In the hand-movement condition, which also made use of visual-spatial stimuli, the observed p-values reached marginal significance ($p$ = .06). These results provide additional converging evidence that children who use CIs have more difficulty encoding and recalling lists of visual-spatial stimuli than their peers, when these stimuli lend themselves to being rapidly encoded in memory using highly familiar verbal labels.

Neither of these two previous studies, however, investigated the performance of children with cochlear implants on visual or visual-spatial memory tasks that discourage linguistic labeling strategies. In light of previous research on nonimplanted hearing-impaired children, it is conceivable that children with cochlear implants would perform at least as well as, or perhaps even better than, normal-hearing children on memory tasks that rely more exclusively on visual and visual-spatial processing.

Moreover, Cleary et al. (2001) and Dawson et al. (2002) were both primarily concerned with memory for *sequences* of temporally ordered stimuli. Neither study assessed implanted children's memory for "static" visual-spatial stimuli. The present study addressed pediatric cochlear implant users' memory for static visual-spatial stimulus configurations by measuring memory performance for visual-spatial dot-pattern stimuli. If the difficulties displayed by the implanted children are specific to temporal sequences of visual-spatial information and ordered recall, we would expect to observe normal levels of

[3] In the tone list memory condition, a statistically reliable difference was not obtained, probably due to the larger within-group variability observed in this condition relative to the other conditions.

performance in pediatric cochlear implant users on this visual/visual-spatial memory task which does not require the encoding of sequential order information.

The previous studies reviewed above also did not compare memory skills for simpler versus more visually complex visual-spatial stimuli. The present study addressed this issue by examining recognition memory for briefly studied, previously unfamiliar faces. A number of published studies in the literature on non-implanted hearing-impaired individuals have reported that face processing skills measured by discrimination and memory tasks, are enhanced in children and adults who are fluent users of manual language, relative to non-signing individuals (e.g., Arnold & Mills, 2001; Arnold & Murray, 1998; Bellugi et al., 1990). This "visual advantage" appears not to be specific to human faces, per se, but rather to the visual processing of complex visual objects that are encountered frequently along with variation among members in the visual category. The unavailability of simple and familiar verbal labels to uniquely identify these perceptually similar category members, it has been argued, leads to a situation in which visual/visual-spatial processing skills are heavily drawn upon during discrimination and memory tasks involving members of such categories.

The hearing-impaired children in the present study were not, however, fluent users of a manual-only language, making unclear whether it is reasonable to expect a "visual advantage" in this population. Because these children use a spoken linguistic system (either alone, or in combination with manual signs), it may be the case, that they, like normal-hearing children, attempt linguistic coding even when visual/visual-spatial stimuli do not lend themselves to this form of encoding, and, indeed even when it is disadvantageous to try to use linguistic labels (e.g., Brandimonte, Hitch, & Bishop, 1992; Carmichael, Hogan, & Walter, 1932). Implanted children would then be unlikely to perform measurably better on a face memory task than their age-matched normal-hearing peers.

In summary, the present investigation was designed to address several unresolved issues regarding visual/visual-spatial memory skills in pediatric cochlear implant users. More specifically, the performance of children with CIs was compared against that of normal-hearing age-matched peers, for their ability to remember dot pattern configurations and recently studied unfamiliar faces. These particular tasks were chosen because linguistic coding strategies are unlikely to be helpful in carrying out these visual/visual-spatial tasks. If the previously reported memory difficulties of pediatric cochlear implant users are due primarily to their less well-developed linguistic encoding strategies rather than to some more general, modality non-specific aspect of memory function, then the performance of implanted children on these visual/visual-spatial memory tasks is predicted to be at least as good as that of normal-hearing children.

## Method

### Participants

**Cochlear Implant Group.** Thirty-one eight and nine-year-old children with cochlear implants took part in the present study. These children were a subset of 45 implanted children who took part in a larger project at the Central Institute for the Deaf in St. Louis, Missouri entitled "Cochlear Implants and Education of the Deaf Child" in the summer of 2000. The present subset of 31 children completed all of the experimental tasks included in this report, and were also able to accurately identify all items in the set of four auditory stimuli used in an auditory list memory task. The performance of these children on the list memory task has been previously detailed in Experiment 3 of Cleary et al. (2001). Only a brief description of the children's list memory performance as it relates to the visual memory measures will be provided in the present report.

The mean age of the children at onset of deafness was 0.2 years with 26 of the 31 children believed to have been deaf since birth. The average duration of deafness prior to implantation was 2.72 years (range, 0.75-4.67 years). The children had used a cochlear implant for an average of 5.78 years at time of testing (range, 4.3-6.9 years). The group included children who use primarily oral communication methods as well as children who use total communication methods (i.e., a combination of speech and sign). The mean chronological age of the group was 8.7 years (range, 7.92-9.91 years). The group included 15 female and 16 male children.

All implanted children were using either a Nucleus 22 or Nucleus 24 device manufactured by Cochlear Corporation. The mean number of active electrodes per implant was 18.7 electrodes (range, 12-22 electrodes). On average, this group of pediatric CI users scored about 1.5 years below their expected language age given their chronological age, as measured by their performance on the Test for Auditory Comprehension of Language Revised (TACL-R, Carrow-Woolfolk, 1985), which was administered using both speech and sign to all children within one week of the present testing (Geers, Nicholas, & Sedey, 2003).

**Normal-Hearing Comparison Group.** Thirty-one age- and gender-matched normal-hearing children were tested at Indiana University-Bloomington in the spring and summer of 2001. Each child passed a simple pure-tone hearing screening administered at 250, 500, 1000, 2000, and 4000 Hz. A response was required at 20dB HL for frequencies of 1000 Hz and above, and at 20 or 25 dB HL for frequencies below 1000 Hz. Each ear was screened separately. Each normal-hearing child was also judged to display receptive language skills appropriate for his/her age, as determined by administering the Peabody Picture Vocabulary Test (PPVT-III, Form A, Dunn & Dunn, 1997).

## Materials and Procedure

The published materials for two subtests of the Children's Memory Scale (CMS, Cohen, 1997) were used as described in the test manual. The two CMS measures used were the Dot Locations subtest and the Faces subtest, as described below. An examiner experienced with working with young hearing-impaired children administered the tasks to the children with cochlear implants. Task instructions were provided using total communication methods to children who used this form of communication. The normal-hearing children were tested at Indiana University by trained graduate research assistants.

**Dot Locations Subtest.** The Dot Locations task from the CMS involves showing the child a picture of six large identical blue "dots" inside a large white rectangle (see Figure 1). The child is told to remember the placement of the dots. After five seconds have passed, the pattern is hidden and the child must immediately reproduce the pattern by placing six plastic disks on a 3 by 4 grid. No constraints are placed on the order in which the child can place the disks on the grid during the recall procedure, and no feedback is given regarding performance. The disks are then removed and the child is shown the same pattern of dots again for 5 seconds, and then asked to reproduce it. This process repeats one additional time, for a total of three "practice/learning" trials. The child is then shown a new pattern of (red) dots and asked to remember and reproduce this new pattern. This serves as a distracter trial. Finally, in a delayed recall trial, the child is again given the disks but is shown no pattern and is asked to recall the pattern of blue dots that was practiced three times. The child receives one point for each disk correctly placed at the right location on the grid. The distracter trial is not scored, but the three practice trials and single delayed recall trial are each worth 6 points, for a summed "Total Score" worth a maximum of 24 points.

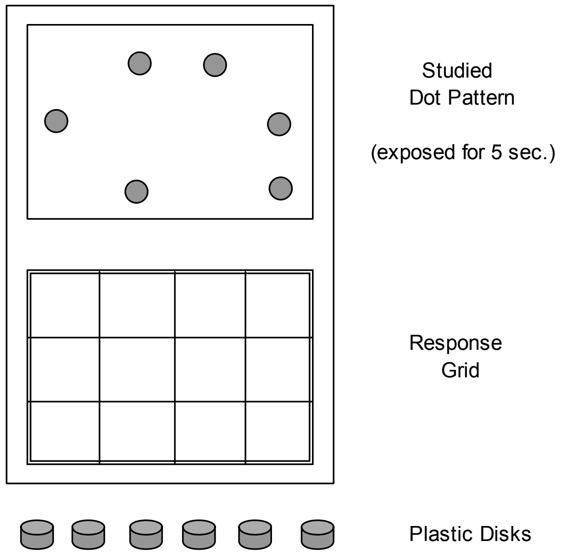


**Figure 1.** Schematic illustration of a stimulus plate and response grid for the Dot Locations task. (Not an actual stimulus.)

**Faces Subtest.** The Faces Subtest of the CMS involves showing the child a series of 12 unfamiliar faces. Each photographed face is shown to the child for approximately 2 seconds (see Figure 2). The child is asked to look closely at each face and to remember what the person looks like, because, as the child is warned, later, he/she will be asked to look at some more photos and to decide if particular faces are ones the child was asked to remember. Immediately following presentation of the 12 "study faces," the child is shown a series of 36 test faces. The child is told to respond "yes" if the face is one he/she remembers seeing during the study portion of the test, and "no" if the face is a new one that he/she was not asked to remember. During test, 18 new faces are mixed into the set of 12 studied faces with six of the studied faces shown twice, for a total of 18 "old" faces. One point is awarded for each correct response, for a maximum possible score of 36 points. No feedback is provided regarding performance on individual trials.

Each face stimulus consists of an oval cutout showing the photographed face of an adult or school-age child. A wide variety of ages and ethnic backgrounds are represented in the photographs. The cutout limits the view of the individual's hair and clothing. No persons with glasses, beards, or mustaches are included.

**Subtest Selection Issues.** The 1997 version of the CMS provides two forms of each subtest, one for children ages 5;0-8;11 and one for children ages 9;0-12;11. In this initial study, we decided to use the test forms designed for the younger age group for both the 8- and 9 year-old children in our study. (There were 22 eight-year-olds and 9 nine-year-olds in each subject group.) Because the mean age of the children tested was 8.7 years, we judged that this choice was reasonable, given that we wished to administer the same form of each test to both the 8- and 9-year-old children. We also based this decision on advice that the longer, more difficult forms, designed for the older age group, might be too difficult for some pediatric cochlear implant users to complete.

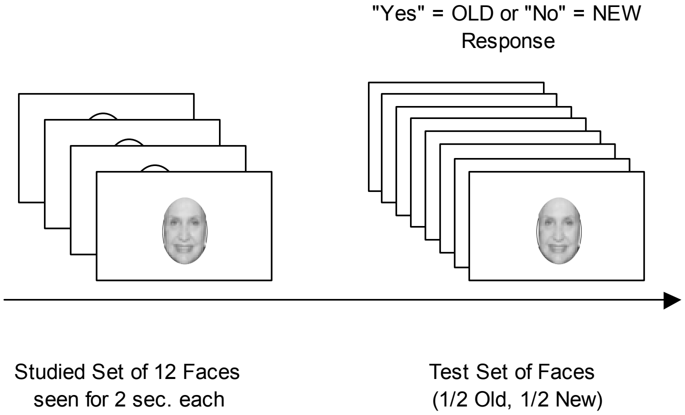


**Figure 2.** Schematic illustration of sample stimulus plate and procedure for the Face recognition task. (Not an actual stimulus.)

**List Memory Measures.** The list memory (or "memory span") measures gathered from these children have been described extensively elsewhere (Cleary, Pisoni, & Geers, 2001). Briefly, as described in the Introduction, each child was tested on his/her ability to reproduce sequences presented under three different conditions. A response box consisting of four backlit colored response buttons was used by the child to make all of his/her responses. In one condition, sequences of colored lights were presented on the response box. In a second condition, the colored light sequences were shown synchronously with matched recorded auditory color-names presented over a loudspeaker. In a third condition, only the auditory color-names were presented with no colored lights. An adaptive testing procedure was used to lengthen the sequence if the child correctly reproduced the list on a given trial, or to shorten the sequence if an incorrect response was detected. The test sequences were different from trial to trial and 20 lists were tested in each of the three counterbalanced conditions. Each child's score in each of the three conditions was tabulated by adding up the proportion of lists correctly reproduced at each list length tested, starting with a list length of 1 item.

**Task Administration.** The children first completed the dot locations task then the face recognition task. The list memory measures were collected during the same week as the dot and face measures for the children with cochlear implants and before the dot and face measures during the same testing session for the children with normal-hearing. The children were thanked and praised for their participation after the test session. The implanted children and their families received tickets for organized activities planned for the afternoon of each day of testing. The normal-hearing children received $6 and a T-shirt or cap for their participation.

## Results and Discussion

### Dot Locations Subtest

As shown in Figure 3, the CI group displayed no problems with the Dot Locations subtest relative to the published test norms for normally-developing eight-year olds. Indeed, the CI group's mean score of 21 points on this subtest did not differ statistically from the expected 50th percentile score for normally-developing children as provided by the test norms (*p* = .95). However, as can also be seen in Figure 3, the normal-hearing comparison group scored significantly higher, on average, than the test norms for eight-year olds on this same task (*t*(30) = 3.784, *p* = .001, 50th percentile score = 21 points, NH group mean = 22.5 points).

In general, the score distributions obtained for this task were quite negatively skewed, with a substantial number of scores at ceiling, particularly for the NH group. We nevertheless also conducted statistical tests between the means obtained from the two groups of children. The CI and NH groups differed significantly from each other on the Dot Locations task (*t*(60) = 2.440, *p* = .018), with the normal-hearing group having a significantly higher mean score despite the ceiling effect which limited the upper range of possible scores for these participants.

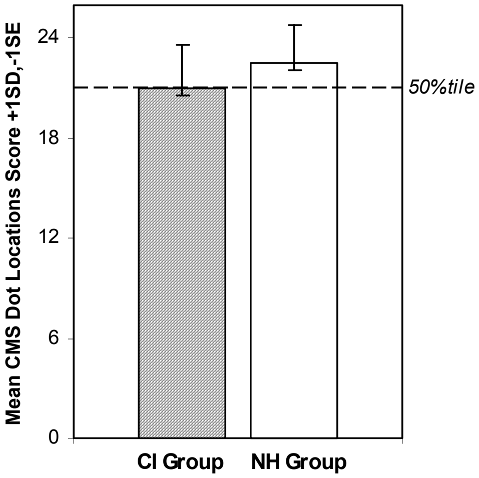


**Figure 3.** Mean performance for each group of children on the Dot Locations task. Error bars in the positive direction indicate one standard deviation. Error bars in the negative direction indicate one standard error. The dashed line shows the 50th percentile score predicted for 8-year-old normally-developing children as provided by the Children's Memory Scale test manual (Cohen, 1997). The maximum score possible on this task was 24 points.

## Faces Subtest

On the face memory task, children with cochlear implants obtained a lower mean score, as shown in Figure 4, than predicted by the test norms provided for the Faces subtest. Although this difference was small, the difference reached statistical significance when a one-sample t-test was applied, testing the group mean against the standardized 50th percentile score of 30 points for normally-developing eight-year olds (*t*(30) = 2.167, *p* = .038). In contrast, the normal-hearing group's mean score on the Faces subtest did not differ significantly from the published 50th percentile score for eight-year olds (*p* = .70).

The distributions of scores obtained for the Face memory task were near-normally distributed between chance and ceiling performance, for both groups of children. Unlike the Dot Locations task, the Face memory task yielded few scores at ceiling for either group of children. The mean scores of the implanted group and the normal-hearing group also differed significantly from each other on the Faces subtest, with the normal-hearing group again scoring significantly higher (*t*(60) = 2.048, *p* = .045).

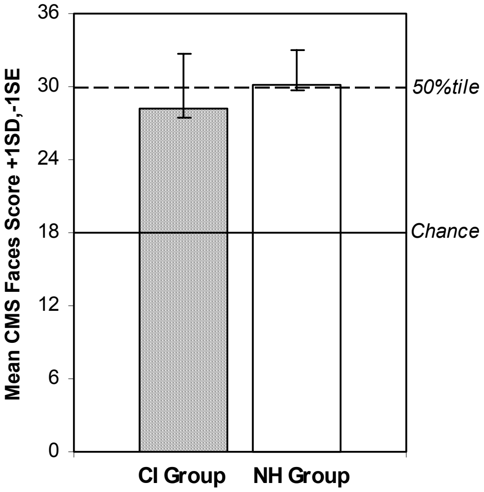


**Figure 4.** Mean performance for each group of children on the Faces task. Error bars in the positive direction indicate one standard deviation. Error bars in the negative direction indicate one standard error. The solid line shows the score predicted for chance performance (random guessing). The dashed line shows the 50th percentile score predicted for 8-year-old normally-developing children as provided by the Children's Memory Scale test manual (Cohen, 1997). The maximum score possible on this task was 36 points.

## Correlational Analyses

Intercorrelations among the memory measures and a variety of demographic characteristics were also examined. Neither chronological age nor gender contributed significantly to the within-group variability in memory scores in either group of children. Characteristics of the CI users' hearing impairment were then examined in relation to memory performance. Although the group of pediatric cochlear implant users was relatively homogeneous with respect to age at onset of deafness, duration of deafness, and length of implant use, sufficient within-group variability existed in the last two variables to examine the degree to which these might be correlated with memory skill. A range of communication methods were also present within the CI group, such that it was possible to look for relationships between memory performance and degree of exposure to an auditory/oral educational setting.

As shown in Table 1, longer periods of auditory deprivation were weakly and nonsignificantly associated with lower scores on the dot locations task, whereas greater experience with an implant was weakly associated with higher scores on this same task. However, no such relationship was observed between these two demographic characteristics and face memory scores. In contrast, greater emphasis on auditory/oral methods of communication showed no relationship with dot location scores, but exhibited a correlation of +.40 with face memory scores. This last finding suggests that skills acquired through exposure to spoken language may be drawn upon by children with CIs during face encoding (see also Bergeson & Pisoni, 2004 for discussion). The data, however, are not consistent with some previous reports that greater exposure to manual language environments leads to enhanced face processing skills in non-implanted hearing-impaired populations (e.g., Arnold & Mills, 2001; Arnold & Murray, 1998; Bellugi et al. 1990).

| *N* = 31 | **Memory for Dot Locations** | **Memory for Faces** |
|---|---|---|
| **Duration of Deafness** | *r* = -.30, *p* = .11 | *r* = +.06, *ns* |
| **Duration of Implant Use** | *r* = +.35, *p* = .06 | *r* = -.01, *ns* |
| **Communication Mode Score** | *r* = +.11, *ns* | *r* = +.40, *p* = .03 |

**Table 1.** Correlations within the CI group between visual/visual-spatial measures and characteristics of the children's hearing impairment. Communication mode scores were assigned to each child based on the degree to which the child's educational environment emphasized auditory/oral methods. Higher scores correspond to a greater emphasis on auditory/oral methods of communication. P-values not adjusted for multiple comparisons.

Next, intercorrelations among the different memory measures were examined. Children's scores on the dot location and face memory tasks were not correlated with each other in either subject group. The restricted range of scores for the dot locations task may have contributed, in part, to this lack of correlation. It is also the case, however, that these two visual memory tasks differed considerably from each other, specifically in the complexity of the stimulus materials, and in the nature of the recall response required, and thus, strong correlations between the tasks were not necessarily expected.

Correlations between the list memory tasks and the two visual/visual-spatial memory measures were also examined. These values are shown for the CI group in Table 2. No relationship was observed in the CI group between any of the three list memory conditions and memory for dot locations. For the face memory task, however, a positive correlation (*r* = +.40) was observed with list memory scores in the auditory-only condition. When the list memory task included a visual component (i.e., in the lights-only and auditory-plus-lights conditions), no correlation was observed. This finding suggests that face recognition, unlike memory for dot configurations, may draw upon some skills that are better developed in hearing-impaired children who also do well on auditory list memory tasks.

| *N* = 31 | **Memory for Dot Locations** | **Memory for Faces** |
|---|---|---|
| **List Memory, Lights-Only** | *r* = +.02, *ns* | *r* = +.04, *ns* |
| **List Memory, Auditory-Only** | *r* = +.13, *ns* | *r* = +.40, *p* = .03 |
| **List Memory, Auditory-plus-Lights** | *r* = +.09, *ns* | *r* = +.19, *ns* |

**Table 2.** Correlations within the CI group between visual/visual-spatial measures and list memory performance tested under three different presentation conditions. P-values not adjusted for multiple comparisons.

Within the normal-hearing group of children, no relations were between either of the visual/visual-spatial memory measures and any of the list memory measures. This result is consistent with the hypothesis that memory for sequential order is a fundamentally different skill from memory for individual static items. The failure to find a correlation in the normal-hearing group between auditory-only list memory scores and face memory scores does not, however, correspond well with the results found in the implant group. More specifically, given that the hearing-impaired children who scored higher on the auditory-only list memory task (and thus, by extension, scored more similarly to normal-hearing children) also tended to score higher on the face memory task, one would predict a similar pattern of findings in the normal-hearing group. This was not observed, however, suggesting that the two groups of children may have encoded these images of static faces using different strategies. For the normal-hearing children, we did not find evidence that face encoding draws upon verbal-encoding resources used in the auditory-only list memory task.

## General Discussion

The obtained pattern of results is not ideal for drawing firm conclusions and further investigation is clearly warranted. The most straightforward result from these data is the finding that the implanted children as a group did somewhat more poorly than expected on the Face recognition task, both as compared to the published norms for slightly younger children, and as compared to age-matched normal-hearing children. The CI group clearly did not show "enhanced" performance on this face processing task, relative to either the published norms or the NH comparison group.

On the Dot Locations task, the implanted children demonstrated levels of performance consistent with that expected from normally-developing children. Relative to published norms, no particular deficit or enhancement of spatial memory skill for simple stimulus patterns was observed. The slightly worse performance of the implanted children on the Dot Locations task as compared to the normal-hearing group must be considered relative to the fact that the normal-hearing children in this study did somewhat better on this task than the published subtest norms (albeit for slightly younger children) would predict.

Assuming that the norms for the CMS subtests are accurate, the present data also suggest that implanted children found the complex visual-spatial stimuli (i.e., faces) more difficult to process relative to the simpler visual-spatial stimulus configurations (i.e., dot patterns) than did normally-developing children. The source of this unexpected difference requires further investigation. The pediatric cochlear implant users may have attempted to encode the face stimuli (but not the dot patterns) using some form of verbal labeling, even though their verbal encoding and rehearsal skills are less well developed than those of age-matched normal-hearing children. The general pattern of correlations supports the notion that auditory/oral skills were associated with higher face memory scores. We observed this relationship even though some prior findings with non-implanted hearing-impaired populations suggest that use of verbal encoding strategies may actually negatively impact memory for unfamiliar faces under some testing situations (e.g., Arnold & Mills, 2001).

The present study also sought to examine the memory performance of pediatric cochlear implant users for static, non-sequenced visual-spatial stimulus configurations, with the aim of comparing these results to previous reports of memory difficulties for temporally ordered visual-spatial patterns (i.e., Cleary et al., 2001; Dawson et al., 2002). We found that neither CMS measure correlated significantly with scores on a list-memory task that used a lights-only stimulus presentation condition, in either group of children. Although this is a "null result," it is consistent with the hypothesis that memory for visual/visual-spatial sequences and memory for static individual items should be considered separately in studies of memory performance. Unexpectedly however, we also found that when the auditory component of the list-memory task was emphasized (i.e., in the auditory-only colornames condition), modest correlations with the CMS face memory task were observed, but only in the CI group. This apparent link between memory success for individual members of spatially complex visual categories and memory for temporally complex auditory verbal sequences should be examined further.

This line of research continues with the goal of further refining the experimental tasks used to examine non-sequential visual/visual-spatial memory in this clinical population. In the present study we assessed the feasibility of using two readily available standardized measures for this purpose, namely, the Dot Locations and Faces subtests from the Children's Memory Scale. Certain limitations and peculiarities of these two measures became apparent in the course of data collection. The dot locations task suffered from ceiling effects, particularly in the case of the normal-hearing children who participated in this study. This is probably due to our use of this subtest with some children who were too old for this form of the test. However, ceiling effect issues for this subtest have since been reported also by other researchers who followed the test directions regarding age more precisely. A critical review of the Children's Memory Scale by Vaupel (2001), for example, noted that although the CMS as a whole is "psychometrically sound" and "well-constructed," "significant floor and ceiling effects exist on several subtests," the dot locations subtest being among these.

Although the Faces subtest yielded more normally-distributed scores and appeared to be able to capture meaningful variability across both groups of children, one aspect of this test struck us as somewhat peculiar. Specifically, as previously described in the Method section, half of the familiarized studied faces are shown twice during testing. Presumably, this characteristic of the test arose due to time and attention constraints on total testing time, or to give the subtest score distribution better psychometric properties, however, the presence of repeated test items seems somewhat questionable.

Another potential drawback of the CMS is that its subtest norms may be overly "coarse" in the sense that norms are offered in 12-month increments—that is, for example, the same norms are provided for children ages 8;0 through 8;11. In contrast, some standardized tests are sensitive enough to offer norms in six-month or four-month increments. Collapsing across a 12-month age range may obscure important developmental changes in visual-spatial memory processing. If we are interested in studying such changes in pediatric cochlear implant users, a more age-sensitive test may be necessary.

The available evidence suggests however, that hearing-impaired children using cochlear implants perform somewhat more poorly than normal-hearing children on list memory tasks and on visual-spatial recognition memory tasks involving complex stimuli. One interesting question that remains is how well do implanted children do relative to hearing-impaired children who are receiving benefit from conventional hearing aids? A recent study by Surowiecki, Sarant, Maruff, Blamey, Busby, and Clark (2002) found that the performance of children with cochlear implants on a battery of visual memory tests did not differ measurably from that of age- and gender-matched children who were using conventional hearing aids to address moderate to profound hearing losses. Moreover, these authors also reported that visual memory scores in these groups of hearing-impaired children (all of whom were enrolled in oral educational settings) tended to positively correlate with individual differences in scores on some language processing tests. Comparisons with test norms for normal-hearing children were not included in Surowiecki et al.'s preliminary report, however, making it difficult to assess how well either group of hearing-impaired children was doing relative to normal-hearing children. Furthermore, as in the present study, ceiling effects were encountered for a number of the measures used.

In summary, the present study found that pediatric cochlear implant users perform at normal levels on memory for dot locations. Small, but statistically reliable differences were found, however, in recognition memory for briefly studied unfamiliar faces: children with cochlear implants did not perform as well as children with normal hearing. Interestingly, we also found a correlation in children with cochlear implants between recognition memory for briefly studied faces and individual variability in list memory for sequences of auditory stimuli. The present set of results suggests that memory difficulties in prelingually-deafened pediatric cochlear implant users may be more pronounced in behavioral tasks that require the processing of sequences of stimuli compared to isolated static stimuli, and for individually presented complex visual stimuli compared to simpler visual stimulus configurations.

## References

Adams, A.M., & Gathercole, S.E. (2000). Limitations in working memory: implications for language development. *International Journal of Language and Communication Disorders, 35,* 95-116.

Arnold, P., & Mills, M. (2001). Memory for faces, shoes, and objects by deaf and hearing signers and hearing nonsigners. *Journal of Psycholinguistic Research, 30,* 185-195.

Arnold, P., & Murray, C. (1998). Memory for faces and objects by deaf and hearing signers and hearing nonsigners. *Journal of Psycholinguistics Research, 27,* 481-497.

Bavelier, D., Tomann, A., Hutton, C., Mitchell, T., Corina, D., Liu, G., & Neville, H. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. *Journal of Neuroscience, 20-RC93,* 1-6.

Bebko, J.M. (1984). Memory and rehearsal characteristics of profoundly deaf children. *Journal of Experimental Child Psychology, 38,* 415-428.

Bellugi, U., O'Grady, L., Lillo-Martin, D., O'Grady, M., vanHoek, K., & Corina, D. (1990). Enhancement of spatial cognition in deaf children. In V. Volterra & C. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children, (pp. 278-298).* New York: Springer Verlag.

Bergeson, T.R. & Pisoni, D.B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence & B.E. Stein (Eds.), *Handbook of Multisensory Processes (pp. 749-772).* MIT Press.

Brandimonte, M.A., Hitch, G.J., & Bishop, D.V. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory and Cognition, 20,* 449-455.

Carmichael, L., Hogan, H.P., & Walter, A.A. (1932). An experimental study of the effect of language on the reproduction of visually perceived forms. *Journal of Experimental Psychology, 15,* 73-86.

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

Cleary, M., Pisoni, D.B., & Geers, A.E. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear & Hearing, 22,* 395-411.

Cohen, M.J. (1997). *The Children's Memory Scale (CMS).* San Antonio, TX: The Psychological Corporation.

Conrad, R. (1972). Short-term memory in the deaf: A test for speech coding. *British Journal of Psychology, 63,* 173-80.

Conrad, R. (1979). *The Deaf Schoolchild.* London, England: Harper & Row.

Dawson, P.W., Busby, P.A., McKay, C.M., & Clark, G.M. (2002). Short-term auditory memory in children using cochlear implants and its relevance to receptive language. *Journal of Speech, Language, and Hearing Research, 45,* 789-801.

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

Furth, H. (1966). *Thinking without language: Psychological implications of deafness.* New York: The Free Press.

Geers, A.E., Nicholas, J.G., & Sedey, A.L. (2003). Language skills of children with early cochlear implantation. *Ear & Hearing, 24(Suppl.),* 46S-58S.

Kelly, R.R., & Tomlinson, K.C. (1976). Information processing of visually presented picture and word stimuli by young hearing-impaired and normal hearing children. *Journal of Speech and Hearing Research, 19,* 628-638.

Mayberry, R.I. (1992). The cognitive development of deaf children: Recent insights. In S.J. Segalowitz & I. Rapin (Eds.), *Handbook of Neuropsychology, Vol. 7 (pp.51-68).* New York: Elsevier Science Publishers.

Montgomery, J.W. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. *Journal of Speech, Language, & Hearing Research, 43,* 293-308.

Pintner, R., & Paterson, D.G. (1917). A comparison of deaf and hearing children in visual memory for digits. *Journal of Experimental Psychology, 2,* 76-88.

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. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear & Hearing, 24(Suppl.),* 106S-120S.

Surowiecki, V.N., Sarant, J., Maruff, P., Blamey, P.J., Busby, P.A., & Clark, G.M. (2002). Cognitive processing in children using cochlear implants: the relationship between visual memory, attention, and executive functions and developing language skills. *Annals of Otology, Rhinology, & Laryngology, 189 (Supplement: VIII Symposium of Cochlear Implants in Children),* 119-126.

Tomlinson-Keasey, C., Smith-Winberry, C. (1990). Cognitive consequences of congenital deafness. *Journal of Genetic Psychology, 151,* 103-115.

Vaupel, C.A. (2001). Cohen, M.J. (1997). The Children's Memory Scale (CMS). San Antonio, TX: The Psychological Corporation (Review). *Journal of Psychoeducational Assessment, 19,* 392-400.

Waters, G.S., & Doehring, D.G. (1990). Reading acquisition in congenitally deaf children who communicate orally: Insights from an analysis of component reading, language, and memory skills. In T.H. Carr & B.A. Levy (Eds.), *Reading and its development (pp. 323-373).* San Diego: Academic Press.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Perceptual Learning and Nonword Repetition Using a Cochlear Implant Simulation[1]

**Rose A. Burkholder, David B. Pisoni, and Mario A. Svirsky**[2]

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

[1] This work was supported by NIH-NIDCD R01 Research Grants DC00111 and DC03937, and Training Grant T32 DC00012 as well as the American Hearing Research Foundation. We are grateful for the time and support from past and present Speech Research Laboratory members Brianna Conrey, Caitlin Dillon, Kuai Hinojosa, and Jeff Reynolds who participated in the piloting of this experiment. We also thank Shivank Sihna and Luis Hernandez for their technical assistance and expertise on this project.

