RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Executive Function, Cognitive Control and Sequence Learning in Deaf Children with Cochlear Implants[1]

David B. Pisoni, Christopher M. Conway,[2] William Kronenberger,[3] Shirley C. Henning[3] and Esperanza M. Anaya

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] The research described in this chapter was supported by NIH-NIDCD Research Grant DC00111 and NIH-NIDCD Training Grant DC00012.

[2] Department of Psychology, Saint Louis University, St. Louis, MO 63103.

[3] Indiana University School of Medicine, Indianapolis, IN, 46202

Executive Function, Cognitive Control and Sequence Learning in Deaf Children with Cochlear Implants

Abstract. The bulk of clinical research on cochlear implants (CIs) has been intellectually isolated from the mainstream of current research and theory in neuroscience, cognitive psychology and developmental neuropsychology. As a consequence, the major clinical research issues have been narrowly focused on speech and language outcomes and efficacy of cochlear implantation as a medical treatment for profound hearing loss. Little basic or clinical research in the past has investigated the underlying neurobiological and neurocognitive bases of the individual differences and variability in the effectiveness of CIs. Many of the deaf children with CIs may have comorbid disturbances and/or delays in several basic neurocognitive processes that subserve information processing systems used in spoken language processing. In this chapter, we report new findings on executive function and implicit visual sequence learning in deaf children with CIs. The results of these experiments suggest that differences in neural reorganization of multiple brain systems resulting from a period of profound deafness and language delay may be responsible for the enormous variability observed in speech and language outcome measures following implantation.

Introduction

Our long-term goal is to understand and predict the enormous variability in speech and language outcomes in deaf children who have received CIs as a treatment for profound deafness. As noted in both of the NIDCD consensus statements on CIs in 1988 and 1995, individual differences and variability in speech and language outcomes are significant clinical problems that have not been addressed adequately in the past. Little, if any, progress has been made in understanding the neurobiological mechanisms and neurocognitive processes that are responsible for the variability observed in speech and language outcomes following cochlear implantation.

Most of the past work on CIs has been concerned primarily with documenting the “efficacy” of cochlear implantation as a medical treatment for profound deafness, focusing research efforts on demographic, medical and educational variables as predictors of outcome and benefit. With the average age of implantation steadily decreasing because of the widespread use of universal newborn hearing screening, the ability to reliably predict the “effectiveness” of CIs from behavioral measures obtained from infants and young children prior to implantation becomes critical to providing appropriate habilitation following implantation. Understanding sources of variability in speech and language outcomes after cochlear implantation is a complex and challenging problem requiring multidisciplinary research efforts from scientists with backgrounds in neuroscience, cognitive psychology, and developmental neuropsychology.

A Neurocognitive Approach to Individual Differences in Outcomes

To understand, explain and predict variability in outcome and benefit, it is necessary to situate the problem of individual differences in a much broader theoretical framework that extends well beyond the narrow clinical fields of audiology and speech pathology and fully acknowledges variability in brain-behavior relations as a natural consequence of biological development of all living systems (Sporns, 1998). Development involves interactions over time between the biological state of the individual (genetics, biological structures and characteristics) and specific environmental experiences (sensory input, external events that influence development such as toxic exposures). There is an increasing recognition that outcomes are not exclusively genetically or environmentally predetermined. Environmental experience allows complex biological systems to self-organize during the process of development (Thelen & Smith, 1994). Alteration in early auditory experience by electrical stimulation through a CI supports a process of neurobiological reorganization that draws on and influences multiple interacting neurocognitive processes and domains and is not isolated to only hearing and speech-language outcomes.

The enormous variability observed in a wide range of speech and language outcome measures following cochlear implantation may not be unique to this particular clinical population at all but may reflect instead more general underlying sources of variability observed in speech and language processing in healthy typically-developing normal-hearing (NH) children as well as adults (Cicchetti & Curtis, 2006). Moreover, because of the important contributions of learning and memory to the development of spoken language processing, it is very likely that the sources of the individual differences observed in speech and language outcomes in deaf children with CIs also reflect variation in the development of domain-general neurocognitive processes, processes that are involved in linking and coordinating multiple brain systems together to form a functionally integrated information processing system (Ullman & Pierpont, 2005).

In order to investigate the sources of variability in performance and understand the neural and cognitive processes that underlie variation in outcome and benefits following implantation, it is necessary to substantially broaden the battery of outcome measures to assess a wider range of behaviors and information processing skills beyond just the traditional clinical audiological speech and language assessment measures that have been routinely used in the past by researchers working in CIs. Furthermore, it is also important to recognize that a child’s failure to obtain optimal benefits and achieve age-appropriate