[2] Indiana University School of Medicine and Purdue University Departments of Biomedical and Electrical Engineering

# Perceptual Learning and Nonword Repetition Using a Cochlear Implant Simulation

**Abstract.** This study examined the effects of perceptual learning on the nonword repetition performance of normal-hearing adult listeners who were exposed to degraded auditory signals that were designed to simulate the auditory input of a cochlear implant. Twenty normal-hearing listeners completed a nonword repetition task, using an 8-band, frequency-shifted cochlear implant simulation strategy, both before and after training on open- and closed-set word recognition tasks. Feedback was provided during the training tasks. The nonword responses obtained from each participant were digitally recorded and then played back to four different normal-hearing listeners. These listeners rated the nonword repetition accuracy in comparison to the original unprocessed target stimuli using a 7-point scale. The mean nonword accuracy ratings were significantly higher for the nonwords repeated after the period of training with the processed stimuli than for nonwords repeated prior to training. These results suggest that the word recognition training tasks facilitated auditory perceptual learning that generalized to the production of novel, nonword auditory stimuli. In addition, adaptation and learning from the degraded auditory stimuli produced by a cochlear implant simulation can be achieved even in a very difficult perceptual-motor task such as nonword repetition which involves both speech perception and production. Our nonword repetition findings extend previous reports of adaptation and perceptual learning with cochlear implant simulations which showed that word and sentence recognition scores and vowel perception improved in normal-hearing listeners after experience with an acoustic simulation of a cochlear implant.

## Introduction

The ability to recognize speech from the degraded auditory input provided by a cochlear implant has been shown to be related to pediatric cochlear implant users' digit spans (Cleary, Pisoni, & Geers, 2001; Pisoni & Cleary, 2003) and nonword repetition skills (Cleary, Dillon, & Pisoni, 2002; Dillon, Burkholder, Cleary, & Pisoni, in press). In addition, memory processes carried out during the digit span task, such as rehearsal speed, have been found to be directly related to nonword repetition in deaf children using cochlear implants (Dillon, Cleary, Pisoni, & Carter, 2004). By using overt speaking rate as a measure of subvocal verbal rehearsal speed, Dillon and colleagues found that children who were able to subvocally rehearse faster received higher nonword repetition ratings than children who rehearsed verbal information slowly.

These findings indicate that speech perception abilities and memory performance are intimately related to one another in this clinical population, especially when the tasks require the encoding and manipulation of auditory stimuli in tasks such as nonword repetition and digit span recall. Recently, the relationships between memory span and speech perception skills have also been examined in normal-hearing listeners under auditory conditions that are designed to simulate what deaf users of cochlear implants may hear (Burkholder, Pisoni, & Svirsky, this issue; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000). Cochlear implant simulation studies provide a unique opportunity to investigate how memory and speech perception are affected when normal-hearing listeners are exposed to novel auditory stimuli.

Acoustic simulations of cochlear implants are able to emulate how speech is processed by the devices by filtering the speech signal into bands or adjacent frequency bins and then identifying the amount of spectral energy within bins that spans the entire range of speech sounds. The electrical pulses directed to the electrodes in a cochlear implant are modulated by the output of each of the adjacent frequency bins. This process is typically mimicked in acoustic models of cochlear implants through filtered noise band modulation carried out by the output of the simulator's frequency bins which corresponds to the spectral bands of the noise. However, in some acoustic models of cochlear implants, the frequency of the analysis filters creating the frequency bins that modulate the noise bands is exceeded by the frequency of the noise bands (Kaiser & Svirsky, 2000). This mismatch in frequencies causes an upward shift in the sound's frequency. This simulation technique is designed to mimic the natural basalward frequency shift that results when the cochlear implant's electrode placement is unable to reach tonotopic areas deep within the apical end that are tuned for low frequency sounds. This often causes a mismatch and a frequency shift between what sound is being detected through the implant and what sound is actually perceived.

Cochlear implant simulation studies have previously been used in normal-hearing listeners to investigate the effects of degraded auditory stimuli on digit span performance (Burkholder et al., this issue; Eisenberg et al., 2000). Results of these studies have shown that digit span significantly declines in both normal-hearing adults and children when the stimuli are presented in degraded auditory conditions meant to simulate a cochlear implant. In addition, cochlear implant simulation studies have provided a unique opportunity to determine whether speech perception abilities in degraded auditory conditions and performance on memory tasks conducted under degraded auditory conditions are related in normal-hearing listeners as they are in deaf children using cochlear implants. For instance, using models of 4- and 8-channel cochlear implants, Eisenberg and colleagues (2000) found modest correlations between normal-hearing adults' and children's performance on word and sentence recognition scores and their digit spans under cochlear implant simulation conditions. Their results are consistent with numerous studies showing a relationship between cochlear implant users' speech perception skills and memory spans.

In addition to demonstrating a link between perceptual abilities and cognitive performance, studies of cochlear implant simulations have also provided a unique opportunity to study perceptual learning and auditory adaptation to severely degraded auditory stimuli. Cochlear implant simulation studies have demonstrated rapid perceptual learning and auditory adaptation after listeners have been trained with a variety of auditory stimuli such as consonants, vowels, words, and sentences (Fu, Shannon, & Gavin, 2002; Rosen, Faulkner, & Wilkinson, 1999). The adaptation to spectrally shifted speech, designed to model a cochlear implant, was examined several years ago by Rosen and colleagues (1999). In their study, they determined that even though initial identification of spectrally shifted speech was very poor, participants' abilities to identify intervocalic consonants and vowels and to recognize words in sentences improved with training and exposure to the simulation.

However, when conducting a word identification task, especially using words embedded in sentences, speech perception performance is heavily influenced by sentence context and the listener's prior lexical knowledge (Miller, Heise, & Lichten, 1951). The identification of vowels and consonants may be relatively free of higher-level contextual influences, but vowels and consonants are not representative of the complexity that characterizes listening to connected speech in everyday circumstances. In order to evaluate perceptual learning for complex auditory patterns that are similar to English words, yet lack contextual or lexical influences, we examined normal-hearing participants' nonword repetition performance under conditions of degraded and spectrally shifted speech, focusing on the specific effects of training on nonword performance.

The nonword repetition task has been used previously to examine speech perception and phonological working memory in deaf children with cochlear implants (see Carter, Dillon, & Pisoni, 2002; Cleary et al., 2002; Dillon et al. in press) as well as phonological working memory in normal-hearing children (Gathercole, Willis, Baddeley, & Emslie, 1994) and adults (Gupta, 2003; Papagno & Vallar, 1995; Service & Craik, 1993). Several recent studies from our laboratory examining the nonword repetition accuracy of deaf children using cochlear implants have shown that nonword repetition is a challenging task for deaf children to complete accurately (Carter et al., 2002; Cleary et al., 2002; Dillon et al., in press). The difficulty of the nonword repetition task renders it appropriate for use in normal-hearing adults listening to a cochlear implant simulation as well. By examining nonword repetition in normal-hearing adults exposed to an acoustic simulation of a cochlear implant, it may be possible to determine the extent of nonword repetition impairment that is the result of degraded auditory input and to estimate the degree of improvement that can be gained through a brief training period.

Performance on the nonword repetition task by deaf children with cochlear implants has been found to be strongly related to a number of other cognitive measures such as digit span, speaking rate, and spoken word recognition (see Cleary et al., 2002; Dillon et al., in press). Similarity, digit span and speech perception are related in normal-hearing adults and children listening to a cochlear implant simulation. Based on these previous results (Eisenberg et al., 2000) and other perceptual learning studies (Burkholder et al., this issue), we would expect that training on word recognition tasks should improve nonword repetition performance in normal-hearing adults listening to a cochlear implant simulation. In addition, we expect that performance in the training tasks would correlate with or predict performance on the nonword repetition task. Thus, the effects of short periods of open- and closed-set word recognition training on nonword repetition accuracy were examined in this study to determine if the training would generalize to unfamiliar nonword stimuli and improve normal-hearing participants' performance on a perceptual-motor task.

## Methods

### Participants

Word recognition and nonword repetition data were collected from 25 normal-hearing adults. Five participants were excluded from the study because they were unable to repeat a sufficient number of nonwords in the nonword repetition task (< 40). The nonword responses of 17 females and 3 males were used in this study. An additional 80 participants were also used to make perceptual ratings of the nonword responses. All participants were undergraduate students at Indiana University who participated in the study to receive course credit for the Psychology class in which they were currently enrolled. Participants reported that they were monolingual native speakers of American English and had no history of language, speech, hearing, or attentional disorders at the time of testing. A brief hearing screening was administered by the first author to determine whether the participants' hearing was within normal limits. Using a standard, portable pure-tone audiometer (Maico Hearing Instruments, MA27) and headphones (TDH-39P), each participant was tested at 250, 500, 1000, 2000, and 4000 Hz at 20 dB first in the right ear and then in the left ear. None of the participants showed any evidence of a hearing loss.

### Stimuli and Materials

Several well known nursery rhymes (i.e. *Twinkle, Twinkle Little Star*; *Jack and Jill*) were used for initial familiarization with the processed speech. The nursery rhymes included in the familiarization phase are included in the Appendix. While listening to these passages in the cochlear implant simulated speech, the participants were provided with the written text of the nursery rhymes so they could read along silently with them.

The training stimuli presented to participants included both open- and closed-set word recognition tasks. A subset of words chosen from the Lexical Neighborhood Test (LNT; Kirk, Eisenberg, Martinez, & Hay-McCutcheon, 1999) was used as the open-set stimuli. The LNT words were chosen from the LNT easy (LNTe), hard (LNTh) and multisyllabic (mLNT) word lists. In addition to the LNT words, words taken from the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) and the Word Intelligibility by Picture Identification (WIPI; Ross & Lerman, 1979) task were used in a closed-set word identification task.

The closed-set word recognition tasks paired recorded words with visual stimuli. Visual materials for the closed-set word recognition task included 20 panels of pictures taken from the PPVT testing booklet that corresponded to the appropriate recorded word. On each panel there were four pictures. One picture was the correct response to the auditory stimulus and three others were foils. The use of this test in the present experiment was not intended to assess vocabulary but rather to provide a moderately challenging closed-set speech perception task.

Another closed-set word recognition task, the WIPI test was also used in the training tasks. The WIPI was designed for use with hearing-impaired children. The WIPI is a more difficult test perceptually, because it contains six pictures instead of four, and the pictures depict phonetically similar minimal pairs of target words (i.e. *spoon* and *moon*; *fox* and *box*). One list of 25 WIPI words was used.

All stimuli used for familiarization and training were recorded digitally in a sound attenuated booth by the first author using an individualized version of a speech acquisition program (SAP; Dedina, 1987; Hernandez, 1995). The stimuli were sampled and digitized at 22,050 Hz with 16-bit resolution and then equated for amplitude using the Level16 software program (Tice & Carrell, 1998). All auditory stimuli were then batch processed using a personal computer equipped with DirectX 8.0, a Sound Blaster Audigy Platinum sound card, and Macro Magic. The signal processing procedure used for the cochlear implant simulation was adapted from the real-time signal processing methods designed by Kaiser and Svirsky (2000).

The signal processing strategy used bandpass filtering with a cutoff frequency of 1200 Hz. Eight filters were then used to simulate the speech processing capabilities of an 8-channel cochlear implant. The output of each filter modulated noise bands of a higher frequency range than the initial analysis filters. This mismatch was designed to model the natural frequency mismatch that occurs when the electrodes of a cochlear implant are shifted more basalward in the cochlea. The basalward shift used in this model was equivalent to a 6.5 mm shift within the cochlea.

The nonword stimuli used in this study were created and processed identically to the training stimuli with the exception that they were recorded by a different female speaker of American English. These nonword recordings were the same stimuli used in previous studies examining nonword repetition in NH children and deaf children using cochlear implants (see Carlson, Cleary, & Pisoni, 1998; Dillon et al., in press). The nonwords included in this study were all phonotactically permissible auditory patterns that could plausibly be real words in English. Forty nonwords originally developed by Gathercole and colleagues (1994) and 20 nonwords developed by Wagner and colleagues (1997) were used. The nonwords ranged in length from one to six syllables. There were ten nonwords used at each syllable length. Table 1 lists a sample of 20 of the nonwords used after training.

| Number of Syllables | Target Nonword Orthography | Target Nonword Transcription |
|---|---|---|
| 2 | ballop | ˈbæ.ləp |
| | prindle | ˈpɹɪn.dl ¯ |
| | rubid | ˈɹu.ˌbɪd |
| | sladding | ˈslæ.diŋ |
| | tafflist | ˈtæ.flɪst |
| 3 | bannifer | ˈbæ.nə.ˌfɚ |
| | berrizen | ˈbɛ.ɹə.ˌzɪn |
| | doppolate | ˈdɑ.pə.ˌle͡ɪt |
| | glistering | ˈglɪ.stɚ.iŋ |
| | skiticult | ˈskɪ.ɾə.ˌkʌlt |
| 4 | comisitate | kə.ˈmi.sə.ˌte͡ɪt |
| | contramponist | kən.ˈtɹæm.pə.ˌnɪst |
| | emplifervent | ɛmˈplɪ.fɚ.ˌvɛnt |
| | fennerizer | ˈfɛ.nɚ.ˌa͡ɪ.zɚ |
| | penneriful | pə.ˈnɛ.ɹə.ˌfʌl |
| 5 | altupatory | æl.ˈtu.pə.ˌtɔ.ɹi |
| | detratapillic | di.ˈtɹæ.ɾə.ˌpɪ.lɪk |
| | pristeractional | ˈpɹɪ.stɚ.ˌæk.ʃə.n̩l |
| | versatrationist | ˈvɚ.sə.ˌtɹe͡ɪ.ʃə.ˌnɪst |
| | voltularity | ˈvɑl.t͡ʃʊ.ˌlɛ.ɹəˌti |

**Table 1.** Sample of nonwords used in the current study (adapted from Gathercole et al., 1994).

The stimuli used in a perceptual ratings portion of the experiment included the original unprocessed target nonwords and digital recordings of the participants' nonword repetition responses to the target stimuli. The nonword responses of the participants were sampled and digitized at 22,050 Hz with 16-bit resolution. Each nonword was segmented and stored individually in a digital file. All sound files were then equated for amplitude.

## Procedure

**Pretraining Nonword Repetition.** Following a familiarization phase of listening to and silently reading along with a series of familiar nursery rhymes, participants were told that they would receive a nonword repetition task in which they would be asked to repeat the nonword that was presented using the same degraded format. The procedure for the nonword repetition was then demonstrated by having participants repeat two nonwords that were in unprocessed form. No participants experienced any difficulty repeating the intact nonwords. To further assist participants in completing the nonword repetition task, the effect of the signal processing on the nonwords was also explicitly demonstrated. Two nonwords were played back first in their unprocessed form and then immediately after under processed conditions.

After this sequence of nonword familiarization, participants completed nonword repetition of 30 different nonwords that were in processed form. Each nonword was played once in random order. The list of 30 nonwords included a combination of stimuli taken from the sets of both Wagner and Gathercole nonwords and included five nonwords at each of six different syllable lengths. All nonword responses were recorded onto digital audiotape (Sony Walkman TCD-D8) via a uni-directional headset cardioid condenser microphone (Audio-Technica ATM75). Because this was a difficult task, the participants were permitted to pass on items they were unable to provide a response to. However, participants were urged to repeat any portion of the nonword that they felt they could.

**Training.** After repeating the first set of nonwords, participants were administered the PPVT (Dunn & Dunn, 1997) and WIPI (Ross & Lerman, 1979) closed-set word recognition tasks. The easier closed-set task, the PPVT, was administered to the participants first in the training session followed by the WIPI. The visual and auditory stimuli used for closed-set word recognition training were presented in nonrandom order, in order to preserve the correct pairing between the picture panels in the testing booklets and the words. Following the presentation of each training word, participants either responded by pointing to the picture of the word they thought they heard or responded verbally. Auditory feedback was then given after each trial to indicate the correct response regardless of whether the correct response had been given or not. The auditory feedback was a repetition of the same utterance of the presented word in its unprocessed form.

After completing the closed-set speech perception tasks, participants were asked to identify words without the aid of any pictures from which to choose. The open-set training words were presented in random order to the participants. The sequence of open-set word recognition tasks began with the LNTe, followed by a combined list of words chosen from the LNTh and mLNT (Kirk et al., 1999). Auditory feedback was administered in the same manner as in the closed-set task. A total of 65 words taken from the LNTs were used for open-set word recognition training. Scoring of these training tasks was conducted online by the experimenter who was present during all sessions.

**Post-training Nonword Repetition.** Following the training sequence, which lasted approximately 35 minutes, participants completed a second nonword repetition task. Thirty different nonwords, 20 from the Gathercole set and 10 from the Wagner set were used. All nonwords ranged from one to six syllables, and there were five different nonwords at each syllable length. The 20 nonwords designed by Gathercole were specifically chosen for the post-training portion of nonword repetition task because they were the same nonwords used previously in the nonword repetition tasks completed by the deaf children with cochlear implants (Cleary et al., 2002). These 20 nonwords were originally selected for the testing of deaf children with cochlear implants, because when a group of normal-hearing children completed the entire nonword repetition task, their performance on these 20 words resulted in the most variation (Carlson et al., 1998).

**Nonword Ratings.** Each participant's nonword utterances were rated for imitation accuracy by a group of four different normal-hearing adults. Each listener first heard the unprocessed target nonword and then heard the nonword response spoken by the participant. A one-second interstimulus interval separated the target nonword stimulus and the participant's response. Both stimuli were presented to listeners through high quality, calibrated headphones (Beyer Dynamic, DT100) at 70dB SPL. All stimulus pairs were presented in two blocks with a different random order used within each block. Each listener rated each nonword repetition twice.

Ratings were made using a 7-point ordinal scale in which a score of "1" indicated that the listener believed the nonword utterance was "not accurate at all" and "totally failed to resemble the target utterance" whereas a score of "7" indicated that the listener believed the utterance was "perfectly accurate, ignoring differences in pitch." To record their perceptual similarity ratings, listeners used a response box with seven numbered buttons. The button box was interfaced to a PC which recorded all the nonword ratings entered by a listener into an identifiable output file.

## Results

A summary of the participants' performance on the word recognition training tasks appears in Table 2. As expected, more words were identified correctly in the closed-set tasks than in the open-set tasks (*t*(18) = 40.56, *p* = .000). Within the closed-set and open-set tasks, differences were also found. Word identification on the WIPI was significantly worse than it was on the easier four-choice PPVT (*t*(18) = 6.33, *p* = .000). In addition, there was a main effect of word difficulty on word recognition accuracy in the open-set LNT tasks (*F*(1, 54) = 24.35, *p* = .000). As expected, words from denser lexical neighborhoods were more difficult to identify than words from sparser lexical neighborhoods.

**Percent Correct Scores on Word Recognition Training Tasks**

| Closed-set Tasks | Mean | Standard Deviation |
|---|---|---|
| PPVT | 85.35 | 8.29 |
| WIPI | 69.60 | 11.27 |
| Open-set Tasks | Mean | Standard Deviation |
| LNT(easy) | 19.15 | 9.78 |
| LNT(hard) | 9.20 | 4.42 |

**Table 2.** Mean percent correct and standard deviations of scores on closed-set and open-set word recognition tasks.

Prior to any analyses of the effects of training on nonword repetition, within-rater reliability was calculated using Cronbach's alpha reliability testing. The reliability of ratings made in the first and second block of trials was determined to be adequate. Alpha values ranged from .65 to .99, with 95% of the raters having reliability coefficients greater than .70. In addition, there was no effect of block number on the average nonword ratings recorded by all the raters (*F*(1, 8318) = .20, *p* = .657). Therefore, only ratings made in the first block were used in the data analyses.

Inter-rater reliability between each group of four listeners' first block of ratings was also assessed. Nineteen out of 20 groups had a reliability rating that exceeded .70, with most groups' alpha value approaching .90. One group of raters had a particularly low reliability rating of .26 and was therefore excluded from the final analysis. This exclusion meant that one of the 20 speakers originally receiving nonword ratings was excluded from the final analysis.

Figure 1 shows the average of all nonword ratings received by the participants. A univariate ANOVA revealed a main effect of the number of syllables in the nonwords on the average nonword rating assigned to each of the participants' utterances (*F*(5, 8318) = 20.645, *p* = .000). In general, nonwords with more syllables received lower ratings. The average nonword ratings received by the speakers for all syllable lengths were concentrated on the lower end of the ratings scale, indicating that this was a particularly difficult perceptual task for them to accomplish accurately.

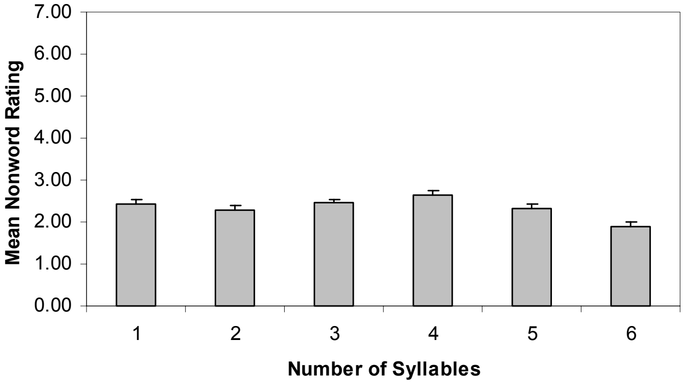


**Figure 1.** Mean nonword rating received by participants at all syllables both before and after training. Error bars represent standard error of the mean.

Several analyses were conducted to determine if there were any effects of training on nonword repetition ratings. An initial analysis including the average ratings of nonwords at all syllable lengths indicated no significant differences in the ratings received in the pre-training and post-training nonword repetition session (*t*(18) = 2.31, *p* > .05). However, when the one- and six-syllable nonwords were omitted from the analysis, post-training nonword ratings were significantly higher than the pre-training nonword ratings (*t*(18) = 4.763, *p* = .000). The six-syllable nonwords were excluded from the final analysis because these nonwords were omitted more frequently during the first block of the nonword repetition task than during the second block.

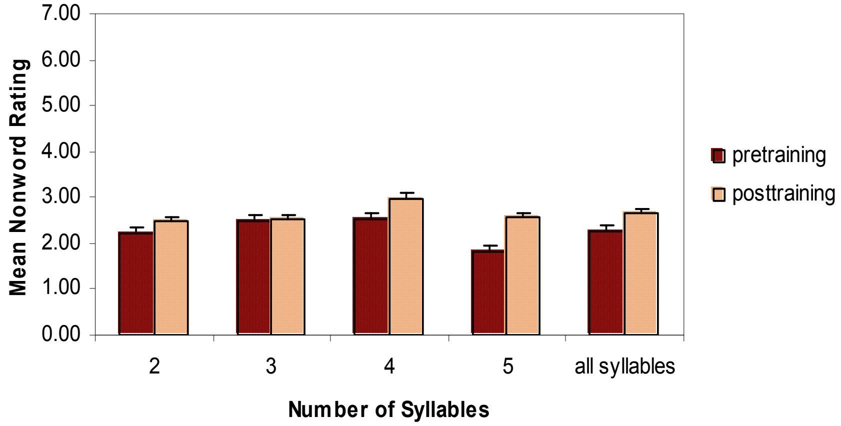


**Figure 2.** Mean nonword ratings received for nonwords repeated both before and after a short period of word recognition training.

In order to balance the data set with respect to difficulty of the nonwords included in the analysis, the one syllable nonwords were also omitted. These exclusions were also desirable because it left the post-training data set with ratings made on the 20 original nonwords that were used previously in samples of pediatric cochlear implant users. Figure 2 shows the average nonword ratings of the 2-, 3-, 4-, and 5-syllable nonwords both before and after training. After the brief period of training, all but three participants received a higher average nonword rating for the remaining nonwords. In addition to the participants' overall performance on the nonword repetition task and response to word recognition tests used during training, we were also interested in determining whether the normal-hearing adults' nonword ratings were related to their word recognition performance and digit spans. Only the identification of processed digits in isolation was correlated to the nonword repetition ratings received by the participants (r = .49, *p* < .05). The failure to obtain the expected pattern of correlations between word recognition scores, and nonword repetition may have resulted from the lack of variance in the nonword ratings received by the speakers.

## Discussion

Accuracy of nonword repetition increased significantly after only a short period of training with the speech processed through a cochlear implant simulation. These results are especially interesting because the nonwords were spoken in a different female voice than the training stimuli. This suggests that not only was the perceptual learning generalizable to novel nonwords but that the training was robust enough to extend to novel stimuli spoken in a voice different from the voice speaking the training stimuli. Specific information regarding the generalizability and potency of training with an acoustic simulation of a cochlear implant was also obtained in this study by showing that nonword repetition performance improved after a short period of training on word identification tasks. Effects of such a brief period of training on the ability to perform a perceptual-motor task using novel, nonword stimuli have not been documented previously in the literature.

However, it is important to note that effects of brief periods of training have not been documented in the same way using words either. A potential drawback to the present study is that spoken *word* recognition has not been tested both before or after such short periods of training with an acoustic model used to simulate a cochlear implant. Any evidence of such rapid effects of training on isolated word identification would provide preliminary support for the idea that improvements in nonword repetition performance could be due to the brief training. Given this lack of comparison data and support, it may be plausible that the improvements that were observed in nonword repetition were partly due to participants' practice in generating a verbal nonword response rather than to improvements in perceiving the auditory input. That is, participants may have performed better on this task due to procedural rather than perceptual learning. To determine more conclusively whether improvements in nonword repetition by these participants were actually due to training, additional nonword repetition data should be collected from a second group of listeners who do not receive any training.

If we accept the finding that the increase in nonword ratings was due to improvements in the participants' skills in accurately perceiving auditory stimuli transformed by the simulation, we do not necessarily know what specific aspect of the processed stimuli they are adapting to most proficiently. It is plausible that the participants were learning and retaining a broad representation about how the cochlear implant simulation alters speech rather than learning representations specific to the indexical properties of one speaker or more precise acoustic-phonetic representations of familiar words. However, both an improvement in phoneme identification and suprasegmental or prosodic knowledge are likely to help participants in the nonword repetition task.

It is difficult to discern exactly what attributes of the cochlear implant simulated speech the normal-hearing listeners were attending to through the current methods of analysis or what properties of the speech signal were most easily learned and reproduced during the nonword repetition task. More detailed analyses such as measuring suprasegmental and segmental accuracy in the nonword repetitions would be useful to help determine more precisely what acoustic-phonetic characteristics of the nonwords the normal-hearing adults were able to perceive and reproduce. Suprasegmental and segmental analyses have previously been conducted on the nonword repetitions of deaf children with cochlear implants (Carter et al., 2002; Dillon et al., 2004). Obtaining these same measures from normal-hearing listeners exposed to a cochlear implant simulation would enable useful comparisons between nonword repetition responses elicited from normal-hearing listeners exposed to an acoustic simulation of a cochlear implant and from deaf children using a cochlear implant. Such comparisons may have implications for how accurately an acoustic simulation of a cochlear implant can model the auditory input from a cochlear implant.

Any comparisons drawn between normal-hearing adults and deaf children completing a nonword repetition task must be considered with some caution at this time because the two populations likely differ in their speech production skills. Normal-hearing adults with no previous history of speech or language disorders are assumed to have normal speech production skills. However, deaf children with cochlear implants often have poor speech intelligibility (Miyamoto, Kirk, Robbins, Todd, Riley et al., 1997; Osberger, Robbins, Todd, & Riley, 1994), which makes it difficult to attribute nonword repetition errors exclusively to poor speech perception abilities. Therefore, it may be difficult to compare the segmental nonword repetition errors of deaf children using cochlear implants and normal-hearing listeners.

However, the results of this study do share some similarities between the nonword repetition results obtained from normal-hearing adults exposed to a cochlear implant simulation and hearing-impaired children using a cochlear implant that are unconfounded by differences in speech production. One common result found between the studies examining nonword repetition in normal-hearing adults exposed to transformed speech and deaf children using a cochlear implant was that nonword repetition accuracy was related to the number of syllables in the nonwords in both groups. Both deaf children using cochlear implants and normal-hearing adults listening to transformed speech were more accurate in repeating nonwords with fewer syllables. These results suggest that adults and children are utilizing similar cognitive processes to complete the nonword repetition task. In addition, under normal listening conditions, both adults and children demonstrate an effect of syllable number on nonword repetition accuracy (Gupta, 2003). The carryover of this effect to nonword repetition under severely degraded auditory conditions may indicate that the basic processes or strategies used to complete the task are retained despite the unusual circumstances in which they are being used.

Although the effect of syllable length on nonword repetition accuracy was found in these adult participants, their nonword ratings showed no relationship to the key speech perception training tasks as the nonword ratings of deaf children with cochlear implants typically have (Cleary et al., 2002; Dillon et al., in press). However, the rating data obtained from normal-hearing adults' nonword repetitions were collected using a different procedure than the rating data previously collected on the deaf children with cochlear implants. One difference was that in this study each listener heard the responses from only one speaker whereas in the studies with deaf children, the listeners heard nonword responses from many children over the entire test block.

In addition, in the studies using pediatric cochlear implant users, the raters were exposed to a combination of multiple speakers based on predetermined speech intelligibility (Cleary et al., 2002; Dillon et al., in press). The speech intelligibility measures ensured that the listeners were exposed to children with a varying range of speech quality. Thus, the methods used to collect ratings in the current study may have failed to appropriately detect a relationship between nonword ratings and performance on the speech perception tasks because of a more restricted range and variance in the ratings and a near floor performance by the normal-hearing adults. To more accurately assess the relations between the nonword repetition ratings and word recognition, a ratings procedure in which different speakers' utterances could be compared to one another should be conducted.

Finally, an additional improvement upon this study would be to conduct it using normal-hearing children rather than adults. Normal-hearing peers of the deaf children using cochlear implants would be a more appropriate control than normal-hearing adults because of the large developmental differences that have been previously shown in speech and memory tasks (Cowan, Saults, Nugent, & Elliot, 1999). We do plan to collect immediate memory span and nonword repetition data from normal-hearing children while listening to an acoustic simulation of a cochlear implant. Such data would be an appropriate comparison to carry out in order to examine how memory processes and perceptual skills are differentially affected by profound deafness and by the unique sensory simulation provided by a cochlear implant which requires a period of auditory adaptation and perceptual learning after implantation.

## References

Burkholder, R.A., Pisoni, D.B., & Svirsky, M.A. (this issue). Effects of a cochlear implant simulation on immediate memory span in normal-hearing adults. In *Research* o*n Speech Perception Progress Report No. 26*. Bloomington, IN: Speech Research Laboratory, Indiana University.

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.

Carter, A., Dillon, C., & Pisoni, D.B. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: suprasegmental analysis. *Clinical Linguistics and Phonetics, 16*, 619-638.

Cleary, M., Pisoni, D.B., & Geers, A. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing, 22*, 395-411.

Cleary, M., Dillon, C., & Pisoni, D. (2002). Imitation of nonwords by deaf children following cochlear implantation. *Annals of Otology, Rhinology, and Laryngology, 111*, 91-96.

Cowan, N., Saults, S., Nugent, L., & Elliott, E. (1999). The microanalysis of memory span and its development in childhood. *International Journal of Psychology, 34*, 353-358.

Dedina, M.J. (1987). SAP: A speech acquisition program for the SRL-VAX. In *Research* o*n Speech Perception Progress Report No. 13* (pp. 331-337). Bloomington, IN: Speech Research Laboratory, Indiana University.

Dillon, C.M., Burkholder, R.A., Cleary, M., & Pisoni, D.B. (in press). Perceptual ratings of nonword repetition responses by deaf children after cochlear implantation: Correlations with measures of speech, language, and working memory. *Journal of Speech Language and Hearing Research*.

Dillon, C.M., Cleary, M., Pisoni, D.B., & Carter, A.K. (2004). Imitation of nonwords by hearing impaired children with cochlear implants: segmental analysis. *Clinical Linguistics and Phonetics, 86*, 39-55.

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

Eisenberg, L., Shannon, R., Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America, 107*, 2704-2709.

Fu, Q., Shannon, R., & Galvin, J. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *Journal of the Acoustical Society of America, 112*, 1664-1674.

Gathercole, S., Willis, C., Baddeley, A., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*, 103-127.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology*, 56A, 1213-1236.

Hernandez, L.R. (1995). Current computer facilities in the Speech Research Laboratory. In *Research on Speech Perception Progress Report No. 13* (pp. 389-393). Bloomington, IN: Speech Research Laboratory, Indiana University.

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 Proceedings of the IXth IEEE-DSP Workshop., Hunt, TX.

Kirk, K.I., Eisenberg, L.S., Martinez, A.S., & Hay-McCutcheon, M. (1999). Lexical neighborhood test: Test-retest reliability and interlist equivalency. *Journal of American Academy of Audiology, 10*, 113-123.

Miller, G.A., Heise, G.A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*, 329-335.

Miyamoto, R., Kirk, K., Robbins, A., Todd, S., Riley, A., & Pisoni, D. (1997). Speech perception and speech intelligibility in children with multichannel cochlear implants. *Cochlear Implant and Related Sciences Update, 52*, 198-203.

Osberger, M., Robbins, A., Todd, S., & Riley, A. (1994). Speech intelligibility of children with cochlear implants. *The Volta Review, 96*, 169-180.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48A,* 98-107.

Pisoni, D.B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*, 106S-120S.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106*, 3629-3636.

Ross, M., & Lerman, J. (1979). A picture identification test for hearing impaired children. *Journal of Speech and Hearing Research, 13,* 44-53.

Service, E., & Craik, F.I.M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language, 32,* 608-623.

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

Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (1997). *Comprehensive Test of Phonological Processes.* Austin, TX: PRO-ED Publishing, Inc.

## Appendix

Stimulus materials used for familiarization with the acoustic model of a cochlear implant.

*Hickory, Dickory, Dock*  
*Hickory, dickory, dock,*  
*The mouse ran up the clock.*  
*The clock struck one,*  
*The mouse ran down.*  
*Hickory, dickory, dock!*

*One, Two, Buckle My Shoe*  
*One, two, buckle my shoe,*  
*Three, four, knock at the door.*  
*Five, six, pick up sticks,*  
*Seven, eight, lay them straight.*  
*Nine, ten, big fat hen.*

*Twinkle, Twinkle*  
*Twinkle, twinkle, little star,*  
*How I wonder what you are!*  
*Up above the world so high,*  
*Like a diamond in the sky.*  
*Twinkle, twinkle, little star,*  
*How I wonder what you are!*

*Jack and Jill*  
*Jack and Jill went up the hill*  
*To fetch a pail of water*  
*Jack fell down and broke his crown,*  
*And Jill came tumbling after*

*Star Light*  
*Star light, star bright*  
*First star I see tonight*  
*I wish I may, I wish I might*  
*Have the wish I wish tonight*

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Some Effects of Early Musical Experience on Sequence Memory Spans[1]

**Adam T. Tierney and David B. Pisoni**

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

[1] This research was supported by NIH-NIDCD Research Grant R01 DC00111 to Indiana University. The author would like to thank David Pisoni and Tonya Bergeson for their guidance with the research and comments on earlier versions of this paper.

# Some Effects of Early Musical Experience on Sequence Memory Spans

**Abstract.** Language and music were once considered completely separate systems, located in the left and right hemispheres of the brain, respectively. However, recent evidence from neuroimaging and behavioral studies has suggested that music and language may be closely linked. For example, syntactic processing and the processing of harmony may share certain neural resources, and some aspects of speech and music may be processed by overlapping working memory systems. If such overlap does exist, early experience with music and musical performance may affect not only the cognitive skills used to process music, but also language skills. The present study tested experienced musicians and three groups of musically inexperienced subjects, gymnasts, Psychology 101 students, and video game players on digit span, word span and sequence memory span tasks. By including skilled gymnasts who began studying their craft by age six, video game players, and Psychology 101 students, we attempted to control for some of the ways skilled musicians may differ from the general population, in terms of gross motor skills and intensive experience in a highly skilled domain form an early age. We found that musicians displayed higher memory spans than the comparison groups on digit span and auditory conditions of the sequence span task, but no differences were found between the four groups on the visual condition of the sequence span task. These results provide further converging support to recent findings showing that musical experience and activity may affect verbal rehearsal, phonological coding, and the allocation of attention in sequence memory tasks.

## Introduction

The idea that exposure to or training in music can make us smarter is perhaps most often associated in the public mind with the so-called “Mozart Effect.” This effect was first reported by Rauscher, Shaw and Ky (1993) who found that exposure to music written by Mozart, as compared to exposure to silence, led to increased scores on the Paper-Folding and Cutting subset of the Stanford-Binet IQ test (Thorndike, Hagen, & Sattler, 1986). Follow-up studies, however, have had a great deal of difficulty replicating the effect. Chabris (1999), in a review of 20 studies of the Mozart Effect, found an average cognitive enhancement of only 1.4 IQ points, while Thompson, Schellenberg, and Husain (2001) found that the cognitive enhancement was not reliable when arousal scores, enjoyment ratings, or subjective mood-arousal ratings were controlled. Nevertheless, the claim that musical experience can lead to cognitive facilitation lives on under a number of different guises. In particular, some researchers, pointing to evidence for a stronger neural link than previously supposed between music and language, argue that musical training facilitates verbal skills, especially verbal memory (Chan, Ho, & Cheung, 1998; Ho, Cheung, & Chan, 2003; Jakobson, Cuddy, & Kilgour, 2003; Koelsch et al., 2002; Patel et al., 1998).

Until recently, one of the only techniques available for studying possible neural overlap between music and language was to examine clinical cases of amusia and aphasia. Marin and Perry (1999) provide a comprehensive overview of these case studies, including several in which the patients suffered from amusia without aphasia, and vice versa. Other cases cited by Marin and Perry include disrupted memory for melodies and disorders of rhythm production with no aphasia, verbal alexia accompanied by preserved music reading and writing, and auditory agnosia restricted to speech sounds that did not impair nonverbal discrimination.

Additional evidence against overlap of music and language processing comes from a substantial body of research showing that music processing takes place primarily in the right hemisphere of the brain, while speech processing takes place primarily in the left. Zatorre and Sampson (1991), for example, asked patients with damage to the right or left anterior temporal lobes to compare two tones separated by six distractor tones, then to indicate if the tones were the same or different. Only the patients with right hemisphere damage showed inhibition on this task. In another study, Penhue, Zatorre, and Evans (1998) had subjects reproduce simple rhythms by tapping and found that the right planum temporale, a part of the auditory association cortex, experienced the greatest activation. Finally, Tervaniemi et al. (2000) used PET to investigate how subjects process musical sequences of "A" major chords interspersed with occasional "A" minor chords, as well as linguistic sequences of standard and occasional deviant phonemes. Their results showed that the deviant chords activated the right superior temporal gyrus most strongly while the deviant phonemes activated the left superior temporal gyrus most strongly.

These studies all support the proposal that music and language are processed by two independent systems, each modality-specific. However, the three studies cited all suffer from a common design problem: they used only non-musicians as subjects. Studies using skilled musicians as subjects have shown greater left hemisphere activation during tonal processing tasks, especially if the task is difficult. For example, Messerli, Pegna, and Sordet (1995) presented musicians and non-musicians with the melodies from eight popular folk songs and then asked both groups of subjects to point to one of eight pictures that best matched the song. Musicians displayed a right ear advantage while non-musicians displayed a left ear advantage. Also, Harris and Silberstein (1999) asked female musicians to memorize ten-note patterns while using EEG to measure their brain activity. The subjects were then presented with a part of the pattern and were asked to determine if any of the notes had been changed. The authors found that musicians showed mainly left temporal activation during this task and the effect was positively correlated with the difficulty of the task and subjects' degree of musical experience.

Thus, research involving skilled musicians has found that music processing is not confined to the right hemisphere, a finding that breathed new life into the question of possible neurobiological linkage between language and music processing. One of the first studies to set foot in this new area of research was reported in Patel, Gibson, Ratner, Besson, and Holcomb (1998). They found that the P600, an ERP correlate of processing of syntactic incongruities, is not specific to language but can also be elicited by chordal sequences containing harmonic incongruities. The amplitude of the P600s was correlated with the degree of violations of expectations in the sentences and chord sequences. Follow-up studies by Maess, Koelsch, Gunter, and Friederici (2001) and Koelsch et al. (2002) using MEG and fMRI respectively, found that processing of harmonic incongruities was also associated with activation in Broca's Area, Wernicke's Area, and the superior temporal sulcus.

While some researchers have investigated whether complex musical and linguistic abilities are processed in part by overlapping neural systems, other researchers have assessed whether musical and linguistic information are stored in a single auditory system within working memory or in two separate subsystems. In an early study, Deutsch (1970) played eight tones for subjects, then asked them to compare the first and last tone to determine if they were the same or different. In some conditions, however, spoken numbers replaced the middle six tones. The interpolated tones greatly impaired subjects' performance, but the spoken numbers had no significant effect on performance. Deutsch concluded that a separate system exists for the storage of tonal pitch in working memory.

Pechmann and Mohr (1992), however, contested Deutsch's conclusions. The authors used a similar experimental paradigm, but presented interpolated tones, spoken numbers, or visual stimuli between the two tones. Subjects included both instrumental musicians and non-musicians. While only the interpolated tones inhibited the musicians' performance, the tones, spoken numbers, and visual stimuli all significantly inhibited non-musicians' performance. The authors speculated that because non-musicians were less efficient in processing the tones, they were more susceptible to interference, even from less related stimuli.

Other researchers have attempted to determine the effects of hearing instrumental and vocal background music on various verbal memory tasks. Salamé and Baddeley (1989), for example, tested 24 non-musicians on recall of sets of nine visually presented digits. While the subjects performed this task, they heard instrumental background music, vocal background music, or silence. The subjects' performance was significantly inhibited by the presentation of instrumental music, but vocal music caused an even greater degree of interference.

However, this research on instrumental music and task interference ignores the fact that music consists of a number of different components. For instance, instrumental music contains significant rhythmic information—if music is stored in a "tonal loop," does it follow that rhythm is also stored there? Saito and Ishio (1998) addressed the question of whether rhythmic information is stored separately in working memory from phonological information. The researchers asked non-musicians to reproduce rhythmic patterns by pressing a key on a computer. During presentation of the rhythmic patterns, subjects were asked to either mouth the vowels "A-E-I-O-U" repeatedly, draw squares as many times as possible, or focus their attention on the rhythms. Performance in the concurrent articulation condition was worse than in the spatial task condition, which was worse than the control condition. The authors concluded that rhythmic information is retained in working memory via the phonological loop.

Taken together, the studies with musicians indicate that music and language may be processed by partially overlapping systems in the brain, both at the level of simple abilities, such as maintaining auditory information in working memory, and at the level of more complex abilities such as syntax and harmony processing. Moreover, research has shown that early experience with music can have effects on brain development that are strong enough to be detectable in the physical brain structure. In 1995, for example, Schlaug, Jancke, Huang, and Steinmetz, using MRI, found that a group of musicians with perfect pitch had a greater disparity between the left and right planum temporale (PT), as compared to non-musicians and musicians with relative pitch. The authors speculated that, given that PET has demonstrated that the PT is part of an area that is involved in music perception, this increase in left-right disparity might indicate increased lateralization of music processing in musicians with absolute pitch, as compared to non-musicians and musicians with relative pitch.

In a follow-up study, Schlaug, Jancke, Huang, Staiger, and Steinmetz (1995b) reported that the midsagittal area of the anterior half of the corpus callosum was larger in a group of musicians who had begun playing piano and stringed instruments by age seven than in a group of non-musicians. Using fMRI, Thomas, Pantev, Wienbruch, Rockstroh, and Taub (1995) found that string musicians had increased cortical representation of the fingers of their left hands, as compared to non-musicians. This finding was correlated with the age at which the person began to play music. Finally, Schneider et al. (2002), using fMRI, found that musicians, as compared to non-musicians, had a significantly greater volume of gray matter in Heschl's gyrus, especially in the anteromedial portion (that is, the primary auditory cortex).

Given the findings that early experience with music can cause measurable neurobiological changes, and the results that music and language processing may be psychologically and neurally linked, it is reasonable to ask what differences in verbal skills musicians might show when compared to non-musicians. Most prior research in this area, although suggestive, did not include appropriate comparison groups, but simply compared subjects with musical experience to subjects without musical experience. Any differences found in verbal skills could therefore be confounded by differences in any structured domain obtained from the early age at which most musicians begin playing. Confounding variables could include socioeconomic status, innate motor skills, or training effects.

The first researcher to attempt to correlate musical and non-musical cognitive skills was McMahon (1982). He found a positive correlation between performance on the Enticknap Picture Vocabulary Test, which requires children to name line drawings of objects on 48 test cards, and a tonal discrimination test designed by the experimenter. Further correlations between verbal and musical skills were discovered by Atterbury (1985), who reported that learning-disabled readers performed significantly worse on tonal discrimination and rhythmic discrimination tasks. Barwick, Valentine, West, and Wilding (1989) also found significant positive correlations between a tonal memory test and reading age, even after adjusting for chronological age and IQ. Finally, Lamb and Gregory (1993) reported significant positive correlations between a pitch discrimination test and tests requiring children to read nonsense words and demonstrate an understanding of rhyme and alliteration.

These correlational studies suggest that basic musical abilities, such as tone, rhythm, and chord discrimination, are related somehow to reading ability. However, the children with better musical skills may also possess better language skills because children with higher socioeconomic status and a more intellectual home environment may be exposed more often to certain aspects of language and music (their parents may read to them more often, for example). In an effort to move beyond mere correlation and determine whether or not early musical training actually affects subjects' non-musical cognitive skills, several researchers have turned to quasi-experimental studies in which subjects with musical experience are directly compared to subjects without musical experience.

In the first study of this kind, Huntsinger and Jose (1991) found that musically experienced children performed better than musically inexperienced children on tonal memory and digit span tasks. Similarly, Chan, Ho, and Cheung (1998) reported that musicians recalled more words from a visually presented 16-word list than did non-musicians. Kilgour, Jakobson, and Cuddy (2000) asked undergraduate musicians and non-musicians to memorize lyrics that were either spoken or sung, one line at a time. The musicians showed better recall of both the sung and spoken lyrics. Munzer, Berti, and Pechmann (2002) found that musicians performed better than non-musicians on a task which required them to decide if the first and last stimuli in a set of speech sounds were the same or different.

Finally, Jakobson, Cuddy, and Kilgour (2003) attempted to establish a theoretical basis for the finding that music training can lead to enhanced verbal recall. The researchers compared musicians and non-musicians on tests of verbal recall and temporal order processing, in which the subjects were asked to determine the order of presentation of two tones or two syllables of equal duration. The correlation between years of musical training and verbal recall scores was strong and significant, as was the correlation between musical training and the composite score for the temporal order tests. Using a multiple-regression analysis, the authors found that the relation between verbal recall scores and years of music training was reduced when temporal-order processing was included in the regression equation. Thus, it appears that the advantage that musicians show in verbal recall tasks may be due in part to their enhanced temporal order processing abilities.

However, one cannot infer causation from these results because it is possible that children with greater cognitive skills in the verbal/auditory domain are simply more likely to be drawn to music and that this predilection could account for the musicians' better performance on verbal memory tasks. In an attempt to establish whether or not the better performance on reading and verbal memory tasks that has been previously found actually results from musical training, several other researchers have performed more extensive, truly experimental studies with randomly selected experimental and control groups. Ho, Cheung, and Chan (2003), for example, found that a year of piano instruction led to improved verbal recall scores, as compared to a group that received no instruction. More recently, Schellenberg (2004) provided children with a year of piano instruction, a year of drama lessons, or no instruction, and found that the children who received piano instruction showed significantly greater increases in IQ than the control groups; this difference was found on all but two of the 12 subtests of the WISC-III IQ test (Wechsler 1991).

It is difficult to say whether these differences in verbal skills are truly due to musical training, because skilled musicians differ from non-musicians in many other ways unrelated to music, such as motor skill and socioeconomic status. One way to attempt to account for some of these possible differences between musicians and non-musicians is to test the verbal skills of a comparison group that has practiced a skilled motor task intensively from an early age. For this reason, we compared the ability of skilled musicians from the Indiana University School of Music to perform a number of short-term memory and working-memory tasks in both the spatial and verbal/auditory domains with that of subjects from three comparison groups, all of whom were unable to read music: (a) gymnasts from the IU Gymnastics Club, (b) Psychology 101 students at IU, and (c) students who played video games for at least 1.5 hours per day. Green and Bavelier (2003) found that video-game players showed facilitated processing on tests of visual attention and rapid temporal processing. It is possible, therefore, that video-game players would show differential facilitation to visual and auditory patterns. These four groups were tested on a number of short-term memory and working-memory tasks in both the visual-spatial and verbal-auditory domains to determine what effect, if any, prior early musical training and experience had on their ability to encode and recall information from immediate memory.

## Methods

### Subjects

Forty-five students participated in the study. A summary of the demographic characteristics of the participants is given in Table 1. By sending e-mails to the distribution list of the IU Gymnastics Club, we recruited ten gymnasts whom could not read music and started gymnastics when they were 4.5 years of age. They practiced an average of 4.5 hours per week. We also recruited 12 experienced music students who were piano performance majors by sending e-mails to the distribution list for all of the undergraduate and graduate piano performance majors. The mean age at which the pianists began playing the piano was 7 years. The average number of hours they spent practicing every week was 24.18. Twelve students were unselected undergraduates who could not read music and were enrolled in beginning psychology courses at IU. The psychology students were contacted via a list of students who indicated on a questionnaire at the end of class that they wished to participate in paid psychology experiments. Finally, we recruited 11 students who played video games for at least 1.5 hours a day, and were unable to read

| | Age | SAT Verbal | SAT Math | Num. Males | Num. Females |
|---|---|---|---|---|---|
| **Gymnasts** | 19.9 (SD 1.4) | 536 (SD 36) | 533 (SD 39) | 0 | 10 |
| **Musicians** | 21.1 (SD 2.3) | 650 (SD 51) | 638 (SD 75) | 7 | 5 |
| **Non-musicians** | 19.4 (SD 1.2) | 530 (SD 105) | 540 (SD 96) | 2 | 10 |
| **Video gamers** | 22.3 (SD 3.1) | 540 (SD 85) | 576 (SD 85) | 11 | 0 |

**Table 1.** Demographic data for gymnasts, musicians, Psychology non-musicians and video game players.

music. The difference in the ages between the four groups at the time of testing was not statistically significant ($p > .05$). All subjects were asked to report their Math and Verbal SAT scores, so we could assess any differences in academic and intellectual abilities between the groups. All subjects were paid $15 for their time.

## Stimulus Materials

The subjects completed several short-term memory and working memory tasks including a digit span task, a word span task, and a word familiarity test. For the digit span task, ten spoken digits were acquired from the Texas Instruments 46-Word (TI46) Speaker-Dependent Isolated Word Corpus (Texas Instruments, 1991). Test words for the word span tasks were obtained from a digital speech database (Goh & Pisoni, 2003) containing the 300 monosyllabic English words from the Modified Rhyme Test (House, Williams, Hecker & Kryter, 1965) and from phonetically balanced word lists. Sixty-six "easy" consonant-vowel-consonant (CVC) words and 66 "hard" CVC words, whose lexical difficulty was computed according to the neighborhood activation model (Luce & Pisoni, 1998), recorded by a single male voice were chosen so that the two sets consisted of different neighborhood density and neighborhood frequency but did not differ on word frequency and the number of intra-set neighbors.

The test words for the word-familiarity (FAM) vocabulary test were taken from Stallings, Kirk, Chin, and Gao (2002), a shortened version of the FAM test originally developed by Lewellen, Goldinger, Pisoni, and Greene (1993). The 150-word stimulus list consisted of 50 high-familiarity words, 50 medium-familiarity words, and 50 low-familiarity words. The familiarity scores were based on normed familiarity ratings of 20,000 words collected originally by Nusbaum, Pisoni, and Davis (1984).

## Experimental Design and Procedures

**Digit Span Task.** Subjects were asked to recall lists of spoken digits, presented over headphones. The list length was increased by one item after every two trials, starting at a list length of four and ending with a list length of ten, with a total of 14 trials for each test. For the first test, Forward Digit Span, subjects were asked to recall and write down the digits as they were given. For the second test, Backward Digit Span, subjects were asked to recall and write down the digits in the reverse order.

**Word Span Task.** Subjects were asked to recall lists of spoken words, presented over headphones. The list length was once again increased by one item after every two trials, starting at a list length of three words and ending with a list length of eight, for a total of 12 trials for each test. Half of the subjects from each group completed the "easy" word span test first and the other half completed the "hard" word span test first.

**Word Familiarity Ratings.** This is a test of subjective familiarity with easy, medium difficulty, and hard words. Fifty words from each category were presented randomly on a computer screen, one at a time, for a total of 150 words. The subjects were asked to judge their familiarity with the word by pressing the appropriate number key on the keyboard, from "1" for least familiar to "7" for most familiar.

**Sequence Reproduction Span Task.** This task used a modified version of the Simon Memory Game, a round box with four colored panels (Karpicke & Pisoni, in press; Pisoni & Cleary, 2004). In the first condition, visual-only (VO), the colored panels lit up, one at a time, in a sequence. The subjects were then asked to reproduce the sequence of colored lights by pressing the appropriate panels on the box. In the second condition, auditory-only (AO), the subjects heard a sequence of color names over their headphones. They were asked to reproduce the sequence by pressing the appropriate panels on the box. In the final condition, audiovisual (AV), the panels on the box were illuminated and the subjects simultaneously heard names of colors that corresponded to the colors of the lights. In each of the three presentation conditions, the sequence length started out at one item, then increased by one whenever the subjects reproduced a particular sequence length correctly twice in a row. If the subjects reproduced a sequence incorrectly, the sequence length decreased by one item. Each condition (VO, AO, and AV) consisted of twenty trials.

All subjects first completed the forward and backward digit span tasks, followed by the easy and hard word span tasks. After a five-minute break, the subjects then completed the word familiarity task and the Simon sequence memory task.

## Results

### Demographic Data

A series of one-way ANOVAs were performed on the subject demographic data. SAT scores were reported by 67% of musicians, 72% of gymnasts, 80% of Psychology students, and 63% of video game players. Mean SAT verbal scores were found to be significantly higher for the musicians than for the gymnasts, Psychology students, and video game players ($p$ = .003, .002, and .003, respectively.) Mean SAT math scores were significantly higher for the musicians as compared to the gymnasts ($p$ = .014).

### Forward and Backward Digit Spans

Due to a corrupted file, the digit span data for one of the gymnast subjects was lost; therefore, the N for the gymnasts on the digit span tasks was 9 not 10. Four scores were computed to measure digit spans. “High” was the longest span which the subjects reached without making any errors. After making an error, 0.5 was added for every full list repeated correctly, resulting in the “Strict” score. For the “Absolute” score, the number of items in each list recalled correctly was summed, and finally, for the “Total” score, the total number of items recalled correctly was computed. Only the results of the “Total” scores will be reported here because the other scores did not vary widely enough to show consistent effects. Figure 1 shows mean scores for the four subject groups in both the forward and backward digit span recall conditions.

The two digit span scores were entered into a repeated measures ANOVA, with group as a between-subjects variable and type of span (forward versus backward) as a within-subjects variable. Main effects of span [$F(1,40) = 61.73$, $MS_e = 3205.004$, $p < .001$] and group [$F(3, 40) = 4.994$, $MS_e = 935.600$, $p = .005$] were found. Post-hoc Bonferroni tests revealed that musicians significantly outperformed the gymnasts on the forward digit span ($p$ = .011) and significantly outperformed the gymnasts and the P101 non-musicians on the backward digit span ($p$ = .049 and .014, respectively). No other significant differences in performance were found between the video game players and the remaining three groups for either forward or backward span.

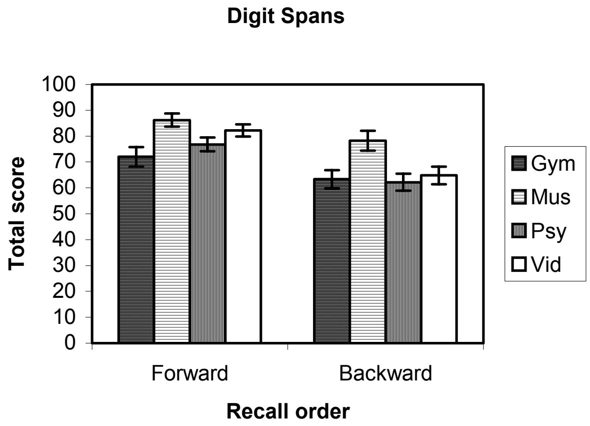


**Figure 1.** Mean total span scores for the four groups in both the forward and backward digit span recall conditions. Error bars are standard error.

## Word Span

Scoring for the word span task was identical to the procedure used in the digit span task. Figure 2 shows mean "Total" scores for the four subject groups in both of the conditions (easy and hard word span). The word span scores were entered into a repeated measures ANOVA, with group as a between-subjects variable and type of span (hard vs. easy) as a within-subjects variable. A significant interaction between span and group was found [F(3,40) = 3.072, $MS_e$ =87.517, p = .039.] No main effects of group or span type were found. Post-hoc analysis revealed that musicians performed significantly better than the video game players on the hard word span condition (*p* = .014); none of the other differences reached significance.

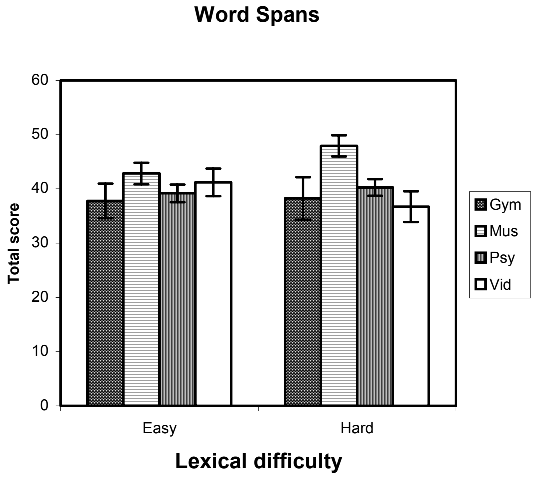


**Figure 2.** Total items recalled correctly for the four groups of subjects for the two types of word span. Error bars are standard error.

## Word Familiarity Ratings

Figure 3 shows mean familiarity ratings for the three subject groups in each of the three different word familiarity conditions (high-familiarity, medium-familiarity, and low-familiarity). The FAM ratings were entered into a repeated measures ANOVA, with group as a between-subjects variable and word familiarity (low, medium, and high) as a within-subjects variable. A main effect of familiarity [$F(2,60) = 767.037$, $MS_e = 168.290$, $p < .001$] was found. Post-hoc analysis revealed no significant differences between the four groups. Musicians were not more likely to rate words of varying difficulty as more familiar than the three comparison groups.

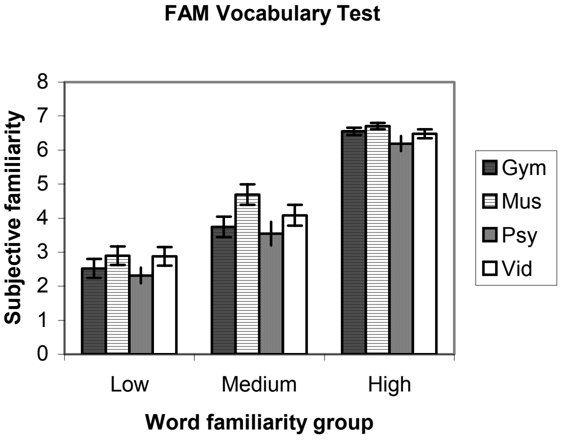


**Figure 3.** Mean subjective familiarity ratings for the four groups in the FAM rating task. Error bars are standard error.

## Simon Sequence Memory Span

Four different memory span scores were computed: ALL was the longest sequence which the subjects reproduced correctly 100% of the time, HALF was the longest span which the subjects reproduced correctly 50% of the time, ONCE was the longest span which the subjects reproduced correctly at least once, and WEIGHT was a weighted sum of the percentage of correct responses at each list length (see Pisoni & Cleary, 2004). Analyses were carried out on each of the four different span methods. A similar pattern of results was found for each type of score, so only the results of the weighted score will be reported here. Figure 4 shows mean weighted span scores for the four subject groups in each of the three different presentation conditions (VO, AO, and AV).

The Simon weighted span scores were entered into a repeated measures ANOVA, with group as a between-subjects variable and presentation condition (VO, AO, and AV) as a within-subjects variable. Main effects of Simon condition [$F(2,82) = 16.88$, $MS_e = 4.307$, $p < .001$] and group [$F(3, 41) = 4.565$, $MS_e = 4.548$, $p = .008$] were found. A significant interaction between condition and group [F(6, 82) = 3.803, $MS_e = .970$, $p = .002$] was also found. None of the pairwise differences in VO were significantly different. For the AO condition, the musicians (mean score 6.48, *SD* =.92) outperformed the gymnasts (mean score 5.43, *SD* =.63, $p = 0.006$), the psychology students (mean score 5.02, *SD* =.33, $p < 0.001$), and the video game players (mean score 5.29, SD =.65, $p = .001$). No significant difference in performance was found between groups for the AV condition, although there was a numerical trend for the musicians to display longer spans in this condition too.

Post hoc Bonferroni tests revealed that, overall, subjects did not perform better on the AO condition than on the VO condition ($p = .11$). However, they did show significantly better performance on the AV condition as compared to the VO condition ($p < .001$) and the AO condition ($p < .001$). Further analysis revealed that the difference between the AO and AV conditions was only significant for the psychology subjects ($p = .002$). The difference between the VO and AO conditions was significant only for the musicians (*p* = .011). Thus, among the four groups of subjects only the musicians were able to reproduce auditory patterns better than visual patterns.

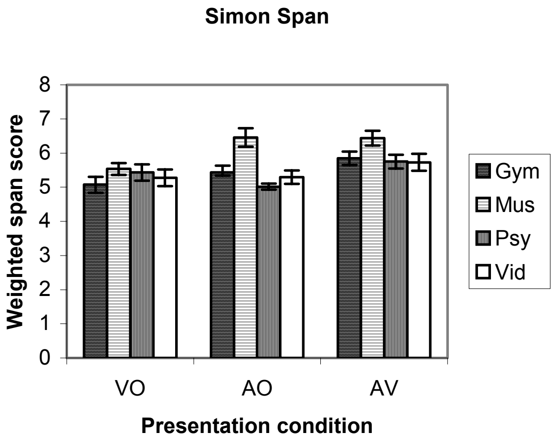


**Figure 4.** Weighted span scores for the four groups of subjects in the three different conditions (VO, AO, and AV) of the Simon Span task. Error bars are standard error.

## Correlations

A summary of the intercorrelations between the four different tests can be found in Table 2. Simon VO, AO, and AV conditions were all significantly correlated with one another. The AO and AV conditions also correlated significantly with the Backward Digit Span measure, but not with the Easy and Hard Word Span measures or the Forward Digit Span measure. The Digit Span and Word Span tasks were all significantly correlated with one another.

| | Digit Span | | Word Span | | FAM Familiarity Test | | | Simon Span | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | FDS | BDS | Easy | Hard | FAM-low | FAM-mid | FAM-high | VO | AO | AV |
| Forward Digit Span | 1 | | | | | | | | | |
| Backward Digit Span | .634** | 1 | | | | | | | | |
| Easy Word Span | .630** | .428** | 1 | | | | | | | |
| Hard Word Span | .432** | .483** | .557** | 1 | | | | | | |
| Familiarity test--low | .137 | .193 | .002 | .131 | 1 | | | | | |
| Familiarity test--mid | .256 | .362* | .045 | .182 | .868** | 1 | | | | |
| Familiarity test--high | .165 | .274 | .060 | .063 | .587** | .809** | 1 | | | |
| Simon VO | .213 | .236 | .213 | .292 | -.151 | -.147 | -.148 | 1 | | |
| Simon AO | .292 | .428** | .054 | .134 | .062 | .183 | .173 | .352* | 1 | |
| Simon AV | .274 | .351* | .205 | .192 | -.179 | -.041 | -.048 | .505** | .665** | 1 |

**Table 2.** Pearson correlations. A (*) signifies that the correlation is significant at the .05 level; (**) signifies that the correlation is significant at the .01 level.

## Discussion

Skilled musicians performed significantly better than the gymnasts, introductory psychology students, and video game players on the auditory-only (AO) condition of the Simon memory span. No significant differences were found on the visual-only (VO) condition or the audio-visual (AV) condition of this task. These new findings on sequence memory span suggest that skilled musicians display a greater capacity for reproducing randomized auditory sequences than non-musicians. The present findings on sequence memory span provide an excellent counterargument to those who would claim that any cognitive differences between skilled musicians and non-musicians might simply be due to general intelligence, or to global differences in academic achievement. Effects of general intelligence and education are not likely to be limited to a single domain or sensory modality (Gardner 1993). Moreover, the facilitation of musician performance in the AO condition cannot be due to differences in motor skills, because the output for each of the three conditions involved exactly the same response demands: pressing a sequence of buttons on the Simon box. The only difference between the three conditions was the modality of the input signals.

The data obtained from the skilled musicians for the three Simon conditions also showed patterns that were different from those found in previous Simon experiments that used psychology students from the general university population. Pisoni and Cleary (2004) reviewed several studies that tested normal-hearing adults and normal-hearing 8- and 9-year-old children on the AO, VO, and AV conditions of the Simon memory span. Children performed worse overall than adults, but both groups showed two significant effects. First, both groups of subjects displayed a "modality effect: performance was better on AO than on VO. Secondly, both groups of subjects exhibited a "redundancy gain": performance was better in the AV condition than in either the VO or AO conditions. Both findings suggest that sequence memory spans are not only sensitive to input modality but also display cross-modal interactions between auditory and visual signals.

In contrast to Pisoni and Cleary's earlier findings with adults and normal-hearing children, in the present study skilled musicians showed a strong redundancy gain when the AV condition was compared to the VO condition, but no redundancy gain when the AV condition was compared to the AO condition. Of the four groups tested in this experiment, only the skilled musicians exhibited a significant modality effect, with performance on AO greater than performance on VO, although the gymnasts showed a trend in the same direction. When performance was averaged across all four groups, we did find an overall redundancy gain (that is, performance on AV was better than on VO and AO), but we found no significant modality effect. These differences may be due to the small number of subjects that we were able to recruit for this study. Pisoni and Cleary (2004) did find a strong modality effect, but they used 48 subjects in the study.

Although we replicated the redundancy gains reported earlier by Pisoni and Cleary (2004) and found a trend towards a modality effect, skilled musicians displayed little benefit from the redundant visual information presented in the AV condition as compared with the AO condition. The other three groups of subjects all showed facilitation of performance in the AV condition when compared to the AO condition, though this trend reached significance only for the psychology subjects. The auditory information was so salient to the musicians that adding redundant visual information provided little additional benefit. Alternatively, the musicians may have used a different encoding strategy in the AO condition than the other groups, which relied more heavily on the temporal coding of auditory information in these sequential patterns.

It is possible that the lack of a redundancy gain for musicians when AO was compared to AV was due to a tendency of musicians to selectively focus their attention on auditory patterns, even random auditory patterns, to the exclusion of other sources of sensory information. Future studies could test this hypothesis directly by decoupling the auditory and visual input in two "A–V" selective attention conditions. In one condition, the subjects would be asked to reproduce the visual pattern (i.e., VO) and selectively ignore irrelevant auditory input (that is, the names of colors that are not congruent with the colored lights). In another condition, the subjects would be asked to reproduce the auditory patterns (i.e., AO) and selectively ignore incongruent visual input. If musicians' attention is controlled or modulated by the auditory input, their performance should be facilitated in the A-V condition in which they are asked to reproduce the auditory input and ignore the visual input, but inhibited in the V-A condition in which they are asked to reproduce the visual input and selectively ignore the auditory input. In a recent study on automaticity in skilled musicians, Stewart, Walsh, and Frith (2004) asked subjects to read numbers from 1 to 5 embedded in five different musical notes; the numbers indicated which finger subjects should use to press a key. Skilled pianists, but not musically inexperienced subjects, were facilitated when the irrelevant musical stimuli were congruent with the numbers, but showed interference when the musical and numerical stimuli were not congruent. These results demonstrate that musical experience can affect the automaticity of specific information processing tasks; just as musicians in this study were not able to ignore the irrelevant musical notation, it is possible that musicians in an A–V Simon span condition might not be able to ignore irrelevant auditory information.

The musicians' long experience and activities in listening to and actively playing music appears to facilitate verbal auditory processing of temporal sequences. It is not clear, however, whether this effect is due to enhanced encoding strategies during early perceptual analysis or are the result of differences in the musicians' use of the central executive, that is, to enhanced manipulation of auditory-verbal information actively maintained in short-term memory. The differences in performance found between the groups could also be due to storage or processing activities. The digit span results can be used to try to further pinpoint the exact locus of the musicians' verbal memory facilitation. Musicians outperformed the gymnasts on the forward digit span. The advantage shown by musicians was slightly greater for the backward digit span task: musicians performed significantly better than both the gymnasts and the psychology subjects. There was, however, no interaction between span and group.

Because there was no interaction between span and group in the digit span task, it is possible that the musicians' improved performance on the Simon AO and digit span tasks may simply be due to more efficient phonological encoding, rather than active manipulation of information in working memory. Data from neuro-imaging studies have suggested that backward digit span is a measure of working memory span rather than passive short-term memory. Several recent fMRI studies have revealed that storage of information is associated with increases in regional cerebral blood flow (rCBF) in ventrolateral frontal cortex, which is found in both short-term memory tasks and working memory tasks. Manipulation and active processing of information, on the other hand, is associated with increases of rCBF in dorsolateral frontal cortex, which is generally found only in tasks thought to require the use of the central executive to actively manipulate information maintained in short-term memory (Owen, 2000). Researchers have found that the backward digit span tasks lead to activation of both ventrolateral and dorsolateral prefrontal cortex, while the forward digit span leads to activation of ventrolateral prefrontal cortex only (Owen, Lee, & Williams, 2000). Factor analysis performed on studies examining subjects' performance on various tasks thought to tap into working memory and short-term memory has provided support for the dissociation between these two memory processes, and has revealed that working memory more often correlates with measures of reading comprehension such as Verbal SAT scores (LaPointe & Engle, 1990; Cantor, Engle, & Hamilton, 1991; Kail & Hall, 2001).

The results of this study suggest that musicians show enhanced encoding of information, rather than enhanced ability to actively manipulate information. It is still not clear, however, whether the Simon sequence memory span task, with which the largest effects of musical experience are found, is solely a test of encoding or if it also involves a significant processing and manipulation component. The AO condition, for example, requires subjects not only to maintain the list of color names in short-term memory but also to convert the color names into the corresponding locations on the button box and execute the motor pattern necessary to activate the buttons. Correlational analysis on the results of this study revealed that scores on both the AO and AV conditions correlated significantly with the backward digit span task, but not with the forward digit span task.

Several differences in performance were found between the four groups on the word span task. Musicians performed significantly better than the video game players on the hard word span task, and a trend was found for better performance by the musicians than the gymnasts on the word span task. However, no significant effects were found on the easy word span task. The difficulty in finding any differences between musicians and non-musicians on the words span tasks may be due to the differences in performance between open-set and closed-set tasks (Sommers, Kirk & Pisoni, 1997); it is possible that the musicians were facilitated only on matching of auditory stimuli to stored patterns, and that when the lexicon must be accessed this advantage vanishes. On the other hand, the difficulty in finding effects may be due to the small number of subjects, and increasing the power of the study might bring the difference between musicians and psychology subjects to significance. Subjects did not show a decrement in performance on the "hard" word span task as compared to the "easy" word span task, a somewhat anomalous result. Goh and Pisoni's (2003) study, which did find a decrement in performance for the "hard" words as compared to the "easy" words, used 56 subjects and counterbalanced the conditions. The current study, however, used 45 subjects and always presented subjects with the "easy" condition first, followed by the "hard" condition. A practice effect may, therefore, have led subjects to perform better on the "hard" condition than they would have in an experiment with counterbalanced conditions.

The results obtained from the comparison group of students who began practicing gymnastics from an early age suggest that the differences between musicians and non-musicians on the span tasks are not the result of gross motor skills or knowledge of a skilled domain from an early age. Future examinations of the effects of early music experience would do well to include multiple comparison groups drawn from other specialized subsets of the population. A 1997 survey of 663 children taking private piano lessons, for example, revealed that the children's parents were mainly college-educated, professional, upper to upper-middle income, Caucasian suburbanites (Duke, Flowers, and Wolfe, 1997). If proposed studies of the cognitive differences between musicians and non-musicians do not attempt to take into account the many extramusical ways in which musicians differ from the general population, it is difficult if not impossible to determine precisely whether or not the results are in fact due to the subjects' selective experience and activity with music.

Several important differences between the skilled musicians and the three comparison groups remain, and the number of subjects was too small to adjust for even those differences that could be measured. The most prominent difference between the subject groups was found in their SAT scores. Musicians had significantly higher SAT scores than the gymnasts and psychology students, both for the Math and Verbal sections. (The effect was larger for the Verbal scores than it was for the Math scores.) This result offers the greatest challenge to the claim that the Simon AO results reflect a difference that stems primarily from the intensive experience that the musicians have had with auditory stimuli from an early age. One could argue, then, that the musicians have been exposed to better quality education and, perhaps, a more intellectual home environment, that this difference has increased the amount of time they have spent on homework and reading, and that this experience with processing visually presented verbal stimuli (that is, reading) has somehow fine-tuned their ear for speech. The differences in SAT scores may also reflect the very selective nature of the School of Music at Indiana University, which is more competitive in admission than the College of Arts and Sciences. According to this argument, the differences in auditory performance between the musicians and non-musicians are not the result of musical training or selective experience with sound and auditory patterns but instead stem from the musicians' higher socioeconomic status and greater educational opportunities.

This hypothesis, however, does not account for the selective failure to find any difference at all between the four groups on the VO condition. If overall educational achievement is responsible for the musicians' enhanced auditory memory spans, then their visual memory spans in VO should have been improved as well. Many academic fields, especially mathematics, involve the use of visual-spatial processing; therefore, the difference in quantitative SAT scores should go hand-in-hand with differences in visual working memory capacity. The argument that these differences are due to educational history could be addressed in a follow-up study with psychology students whose SAT scores closely matched to those of the musicians. The group of video game players was included in an attempt to find subjects who would display precisely this pattern of performance, but they did not perform significantly better on the VO condition than the other three groups of subjects.

In conclusion, skilled musicians displayed significantly longer auditory reproductive sequence memory spans than three groups of non-musicians: gymnasts, introductory psychology students, and video game players, all who were unable to read music. This effect was highly selective in nature and was observed only for the AO condition, and not for the VO and AV conditions. Musicians also performed significantly better on Forward and Backward Digit Span tasks than the other three comparison groups. Analyses of the digit span measures revealed no significant interaction between span and group. Thus, it appears that early experience with music facilitates performance on memory span tasks and because it affects both forward and backward digit spans, it is therefore more likely to be related to early encoding and verbal rehearsal-maintenance rather than active manipulation or transformation of pattern information already in short-term memory.

## References

Atterbury, B. (1985). Musical differences in learning-disabled and normal-achieving readers, aged seven, eight, and nine. *Psychology of Music*, *13*, 114-123.

Barwick, J., Valentine, E., West, R., & Wilding, J. (1989). Relations between reading and musical abilities. *British Journal of Educational Psychology*, *59*, 253-257.

Cantor, J., Engle, R.W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*, 229-246.

Chabris, C.F. (1999). Prelude or requiem for the 'Mozart effect'? *Nature*, *400,* 826-827.

Chan, S., Ho, Y., & Cheung, M. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*, 1604-1605.

Duke, R., Flowers, P., & Wolfe, D. (1997). Children who study piano with excellent teachers in the United States. *Bulletin of the Council for Research in Music Education*, *132*, 51-84.

Gardner, Howard. (1993). *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books.

Goh, W., & Pisoni, D. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly Journal of Experimental Psychology: Section A*, *56*, 929-954.

Green, S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534-537.

Harris, P., & Silberstein, R (1999). Steady-state visually evoked potential (SSVEP) responses correlate with musically trained subjects' encoding and retention Phases of musical working memory task performance. *Australian Journal of Psychology*, *51*, 140-146.

Ho, Y.C., Cheung, M.C. & Chan, A.S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*, 439-450.

House, A., Williams, C., Hecker, M., and Kryter, K. (1965). Articulation-testing methods: consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America* 37, 158-166.

Huntsinger, C., & Jose, P. (1991). A test of Gardner's modularity theory: A comparison of short-term memory for digits and tones. *Psychomusicology*, *10*, 3-18.

Jakobson, L., Cuddy, L. & Kilgour, A. (2003). Time tagging: A key to musicians' superior memory. *Music Perception*, *20*, 307-313.

Kail, R., & Hall, L.K. (2001). Distinguishing short-term memory from working memory. *Memory and Cognition*, *29*, 1-9.

Karpicke, J. & Pisoni, D.B. (in press). Using immediate memory span to measure implicit learning. *Memory and Cognition.*

Kilgour, A., Jakobson, L., & Cuddy, L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory and Cognition*, *28*, 700-710.

Koelsch, S., Gunter, T., Cramon, D., Zysset, S., Lohmann, G., & Friederici, A. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, *17*, 956-966.

Lamb, S., & Gregory, A. (1993). The relationship between music and reading in beginning readers. *Educational Psychology*, *13*, 19-28.

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.

Lewellen, M., Goldinger, S., Pisoni, D., & Greene, B. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, *3*, 316-330.

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

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, *4*, 540-545.

Marin, O., & Perry, D. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The Psychology of Music*. Academic Press.

McMahon, O. (1982). A comparison of language development and verbalization in response to auditory stimuli in pre-school age children. *Psychology of Music*, *12*, 94-105.

Messerli, P., Pegna, A., & Sordet, N. (1995). Hemispheric dominance for melody recognition in musicians and non-musicians. *Neuropsychologia*, *9*, 97-113.

Munzer, S., Berti, S., & Pechmann, T. (2002). Encoding of timbre, speech, and tones: Musicians vs. non-musicians. *Psychologische Beitrage*, *44*, 187-202.

Nusbaum, H., Pisoni, D., & Davis, C. (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.

Owen, A. (2000). The role of the lateral frontal cortex in mnemonic processing: The contribution of functional neuroimaging. *Experimental Brain Research*, *133*, 33-43.

Owen, A., Lee, A., & Williams, E. (2000). Dissociating aspects of verbal working memory within the human frontal lobe: Further evidence for a "process-specific" model of lateral frontal organization. *Psychobiology*, *28*, 146-155.

Patel, A., Gibson, E., Ratner, J., Besson., M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neurosci*ence, *10*, 717-33.

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory and Cognition*, *20*, 314-320.

Penhue, V.B., Zatorre, R., & Evans, A. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *Journal of Cognitive Neuroscience*, *19*, 752-765.

Pisoni, D.B. & Cleary, M. (2004). Learning, memory and cognitive processes in deaf children following cochlear implantation. In F.G. Zeng, A.N. Popper & R.R. Fay (Eds.), *Springer Handbook of Auditory Research: Auditory Prosthesis, SHAR Volume X*. Pp. 377-426.

Rauscher, F.H., Shaw, G.L, & Ky, K.N. (1993). Music and spatial task performance. *Nature, 365*, 611.

Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: Effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research*, *40*, 10-19.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 41*, 107-122.

Schellenberg, E.G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511-514.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267(5198)*, 699-701.

Schlaug, G., Jancke, L., Huang, Y., Staiger, J., & Steinmetz, H. (1995b). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*, 1047-1055.

Schneider, P., Scherg, M., Dosch, H., Specht, H., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*, 688-694.

Sommers, M., Kirk, K., & Pisoni, D. (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.

Stallings, L., Kirk, K., Chin, S, & Gao, S. (2002). Parent word familiarity and the language development of pediatric cochlear implant users. *The Volta Review*, *102*, 237-258.

Stewart, L., Walsh, V. & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*, 183-195.

Tervaniemi, M., Medvedev, S., Alho, K., Pakhomov, S., Roudas, M., Van Zuijen, T., & Naatanen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, *10*, 74-79.

Texas Instruments. (1991). TI 46-word speaker-dependent isolated word corpus (CD-ROM). Gaithersburg: NIST.

Thomas, E., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*, 305-307.

Thompson, W., Schellenberg, G. & Husain, G. (2001). Arousal, mood, and the Mozart Effect. *Psychological Science*, *12*, 248-251.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). *The Stanford-Binet Scale of Intelligence*. Riverside: Chicago.

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

Zatorre, R., & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain*, *114*, 2403-2417.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Behavioral Inhibition and Audiological Outcomes in Prelingually Deaf Children with Cochlear Implants[1]

**David L. Horn,[2] Rebecca A.O. Davis[2] 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 T32 DC00012 and NIH-NIDCD Research Grants R01 DC00064, NIH NIDCD R01 DC00423 to Indiana University. We would like to thank all of the speech/language pathologists and other clinicians who helped in testing the children, Sujuan Gao and Amy Rong QI for their statistical assistance, and Josh Bradley for his comments on an earlier version of this paper. These findings were presented in a podium session at the International Cochlear Implant Conference in Indianapolis, IN, May 2004. This is a draft of a paper to be submitted for publication.

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

# Behavioral Inhibition and Audiological Outcomes in Prelingually Deaf Children with Cochlear Implants

**Abstract.** Cochlear Implants (CIs) enable many prelingually deaf children to acquire spoken language skills although individual speech and language outcomes are quite varied. Differences in a range of underlying cognitive skills may explain a portion of this variance. The aim of this study was to determine if variation in behavioral inhibition skills of prelingually deaf children were related to speech perception, language, vocabulary knowledge, and speech intelligibility outcome measures after implantation. We conducted a retrospective analysis of longitudinal data collected from prelingually and profoundly deaf children who used CIs. A continuous performance task that did not require any auditory processing was used to measure behavioral inhibition skills. Audiological outcomes based on a battery of speech and language measures were obtained from children after 1, 2, and 3 years of CI use. Compared to published norms, the children in our sample performed in the low-normal range and delay task performance improved as a function of chronological age as well as length of CI use. A correlational analysis revealed several significant associations between behavioral inhibition and receptive and expressive language, vocabulary knowledge, and speech intelligibility scores. Most of these relations remained significant even when the effects of length of CI use and chronological age were partialled out. These findings suggest that speech and language processing skills are closely related to the development of verbal encoding skills, subvocal rehearsal skills, and verbally mediated self-regulatory skills in prelingually deaf children with CIs.

## Introduction

Cochlear Implants are sensory aids that have been approved for the treatment of profound prelingual deafness in children as young as one year of age (US FDA, 2000). Numerous studies of audiological outcomes in deaf children with CIs have consistently demonstrated benefit of a CI on the development of oral language abilities such as speech perception, expressive language, vocabulary knowledge, and speech intelligibility (Miyamoto, Svirsky, Kirk, Robbins, Todd, & Riley, 1997; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000; Tyler, Fryauf-Bertschy, Kelsay, Gantz, Woodworth, & Parkinson, 1997). Despite the impressive results of CIs in many deaf children, the significant individual differences in audiological outcomes of prelingually deaf children remain a challenging clinical and theoretical problem (Blamey, Sarant, Paatsch, Barry, Bow, Wales, Wright, Psarros, Rattigan, & Tooher, 2001; Pisoni, Cleary, Geers, & Tobey, 2000; Sarant, Blamey, Dowell, Clark, & Gibson, 2001). Reported outcomes vary a great deal. Some children obtain age-appropriate speech while others vary from children who obtain age-appropriate speech and language skills with their CI to others who obtain little benefit other than an awareness of sound. Studies which have examined predictors of CI performance in children have reported effects of early implantation, communication mode, device type, and dynamic range (Geers, 2003; Geers, Brenner, & Davidson, 2003; Geers, Nicholas, & Sedey, 2003; Kirk, 2000; Tobey, Geers, Brenner, Altuna, & Gabbert, 2003). However, a significant portion of outcome variance still remains unexplained by these demographic and medical factors (Pisoni et al., 2000; Sarant et al., 2001).

Recently, a number of investigators have suggested that cognitive and behavioral factors may play a role in acquiring spoken language skills with a cochlear implant (Dawson, Busby, McKay, & Clark, 2002; Knutson, Ehlers, Wald, & Tyler, 2000a; Knutson, Ehlers, Wald, & Tyler, 2000b; Pisoni, 2000). Individual differences in attention, learning, memory, and executive function of deaf children with cochlear implants may explain a portion of the variance in audiological outcomes following implantation with a CI. Recent studies of working memory have demonstrated close relations between deaf children's ability to encode, temporarily store, and reproduce, sequential stimuli such as digit lists and their performance on a wide range of speech perception, production, and language tests (Burkholder & Pisoni, 2003; Dawson, Busby, McKay, & Clark, 2002; Pisoni, 2000; Pisoni & Cleary, 2003; Pisoni & Geers, 2000; Surowiecki, Sarant, Maruff, Blamey, Busby, & Clark, 2002). Other research has shown that a deaf child's ability to obtain benefit from audiovisual redundancy in a sequence memory task is closely related to their spoken language skills with a CI (Cleary, Pisoni, & Geers, 2001; Lachs, Pisoni, & Kirk, 2001).

Compared to the recent literature on working memory, we know relatively little about the relations between CI outcomes and other cognitive functions. Several studies of sustained visual attention in deaf children with CIs have demonstrated an effect of CI experience, suggesting that this cognitive ability is closely tied to auditory and verbal language experience (Mitchell & Quittner, 1996; Quittner, Smith, Osberger, Mitchell, & Katz, 1994; Smith, Quittner, Osberger, & Miyamoto, 1998; Horn, Davis, Pisoni, & Miyamoto, under review). Only one of these studies actually examined relations between measures of sustained attention and audiological outcomes and this study did not find any significant correlations although children in this study had only used their CI for one year (Horn et al., under review). We know even less about executive functions in deaf children with cochlear implants. A recent study by Surowiecki et al. (2002) examined the relations between several measures of executive functions and CI outcomes in deaf children and did not find significant correlations, although the authors did not control for length of CI experience.

Knutson and colleagues investigated the relations between behavioral factors and audiological outcomes in deaf children with CIs (Knutson, Ehlers, Wald, & Tyler, 2000a; Knutson et al., 2000b). They tested children who had used their CIs for 36 months on the Child Behavior Checklist (CBCL) (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986). This standardized parental report measures a variety of behavioral problems and produces composite standardized behavior scores for externalizing, internalizing, and total behavior problems. Externalizing behaviors on the CBCL include attention problems, impulsivity, aggression, and rule-breaking in contrast to internalizing behaviors which include withdrawal, depression, and anxiety. Knutson et al. reported that the externalizing and total problem composites were negatively correlated with performance on a range of speech perception and language tests (Knutson et al., 2000b). Their results suggested that a child's ability to control their behavior, particularly when behavioral inhibition was required, was related to their oral language skills.

The importance of Knutson et al.'s findings is magnified by a number of earlier studies which have found that deaf children and adults who use sign language appear to be more behaviorally impulsive than their peers (Altshuler, Deming, Vollenweider, Rainer, & Tendler, 1976; Chess & Fernandez, 1980; Kelly, Kelly, Jones, Moulton, Verhulst, & Bell, 1993; O'Brien, 1987). Therefore, it is critical to better understand how behavioral control, language development, and cochlear implant use interact during early development. However, we currently know little about the behavioral inhibition skills of deaf children who use CIs.

One aspect of behavioral control that has received some attention in the field of pediatric neuropsychology is response delay. This cognitive skill is thought to be distinct from the executive functions and is crucial for analysis and control of behavior (Barkley, 1997). Response delay includes related skills including inhibition of action, maintenance of a delayed state, and prevention of distraction from extemporaneous stimuli. These skills are critical for self-directed, goal-oriented behaviors. Impairments in behavioral inhibition and response delay are thought to be a central deficit in some types of Attention Deficit Hyperactivity Disorder (Barkley, 1997).

In the only study of response delay skills of deaf children with CIs to date, Mitchell and Quittner (1996) tested children on several continuous performance tasks. One of these tasks, the response delay task, required children to make serial key presses separated in time by a specific length. This length (4 seconds) was not explicitly stated but could be discerned from a visual counter which registered a point each time 4 or more seconds separated the key presses. Over a six minute period, children attempted to accumulate as many points as possible. Mitchell and Quittner found that deaf children's performance on this task was similar to age matched normal hearing children. The authors, therefore, concluded that behavioral inhibition skills of the deaf children with CIs were not impaired. However, the authors did not attempt to control for length of CI use and, therefore, their results cannot rule out the effect of auditory experience on response delay skills. Furthermore, Mitchell and Quittner did not examine relations between response delay skills and any outcome measures that assess oral language skills in this population.

We designed the current study to test the hypothesis that response delay skills would be related to oral language outcomes in deaf children with CIs. We also investigated whether response delay skills improved as a function of CI use when we controlled for chronological age. Finally, we used standardized scores to investigate whether the response delay skills of these children were comparable to a large national sample of normal hearing children. To assess response delay skills, we used a laboratory measure of response delay skills identical to the one employed by Mitchell and Quittner (1996).

## Experiment

### Methods

**Subjects.** All participants were part of an ongoing longitudinal study of the development of oral speech and language in prelingually deaf children with CIs at the Indiana University School of Medicine Cochlear Implant Program. A total of 47 prelingually deaf children (profoundly deaf by 2.5 years old) who received CIs by 9 years of age were included in this retrospective analysis. All children used Nucleus 22 devices. Mean age of implantation was 4.8 years old. Twenty-two children used OC (immersed in a training program using oral communication only) and 25 used TC (a program in which both oral and manual language are used) at the time of testing. Over 65% had an unknown cause of hearing loss, presumed to be congenital. The most common acquired etiology was meningitis. A subset of children (n=18) had pre-implant non-verbal IQ scores (WISC or WPPSI) in their charts. The mean standardized IQ score was 104.1 (sd=15.28).

Children were tested once every 6-12 months from before implantation to three years post-implantation. Interval data were collapsed into one of four intervals: pre-implant, one year post-implant, two years post-implant, and three years post-implant. Not all children were tested at each interval, as is common in clinical populations such as this, creating missing data cells. Missing data occurred for several reasons. First, some children moved away from the Indianapolis area after implantation and were unable to continue with the study. Second, because our clinical participants are given a large number of tests during each visit, they are often too tired or not cooperative enough to complete all testing procedures.

**Procedures.** The Preschool Delay task (Gordon, McClure, & Aylward, 1996) is recommended for use in normal hearing children from 3-5 years of age. The test apparatus consists of a customized mechanical button box with one blue key located centrally below an LED display. The apparatus is capable of running a number of neuropsychological tests and is used to measure response delay and behavioral inhibitory skills (Gordon et al., 1996). The delay task is a 6 minute test in which children are asked to press a key and then wait for some period of time before pressing the button again. Each time the child presses the button after waiting four seconds, they receive a point which is shown on the display screen. The number of points a child receives over 6 minutes is divided by the total number of key presses to compute the "error-ratio" (ER) which is the primary dependent measure on this task. Children receive no feedback regarding their performance other than the point counter. Thus, it is up to the subject to figure out how long they have to wait in order to get a point.

Five speech and language outcome measures were used in the correlational analyses. Open set speech perception was measured using the Phonetically Balanced Kindergarten (PBK) test (Haskins, 1949). Children hear a spoken word and are asked to repeat the word aloud to the examiner. This test is administered using a live voice and is scored by word and phoneme. The words used in this test are phonetically balanced, English monosyllables.

Sentence comprehension was measured with the Common Phrases test (Osberger, Miyamoto, Zimmerman-Philips, Kemink, Stroer, Firszt, & Novak, 1991). The child is asked to repeat a sentence or follow a command and his response is scored as correct or incorrect. Sentences are administered in auditory (CPA), visual (CPV), and auditory plus visual modalities (CPAV) in separate conditions. We analyzed scores from all three presentation formats.

Vocabulary knowledge was assessed with the Peabody Picture Vocabulary (PPVT) Test (Dunn & Dunn, 1997). This widely used test is a closed set, forced choice task. At our center, each vocabulary item is presented live either orally or manually with Signed Exact English depending on the individual child's preferred mode of communication. The child then chooses from four pictures, one of which correctly corresponds to the meaning of the word.

The Reynell Developmental Language Scales 3rd edition (Reynell & Huntley, 1985) was administered to assess both receptive (RDLS-R) and expressive (RDLS-E) language skills. The receptive scales measure 10 different skills including word recognition, sentence comprehension, and verbal comprehension of ideational content. The expressive language scales assess skills such as spontaneous production of speech and picture description. Like the PPVT, the RDLS was administered in the child's preferred mode of communication.

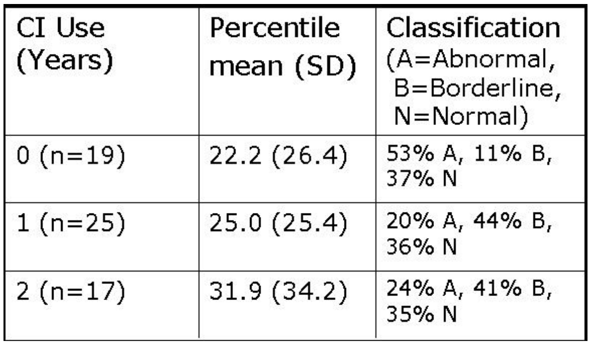
Speech intelligibility was assessed using the Beginner's Intelligibility Test (BIT) developed by Osberger, Robbins, Todd, and Riley (1994). Audio recordings were made of children repeating a list of 10 sentences presented to them in the auditory modality by a clinician. These recordings were then played back to 3 naïve adult listeners who were asked to transcribe what they thought the children were saying. Intelligibility scores were based on the number of words correctly transcribed by the adult listeners.

Each of these tests, including the delay task, were administered in an Otolaryngology clinical setting by licensed health professionals who had received special training in working with deaf children who used CIs.

## Results

**Delay task performance of prelingually deaf children with CIs compared to a normative sample.** Normative data for the delay task performance are available for children aged 3-5 years old. Therefore, many of the children in our sample were too old to derive percentile or standard scores from these norms. Therefore, we calculated normative percentile scores from the CPT raw scores only for children who were 6 years of age or younger at the time of testing. Table 1 shows the percentile means for ER at each interval along with the number of subjects who could be normed at each interval. This subset of our sample performed in the low normal range on this task compared to the normative sample. The highest mean percentile scores were found in children who had used their CI for two years (32nd percentile). However, even after two years of implant use, about 1/3 of the children performed in the "abnormal range" according to the published norms (Gordon et al., 1996). We did not perform statistical analyses using these percentile scores because these scores may not have been representative of the entire sample of children.

**Effects of chronological age and length of CI use.** As mentioned previously, because our subjects were drawn from a larger clinical population enrolled in a long-term longitudinal study of CI outcomes, there were missing data cells for different children at each test interval. A traditional repeated measures ANOVA would, therefore, eliminate data from any child who was not tested at each interval. Such an analysis can lead to skewed results as well as underestimates of variability (Schafer & Graham, 2002). To avoid this problem, we analyzed our data using the SAS Mixed Procedure (Wolfinger & Chang, 1995). The SAS Mixed Procedure utilizes a maximum-likelihood estimation method to create a model without eliminating any participants (Schafer & Graham, 2002). In this manner, systematic selection biases can be avoided and data from all children, even those who were not tested at each interval, could be included in the statistical model.



| CI Use (Years) | Percentile mean (SD) | Classification (A=Abnormal, B=Borderline, N=Normal) |
|---|---|---|
| 0 (n=19) | 22.2 (26.4) | 53% A, 11% B, 37% N |
| 1 (n=25) | 25.0 (25.4) | 20% A, 44% B, 36% N |
| 2 (n=17) | 31.9 (34.2) | 24% A, 41% B, 35% N |

**Table 1.** Mean percentile error ratio on the delay task at each interval of CI use. Clinical classifications based on the normative sample tested by Gordon et al. are provided as well.

The results obtained using the SAS mixed model revealed a significant effect of chronological age on delay ER (*F*(1,46.7) = 11.02, *p*=0.0018). We also found an effect of length of CI use on delay ER (*F*(2,50.1) = 4.66, *p*=0.014), when chronological age was controlled. Our analysis also revealed a significant interaction between chronological age and length of CI use (*F*(2,52.1) = 3.51, *p*=0.037). Figure 1 displays the individual ER scores as a function of chronological age at CI activation and length of implant use. Examination of this figure shows that at pre-implant and 1 year post-implant intervals, delay task performance improved as a function of chronological age. However, at two years after implantation, mean ER decreased with chronological age. Children who were younger at time of implantation (and younger over the three-year testing period) showed more improvement in ER with CI use than children who were older at implantation.

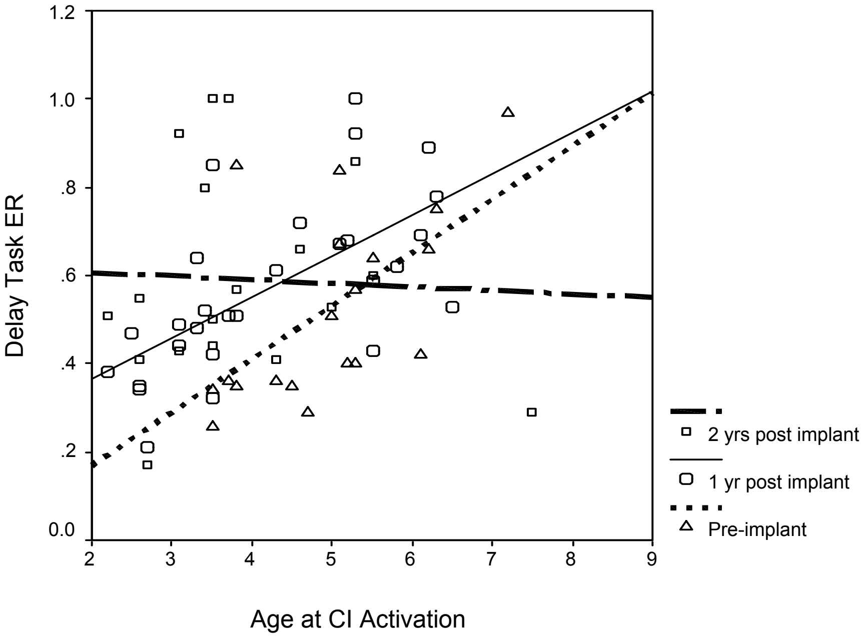


**Figure 1**. Delay task error ratio as a function of chronological age of CI activation and length of CI use.

This interaction should be interpreted with caution, due to a possible selection bias in the older children. As mentioned previously, the Preschool Delay task is clinically recommended for children 3-5 years old. For older children, the Gordon Diagnostic System offers a harder version of the delay task. Some of the older children in this study were unable to complete the harder delay task and, therefore, the easier task was administered. Thus, these delayed children may have been over-represented among our oldest children (particularly the child who was implanted at 7.5 years old and had an ER of 0.2). Nevertheless, the main effect of length of CI use on delay ER demonstrates that, overall, children's performance on this task improved with CI experience.

**Correlations between delay task and speech and language outcome measures** A total of 35 children in our sample had delay task scores obtained after 1 to 3 years of CI use. For children who had delay ER scores at more than one interval, we used only their post-implant year 1 scores for correlation purposes. The number of children who had scores on both the delay task and the outcome measures differed across post-implant intervals and outcome measures. Therefore, we report the sample size of each correlation.

Bivariate correlations were computed between delay task ER and each of the outcome measures at 1, 2, and 3 years post-implantation. A summary of these results is displayed in Table 2 which includes each individual Pearson correlation, p value, and sample size. We found significant and positive correlations between ER and vocabulary (PPVT), expressive language (RDLS-E), receptive language (RDLS-R), and speech intelligibility (BIT) scores at all three post-implant outcome intervals. As shown in Table 2, these correlations were strong and most were significant at a level of $p$<0.01. In contrast, no significant correlations were found between delay task ER and open set speech perception scores (PBK) or sentence comprehension scores (CPA, CPV, CPAV).

| Outcome Measure* | Post-implant yr 1 | Post-implant yr 2 | Post-implant yr 3 |
|---|---|---|---|
| Speech perception | | | |
| PBK words | NS | NS | NS |
| PBK phonemes | NS | NS | NS |
| Sentence comprehension | | | |
| CP auditory | NS | NS | NS |
| CP visual | NS | NS | 0.51 (22) |
| CP audiovisual | NS | NS | NS |
| Language & Vocabulary | | | |
| PPVT | **0.69 (34) ++** | **0.62 (35) ++** | **0.55 (31)** |
| Reyn receptive | **0.60 (32) +** | **0.64 (29) ++** | **0.61 (27) ++** |
| Reyn expressive | **0.54 (32) +** | **0.61 (25)** | 0.46 (27) |
| Speech Intelligibility | | | |
| BIT | **0.42 (30) +** | **0.38 (32)** | 0.40 (24) |

*. All outcome measures were in % correct except for PPVT (raw scores) and Reynell (age equivalent scores in months)

| | |
|---|---|
| NS: | non-significant bivariate correlation (*p*>0.05) |
| Normal face type: | significant bivariate correlation (*p*<0.05) |
| Bold face type: | significant bivariate correlation (*p*<0.01) |
| ++. | significant with length of CI use and chronological age partialled out (*p*<0.01) |
| +. | significant with length of CI use and chronological age partialled out (*p*<0.05) |

**Table 2.** Bivariate and partial correlations between delay task scores and outcome measures.

For those bivariate correlations which were significant, we also conducted a series of partial correlations to control for chronological age and length of CI use at time of delay task administration. The results shown in Table 2 reveals that many of these partial correlations were also significant at the p<0.01 level and several others at the *p*<0.05 level. For the outcome measures obtained after 1 year of use, partial correlations were significant between delay ER and PPVT, RDLS-R, RDLS-E, and BIT measures. For outcomes obtained after 2 or 3 years of CI use, a number of these relations no longer reached statistical significance. This was not surprising since most of the delay ER scores were obtained after 1 year of CI use. For PPVT at year 2, and RDLS-R at years 2 and 3, the partial correlations with delay ER remained significant.

## Discussion

The results of our analysis of the Gordon Delay Task revealed several new findings regarding the behavioral inhibition skills of prelingually deaf children who use CIs. First, a large number of deaf children displayed behavioral inhibition skills which were atypical when compared to the results of a normative sample reported by Gordon et al. (1996). This finding contrasts with the earlier results of Mitchell and Quittner (1996) who tested an older population of children on a similar delay task and found no differences in performance between normal-hearing children and prelingually deaf children with CIs. The difference between the current and previous findings may be due to the fact that the participants in our study were several years younger on average than the children tested by Mitchell and Quittner. Measures obtained from the Gordon Delay Task may not have been sensitive enough to detect differences in behavioral inhibition ability between older deaf and normal hearing children. Furthermore, most of the participants in our study had used their CI for only one year while Mitchell and Quittner included children who had used their implant for as long as six years. It is possible that the deaf children Mitchell and Quittner tested caught up to normal hearing children in their behavioral inhibition skills due to these extra years of CI use. However, additional research needs to be done in this area to better understand this issue.

We also found that delay task performance increased with length of CI use as well as chronological age. These two results suggest that CI use in prelingually deaf children facilitates the development of behavioral inhibition skills. The skills used in the Gordon Preschool Delay Task may indeed include some language related skills as discussed below. In addition, we found an interaction between length of CI use and chronological age suggesting that the older children did not demonstrate the same degree of improvement with CI use as did younger children in our study. This interaction may represent a sampling artifact in our study design.

Finally, our analysis of the delay task revealed several significant correlations between behavioral inhibition skills and vocabulary knowledge, language, and speech intelligibility scores at one year post implantation. Several of these relations also remained significant after two or three years of CI use. One explanation of these findings is that performance on the response delay task reflects behavioral control skills, which play a significant role in real world language learning. Our findings along with the results of Knutson et al. (2000a, b) provide some converging support for this hypothesis, although prospective, longitudinal studies are needed to directly assess the possible causal relationship between behavioral control abilities and oral language development in deaf children with CIs. There is an extensive literature showing that the difficulties faced by children with ADHD, including impulsivity, have a significantly negative impact on scholastic performance (Lamminmaki, Ahonen, Narhi, Lyytinent, & de Barra, 1995) and many children with ADHD also exhibit delays in language and math abilities (Barkley, 1990, 1997).

Another explanation of our current findings is that the relations observed between response delay and language outcomes in our sample of children reflect underlying language related skills such as internal speech, sub-vocal verbal rehearsal, and counting which are used by children as a strategy to perform the delay task. For instance, one possible strategy would be to press the button, then count to oneself for a few seconds, and then press the button again. This sequence would be repeated until the score counter registered a point, indicating that the subject had waited a sufficient length of time. Thereafter, the subject would simply count the required number of seconds before each response, thereby maximizing the ER.

Clearly, such a strategy would require some degree of sophistication in counting, and internal speech mechanisms which utilize subvocal verbal rehearsal strategies. Several related findings in the literature suggest that deaf children may not be able to take advantage of subvocal rehearsal strategies during tasks in which it would be beneficial to do so. For example, in a recent study, Pisoni & Cleary (2003) measured the immediate memory capacity for spoken lists of digits in normal hearing children and in deaf children with cochlear implants and reported a main effect of hearing status and an interaction between hearing status and order of recall. The normal hearing children showed longer digit spans overall compared to deaf children with CIs. When asked to recall the sequence in the same order as presented (forward digit span) the normal hearing children showed greater digit spans compared to the condition that required recall of the sequence in the backwards order (backward digit span). This effect in normal hearing children was taken as evidence that normal hearing children were able to benefit from subvocal verbal rehearsal strategies to maintain a longer digit sequence in working memory when temporal sequence did not have to be manipulated. In contrast, the deaf children showed significantly smaller differences between spans obtained in the forward and backward conditions, suggesting that these children did not utilize subvocal rehearsal strategies successfully.

Other studies have demonstrated that the atypical recall capacity of deaf children with CIs is not limited to tasks involving lists of auditory stimuli. Indeed, immediate recall for sequences of visual stimuli which can be verbally encoded such as colored lights (Cleary et al., 2001; Dawson et al., 2002) have also been found to be atypical compared to age matched, normal hearing peers. This finding lends further support to the hypothesis that deaf children with cochlear implants are atypical in their capacity to encode, manipulate and store stimuli which can be represented phonologically or subvocally (Pisoni & Cleary, 2002).

Finally, there is now strong evidence that subvocal rehearsal abilities are related to open set speech perception, vocabulary knowledge, expressive and receptive language, speech intelligibility, and speaking rates in prelingually deaf children with CIs (Burkholder & Pisoni, 2003; Dawson et al., 2002; Pisoni & Cleary, 2002; Pisoni & Geers, 2000). The present results obtained with the response delay task suggest that subvocal rehearsal abilities are closely tied to other traditional measures of CI benefit. However, relations between delay ER and other process measures known to reflect subvocal rehearsal (such as forward-backward digit span) need to be investigated in future studies of deaf children with CIs.

In summary, the present results provide several new findings regarding the relations between the cognitive processes involved in behavioral inhibition and audiological outcomes in prelingually deaf children with CIs. A period of early auditory deprivation prior to implantation may have specific cognitive effects on deaf children, which may impact their ability to successfully acquire an oral language with a CI. However, further work is needed in order to understand precisely how early auditory deprivation affects the development of behavioral inhibition and how language skills and response delay skills are related. For instance we do not know whether auditory experience has an effect on behavioral inhibition skills through remodeling of language processes per se, or whether there are more specific effects of audition on cortical areas responsible for self-regulatory behavior mediated by pre-frontal cortex, anterior and posterior cingulated gyrus, and other areas (Barkley, 1997; Rubia, Overmeyer, Taylor, Brammer, Williams, Simmons, Andrew, & Bullmore, 2000). Future research on neuropsychological functions of deaf children who use CIs is needed to answer this and other related questions. Furthermore, neuropsychological testing of these children may prove useful as a clinical tool in assessing outcomes and benefit from a CI in this unique population of children.

## References

Achenbach, T., & Edelbrock, C. (1983). Manual for the child behavior checklist and revised child behavior profile. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T., & Edelbrock, C. (1986). Manual for the child behavior checklist-teacher report form. Burlington: University of Vermont: Department of Psychiatry.

Altshuler, K., Deming, W., Vollenweider, J., Rainer, J., & Tendler, R. (1976). Impulsivity and profound early deafness: a cross cultural inquiry. *American Annals of the Deaf, 121*, 331-345.

Barkley, R. (1990). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment.* New York: Guilford Press.

Barkley, R. (1997). *ADHD and the nature of self control*. New York, NY: The Guilford Press.

Blamey, P., Sarant, J., Paatsch, L., Barry, J., Bow, C., Wales, R., et al. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech Language Hearing Research, 44*, 264-285.

Burkholder, R., & Pisoni, D. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85,* 63-88.

Chess, S., & Fernandez, P. (1980). Impulsivity in rubella deaf children: a longitudinal study. *American Annals of the Deaf, 125*, 505-509.

Cleary, M., Pisoni, D., & Geers, A. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing, 22*, 395-411.

Dawson, P., Busby, P., McKay, C., & Clark, G. (2002). Short-term auditory memory in children using cochlear implants and its relevance to receptive language. *Journal of Speech Language and Hearing Research, 45,* 789-801.

Dunn, L., & Dunn, L. (1997). *Peabody picture vocabulary test, 3rd edition*. Circle Pines, MN: American Guidance Service.

Geers, A. (2003). Predictors of reading skill development in children with early cochlear implantation. *Ear and Hearing, 24,* 59S-68S.

Geers, A., Brenner, C., & Davidson, L. (2003). Factors associated with development of speech perception skills in children implanted by age five. *Ear and Hearing, 24,* 24S-35S.

Geers, A., Nicholas, J., & Sedey, A. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24,* 46S-58S.

Gordon, M., McClure, F., & Aylward, G. (1996). *The Gordon diagnostic system (GDS) instruction manual & interpretive guide (3 ed.)*. Dewitt, NY: Michael Gordon, PhD.

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

Horn, D.H., Davis, R.A.O., Pisoni, D.B., Miyamoto, R.T (under review). Development of visual attention skills in prelingually deaf children who use cochlear implants.

Kelly, D., Kelly, B., Jones, M., Moulton, N., Verhulst, S., & Bell, S. (1993). Attention deficits in children and adolescents with hearing loss. A survey. *American Journal of Diseases of Children, 147*, 737-741.

Kirk, K. (2000). Cochlear implants: new developments and results. *Opinion in Otolaryngology Head and Neck Surgery, 8,* 415-420.

Knutson, J., Ehlers, S., Wald, R., & Tyler, R. (2000a). Psychological predictors of pediatric cochlear implant use and benefit. *Annals of Otology Rhinology Laryngology Supplement, 185*, 100-103.

Knutson, J., Ehlers, S., Wald, R., & Tyler, R. (2000b). Psychological consequences of pediatric cochlear implant use. *Annals of Otology Rhinology Laryngology Supplement, 185*, 109-111.

Lachs, L., Pisoni, D., & Kirk, K. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: a first report. *Ear and Hearing, 22,* 236-251.

Lamminmaki, T., Ahonen, T., Narhi, V., Lyytinent, H., & de Barra, H. (1995). Attention deficit hyperactivity disorder subtypes: Are there differences in academic problems? *Developmental Neuropsychology, 11,* 297-310.

Mitchell, T., & Quittner, A. (1996). Multimethod study of attention and behavior problems in hearing-impaired children. *Journal of Clinical Child Psychology, 25,* 83-96.

Miyamoto, R., Svirsky, M., Kirk, K., Robbins, A., Todd, S., & Riley, A. (1997). Speech intelligibility of children with multichannel cochlear implants. *Annals of Otology Rhinology Laryngology Supplement, 168,* 35-36.

O'Brien, D. (1987). Reflection-impulsivity in total communication and oral deaf and hearing children: a developmental study. *American Annals of the Deaf, 132,* 213-217.

Osberger, M., Miyamoto, R., Zimmerman-Philips, S., Kemink, J., Stroer, B., Firszt, J., et al. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus-22 channel cochlear implant system. *Ear and Hearing, 12,* S66-80.

Osberger, M., Robbins, A., Todd, S., & Riley, A. (1994). Speech intelligibility of children with cochlear implants. *Volta Review, 96*, 169-180.

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

Pisoni, D., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24,* 106S-120S.

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

Pisoni, D., & 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 Supplement, 185*, 92-93.

Pisoni, D.B. & Cleary, M. (2002). Some new findings on learning, memory and cognitive processes in deaf children following cochlear implantation. In F.G. Zeng, A.N. Popper & R.R. Fay (Eds*.), Springer Handbook of Auditory Research: Auditory Prosthesis, SHAR Volume X.*

Quittner, A., Smith, L., Osberger, M., Mitchell, T., & Katz, D. (1994). The impact of audition on the development of visual attention. *Psychological Science, 5,* 347-353.

Reynell, J. K., & Huntley, M. (1985). *Reynell Developmental Language Scales (2nd ed.)*. Windsor, UK: NFER-Nelson.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S.C.R., Simmons, A., Andrew, C., Bullmore, E.T. (2000). Functional frontalisation with age: mapping neurodevelopmental trajectories with fMRI. *Neuroscience and Biobehavioral Review, 24*, 13-9.

Sarant, J., Blamey, P., Dowell, R., Clark, G., & Gibson, W. (2001). Variation in speech perception scores among children with cochlear implants. Ear and Hearing, 22(1), 18-28.

Schafer, J., & Graham, J. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*, 147-177.

Smith, L., Quittner, A., Osberger, M., & Miyamoto, R. (1998). Audition and visual attention: the developmental trajectory in deaf and hearing populations. *Developmental Psychology, 34,* 840-850.

Surowiecki, V., Sarant, J., Maruff, P., Blamey, P., Busby, P., & Clark, G. (2002). Cognitive processing in children using cochlear implants: the relationship between visual memory, attention, and executive functions and developing language skills. *Annals of Otology Rhinology Laryngology Supplement, 189*, 119-126.

Svirsky, M., Robbins, A., Kirk, K., Pisoni, D., & Miyamoto, R. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science., 11,* 153-158.

Tobey, E., Geers, A., Brenner, C., Altuna, D., & Gabbert, G. (2003). Factors associated with development of speech production skills in children implanted by age five. *Ear and Hearing, 24*, 36S-45S.

Tyler, R., Fryauf-Bertschy, H., Kelsay, D., Gantz, B., Woodworth, G., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology Head and Neck Surgery, 117,* 180-187.

US Food and Drug Administration Approval Report (2000). Found on http://www.fda.gov/cdrh/pma/pmaaug00.html

Wolfinger, R., & Chang, M. (1995). *Comparing the SAS GLM and MIXED procedures for repeated measures.* Paper presented at the Proceedings of the Twentieth Annual SAS Users Group International Conference, Orlando, Florida.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Sequence Learning as a Predictor of Audiological Outcomes in Deaf Children with Cochlear Implants[1]

**David B. Pisoni[2] and Rebecca A. O. Davis[2]**

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

[1] This work was supported by NIH-NIDCD R01 Research Grants DC00111 and DC03937 to Indiana University.

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

# Sequence Learning as a Predictor of Audiological Outcomes in Deaf Children with Cochlear Implants

**Abstract.** The present study extends earlier research carried out by Cleary and Pisoni (2001) who found that measures of auditory sequence learning, using the Simon memory game procedure, were related to open-set spoken word recognition and language comprehension. Sequence learning scores were obtained from a new group of profoundly deaf children with cochlear implants. Two different measures of sequence learning were computed. One measure, Simon auditory redundancy gain, was used to assess the benefit of redundant auditory information on the reproduction of visual sequences of colored lights. A second measure, Simon learning improvement, was used to assess the increase in sequence learning observed over time after a period of implant use. Both learning measures were found to be correlated with traditional audiological measures of outcome and benefit. The auditory gain measure was significantly correlated with two measures of language comprehension on the Common Phrases test. Vocabulary knowledge on the PPVT was found to be correlated with the improvement in sequence learning over time. Taken together with the earlier findings reported by Cleary and Pisoni (2001), the present results suggest that differences in learning may contribute an additional source of variance to traditional measures of speech and language outcomes in this clinical population. Measures of learning and memory may therefore provide important new insights into the underlying cognitive and neurobiological factors that are responsible for the individual differences and enormous variation in a range of clinical speech and language outcome measures that are routinely obtained from deaf children who have received cochlear implants as a treatment for profound hearing loss.

## Introduction

Many of the traditional methods for measuring working memory span and the capacity of immediate memory use recall tasks that require a subject to repeat back a sequence of test items using an overt articulatory-verbal motor response (Dempster, 1981). Because deaf children with cochlear implants may also have delays and/or disorders in speech motor control and phonological development, it is possible that any differences in performance between deaf children and age-matched normal-hearing children using memory span tasks could be due to the nature of the response requirements used during retrieval and output. Differences in articulation and speech motor control could magnify other differences in encoding, storage, rehearsal or retrieval processes.

To eliminate the use of an overt articulatory-verbal response, we developed a new experimental methodology to measure immediate memory span in deaf children with cochlear implants based on Milton-Bradley's Simon, a popular memory game. Figure 1 shows a display of the apparatus which was modified so it could be controlled by a PC. In carrying out the procedure, a child is asked to simply "reproduce" a stimulus pattern by manually pressing a sequence of colored panels on the four-alternative response box. In addition to eliminating the need for a verbal response, the Simon methodology permitted us to manipulate the stimulus presentation conditions in several systematic ways while holding the response format constant. This particular property of the experimental procedure was important because it provided us with a novel way of measuring how auditory and visual stimulus dimensions are analyzed and processed alone and in combination and how these stimulus manipulations affected measures of memory span. The Simon memory game apparatus and methodology also offered us an opportunity to study learning processes, specifically, sequence learning and the relations between memory and learning using the same identical experimental procedures and response demands.

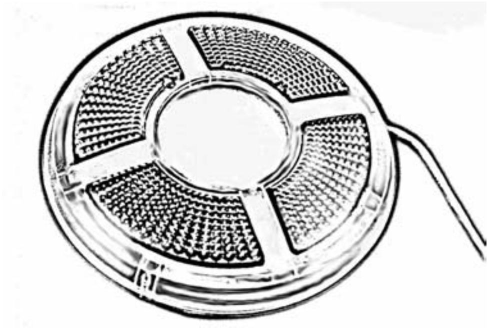


**Figure 1.** The memory game response box based on the popular Milton Bradley game "Simon."

In our initial studies with the Simon apparatus, three different stimulus presentation formats were employed (Pisoni & Cleary, 2004). In the first condition, the target sequences to be reproduced consisted only of spoken color names (A). In the second condition, sequences of colored lights (L) were presented in the visual modality. In the third presentation condition, the spoken color names were presented simultaneously with matching colored lights (A+L).

Forty-five hearing-impaired children with cochlear implants were tested using the Simon memory game apparatus. Thirty-one of these children were able to complete all six conditions included in the testing session. They also were able to identify the recorded color-name stimuli used in this task when these items were presented alone in isolation. Thirty-one normal-hearing children who were matched in terms of age and gender with the group of children with cochlear implants were also tested. Finally, 48 normal-hearing adults were recruited to serve as an additional comparison group (see Pisoni & Cleary, 2004).

Of the six conditions tested, three measured the children's immediate memory skills and three measured the children's sequence learning skills. In the immediate memory task, the temporal sequences systematically increased in length as the subject progressed through successive trials in the experiment. Within each condition, the 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 was increased by one item in length. If a list was incorrectly reproduced, the next trial used a list that was one item shorter in length. This adaptive tracking procedure is similar to methods used in psychophysical testing (Levitt, 1970). Sequences used for the Simon 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. We computed a weighted memory span score for each child by finding the proportion of lists correctly reproduced at each list length and averaging these proportions across all list lengths.

A summary of the results from the Simon immediate memory task for the three groups of subjects is shown in Figure 2. The normal-hearing adults are shown in the left panel, the normal-hearing aged-matched children are shown in the middle panel and the children with cochlear implants are shown in the right panel. Within each panel, the scores for auditory-only presentation (A) are shown on the left, scores for lights-only presentation (L) are shown in the middle and scores for the combined auditory and lights presentation condition (A+L) are shown on the right.

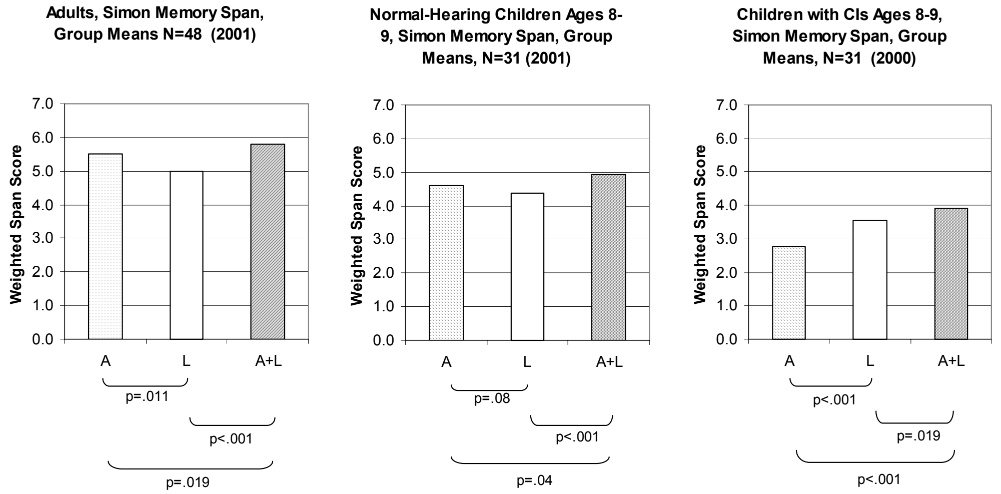


**Figure 2.** Mean sequence memory spans in each of the three presentation conditions using the "Simon" memory game (Redrawn from Pisoni & Cleary, 2004).

Examination of the Simon memory span scores for the normal-hearing adults shown in the left-hand panel of Figure 2 reveals several findings that can serve as a benchmark for comparing and evaluating differences in performance of the two groups of children. First, we found a "modality effect" for presentation format. Auditory presentation (A) of sequences of color names produced longer immediate memory spans than visual presentation (L) of sequences of colored lights ($p$ < .02). Second, we found a "redundancy gain." When information from the auditory and visual modalities was combined together and presented simultaneously (A+V), the memory spans were longer compared to presentation using only one sensory modality ($p$ < .02 for A and $p$ < .001 for L).

The modality effect and the redundancy gain demonstrate that the Simon memory game procedure is a valid and potentially useful experimental methodology for measuring immediate memory span in normal-hearing adults because it reveals subtle differences in the sensory modality used for presentation of the stimulus patterns. As in other studies of verbal short-term memory, longer Simon memory spans were found for auditory stimuli compared to visual stimuli in the normal-hearing adults, suggesting the active use of phonological coding and verbal rehearsal strategies (Penny, 1989; Watkins, Watkins & Crowder, 1974). In addition, the Simon memory spans reflected cross-modality redundancies between stimulus dimensions when the same information about a stimulus pattern was correlated and presented simultaneously to more than one sensory modality. This latter finding demonstrates that adults are not only able to combine and integrate redundant sources of stimulus information across different sensory modalities, but the consequence of this integration and redundancy gain is an increase in immediate memory capacity when the stimulus dimensions are correlated in the auditory and visual modalities.

The middle panel of Figure 2 shows the results of the three presentation conditions for the group of normal-hearing 8-and 9-year old children who were age-matched to the group of deaf children with cochlear implants. Overall, the pattern of the Simon memory span scores is similar to the findings obtained with the normal-hearing adults shown in the left-hand panel of Figure 2 although several differences were observed. First, the absolute memory spans for all three presentation conditions were lower for the normal-hearing children than the memory spans obtained from the adults. Second, while the modality effect found with the adults was also present in these data, it was smaller in magnitude and was only marginally significant, suggesting possible developmental differences in the rate and efficiency of verbal rehearsal between adults and children in processing auditory and visual sequential patterns like those used in this task. Third, the cross-modal "redundancy gain" observed with the adults was also found with the normal-hearing children although it was also smaller in magnitude ($p$ <.04 A and $p$ <.001 for L). Again, these differences may simply be due to age, maturation and development.

The Simon memory spans for the deaf children with cochlear implants are shown in the right-hand panel of Figure 2 for the same three presentation conditions. Examination of the pattern of these memory spans reveals several striking differences from the memory spans obtained for the normal-hearing children. First, the memory spans for all three presentation conditions were consistently lower overall than the spans from the corresponding conditions obtained for the normal-hearing children. Second, the modality effect observed in both the normal-hearing adults and normal-hearing children was reversed for the deaf children with cochlear implants. The memory spans for the deaf children were longer for visual-only presentation than auditory-only presentation and this difference was highly significant ($p < .001$). Third, although the cross-modal "redundancy gain" found for both the adults and normal-hearing children was also observed for the deaf children and was statistically significant for both conditions ($p < .001$ for A and $p < .02$ for L), the absolute size of the redundancy gain was smaller in magnitude than the gain observed with the normal-hearing children.

The results obtained for the visual-only presentation conditions are of particular theoretical interest because the deaf children with cochlear implants displayed shorter memory spans than the normal-hearing children. This finding adds additional support to the hypothesis that phonological recoding and verbal rehearsal processes in working memory play important roles in perception, learning and memory in these children (Pisoni & Cleary, 2003). Capacity limitations of working memory are closely tied to speed of processing information even for visual patterns which can be rapidly recoded and represented in memory in a phonological or articulatory code for certain kinds of sequential processing tasks. Verbal coding strategies may be mandatory in memory tasks that require immediate serial recall of temporal patterns that preserve item and order information (Gupta & MacWhinney, 1997). Thus, although the visual patterns were presented using only sequences of colored lights, both groups of children appeared to recode these sequential patterns using verbal labels and verbal coding strategies to create stable phonological representations in working memory for maintenance and rehearsal prior to response output.

The deaf children also showed much smaller redundancy gains under the multi-modal presentation conditions, which suggests that in addition to differences in working memory and verbal rehearsal, their information processing skills and abilities to perceive and encode complex multi-dimensional stimuli are atypical and compromised relative to age-matched normal-hearing children. The smaller redundancy gains observed in these deaf children may also be due to the reversal of the typical modality effects observed in studies of working memory that reflect verbal coding of the stimulus materials. The modality effect in short-term memory studies is generally thought to reflect phonological coding and verbal rehearsal strategies that actively maintain temporal order information of sequences of stimuli in immediate memory for short periods of time (Watkins et al., 1974). Taken together the present findings demonstrate important differences in both attention and memory processes in this clinical population. These basic differences in information processing skills may be responsible for the wide variation in speech and language outcomes observed in deaf children following cochlear implantation.

## Simon Learning Spans

The initial version of our Simon memory game used novel sequences of color names and/or colored lights (Pisoni & Cleary, 2004). All of the sequences were generated randomly on each trial in order to prevent any learning. Our primary goal in this research was to obtain estimates of working memory capacity for temporal patterns that were not influenced by sequence repetition effects or idiosyncratic coding strategies that might increase memory capacity from trial to trial. Each test sequence was novel and was created by a random number generator so that the structure of a sequence of stimuli was always different and varied from trial to trial during the course of the experiment. If a subject correctly reproduced a pattern at a given length twice in a row, the adaptive testing algorithm in the experimental control program automatically increased the length of the sequence by one item on the next trial and then generated an entirely new temporal sequence of stimuli that was different from the sequence presented on the previous trial. This procedure was used throughout the entire experiment to obtain estimates of immediate memory capacity. Thus, there was no basis for any new learning to take place and we can use the measures of Simon memory span as estimates of capacity of immediate memory for sequences of highly familiar stimuli such as color names.

In addition to measuring immediate memory capacity, we have also used the Simon memory game procedure to study sequence learning and to investigate the effects of long-term memory on coding and rehearsal strategies in working memory (Cleary & Pisoni, 2001). To accomplish this goal and to be able to directly compare the gains in learning and the increases in working memory capacity to our earlier Simon memory span measures, we examined the effects of sequence repetition on immediate memory span by simply repeating the same pattern again if the subject correctly reproduced the sequence on a given trial. In the Simon learning condition, the same stimulus pattern was repeated on each trial for an individual subject and the sequences gradually increased in length by one item after each correct response until the subject was unable to correctly reproduce the pattern. This change in the methodology provided an opportunity to study learning based on simple repetition and to investigate how repetition of the same pattern affects the capacity of immediate memory.

Figure 3 displays a summary of the results obtained in the Simon learning conditions that investigated the effects of sequence repetition on memory span for the same three presentation formats used in the earlier conditions, auditory-only (A), lights-only (L) and auditory+lights (A+L). The weighted memory span scores for the sequence learning conditions are shown on the right-hand side of each panel in this figure; the corresponding set of memory span scores obtained earlier under random presentation format for the same three presentation conditions are shown on the left-hand side of each panel. The data for the normal-hearing adults are shown in the left panel, the data for the normal-hearing 8-and 9-year old children are shown in the middle panel and the data for the deaf children with cochlear implants are shown in the right panel.

Examination of the two sets of memory span scores shown within each panel reveals several consistent findings. First, repetition of the same stimulus sequence produced large learning effects for all three groups of subjects. This repetition effect can be seen clearly by comparing the three scores on the right-hand side of each panel to the three scores on the left-hand side. For each of the three groups of subjects, the learning span scores on the right were higher than the memory span scores on the left. Repetition of a stimulus pattern increased immediate memory span capacity, although the magnitude of the learning effects differed systematically across the three groups of subjects. The memory spans observed for the adults in the learning condition were about twice the size of the memory spans observed when the sequences were generated randomly from trial to trial. Although a repetition effect was also obtained with the deaf children who use cochlear implants in the right panel, the size of their repetition effect was about half the size of the repetition effect found for the normal-hearing children shown in the middle panel of Figure 3.

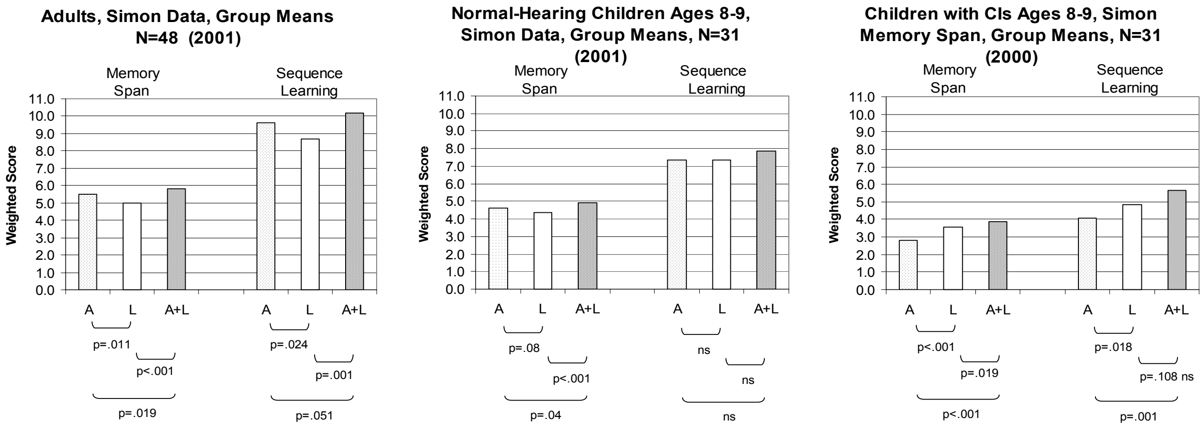


**Figure 3.** Mean immediate memory spans and sequence learning scores in each of the three conditions tested using the "Simon" memory game (Redrawn from Pisoni & Cleary, 2004).

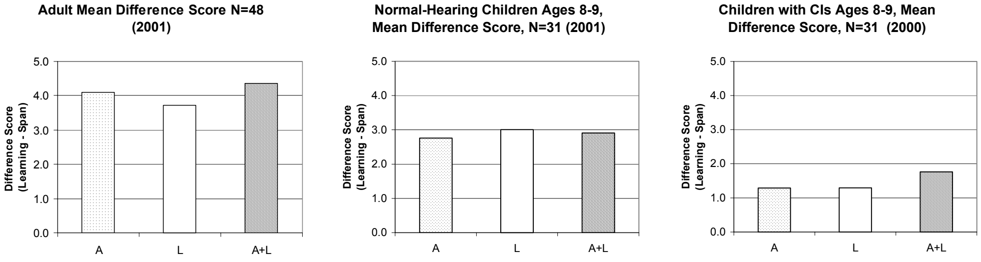


**Figure 4.** Different scores between memory and learning scores for each of the three conditions (A, L, A+L) for the three groups of participants tested using the "Simon" memory game.

Second, the rank ordering of the three presentation conditions in the sequence learning conditions was similar to the rank ordering observed in the memory span conditions for all three groups of subjects. The repetition effect was largest for the A+L conditions for all three groups. For both the normal-hearing adults and normal-hearing children, we also observed the same modality effect in learning that was found for immediate memory span. Auditory presentation was better than visual presentation. And, as before, the deaf children also showed a reversal of this modality effect for learning. Visual presentation was better than auditory presentation.

To assess the magnitude of the repetition learning effects, we computed difference scores between the learning and memory conditions by subtracting the memory span scores from the learning span scores for each subject. The average difference scores for the three groups of subjects are shown in Figure 4, while the data for individual subjects in each group for the three presentation formats are displayed in Figure 5. Inspection of these distributions in Figure 5 reveals a wide range of performance for all three groups of subjects. While most of the subjects in each group displayed some evidence of learning in terms of showing a positive repetition effect, there were a few subjects in the tail of the distribution who either failed to show any learning at all or showed a small reversal of the predicted repetition effect. Although the number of subjects who failed to show a repetition effect was quite small in the adults and normal-hearing children, about one-third of the deaf children with cochlear implants showed no repetition learning effect at all and failed to benefit from having the same stimulus sequence repeated on each trial.

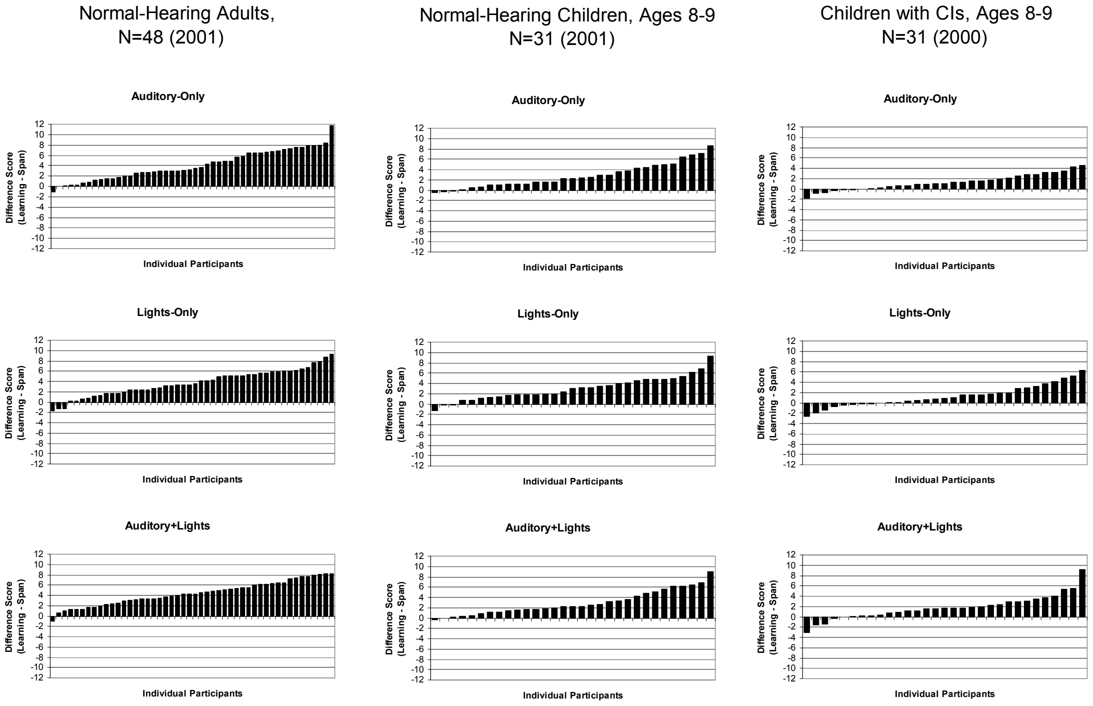


**Figure 5.** Difference scores for individual subjects showing sequence learning score minus his working memory span score. Data for the auditory-only (A) condition is shown on the top, lights-only (L) condition in the middle, and auditory-plus-lights (A+L) condition on the bottom. Data from normal-hearing adults are shown on the left, scores for normal-hearing 8- and 9-year-old children in the center, and scores for 8- and 9-year-old cochlear implant users on the right (Redrawn from Pisoni & Cleary, 2004).

To study the relations between sequence learning and speech and language development in these children, Cleary and Pisoni (2001) computed a series of correlations between the three learning scores obtained from the Simon learning task and several of the traditional audiological outcome measures of benefit that were obtained from these children as part of the larger CID project (see Geers, Nicholas & Sedey, 2003). None of the demographic variables were found to be correlated with any of the Simon sequence learning scores. However, moderate positive correlations were obtained for three measures of spoken word recognition, the WIPI, BKB sentences and the LNT ($r$ = +.36, $p$ = <.05) and the auditory-only Simon learning condition. Moreover, the auditory-only Simon learning span was also found to be correlated with the TACL-R measure of receptive language ($r$ = +.42, $p$ = <.05) as well as the backwards WISC digit span ($r$ = +.43, $p$ = <.05). Thus, Simon learning in the auditory-only condition was positively correlated with outcome measures that involve more complex cognitive processing activities that reflect "executive functions" and "controlled attention" (Engle, Kane & Tuholski, 1999). Performance on the TACL-R reflects the ability to comprehend subtle morphological and syntactic distinctions. Similarly, performance on the backward digit span task assesses the ability to explicitly manipulate the serial order of items actively maintained in working memory. Both of these measures, along with measures of open-set word recognition on the LNT, assess the storage and maintenance of verbal items in short-term memory and the subsequent processing operations of working memory and controlled attention.

In the current study, we assessed the relations between measures of sequence learning in children with cochlear implants and several speech and language outcome measures with a different group of deaf children who use cochlear implants. The initial studies on sequence learning by Cleary and Pisoni (2001) were carried out with 8- and 9-year old deaf children as part of the large scale CID project directed by Ann Geers (see Geers et al., 2003). The present study used deaf children from the IU School of Medicine. These children spanned a wider age range than the children used in the earlier study. We also examined two new measures of sequence learning.

## Methods

### Participants

Participants in the present study were 21 children who experienced a profound hearing loss before the age of 36 months and who received a cochlear implant before 9 years of age. Children were tested once every six months to a year for five years on a battery of clinical tests that are used to assess speech and language benefit after implantation. A summary of the demographics is provided in Table 1.

| Communication Mode | Age at Implantation | Age of Onset of Deafness | Pure Tone Average (unaided) | Pure Tone Average (cochlear implant) |
|---|---|---|---|---|
| 17 OC<br>9 TC | 39.42 (14.1) | 6.19 (9.6) | 111.9 (5.9) | 35.9 (6.2) |

**Table 1.** Demographic characteristics of the participants. Standard deviations are shown in parenthesis.

### Procedures

To measure sequence learning in this group of children, we used the same Simon memory game methodology that was employed in our earlier studies (Cleary & Pisoni, 2001; Cleary, Pisoni & Geers, 2001; Cleary, Pisoni & Kirk, 2002; Pisoni & Cleary, 2004). This methodology used a customized response box to investigate the effects of different presentation formats on immediate recall. Participants were presented with sequences of color names or colored lights under two conditions, visual-alone (V) and auditory+visual (AV). The child was asked to reproduce the stimulus pattern by pressing a sequence of colored response panels on the four-alternative response box using a manual response. The dependent measure of performance was the child's immediate memory span, defined as the longest length sequence he/she could correctly reproduce.

In the sequence learning condition, the stimuli on the Simon were arranged in temporal patterns that systematically increased in length using an adaptive staircase procedure as the subject successfully progressed through a block of trials. If the participant reproduced a pattern correctly, the same pattern was repeated again, but was increased in length by one item. If the child made an error, the same pattern was repeated again, but the sequence was reduced in length by one item. The Simon learning procedure spans in this study were obtained from all children under two presentation conditions: auditory+visual (AV) and visual-alone (V). All the children were tested at the Indiana University School of Medicine by a speech language pathologist or audiologist who was trained in testing deaf children with cochlear implants.

**Dependent Measures**

Two measures of learning were examined in this study. The first measure, Simon redundancy gain, was computed by subtracting the V weighted span from the AV weighted span on the Simon learning task in the first interval the child was tested. The difference in performance between the AV and V conditions can be thought of as a measure of how much gain the child received from the addition of redundant auditory information to the visual pattern. Because of the way we selected children for this analysis, length of cochlear implant use and chronological age were confounded. However, to control for these difference, length of cochlear implant use and chronological age were treated as covariates in the statistical analyses.

A second measure of learning, Simon learning improvement, was computed by subtracting the Simon learning weighted span from the first interval the child was tested (for both V and AV conditions) from the span obtained in the last interval the child was tested, and dividing by the total number of years between the scores. This measure was designed to assess changes in sequence learning over time with a cochlear implant, while eliminating any baseline differences. Unlike the first measure, which was used to assess the contribution of redundant auditory information on visual sequence learning, the second measure allowed us to examine the changes in memory and learning over time after a period of cochlear implant use.

**Outcome Measures**

To examine the relationship between these two measures of learning and the children's speech and language outcomes with their implant, we first performed a series of simple bivariate correlations with several traditional speech and language outcome measures. We looked at open-set word recognition (PBK words), sentence comprehension (Common Phrases A, V and AV), vocabulary knowledge (PPVT), language development (RDLS and CELF), and speech intelligibility (BIT). In each of these analyses, the outcome measures were from the first interval the child was tested using the Simon learning procedure.

## Results

**Simon Redundancy Gain**

Pearson correlations were used to assess the relations between redundancy gain and the clinical outcome measures. A summary of these correlations is provided in Table 2. A moderate correlation was found with the Common Phrases auditory alone scores ($r$ = +0.62, $p$ = .02). After controlling for age and length of use using partial correlations, the relationship between auditory redundancy gain and auditory comprehension as measured by the Common Phrases A-alone test was still reliable, indicating that this finding is not attributable to the age or device use of the children.

| Clinical Outcome Measures | Correlations with Simon Learning Redundancy Gain Scores |
|---|---|
| CPA (n=13) | 0.62 (*p*=0.02) |
| CPAV (n=12) | 0.56 (*p*=0.06) |
| CPV (n=16) | 0.17 (ns) |
| PBK words (n=19) | 0.23 (ns) |
| PBK phonemes (n=19) | 0.33 (ns) |
| PPVT AE (n=20) | -0.40 (ns) |

**Table 2.** Correlations between Simon redundancy gain scores and outcome measures.

## Simon Learning Improvement

Correlational analyses also revealed that the learning improvement measure was related to the vocabulary knowledge of the child at the time of first testing using the Simon memory game, although the relationship was in different directions for the AV and V conditions (see Table 3). The amount of auditory+visual improvement in learning over time was positively related to the child's initial vocabulary knowledge (*r* = +0.55, *p* = 0.04), while the amount of longitudinal visual-only gain was negatively related (*r* = -0.64, *p* = 0.01). The learning effect also remained reliable after performing partial correlations controlling for age and device use. This pattern suggests that greater vocabulary knowledge is associated with better sequence learning skills. Higher PPVT vocabulary scores were associated with increases in AV span and decreases in V span scores.

| Clinical Outcome Measures | Correlations with Simon Learning Improvement Measure (AV) | Correlations with Simon Learning Improvement Measure (V) |
|---|---|---|
| CPA (n=11) | 0.09 (ns) | -0.42 (ns) |
| CPAV (n=10) | -0.02 (ns) | -0.09 (ns) |
| CPV (n=14) | -0.23 (ns) | 0.16 (ns) |
| PBK words (n=14) | 0.07 (ns) | -0.16 (ns) |
| PBK phonemes (n=14) | 0.17 | -0.25 (ns) |
| PPVT AE (n=15) | 0.55 (*p*= 0.04) | -0.64 (*p*= 0.01) |

**Table 3.** Correlations between Simon learning improvement scores and outcome measures.

## Intercorrelations Between Measures of Learning

These two measures of learning were also strongly correlated with each other (see Table 4). The correlations remained strong even after partial correlations were performed controlling for age and device use. These two different measures of learning may share a common source of variance related to learning temporal sequences.

| | Simon Learning Redundancy Gain (N=16) |
|---|---|
| Simon Learning Improvement Measure (AV) | +0.62 (*p*=.01) |
| Simon Learning Improvement Measure (V) | -0.63 (*p*=.01) |

**Table 4.** Intercorrelations between the two measures of Simon learning.

## Discussion

The results of the present study reveal that simple measures of sequence learning in deaf children with cochlear implants are associated with changes in several traditional audiological outcome measures of speech and language. Our findings are of interest both clinically and theoretically because they suggest that the individual differences in outcome of children who receive cochlear implants may be due to fundamental learning processes that affect the encoding and retention of information in both short-term and long-term memory. Large improvements in immediate reproductive memory span for sequences of colored lights were obtained by simple repetition of a familiar sequence. Differences in the susceptibility to repetition effects such as these are associated with several outcome measures of speech and language. These initial findings on learning and memory suggest that differences in the development and operation of basic learning mechanisms in this clinical population may contribute an additional unique source of variance to the overall variation observed in a wide range of outcome measures following cochlear implantation. Additional studies of learning and memory in deaf children with cochlear implants are clearly warranted by the finding of this study and the earlier results on sequence learning first reported by Cleary and Pisoni (2001).

## References

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. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing, 22,* 395-411.

Cleary, M., Pisoni, D.B. & Kirk, K.I. (2002). Working memory spans as predictors of word recognition and receptive vocabulary in children with cochlear implants. *Volta Review, 102*, 259-280.

Dempster, F.N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89,* 63-100.

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.* London: Cambridge Press.

Geers, A.E., Nicholas, J.G. & Sedey, A.L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24*, 46S-58S.

Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*, 267-333.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467-477.

Penny, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17*, 398-422.

Pisoni, D.B. & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*, 106S-120S.

Pisoni, D.B. & Cleary, M. (2004). Learning, memory and cognitive processes in deaf children following cochlear implantation. In F.G. Zeng, A.N. Popper & R.R. Fay (Eds*.), Springer Handbook of Auditory Research: Auditory Prosthesis, SHAR Volume X.* (pp. 377-426). New York: Spring-Verlag.

Watkins, M.J., Watkins, O.C. & Crowder, R.G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior, 13*, 430-447.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Speech Perception Skills of Deaf Children with Cochlear Implants[1]

**David B. Pisoni**

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

[1] This work was supported by NIDCD Research Grant DC00111 and Training Grant DC00012 to Indiana University. I would like to thank Miranda Cleary, Caitlin Dillon and Rose Burkholder for their help and advice on all stages of the research described in this paper and Cynthia Clopper for editorial suggestions and advice. I would also like to thank Richard Miyamoto for extending an invitation to me to join his research program on cochlear implants at the Indiana University School of Medicine in Indianapolis. This paper is a longer written version of an invited talk delivered at "From Sound to Sense: 50+ Years of Discoveries in Speech Communication," June 12, 2004 at MIT, a conference to honor Professor Kenneth N. Stevens on his 80th birthday. I wish to extend my congratulations to Ken for his help and advice over the years. It was a great experience working in his group at MIT as a post-doc in 1975-76 and then on sabbatical leaves in 1978-79 and 1985-86. I'm ready for another year at MIT anytime!!

# Speech Perception Skills of Deaf Children with Cochlear Implants

**Abstract.** Cochlear implants work reasonably well in many profoundly deaf adults and children. For a prelingually deaf child, the electrical stimulation transmitted by a cochlear implant represents the introduction of a new sensory modality that provides spectral and temporal information about speech and spoken language. Despite the success of cochlear implants in many deaf children, large individual differences have been reported on a wide range of speech and language outcome measures. This finding is observed in all research centers around the world. Some children do extremely well with their cochlear implant while others derive only minimal benefits after receiving their implant. Understanding the reasons for the variability in outcomes and the large individual differences following cochlear implantation is one of the most important problems in the field today. In this paper, I present a brief summary of recent findings on the speech perception skills of deaf children following cochlear implantation. The results of these studies suggest that in addition to several demographic variables, variation in children's success with cochlear implants also reflects fundamental differences in rapid phonological coding and verbal rehearsal processes used in working memory.

## Introduction

A cochlear implant (CI) is a surgically implanted electronic device that functions as an auditory prosthesis for patients with severe to profound sensorineural hearing losses. The device provides electrical stimulation to the surviving spiral ganglion cells of the auditory nerve while bypassing the damaged hair cells of the inner ear to restore hearing. A cochlear implant provides both adults and children with access to sound and sensory information from the auditory modality. All of the cochlear implants in use today consist of an internal multiple electrode array and an external processing unit. The external unit consists of a microphone that picks up sound energy from the environment and a signal processor that codes frequency, amplitude and time and compresses the signal to match the narrow dynamic range of the ear. Cochlear implants provide temporal and amplitude information. Depending on the manufacturer, several different place coding techniques are used to represent and transmit frequency information in the signal. Since the approval of cochlear implants by the FDA as a treatment for profound deafness, over 60,000 patients have received cochlear implants at centers all over the world (Clarke, 2003).

For a postlingually deaf adult with a profound hearing loss, a cochlear implant provides a transformed electrical signal to an already fully developed auditory system and intact mature language processing system. These patients have already acquired spoken language under normal listening conditions so we know their central auditory system and brain are functioning normally. For a congenitally deaf child, however, a cochlear implant provides novel electrical stimulation through the auditory sensory modality and an opportunity to perceive speech sounds and develop spoken language for the first time after a period of auditory deprivation. Congenitally deaf children have not been exposed to speech and do not develop spoken language normally. Although their brain and nervous system continue to develop in the absence of normal auditory stimulation, recent findings suggest that some cortical reorganization has already taken place during the period of sensory deprivation before implantation and that several aspects of speech and language skills after implantation may develop in an atypical fashion. Both peripheral and central differences in neural function are likely to be responsible for the wide range of variability observed in outcome and benefit following implantation.

For the last 12 years, I have been studying congenitally deaf children who have received cochlear implants. These children are the most interesting and theoretically important clinical population to study because they have been deprived of sound and auditory stimulation at a very early point in neural and cognitive development. After implantation, their hearing is restored with electrical stimulation that is designed to simulate the response of a healthy cochlea to speech and other auditory signals. Aside from the obvious clinical benefits of cochlear implantation as a method of treating profound prelingual deafness in children, this clinical population also provides a unique opportunity to study the effects of auditory deprivation on the development of speech perception and language processing skills and to assess the effects of restoration of hearing via artificial electrical stimulation of the nervous system.

In some sense, one can think of research on this clinical population as the modern-day analog of the so-called "forbidden experiment" in the field of language acquisition, except that in this case after a period of sensory deprivation has occurred, hearing is restored via medical intervention and children begin to receive exposure to sound and stimulation through the auditory modality. Under these conditions, it is possible to study both the consequences of a period of auditory deprivation on speech and language development as well as the effects of restoring hearing using artificial electrical stimulation of the auditory nerve.

While cochlear implants work well for many profoundly deaf children, they do not always provide benefits to all children who receive them. Numerous studies have shown that the variation in audiological outcomes and benefits following cochlear implantation is enormous. Some children do extremely well with their cochlear implants and display near-typical speech perception and language skills on a wide range of traditional clinical speech and language tests when tested under quiet listening conditions in the laboratory. In contrast, other children struggle for long periods of time after they receive their cochlear implant and often never achieve comparable levels of speech and language performance or verbal fluency.

Why are some children doing so well with their CIs while others struggle and perform more poorly? How can we explain these differences in outcome and benefit? What underlying factor or set of factors are responsible for these large differences in performance on a wide range of behavioral tests? If we can identify the reasons why a good child is performing so well, we should be able to use this basic knowledge to help poorer performing children improve their speech and language skills and achieve their potential to derive optimal/maximal benefit from their CIs.

What do we know about outcome and benefit in deaf children with CIs? Table I lists seven key findings observed universally at all implant centers around the world. The findings indicate that several demographic and medical factors contribute to outcome and benefit following implantation. In addition to the enormous variability observed in the speech and language outcome measures, several other findings have been consistently reported in the clinical literature on cochlear implants in deaf children. An examination of these findings provides some preliminary insights into the possible underlying cognitive and neural basis for the variability in outcome and benefit among deaf children with cochlear implants. When these contributing factors are considered together, it is possible to begin formulating some specific hypotheses about the reasons for the enormous variability in outcome and benefit.

Almost all of the clinical research on cochlear implants has focused on the effects of a small number of demographic variables using traditional outcome measures based on assessment tools developed by clinical audiologists and speech pathologists. Although rarely discussed explicitly in the literature, these behaviorally-based clinical outcome measures of performance are the final product of a large number of complex sensory, perceptual, cognitive and linguistic processes that contribute to the observed variation among cochlear implant users. Until recently, little if any research focused on the underlying information processing mechanisms used to perceive and produce spoken language in this clinical population. Our investigations of these fundamental neurocognitive and linguistic processes have provided some new insights into the basis of individual differences in profoundly deaf children with cochlear implants.

1. **Large Individual Differences**
2. **Age of Implantation (Sensitive Periods)**
3. **Effects of Early Experience (Oral vs. TC)**
4. **Cross-Modal Plasticity**
5. **Links Between Perception & Production**
6. **No Preimplant Predictors of Outcome**
7. **Abilities Emerge after Implantation (Learning)**

**Table I.** Seven key findings on cochlear implants in deaf children.

Age at implantation is another factor that has been shown to influence all outcome measures of performance. Children who receive an implant at a young age do much better on a whole range of outcome measures than children who are implanted at an older age (Kirk, 2000). Length of auditory deprivation or length of deafness is also related to outcome and benefit. Children who have been deaf for shorter periods of time before implantation do much better on a variety of clinical measures than children who have been deaf for longer periods of time. Both findings demonstrate the contribution of sensitive periods in sensory, perceptual and linguistic development and serve to emphasize the close links between neural development and behavior, especially the development of hearing, speech and language (Konishi, 1985; Marler & Peters, 1988).

Early sensory and linguistic experience and language processing activities after implantation have also been shown to affect performance on a wide range of outcome measures. Implanted children who are immersed in Oral-only communication environments do much better on clinical tests of speech and language development than implanted children who are enrolled in Total Communication programs (Kirk, Pisoni, & Miyamoto, 2000). Oral communication approaches emphasize the use of speech and hearing skills and actively encourage children to produce spoken language to achieve optimal benefit from their implants. In contrast, total communication approaches employ the simultaneous use of some form of manual-coded English along with speech to help the child acquire language using both sign and spoken language inputs.

Until just recently, clinicians and researchers have been unable to find reliable preimplant predictors of outcome and success with a cochlear implant (see, however, Bergeson & Pisoni, 2004). The absence of preimplant predictors is a theoretically significant finding because it suggests that many complex interactions take place between the newly acquired sensory capabilities of a child after a period of auditory deprivation, properties of the language-learning environment and various interactions with parents and caregivers that the child is exposed to after receiving a cochlear implant. More importantly, however, the lack of preimplant predictors of outcome and benefit makes it difficult for clinicians to identify those children who are doing poorly with their cochlear implant at a time in development when changes can be made to modify and improve their language processing skills.

Finally, when all of the outcome and demographic measures are considered together, the available evidence strongly suggests that the underlying sensory and perceptual abilities for speech and language "emerge" after implantation. Performance with a cochlear implant improves over time for almost all children. Success with a cochlear implant therefore appears to be due, in part, to perceptual learning and exposure to a language model in the environment (Clarke, 2003). Because outcome and benefit with a cochlear implant cannot be predicted reliably from traditional behavioral measures obtained before implantation, any improvements in performance observed after implantation must be due to sensory and cognitive processes that are linked to maturational changes in neural and cognitive development (see Sharma, Dorman & Spahr, 2002).

Although these demographic factors can account for a large portion of the variance in outcomes, there are still substantial gaps in our basic knowledge of how CIs work in the brain. Moreover, several other sources of variability related to the "information processing" capacities of the children have also been found to contribute to outcome. These factors involve the sensory and perceptual encoding of speech, the storage and processing of phonological and lexical information in short-term memory and response output. There are also several linguistic factors that reflect the use of common phonological representations in a range of behavioral tasks that are routinely used to assess speech and language outcome and measure benefit following cochlear implantation.

One of the primary research questions deals with the nature of the variation in outcome and the reasons for the large individual differences in effectiveness of CIs. If we can structure, shape, modify and reorganize how the central auditory, cognitive and linguistic processes are working in these children, perhaps we can provide more benefit to the low-performing children and explain the observed variability in outcomes.

We are not only interested in whether deaf children with CIs are able to "hear" via their CI (i.e., detection and discrimination). We are also interested in what they are able to do with the sensory information they do hear. How do they go from continuous waveforms to discrete words? How do these children construct phonological representations of the input signals they hear? What kinds of linguistic representations do they create? Are their phonological and lexical representations "fully specified" like normal-hearing typically developing children or are they "underspecified" reflecting more primitive coarse-coding strategies and the use of broader phonetic categories? What can deaf children with CIs do with these representations? How are these representations used in speech perception, word recognition and spoken language comprehension? How are they used in speech production? And how are they used in other linguistic tasks like sentence comprehension, language production, and reading?

To investigate individual differences and the sources of variation in outcome, we began by analyzing a set of data from a longitudinal project on cochlear implants in children (see Pisoni et al., 1997; 2000). Our first study examined the exceptionally good users of cochlear implants—the so-called Stars. These are the children who did extremely well with their cochlear implants after only two years of implant use. The Stars are able to acquire spoken language quickly and easily and appear to be on a developmental trajectory that parallels normal-hearing children although delayed a little in time (see Svirsky et al., 2000). The theoretical motivation for studying the exceptionally good children was based on an extensive body of research on "expertise" and "expert systems" theory (Ericsson & Smith, 1991). Many important new insights have come from studying expert chess players, radiologists, spectrogram readers, like Victor Zue, and other individuals who have highly developed domain-specific skills.

## Analysis of the Stars

Using an extreme groups design, we initially sorted a large sample of deaf children who used CIs (N=160) into two categories based on their PBK phoneme scores after two years of implant use. The PBK test is an open-set test of spoken word recognition that requires the child to repeat back isolated English words spoken by an examiner using live-voice presentation (Haskins, 1949). The PBK test is considered to be very difficult for most deaf children who have received CIs and is considered to be the "gold standard" of outcome performance. The Stars consisted of children who scored in the top 20 percent of the distribution on the PBK test; the low-performers were the children in the bottom 20 percent of the distribution. We then examined the children's performance on several different speech and language outcome measures over a six year period (see Pisoni, et al., 1997; 2000). In this section, we describe the results of the speech feature perception and spoken word recognition tests obtained from this study. Other findings are reported in our earlier papers.

**Speech Feature Perception.** Measures of speech feature perception for consonants and vowels were obtained for both groups of subjects with the Minimal Pairs Test (Robbins et al., 1988). This test uses a two-alternative forced-choice picture pointing procedure. The child hears a single word spoken in isolation on each trial by the examiner using live-voice presentation and is required to select the picture that corresponds to the test word.

A summary of the consonant perception results for both groups of children is shown in Figure 1. Percent correct perception is displayed separately for manner, voicing and place of articulation as a function of implant use in years. Data for the Stars are shown by the filled bars; data for the low performers are shown by the open bars. Chance performance on this task is 50% correct as shown by the solid line. A second horizontal line is also shown in this figure at 70% correct corresponding to scores that are significantly above chance using the binominal distribution.

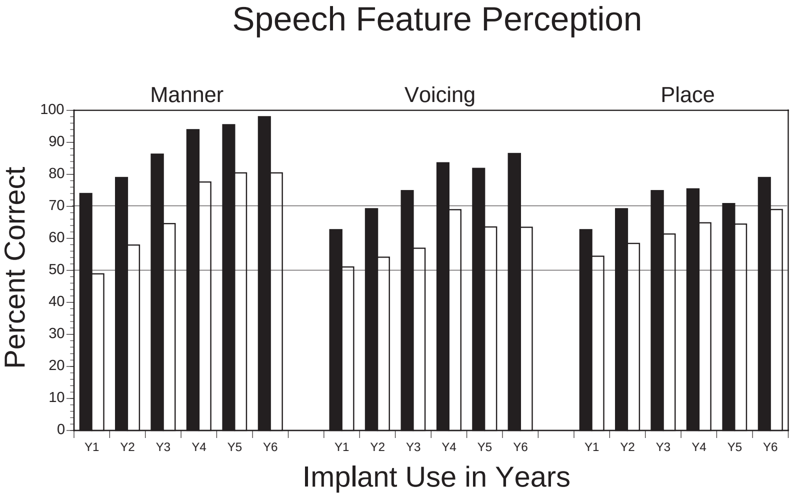


**Figure 1.** Percent correct recognition on the Minimal Pairs Test (MPT) for manner, voicing and place as a function of implant use. The Stars are shown by filled bars, the low performers are shown by open bars.

Inspection of the results for the minimal pairs test obtained over a period of six years of implant use reveals several consistent findings. First, performance by the Stars was consistently better than performance by the low performers for every comparison across all three consonant features. Second, perception improved over time with implant use for both groups overall, although the increases were primarily due to improvements in perception of manner and voicing by the Stars. At no interval did the mean scores of the low performers exceed chance performance on perception of voicing and place features. Although increases in performance were observed over time for this group, their perception scores never reached the levels of performance observed with the Stars, even for the gross manner contrasts that eventually exceeded chance performance in years 4, 5 and 6.

The results of the minimal pairs test indicate that both groups of children have difficulty perceiving fine phonetic details of isolated spoken words even in a two-choice closed-set testing format. The Stars were able to discriminate differences in manner of articulation after one year of implant use and they showed consistent improvements in performance over time for both manner and voicing contrasts but they still had more difficulty reliably discriminating differences in place of articulation, even after five years of implant use. In contrast, the low performers were just barely able to discriminate differences in manner of articulation after four years of implant use and they were unable to reliably perceive differences in voicing and place of articulation even after five or six years of use.

The pattern of results shown in Figure 1 suggests that both groups of children are encoding spoken words using “coarse” phonetic representations that contain much less fine-grained phonetic detail than normal hearing children typically use. Their lexical representations appear to be “underspecified” compared to the representations that normal-hearing children typically use. The Stars are able to reliably discriminate manner and to some extent voicing much earlier after implantation than the low performing children. They also display consistent improvements in speech feature perception over time. These speech feature perception skills are assumed to place initial constraints on the basic sensory information that can be used for subsequent word learning and lexical development. It is very likely that if a child cannot reliably perceive differences between pairs of spoken words that are acoustically similar under these relatively easy forced-choice test conditions, he/she will also have a great deal of difficulty recognizing words in isolation with no context or retrieving the phonological representations of these sound patterns from memory for use in simple speech production tasks such as imitation or immediate repetition. We would also expect these children to have a great deal of difficulty in recognizing and imitating nonwords that have no lexical representations.

**Open-Set Word Recognition.** Two word recognition tests, the Lexical Neighborhood Test (LNT) and the Multi-syllabic Lexical Neighborhood test (MLNT), were used to measure open-set word recognition skills in both groups of subjects (Kirk, Pisoni & Osberger, 1995). Both tests use words that are familiar to normal-hearing preschool age children. The LNT contains monosyllabic words, the MLNT contains polysyllabic words. Each test uses two different sets of words in order to measure lexical discrimination and provide details about how the lexical selection process is carried out. Half of the items in each test consist of lexically “easy” words and half consist of lexically “hard” words. The differences in performance on these two types of items in each test provide an index of how well a child is able to make fine phonetic discriminations among phonetically similar words. Differences in performance between the LNT and the MLNT provide a measure of the extent to which the child is able to make use of word length cues to recognize and access words from the mental lexicon. The items on both tests were presented in isolation one at a time by the examiner using live voice auditory-only presentation. A mesh screen was used to cover the talker’s face. The child was required to repeat back the test item immediately after it was presented by the examiner on each trial.

Figure 2 shows percent correct word recognition on the LNT and the MLNT for the Stars as a function of years of implant use. Scores for the "easy" and "hard" words are shown separately for each test by year. Within each year of use, the results for the LNT are shown on the left while the MLNT results are shown on the right. Several important differences are shown here that provide some insight into the task demands and processing operations used in open-set word recognition tests. First, the Stars consistently demonstrated higher levels of word recognition performance on both the LNT and the MLNT than the low performers. These differences were present across all six years but they are most prominent during the first three years after implantation. Word recognition scores for the low performers on both tests were low and close to the floor. Normal-hearing children typically display ceiling levels of performance on both of these tests by age 4 (see Kluck et al., 1997).

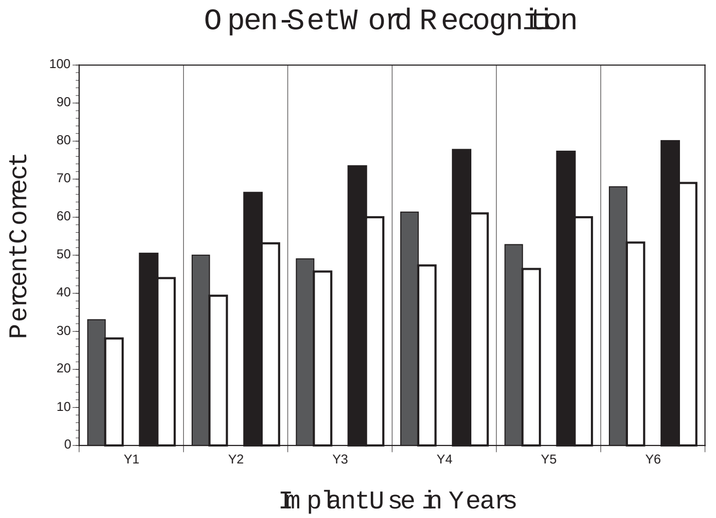


**Figure 2.** Percent correct word recognition performance of the Stars on the Lexical Neighborhood Test (LNT) and the Multisyllabic Lexical Neighborhood Test (MLNT) as a function of implant use and lexical difficulty. "Easy Words" are shown by filled bars, "Hard Words" are shown by open bars.

Two other theoretically important findings are also shown in this figure. First, the Stars displayed evidence of a word length effect at each testing interval. Recognition was always better for the long words than short words. This pattern was obscured by a floor effect for the low performers who were unable to do this open-set task at all during the first three years. The presence of a word length effect for the Stars demonstrates that they recognize words "relationally" in the context of other words that they have in their lexicon (Luce & Pisoni, 1998). If these children were simply recognizing words in isolation, feature-by-feature or segment-by-segment, without reference to other words they already know and can access from their lexicons, we would expect that performance would be worse for longer words than shorter words because longer words contain more information. The pattern of findings shown in Figure 2 is exactly the opposite of this prediction and parallels earlier results obtained with normal-hearing adults and children (Luce & Pisoni, 1998; Kirk et al., 1995; Kluck et al., 1997). Longer words are easier to recognize than shorter words because they are more distinctive and discriminable and therefore less confusable with other phonetically similar words. The present findings suggest that the Stars are recognizing words based on their knowledge of other words in the language and generalizations about sublexical patterns of words in their lexicons using processing strategies that are similar to those used by normal-hearing listeners.

Additional support for the role of the lexicon and the use of lexical knowledge of words in open-set word recognition is provided by a second finding shown in this figure. The Stars also showed a consistent effect of "lexical discrimination" for the words on both tests. They recognized lexically "easy" words better than lexically "hard" words according to predictions of the NAM model of spoken word recognition (see Luce & Pisoni, 1998). The NAM model assumes that words are organized and stored in an acoustic-phonetic similarity space in long-term lexical memory and that words are recognized in the context of other phonetically similar words by processes of bottom-up acoustic-phonetic activation followed by top-down lexical competition. The difference in performance between "easy" words and "hard" words is present for both tests but it is larger and more consistent over time for the words on the MLNT test. The low performers did not show the same pattern of sensitivity to lexical competition among the test words.

The differences in performance observed between these two groups of children on both open-set word recognition tests are not at all surprising and were anticipated because these two extreme groups were initially created based on their PBK scores, another open-set word recognition test. The overall pattern of the results is theoretically important because the findings obtained with these two open-set word recognition tests demonstrate that the skills and abilities used to recognize isolated spoken words are not specific to the words used on the PBK test. The initial differences between the two groups were replicated with two other open-set word recognition tests using different sets of words.

**Correlations Among Outcome Measures**. We also found that the speech feature perception and spoken word recognition scores from the Stars along with several other outcome measures such as sentence comprehension, vocabulary knowledge, expressive and receptive language and speech intelligibility were all highly inter-correlated with each other (Pisoni et al., 2001). This pattern of results suggests the existence of a common underlying source of variance. We have suggested that this source of variance is related to the rapid construction of phonological representations in speech perception and the use of a set of common phonological processing skills across a wide range of different behavioral tests that are routinely used to measure outcome and assess benefit following cochlear implantation.

The speech perception and word recognition data obtained from these children were based on traditional audiological outcome measures that were collected as part of their annual clinical assessments. The scores on these tests are "endpoint measures" of the child's behavior that reflect the final product of a long series of cognitive, perceptual, and linguistic analyses of the input signal. Process measures of performance that assess what a child does with the sensory information provided by his/her cochlear implant were not obtained in the standard research protocol used in our longitudinal study of the Stars so it was impossible to measure information processing capacity, processing speed or learning. To determine if fundamental differences in immediate memory capacity, processing speed and learning might be responsible for the variation in performance, we obtained several new process measures of performance on another group of children with CIs.

## New Process Measures of Performance

**Immediate Memory Capacity.** The information processing capacity of immediate memory was measured using forward and backward WISC digit spans in a large group (N=176) of 8- and 9- year old children who had used their cochlear implants for at least five years (see Pisoni & Cleary, 2003). We found that their forward digit spans which are assumed to reflect early phonological coding and verbal rehearsal skills were "atypical" compared to a group of age-matched normal-hearing children (N=44). Both forward and backward digit spans were shorter for children with CIs than normal-hearing children suggesting limitations on processing capacity of immediate memory. We also found that children who were enrolled in oral communication (OC) programs had longer forward digit spans than children enrolled in total communication (TC) programs. This result demonstrates significant effects of early linguistic experience and activity-dependent learning on information processing capacity of immediate memory.

**Verbal Rehearsal Speed.** Estimates of the deaf children's verbal rehearsal speed were also obtained in this study by measuring speaking rates from spoken sentences based on sentence durations (see Pisoni & Cleary, 2003). The results of this analysis revealed slower verbal rehearsal speeds. As shown in other typical-developing populations, verbal rehearsal speed was also found to be strongly correlated with immediate memory capacity (i.e., forward digit spans) as well as several independent measures of spoken word recognition in isolation and sentence context and a measure of speech intelligibility (see Pisoni & Cleary, 2004).

**Scanning of STM.** Several novel measures of speech timing during the digit span recall task were also used to investigate retrieval and scanning of phonological information in STM and speed of articulation at output (see Burkholder & Pisoni, 2003). We found that scanning of familiar digits was significantly slower for the children with CIs, suggesting less robust perceptual encoding and active maintenance of phonological representations of digits in STM. In addition, we found that early linguistic experience also affected the scanning rates. Children who were immersed in oral (OC) educational programs displayed faster scanning rates for items in STM than children who were in total communication (TC) programs.

**Sequence Memory Spans.** Using a novel experimental methodology, based on Milton-Bradley's Simon memory game, we also measured reproductive memory spans for auditory (A), visual (V) and auditory+visual (AV) sequences (see Cleary, Pisoni & Geers, 2001; Pisoni & Cleary, 2004). Sequences of spoken color names (A) or colored lights (V) or color names combined simultaneously with colored lights (A+V) were presented randomly on each trial using the Simon game box. Subjects were asked to reproduce the sequence by pressing the appropriate colored panels on the Simon box. With an adaptative testing algorithm, we obtained several measures of the longest sequence a subject could correct reproduce. Across all three presentation conditions, we found that deaf children with CIs displayed shorter reproductive memory spans than normal-hearing children and adults. More interestingly, they also showed smaller redundancy gains in the A+V multimodal presentation condition compared to A-only and V-only unimodal presentation conditions. Also, as reported in the memory literature, the normal-hearing children and adults both showed a "modality effect," that is, they had longer memory spans for auditory sequences of color names compared to visual sequences of colored lights. In contrast, the deaf children with CIs displayed a reversal of the "modality effect" found with the normal-hearing children and adults. Their memory spans for visual sequences spans were actually longer than their spans for auditory sequences suggesting differential effects of cross-modal plasticity and reorganization as a consequence of a period of deafness early in life before implantation.

**Sequence Learning Spans.** The Simon memory game methodology was also used to study simple sequence learning and repetition effects under the same three presentation conditions used in the memory experiments (see Pisoni & Cleary, 2004). In the learning experiment, if a sequence was reproduced correctly on a given trial, the same sequence was repeated again on the next trial but the length of the sequence was increased by one item chosen randomly from the set of four colors. We found that the sequence learning spans were much shorter for children with CIs than children with normal hearing across all three presentation conditions. More importantly, about one third of the deaf children with CIs failed to show any sequence repetition effects or learning at all. This result was observed across all three presentation conditions indicating the presence of significant differences in simple sequence learning based on pattern repetition that are not modality specific in nature. These findings suggest that a period of sensory deprivation due to deafness at early stages of development produces effects on the central nervous system and the cortical areas that are involved in learning, memory and cognitive processes used to encode and maintain sensory inputs from both visual and auditory modalities.

## Phonological Decomposition and Sublexical Knowledge

**Nonword Repetition Task.** To understand the linguistic factors that are responsible for the variation in speech and language outcome measures following implantation, we have examined the use of sublexical phonological knowledge in speech perception with data obtained from a nonword repetition task (Cleary, Dillon et al., 2002; Carter et al., 2002; Dillon et al., 2004a, b). In this task, children are asked to listen to a novel nonsense word that conforms to English phonology and immediately repeat it back to the experimenter. The children are told in advance that the stimuli will be unfamiliar "funny-sounding" words and they should just try to say back whatever they hear on each trial to the best of their ability. Their verbal responses were recorded on DAT tape for later linguistic analysis and playback. Although nonword repetition appears at first glance to be a simple information processing task, in actuality it is a very complex linguistic task that requires the child to perform well on each of the individual component processes including: speech perception, phonological encoding and decomposition, active verbal rehearsal in working memory, retrieval and phonological reassembly and finally phonetic implementation and speech production. One motivation for studying the nonword repetition skills of children with CIs was the assumption in the clinical literature that hearing-impaired children display great difficulty in open-set tests of word recognition because they do not know the meanings of the stimulus words that are used on the test (see Kirk et al., 1995). A second reason was that nonword repetition performance by normal-hearing children has been found to be strongly related to variation in vocabulary development and other language learning milestones (Gathercole & Baddeley, 1993; Gathercole et al., 1999).

**Linguistic Analysis and Perceptual Ratings of Nonword Responses.** Twenty nonword patterns were presented to a large group of 8- and 9-year old deaf children (N=88) with cochlear implants (see Dillon et al., 2004a, b). Responses were then phonetically transcribed by two trained linguists. The nonword responses were also played back to normal-hearing listeners who were asked to make perceptual similarity ratings of each response following presentation of the original nonword pattern. Several scores were computed for each child to measure correct repetition of the consonants and vowels in each nonword stimulus as well as perceptual ratings. These scores were then correlated with several independent measures of performance on the individual component skills. A summary of the correlations of these nonword repetition scores with open-set word recognition (LNT and MLNT), Forward Digit Span, Speech Intelligibility, Speaking Rate, Word Attack, and Rhyme Errors is given in Table II. The Word Attack and Rhyme Errors were obtained from visual reading tasks that were specifically designed to measure single word reading. As shown here, the transcription scores for both consonants and vowels as well as the perceptual ratings were all strongly correlated with these separate component measures. The pattern and consistency of these correlations suggests the use of common phonological representations and analysis skills across a wide range of language processing tasks. The results obtained with the nonword repetition task also suggest an independent and autonomous sublexical level of phonological analysis that is separate and distinct from the lexical level. Like normal-hearing children, deaf children with cochlear implants are able to reproduce and imitate novel word-like patterns. Many of these children were able to decompose unfamiliar nonsense words into smaller linguistically significant segments and to then rapidly retrieve and reassemble these segments into motor outputs for speech production without relying on lexical representations in long-term memory.

| | Percent Correct Consonants (N=76) | Percent Correct Vowels (N=76) | Mean Perceptual Accuracy Rating (N=76) |
|---|---|---|---|
| **LNT easy words** | +.83*** | +.78*** | +.76*** |
| **LNT hard words** | +.85*** | +.71*** | +.70*** |
| **MLNT** | +.77*** | +.74*** | +.77*** |
| **Forward Digit Span** | +.60** | +.62** | +.76*** |
| **Speech Intelligibility** | +.91*** | +.88*** | +.87*** |
| **Speaking Rate** | -.84*** | -.81*** | -.85*** |
| **Word Attack** | +.75*** | +.72*** | +.78*** |
| **Rhyme Errors** | -.63** | -.68** | -.54* |

*$p$<.05, **$p$<.01, ***$p$<.001

**Table II.** Partial correlations between nonword repetition scores and several speech and language outcome measures (controlling for performance IQ, age at onset of deafness and communication mode).

## Discussion and Conclusions

What is the common factor that links these diverse sets of findings together? We suggest that the development and efficient use of phonological processing skills is a significant contributor above and beyond the traditional demographic and medical variables that have been shown to affect outcome and benefit following cochlear implantation. Phonological analysis involves the rapid encoding and decomposition of speech into sequences of discrete meaningless phonetic segments or “particles” and the assignment of structural descriptions to these sound patterns that reflect the linguistically significant sound contrasts of the words in the ambient target language. For many years, both clinicians and researchers have considered open-set tests of spoken word recognition performance to be the “gold standard” of outcome and benefit in both children and adults who have received CIs. The reason open-set tests have achieved this privileged status in the field is because they load heavily on several component processes including speech perception, verbal rehearsal and maintenance, retrieval of phonological representations from STM, and phonetic implementation strategies required for speech production, motor control and response output. All of these subprocesses rely on highly automatized phonological processing skills involving analysis and decomposition of the input signal into familiar linguistic units, like features or segments, and the reassembly and synthesis of these units into sequences of motor commands and gestures for output.

When prelingually deaf children receive a cochlear implant as a treatment for their profound hearing loss, they do not simply have their hearing restored at the auditory periphery. More significantly, after implantation they begin to receive substantial auditory stimulation to specialized areas of their central nervous system that are critical for the development of spoken language and specifically for the development of phonological processing skills that are used to rapidly encode and process speech signals. Our recent findings on speech perception and phonological decomposition in deaf children with cochlear implants suggest that in addition to the traditional demographic and medical variables that are able to predict some proportion of the variance in traditional audiological measures of outcome and benefit, there are several additional sources of variance that reflect the contribution of basic information processing skills commonly used in a wide range of language processing tasks, specifically those which rely on rapid phonological encoding of speech and verbal rehearsal strategies in working memory. Thus, some proportion of the variability and individual differences in outcome following cochlear implantation is related to central auditory, cognitive and linguistic factors that reflect how the initial sensory information transmitted by the cochlear implant is subsequently encoded and processed and how it is used by the listener in specific behavioral tasks that are routinely used to measure speech and language outcomes and assess benefit.

## References

Bergeson, T. & Pisoni, D.B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence & B.E. Stein (Eds.), *Handbook of Multisensory Integration*, Cambridge: MIT Press, 749-772.

Burkholder, R. & Pisoni, D.B. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85*, 63-88.

Carter, A.K., Dillon, C.M. & Pisoni, D.B. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: Suprasegmental analyses. *Clinical Linguistics & Phonetics, 16,* 619-638.

Clarke, G. (2003). *Cochlear Implants: Fundamentals and Applications*. New York: Springer-Verlag.

Cleary, M., Dillon, C.M. & Pisoni, D.B. (2002). Imitation of nonwords by deaf children after cochlear implantation: Preliminary findings. *Annals of Otology, Rhinology, & Laryngology Supplement-Proceedings of the 8th Symposium on Cochlear Implants in Children, 111*, 91-96.

Cleary, M., Pisoni, D.B. & Geers, A.E. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear & Hearing, 22*, 395-411.

Dillon, C.M., Cleary, M., Pisoni, D.B. & Carter, A.K. (2004a). Imitation of nonwords by hearing-impaired children with cochlear implants: Segmental analyses. *Clinical Linguistics and Phonetics, 18,* 39-55.

Dillon, C.M., Pisoni, D.B., Cleary, M., & Carter, A.K. (2004b). Nonword imitation by children with cochlear implants: Consonant analyses. *Archives of Otolaryngology -Head & Neck Surgery, 130,* 587-591.

Ericsson, K.A., & Smith, J. (1991). *Toward a General Theory of Expertise: Prospects and Limits*. New York, NY: Cambridge University Press.

Gathercole, S.E. & Baddeley, A.D. (1993). *Working Memory and Language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

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

Kirk, K.I. (2000). Challenges in the clinical investigation of cochlear implant outcomes. In J.K. Niparko, K.I. Kirk, N.H. Mellon, A.M. Robbins, B.L. Tucci & B.S. Wilson (Eds.), *Cochlear Implants: Principles and Practices*, Philadelphia, PA; Lippincott, Williams & Wilkins, 225-259.

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.B. Waltzman & N.L. Cohen (Eds.), *Cochlear Implants*, New York: Thieme, 252-254.

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

Kluck, M., Pisoni, D.B. & Kirk, K.I. (1997). Performance of normal-hearing children on open-set speech perception tests. *Progress Report on Spoken Language Processing #21*, Indiana University, Department of Psychology, Bloomington, IN.

Konishi, M. (1985). Birdsong: From behavior to neuron. *Annual Review of Neuroscience, 8*, 125-170.

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

Marler, P., & Peters, S. (1988). Sensitive periods for song acquisition from tape recordings and live tutors in the swamp sparrow, Melospiza georgiana. *Ethology, 77*, 76-84.

Pisoni, D.B. & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear & Hearing, 24*, 106S-120S.

Pisoni, D.B. & Cleary, M. (2004). Learning, memory and cognitive processes in deaf children following cochlear implantation. In F.G. Zeng, A.N. Popper & R.R. Fay (Eds), *Handbook of Auditory Research: Auditory Prosthesis, SHAR Volume X* (Pp. 377-426*),* Springer.

Pisoni, D.B., Svirsky, M.A., 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.

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 prelingually deaf children: Some new process measures of performance. *Volta Review, 101,* 111-164.

Robbins, A.M., Renshaw, J.J., Miyamoto, R.T., Osberger, M.J., & Pope, M.L. (1988). Minimal pairs test. Indianapolis, IN: Indiana University School of Medicine.

Sharma, A., Dorman, M.F. & Spahr, A.J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: Implications for age of implantation. *Ear & Hearing, 23*, 532-539.

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.

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**
Progress Report No. 26 (2003-2004)
*Indiana University*

# Multimodal Sentence Intelligibility and the Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals: A First Report[1]

**Brianna L. Conrey**

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

[1] This work was supported by NIH NIDCD R01 research grant DC00111 and NIH NIDCD T32 training grant to Indiana University. The author was also supported by an NSF Graduate Research Fellowship. I would like to thank David Pisoni for his help and advice through all stages of this project. I would also like to thank Luis Hernandez for technical support and Sara Phillips for assistance with data collection.

# Multimodal Sentence Intelligibility and the Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals: A First Report

**Abstract.** The ability to perceive and understand visual-only speech and the benefit experienced from having both auditory and visual signals available during speech perception tasks varies widely in the normal-hearing population. At the present time, little is known about the underlying neural mechanisms responsible for this variability or the possible relationships between multisensory speech perception abilities and performance on other perceptual or cognitive tasks. Previous studies have hypothesized that lipreading ability and auditory-visual (AV) benefit measures might be positively correlated with the ability to detect asynchronies in AV signals. Good integrators might be more attuned to detailed temporal relationships between auditory and visual information. However, this hypothesis has not been explicitly tested in normal-hearing individuals. In the present investigation, 30 normal-hearing participants were given a modified clinical test of sentence intelligibility (the CUNY sentences) under auditory-only, visual-only, and auditory-visual (AV) presentation conditions. The same participants also performed an AV asynchrony detection task using both speech and nonspeech multimodal stimuli that varied over a range of temporal asynchronies. The results suggest a relationship between auditory-only, visual-only, and AV sentence intelligibility measures and the ability to detect AV asynchrony in speech and nonspeech signals. Implications for AV integration in speech perception are discussed.

## Introduction

The ability to detect temporal asynchrony between auditory and visual signals varies considerably among individuals (e.g., Conrey & Pisoni, 2003, this volume; Grant & Seitz, 1998; McGrath & Summerfield, 1985). For example, McGrath and Summerfield reported that although some individuals are consistently able to detect auditory-visual (AV) asynchronies when the auditory signal precedes the visual signal by as little as 30 ms or when the visual signal precedes the auditory signal by as little as 70 ms, others require as much as 210 ms of auditory lead or 330 ms of visual lead.

Previous reports in the literature have suggested that individuals who are highly accurate at understanding visual-only speech (i.e., "good lipreaders") may also be better at detecting AV asynchronies than poor lipreaders, because good lipreaders might be more attuned to detailed temporal relationships between auditory and visual information (Grant & Seitz, 1998; McGrath & Summerfield, 1985; Pandey, Kunov, & Abel, 1986). In their 1985 study, McGrath and Summerfield presented participants with asynchronous AV sentences in which the auditory signal was only the fundamental frequency of the speaker's voice. They reported that good lipreaders displayed significantly decreased speech perception accuracy scores with auditory delays of 80 ms, but the performance of average and poor lipreaders did not decrease significantly until the auditory signal was delayed by 160 ms. This finding suggests that good lipreaders were less able to effectively integrate auditory and visual information when the two sources of information were asynchronous. In a follow-up experiment, McGrath and Summerfield used a three-interval forced-choice procedure and asked participants to identify the target synchronous stimulus from a pair of AV nonspeech approximations of CV syllables. They reported that higher lipreading scores had a moderate association with lower thresholds for detecting AV asynchrony, but this association did not reach statistical significance, perhaps because of the small number of subjects.

In another study, Pandey et al. (1986) also tested experienced and inexperienced lipreaders on the perception of AV sentences in which the auditory signal was delayed and mixed with multitalker babble. At a SNR of 0 dB, inexperienced lipreaders showed decreased word identification accuracy when the auditory signal was delayed by 300 ms relative to the visual signal. At a SNR of -10 dB, their performance declined from the synchronous condition when the auditory signal was delayed by 180 ms. For experienced lipreaders, word-identification performance deteriorated significantly by 160 ms at an SNR of -5 dB. Unfortunately, because the two groups were not tested at the same asynchrony levels or SNRs, it is difficult to form any firm conclusions about the effects of lipreading experience on integration in asynchronous AV sentences based on this study.

More recently, Grant and Seitz (1998) tested hearing-impaired adults on several measures of AV integration. They found no consistent relationship between high lipreading scores and the effects of auditory delays on AV sentence intelligibility scores. However, they also computed a measure of AV sentence benefit for synchronous sentences. This measure compared the actual improvement in AV sentence intelligibility over auditory-only sentence intelligibility with the maximum potential improvement over auditory-only sentence intelligibility. Grant and Seitz reported that AV sentence benefit scores for synchronous sentences were negatively correlated with sentence intelligibility scores at the maximum level of auditory-visual asynchrony tested. In other words, participants who were better at integrating auditory and visual information in the synchronous condition tended to have lower intelligibility scores for asynchronous auditory and visual information and thus showed greater sensitivity to the temporal coherence of auditory and visual signals. They also found that the integration score obtained from the asynchronous AV sentences was a good predictor of the AV sentence benefit score.

Although the Grant and Seitz (1998) study demonstrated a relationship between measures of AV integration and sentence intelligibility in asynchronous conditions, their study and the other studies reviewed above leave many questions unexplored. Grant and Seitz only used hearing-impaired individuals as participants. These individuals may have used fundamentally different AV integration strategies than normal-hearing individuals because of their hearing loss (see Bergeson & Pisoni, 2004). Grant and Seitz did not report correlations for raw auditory-only, visual-only, or AV sentence intelligibility scores or for several other measures of AV benefit. In addition, their study did not directly address the potential relationship between auditory, visual, or AV speech perception abilities and the simple detection of AV asynchrony in both speech and nonspeech signals. Previous research in our lab has found that several measures of AV asynchrony detection do not differ for speech or nonspeech signals (Conrey & Pisoni, this volume). Finally, only auditory-leading asynchronies were tested by both Grant and Seitz (1998) and Pandey et al. (1986), although the detection of auditory-leading asynchronies is known to be more accurate than the detection of visual-leading asynchronies (Conrey & Pisoni, 2003; Grant, van Wassenhove, & Poeppel, 2003; McGrath & Summerfield, 1985).

To begin to investigate some of these questions, we tested normal-hearing adults on their ability to detect AV asynchronies in both speech and nonspeech signals. We also measured the performance of these participants on a routine clinical sentence intelligibility task, the CUNY sentences, under auditory-only, visual-only, and AV presentation conditions. Based on previous research (Grant & Seitz, 1998; McGrath & Summerfield, 1985; Pandey, Kunov, & Abel, 1986), we hypothesized that participants with better lipreading skills and/or better AV benefit scores would be better (i.e., more accurate) at detecting AV asynchronies in both speech and nonspeech multimodal signals.

## Methods

### Participants

Thirty-nine Indiana University undergraduates participated in the study, which took about an hour to complete. Data from nine subjects were discarded for the following reasons: three participants did not follow directions on one or the other of the asynchrony detection task conditions (they reversed their response hand), and six participants responded "synchronous" more than 50% of the time at all asynchrony levels and their data could not be fit with the same curve-fitting procedures used for the other participants. The remaining 30 participants included 25 females and 5 males who ranged in age from 18 to 22 years (average = 19.67, *SD* = 1.03). Thirteen participants received course credit in an introductory psychology course, and 17 participants were paid $10 for their services. No significant differences between paid and unpaid participants were found on subsequent analyses, so results from the two groups were pooled in this report.

### Procedure

Each participant completed the full-face (FF) and nonspeech (NS) conditions from the asynchrony detection task described in Experiment 1 of Conrey and Pisoni (2003, this volume). The participants also completed a modified version of the City University of New York (CUNY) Sentences Test (Boothroyd, Hannin, & Hnath, 1985), which is used clinically to assess auditory-only (A-only), visual-only (V-only), and auditory-visual (AV) speech perception skills in hearing impaired populations. Participants always completed the CUNY sentences first, followed by the FF and NS asynchrony detection tasks. Because the CUNY sentences are presented in the order A-only, then V-only, then AV in the clinic, they were always presented in that order in this experiment. However, the order of the FF and NS conditions of the asynchrony detection task was counterbalanced across participants. The visual stimuli were presented on an Apple Macintosh G4 computer. Auditory stimuli were presented over Beyer Dynamic DT headphones at 70 dB SPL.

**Asynchrony detection.** Two types of stimuli were used, speech and nonspeech. For the full-face speech (FF) condition, 10 familiar English words spoken by a female speaker of American English were chosen from the Hoosier Audiovisual Multitalker Database (Lachs & Hernandez, 1998; Sheffert, Lachs, & Hernandez, 1996). This database contains digitized AV movies consisting of single talkers speaking isolated monosyllabic words. For the nonspeech (NS) condition, the stimulus was a red circle paired with a 2000-Hz tone. On each trial, the participants were presented with an AV stimulus and were asked to indicate whether it was synchronous ("in sync") or asynchronous ("not in sync"). There were 25 levels of asynchrony, ranging at every 33 ms from the auditory leading the visual signal by 300 ms (A300V ms) to the visual leading the auditory signal by 500 ms (V500A). The stimuli were blocked by condition (FF or NS). In each condition, a participant received 10 trials at each asynchrony level, for a total of 250 trials. PsyScope version 1.5.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) was used for stimulus presentation and response collection. For a more detailed description of the stimuli and the procedures used in this task, please refer to Conrey and Pisoni (this volume).

**CUNY sentences.** As administered clinically, the CUNY sentences consists of three sets of 12 meaningful English sentences, with one set each being presented in the A-only, V-only, then AV conditions. The observer watches and/or listens to a sentence and then is asked to repeat the sentence out loud. The clinician scores the test online based on words correctly perceived.

The use of the CUNY sentences in the present study differed in several ways from the clinical applications. First, we used normal-hearing young adults as participants. In order to avoid "ceiling effects" in the A-only and AV conditions, the auditory signal was transformed to make the test more difficult. Specifically, following the methods for locally time-reversed speech described by Saberi and Perrott (1999), the auditory signal was divided into 80 ms long segments. Each segment was time-reversed, and then the segments were recombined in their original order. Phenomenologically, this transformation produces a duplex percept in which clicking sounds and speech are perceived simultaneously. Eighty ms was chosen for the reversal interval because earlier pilot testing in our lab found that this level was more difficult than the 50 ms interval, at which participants performed at near-ceiling levels, and less difficult than the 100 ms interval, which was nearly impossible for some participants. The 80-ms intervals seemed likely to produce a range of variability in results without unduly frustrating participants.

Another difference in the present study was that the participants responded by typing their responses into the computer rather than reporting them orally to the experimenter. A pilot test of seven participants in our lab showed no significant differences between these two response modes within subjects ($F(1, 6) < 1$), and most pilot participants indicated on a questionnaire that they were more comfortable responding in the written format.

As in the clinical test, before beginning each condition—A-only, V-only, and AV—the participants were given three example trials in which they viewed and/or heard sentences but were not required to respond. During the test itself, each sentence was presented only once and the participant was asked to respond by typing what he or she thought the speaker said.

The sentences were all spoken by the same female talker from the Bernstein-Eberhardt database, and they were digitized by Theresa Hnath-Chisolm and her graduate students at the University of South Florida. In the present experiment, the sentences were presented using SuperCard running a program created in SuperEdit for the MacIntosh.

## Results

### Asynchrony Detection

The methods of analysis for the asynchrony detection task were as described in Conrey and Pisoni (this volume). Specifically, the proportion of times each participant responded "in sync" was calculated for every asynchrony level in the FF and NS conditions. The proportion "in sync" responses for each participant were then fitted with a Gaussian curve for the FF and NS conditions, using the program Igor Pro 4.05A Carbon (copyright 1988-2002, WaveMetrics, Inc.). The mean of the curve was taken to be the mean point of synchrony (MPS), or the point in the range of asynchrony levels that was most likely to result in a judgment of "in sync." The width of the curve, as measured by the full width at half maximum (FWHM), represented the range of asynchronies over which stimuli were judged to be synchronous more than half the time. The auditory-leading and visual-leading endpoints were the limits of the synchrony window and were calculated by subtracting (auditory-leading) or adding (visual-leading) half of the FWHM to the MPS. The averages of these measures are presented in Table 1. All statistical tests were performed on individual subject data and then averaged for descriptive purposes.

| | | Primary Measures | | Derived Measures | |
|---|---|---|---|---|---|
| | | MPS | FWHM | A-Lead | V-Lead |
| Asynchrony Conditions | FF | 47 (15) | 357 (61) | -131 (31) | 225 (36) |
| | NS | 47 (43) | 400 (66) | -153 (46) | 247 (61) |

**Table 1.** Mean (standard deviation) curve fits for the AV asynchrony detection task. All numbers are in milliseconds. FF = full-face condition; NS = nonspeech condition; MPS = mean point of synchrony; FWHM = full width at half maximum; A-Lead = auditory-leading threshold; V-Lead = visual-leading threshold. "Primary measures" came directly from the curve fitting; "derived measures" were calculated based on the primary measures.

**Mean Point of Synchrony (MPS).** The MPSs for the FF and NS conditions were not significantly different (*F*(1, 29) < 1). This result is consistent with the earlier findings reported by Conrey and Pisoni (2003, this volume).

**Full Width at Half Maximum (FWHM).** The FWHM for the NS condition was significantly larger than the FWHM for the FF condition (*F*(1, 29) = 7.690, *p* < .05). Conrey and Pisoni (this volume) found no significant differences between the FF and NS conditions in terms of FWHM. The average difference in this study was about 28 ms, which corresponds to approximately one asynchrony level.

Auditory-leading and visual-leading thresholds. The FF and NS conditions did not differ significantly in terms of the auditory- or visual-leading thresholds of the synchrony window (*F*(1, 29) = 1.547, *F*(1, 29) = 3.606; *p*'s > .05). These results indicate that the difference between the conditions in FWHM was relatively evenly divided between the auditory- and visual-leading sides of the synchrony window.

### CUNY Sentences

Participants' responses were printed out from the output file and scored by comparison with a master list of sentences. Responses were scored using a "whole-word" method, as is typically done when the CUNY sentences are used clinically. Each word was given a score of 1 or 0 points based on whether it was completely correct or had any errors, respectively. The reversal of two letters in a word, as in teh for the, was counted as correct as long as the reversal did not form a new English word. Similarly, one-letter typographical errors that did not result in the formation of a new word were counted as correct responses.

As expected, the participants were worst overall under visual-only presentation condition, with a mean score of 16 words correct (*SD* = 7) out of 102. Auditory-only scores were next, with a mean of 44 (*SD* = 15). Participants were best at the AV condition, with a mean score of 81 (*SD* = 10). The distribution of scores for these conditions is shown in Figure 1.

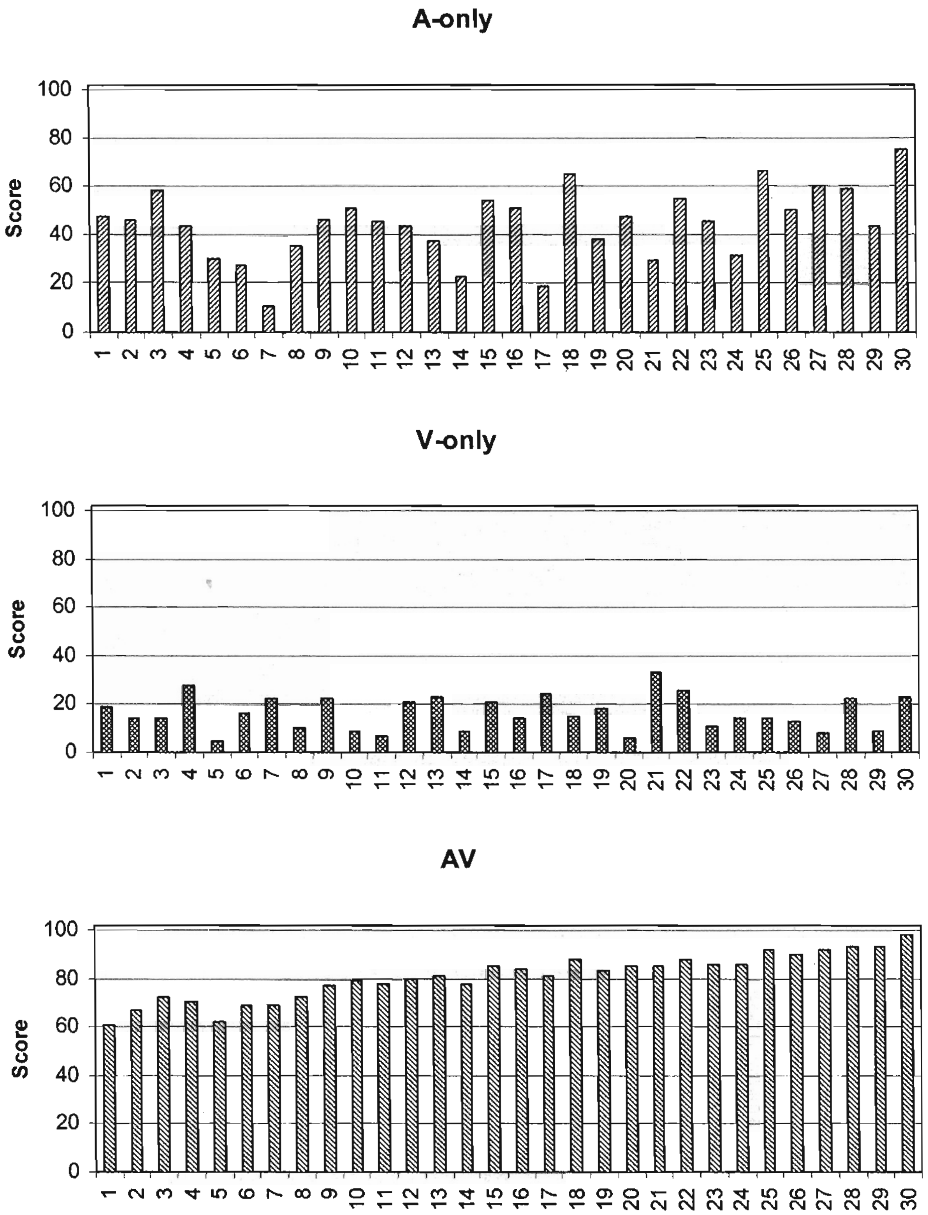


**Figure 1.** CUNY scores for A-only (top panel), V-only (middle panel), and AV (bottom panel) presentation conditions for each of the 30 participants, represented by participant number on the x-axis. Each score is the number of whole words correct out of a possible 102 total words.

Several other measures of performance were derived from these speech intelligibility scores. Auditory benefit was computed as the ratio of the difference between the AV score and the visual-only score (the amount the score improved after the auditory signal was added) compared with the total possible improvement in score with the added auditory signal. This measure was calculated as (AV - V) / (1 - V). Visual benefit was computed as the ratio of the difference between the AV and auditory-only scores compared with the total possible improvement in score with the added visual signal. This measure was calculated as (AV – A) / (1 – A). Finally, a measure of AV gain—the actual benefit resulting from the presence of both auditory and visual signals compared with the benefit expected from the simple sum of the scores for the individual modalities—was calculated as AV / (A + V). This measure and other similar ones have been employed in animal studies of multisensory integration (Stein & Meredith, 1993) and also in research on the BOLD response in fMRI studies (Calvert, Hansen, Iversen, & Brammer, 2001). If the ratio is greater or less than 1, then a supra-additive or subadditive effect has occurred; these types of effect have served as signatures of multisensory interactions in the literature.

### Correlational Analyses

Visual inspection of the individual data revealed that some of the measures used were not normally distributed, and so correlational analyses were performed using Spearman's rho ($r_s$) rather than Pearson's $r$.

**Asynchrony detection.** Among the measures obtained directly from the curve fits, the FWHMs were positively correlated for the FF and NS conditions, $r_s = +.65$, $p < .01$. Also, the MPSs were positively correlated for the FF and NS conditions, $r_s = +.46$, $p < .01$.

Among the measures derived from the MPS and FWHM, the visual-leading thresholds were positively correlated for the FF and NS conditions, $r_s = +.50$, $p < .01$. Also, the FF auditory-leading and visual-leading thresholds were negatively correlated, $r_s = -.47$, $p < .05$. Because auditory-leading thresholds were coded as negative numbers and visual-leading thresholds were coded as positive, this finding indicates that larger ("lower") auditory-leading thresholds were related to larger ("higher") visual-leading thresholds. The relationship between auditory- and visual-leading thresholds for the NS condition was not significant, however ($r_s = +.17$, $p > .05$). Several of the correlations between "primary" and "derived" asynchrony detection performance measures were significant, although this finding may not have much practical significance because the derived measures were calculated using the primary measures. A summary of the intercorrelations is given in Table 2.

| | | | Primary | | | | Derived | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | FF | | NS | | FF | | NS | |
| | | | MPS | FWHM | MPS | FWHM | A-Lead | V-Lead | A-Lead | V-Lead |
| Primary | FF | MPS | — | | | | | | | |
| | | FWHM | .18 | — | | | | | | |
| | NS | MPS | .46** | .30 | — | | | | | |
| | | FWHM | -.04 | .65*** | .24 | — | | | | |
| Derived | FF | A-Lead | .28 | -.85*** | -.01 | -.68*** | — | | | |
| | | V-Lead | .65*** | .82*** | .47* | .41* | -.47** | — | | |
| | NS | A-Lead | 0.37* | -.12 | .67*** | -.47* | .37* | .15 | — | |
| | | V-Lead | .35 | .47** | .81*** | .68*** | -.31 | .51** | .17 | — |

**Table 2.** Correlations among measures of AV asynchrony detection. Abbreviations as in Table 1.
* = $p < .05$; ** = $p < .01$; *** = $p < .001$.

**CUNY sentences.** A-only and AV scores were positively correlated, $r_s = +.53$, $p < .01$. The correlations between V-only and A-only or AV scores were small and not significant. Auditory benefit and visual benefit were positively correlated, however ($r_s = +.85$, $p < .001$). In addition, A-only scores and V-only scores were negatively correlated with AV gain ($r_s = -.71$, $r_s = -.42$; $p < .001$, $p < .05$, respectively). These correlations indicate that larger AV gains were associated with lower unimodal scores. Both A-only and AV scores were positively correlated with auditory benefit ($r_s = +.56$, $r_s = +.97$, respectively; $p$'s < .001), and AV score was also positively correlated with visual benefit, $r_s = +.89$, $p < .001$. Table 3 shows the bivariate correlations among the primary and derived measures from the CUNY sentences task.

| | | Primary | | | Derived | | |
|---|---|---|---|---|---|---|---|
| | | A-Only | V-Only | AV | A-Ben | V-Ben | AV Gain |
| Primary | A-Only | --- | | | | | |
| | V-Only | -.06 | --- | | | | |
| | AV | .53** | .03 | --- | | | |
| Derived | A-Ben | .56*** | .10 | .97*** | --- | | |
| | V-Ben | .16 | -.15 | .89*** | .85*** | --- | |
| | AV Gain | -.71*** | -.42* | -.04 | .01 | .27 | --- |

**Table 3.** Correlations among measures of AV sentence intelligibility (CUNY sentences). A-Ben = auditory benefit, V-Ben = visual benefit. "Derived" means derived from primary scores. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

**Asynchrony detection and CUNY sentences.** Several of the intercorrelations among the asynchrony detection and CUNY sentences measures were also significant. AV scores were negatively correlated with the FWHM for both the FF and NS conditions ($r_s = -.47$, $r_s = -.45$; $p < .01$, $p < .05$, respectively). A-only scores were also negatively correlated with the FWHM for the NS condition, $r_s$ = -.41, $p < .05$. These results indicate that higher AV scores for the FF and NS conditions and higher A-only scores for the NS condition were associated with smaller synchrony windows, or less tolerance for asynchrony. In addition, auditory benefit was correlated with the FWHM for both the FF and NS conditions ($r_s = -.48$, $r_s = -.52$, respectively; $p$'s < .01), and visual benefit was negatively correlated with the FWHM for the FF condition ($r_s = -.39$, $p < .05$). Finally, AV, A-only, auditory benefit, and visual benefit scores were all positively correlated with the auditory-leading threshold for the FF condition ($r_s$ = +.49, $r_s$ = +.39, $r_s$ = +.55, $r_s$ = +.43, respectively; $p$'s < .05). This pattern indicates that higher AV and A-only scores and higher multimodal benefit scores for sentence intelligibility were associated with lower auditory-leading thresholds that were closer to physical synchrony. Table 4 summarizes these results.

| | | | | CUNY Sentences | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Primary | | | Derived | | |
| | | | | A-Only | V-Only | AV | A-Ben | V-Ben | AV Gain |
| Asynchrony Detection | Primary | FF | MPS | .10 | -.29 | .08 | .15 | .08 | .19 |
| | | | FWHM | -.31 | -.04 | -.47** | -.48** | -.39* | .07 |
| | | NS | MPS | -.12 | -.17 | -.09 | -.02 | -.02 | .23 |
| | | | FWHM | -.41* | .20 | -.45** | -.52** | -.35 | .06 |
| | Derived | FF | A-Lead | .39* | -.14 | .49** | .55** | 0.43* | .00 |
| | | | V-Lead | -.10 | -.15 | -.31 | -.27 | -.28 | .05 |
| | | NS | A-Lead | .14 | -.17 | .21 | .30 | .21 | .08 |
| | | | V-lead | -.31 | .14 | -.14 | -.18 | -.12 | .21 |

**Table 4.** Correlations among measures of AV asynchrony detection and AV sentence intelligibility. Abbreviations as in previous tables. * = p < .05; ** = p < .01; *** = p < .001.

## Discussion

The results of this study demonstrate several relations between measures of AV sentence intelligibility and the ability to detect AV asynchrony in both speech and nonspeech signals. Specifically, the participants who obtained higher AV sentence intelligibility scores tended to have smaller windows over which they identified AV signals as synchronous and thus were more accurate at detecting small differences in AV asynchrony in speech and nonspeech signals. In addition, higher auditory and visual sentence benefit scores were also correlated with smaller AV synchrony windows for both speech and nonspeech signals. Finally, higher auditory-only, AV, and auditory and visual benefit scores were correlated with auditory-leading thresholds in the FF condition that were closer to physical synchrony. This pattern was not observed in the NS condition. This result implies that subjects who performed better on AV sentence intelligibility measures were more accurate at identifying the asynchrony in auditory-leading speech, but not in auditory-leading nonspeech signals. This finding is interesting given that the auditory signal typically lags behind the visual signal in natural speech, because the articulators must be positioned before the sound can be made. Perhaps the better AV integrators were able to take advantage of this natural relationship in speech, whereas the lack of expectation about which signal should lead in the artificial nonspeech signal condition prevented the use of natural cues in that condition.

Several researchers have suggested that AV asynchrony detection may have its neural basis in multisensory processing times (Conrey & Pisoni, this volume; Lewald, Ehrenstein, & Guski, 2001). However, in earlier research it was unclear whether AV asynchrony detection was meaningfully related to AV integration measures relevant for speech perception. The results of the present study have revealed strong relations among measures of AV integration and the detection of AV asynchrony in speech and nonspeech signals that provide several new insights into the common cognitive and neural mechanisms relevant for both multisensory temporal sensitivity and speech perception, especially perception of multimodal speech signals. Fundamental differences in neural timing of multimodal sensory events may be linked to the variation observed in lipreading performance and multimodal benefit among normal-hearing individuals.

## References

Bergeson, T. R., & Pisoni, D. B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 749-772). Cambridge, MA: MIT Press.

Boothroyd, A., Hannin, L., & Hnath, T. (1985). A sentence test of speech perception: Reliability set equivalence and short-term learning (internal report RCI 10). New York: City University of New York.

Calvert, G., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14,* 427-438.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Conrey, B. L., & Pisoni, D. B. (2003). Audiovisual asynchrony detection for speech and nonspeech signals. Proceedings of the Audio Visual Speech Processing (AVSP) Workshop, 25-30.

Conrey, B. L., & Pisoni, D. B. (this volume). Detection of auditory-visual asynchrony in speech and nonspeech signals. In *Research on Spoken Language Processing Progress Report No. 26* (pp. 71-94). Bloomington, IN: Speech Research Laboratory, Indiana University.

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., van Wassenhove, V. & Poeppel, D. (2003). Discrimination of auditory-visual synchrony. Proceedings of the Audio Visual Speech Processing (AVSP) Workshop, 31-35.

Lachs, L., & Hernandez, L. R. (1998). Update: The Hoosier audiovisual multitalker database. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 377-388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lewald, J., Ehrenstein, W.H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioral Brain Research, 121*, 69-79.

McGrath, M. & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults, *Journal of the Acoustical Society of America*, *77*, 678-684.

Pandey, C.P., Kunov, H. & Abel, M.S. (1986). Disruptive effects of auditory signal delay on speech perception with lip-reading, *The Journal of Auditory Research*, *26*, 27-41.

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature, 398,* 760.

Sheffert, S. M., Lachs, L., & Hernandez, L. R. (1996). The Hoosier audiovisual multitalker database. In *Research on Spoken Language Processing No. 21* (pp. 578-583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Stein, B. & Meredith, M.A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

# III. Publications

## ARTICLES PUBLISHED:

Bergeson, T.R., Pisoni, D.B., & Davis, R.A.O. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. Volta Review, 103, 347-370.

Burkholder, R.A. & Pisoni, D.B. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85*, 63-88.

Carter, A., Dillon, C., & Pisoni, D. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: Suprasegmental analyses. *Clinical Linguistics and Phonetics, 16,* 619-638.

Cleary, M., Pisoni, D.B. & Kirk, K.I. (2002). Working memory spans as predictors of word recognition and receptive vocabulary in children with cochlear implants. *Volta Review, 102,* 259-280.

Clopper, C.G. & Pisoni, D.B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32,* 111-140.

Clopper, C.G & Pisoni, D.B. (2004). Homebodies and Army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change, 16,* 31-48.

Dillon, C.M., Cleary, M., Pisoni, D.B. & Carter, A.K. (2004). Imitation of nonwords by hearing-impaired children with cochlear implants: Segmental analyses. *Clinical Linguistics and Phonetics, 18*, 39-55.

Dillon, C.M., Pisoni, D.B., Cleary, M., and Carter, A.K. (2004). Nonword imitation by children with cochlear implants: Consonant analyses. *Archives of Otolaryngology -Head & Neck Surgery, 130*, 587-591.

Goh, W.D. & Pisoni, D.B. (2003). Effects of lexical neighborhoods on immediate memory span for spoken words. *Quarterly Journal of Experimental Psychology, 56A*, 929-954.

Herman, R. & Pisoni, D.B. (2002). Perception of "elliptical speech" following cochlear implantation: Use of broad phonetic categories in speech perception. *Volta Review, 102,* 321-347.

Houston, Pisoni, D.B., Kirk, K.I., Ying, E.A. & Miyamoto, R.T. (2003). Speech perception skills of deaf children following cochlear implantation: A first report. *International Journal of Pediatric Otorhinolaryngology, 67*, 479-495.

Houston, D.M., Ying, E.A., Pisoni, D.B. & Kirk, K.I. (2003). Development of pre word-learning skills in infants with cochlear implants. *Volta Review, 103*, 303-326.

Kaiser, A.R., Kirk, K.I., Lachs, L. & Pisoni, D.B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech-Language-Hearing Research, 46,* 390-404.

Lachs, L. & Pisoni, D.B. (2004). Crossmodal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 378-396.

Lachs, L. & Pisoni, D.B. (2004). Specification of crossmodal source information in isolated kinematic displays of speech. *Journal of the Acoustical Society of America, 116,* 507-518.

Lachs, L., Weiss, J.W. & Pisoni, D.B. (2002). Use of partial information by cochlear implant patients and normal-hearing listeners in identifying spoken words: Some preliminary analyses. *Volta Review, 102,* 303-320.

Meyer, T.A., Frisch, S.A., Pisoni, D.B., Miyamoto, R.T., & Svirsky, M.A. (2003). Modeling open-set spoken word recognition in postlingually deafened adults after cochlear implantation: Some preliminary results with the Neighborhood Activation Model. *Otology & Neurotology, 24*, 612-620.

Pisoni, D.B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children following cochlear implantation. *Ear & Hearing, 24,* 106S-120S.

Roediger, H.L., III, McDermott, K.B., Pisoni, D.B. & Gallo, D.A. (2004). Illusory recollection of voices. *Memory, 12,* 586-602.

Sheffert, S.M., Pisoni, D.B., Fellowes, J.M. & Remez, R.E. (2002). Learning to recognize talkers from natural, sinewave and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1447-1469.

Sheffert, S.M. & Shiffrin, R.M. (2003). Auditory "registration without learning." *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 10-21.

Vitevitch, M.S. (2003). Change Deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance 29,* 333-342.

Vitevitch, M.S. (2002). The influences of phonological similarity neighborhoods on speech perception. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 735-747.

Vitevitch, M.S., Pisoni, D.B, Kirk, K.I., Hay-McCutcheon, M. & Yount, S.L. (2002) Effects of phonotactic probabilities on the processing of spoken words and nonwords by postlingually deafened adults with cochlear implants. *Volta Review, 102,* 283-302.

Vitevitch, M.S. & Sommers, M. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31,* 491-504.

Wong, D., Pisoni, D.B., Learn, J., Gandour, J.T., Miyamoto, R.T. & Hutchins, G.D. (2002). Differential cortical activation to monaural speech and nonspeech stimuli: A PET imaging study. *Hearing Research, 166,* 9-23.

## BOOK CHAPTERS PUBLISHED:

Bergeson, T.R. & Pisoni, D.B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence & B.E. Stein (Eds.), *Handbook of Multisensory Processes*. MIT Press. Pp. 749-772.

Cleary, M. & Pisoni, D.B. (2003). Speech Perception. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science Vol. 4*. London, UK: Macmillan. Pp. 163-169.

Houston, D.M. & Pisoni, D.B. (2002). Early speech perception and language development in normal-hearing and deaf infants following cochlear implantation. In K.T. Houston (Ed.), *The Role of Audition in Spoken Language*. Washington, DC: Alexander Graham Bell. Pp. 5-8.

Lachs, L., McMichael, K. & Pisoni, D.B. (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory.* Oxford: Oxford University Press. Pp. 215-235.

Pisoni, D.B. & Cleary, M. (2004). Learning, memory and cognitive processes in deaf children following cochlear implantation. In F.G. Zeng, A.N. Popper & R.R. Fay (Eds*.), Springer Handbook of Auditory Research: Auditory Prosthesis, SHAR Volume X.* Pp. 377-426.

## PROCEEDINGS PUBLISHED:

Bergeson, T.R., Pisoni, D.B., & Davis, R.A.O. (2003). A longitudinal study of audiovisual speech perception by prelingually deaf children with cochlear implants. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 139-142). Adelaide, Australia: Causal Productions.

Bergeson, T.R., Pisoni, D.B., Lachs, L., & Reese, L. (2003). Audiovisual integration of point light displays of speech by deaf adults following cochlear implantation. In M.J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1469-1472). Adelaide, Australia: Causal Productions.

Bergeson, T.R., Pisoni, D.B., & Reynolds, J.T. (2003). Perception of point light displays of speech by normal-hearing adults and deaf adults with cochlear implants. In J.L. Schwartz, F. Berthommier, M.A. Cathiard, & Sodoyer (Eds), *Proceedings of Audio Visual Speech Processing 2003* (pp. 55-60). Grenoble, France: Institut de la Communication Parlée, CNRS UMR 5009, INP Grenoble, Université Stendhal.

Conrey, B.L. & Pisoni, D.B. (2003). Audiovisual asynchrony detection for speech and nonspeech signals. In J.L. Schwartz, F. Berthommier, M.A. Cathiard, & Sodoyer (Eds), *Proceedings of Audio Visual Speech Processing 2003* (pp. 25-30). Grenoble, France: Institut de la Communication Parlée, CNRS UMR 5009, INP Grenoble, Université Stendhal.

Kirk, K.I., Pisoni, D.B. & Lachs, L. (2002). Audiovisual integration of speech by children and adults with cochlear implants. *Proceedings of the 7th International Conference on Spoken Language Processing,* 1689-1692.

## MANUSCRIPTS ACCEPTED FOR PUBLICATION (IN PRESS):

Burkholder, R.A., & Pisoni, D.B. (In press). Working memory capacity, verbal rehearsal speed, and scanning in deaf children with cochlear implants. In P.E. Spencer & M. Marschark (Eds.), *Advances in the Development of Spoken Language by Deaf Children*. Oxford University Press.

Cleary, M., Pisoni, D.B. & Kirk, K.I. (In press). Influence of voice similarity on talker discrimination in normal-hearing children and hearing-impaired children with cochlear implants. *Journal of Speech, Language, and Hearing Research.*

Clopper, C.G., Conrey, B.L. & Pisoni, D.B. (In press). Effects of talker gender on dialect categorization. *Journal of Language and Social Psychology*.

Clopper, C.G. & Pisoni, D.B. (In press). Speech perception, hearing impairment, and linguistic variation. In M.J. Ball (Ed.), *Clinical Sociolinguistics*. Blackwell Publishers.

Clopper, C.G. & Pisoni, D.B. (In press). Some new experiments on perceptual categorization of dialect variation in American English: Acoustic analysis and linguistic experience. In N. Neidzielski (Ed.), *Speech perception in context: Beyond acoustic pattern matching*. Erlbaum.

Clopper, C.G. & Pisoni, D.B. (In press). Perception of dialect variation: Some implications for current research and theory in speech perception. In D.B. Pisoni & R.E. Remez (Eds.), *Handbook of Speech Perception*. Blackwell Publishers.

Clopper, C.G. & Pisoni, D.B. (In press). Perceptual learning of dialects. *Language and Speech*.

Dillon, C.M., Burkholder, R.A., Cleary, M. & Pisoni, D.B. (In press). Nonword repetition by children with cochlear implants: Accuracy ratings from normal-hearing listeners. *Journal of Speech, Language and Hearing Research*.

Karpicke, J & Pisoni, D.B. (In press). Using immediate memory span to measure implicit learning. *Memory & Cognition*.

Lachs, L. & Pisoni, D.B. (In press). Crossmodal source identification in speech perception. *Ecological Psychology*.

Meyer, T.A., Pisoni, D.B., Luce, P.A. & Bilger, R.C. (In press). An analysis of the psychometric and lexical neighborhood properties of the spondaic words. *Journal of the American Academy of Audiology.*

Pisoni, D.B. (In press). Speech perception in deaf children with cochlear implants. In D.B. Pisoni & R.E. Remez (Eds.), *Handbook of Speech Perception*. Blackwell Publishers.

Teoh, S.W., Pisoni, D.B. & Miyamoto, R.T. (In press). Cochlear implantation in adults with prelingual deafness: I. Clinical results. *Laryngoscope*.

Teoh, S.W., Pisoni, D.B. & Miyamoto, R.T. (In press). Cochlear implantation in adults with prelingual deafness: II. Underlying constraints that affect audiological outcomes. *Laryngoscope.*

## Manuscripts Submitted:

Bergeson, T., Pisoni, D.B. & Davis, R.B.O. (Submitted). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear & Hearing.*

Conrey, B.L. & Pisoni, D.B. (Submitted). Detection of auditory-visual asynchrony in speech and nonspeech signals. *Cognitive Brain Research*.

Cummings, K.E., Chin, S.B. & Pisoni, D.B. (Submitted). Peturbation of speech after consumption of alcohol. *Journal of Phonetics*.

Harnsberger, J.D., Wright, R. & Pisoni, D.B. (Submitted). Effects of speaking style on the perceptual learning of novel voices: A first report. *Phonetica*.

Horn, D.L., Davis, R.A.O., Pisoni, D.B. & Miyamoto, R.T. (Submitted). Development of visual attention skills in prelingually deaf children who use cochlear implants. *Ear & Hearing.*

Horn, D.L., Davis, R.A.O., Pisoni, D.B. & Miyamoto, R.T. (Submitted). Behavioral inhibition and clinical outcomes in children with cochlear implants. *Laryngoscope.*

Houston, D.M., Carter, A.K., Pisoni, D.B., Kirk, K.I., Ying, E.A. (Submitted). Name-learning skills of deaf children following cochlear implantation: A first report. *Volta Review.*

McMichael, K.H. & Pisoni, D.B. (Under revision). Effects of talker-specific encoding on recognition memory for spoken sentences. *Memory & Cognition.*

Tierney, A.T. & Pisoni, D.B. (Submitted). Some effects of early musical experience on sequence memory spans. *Neuropsychologia.*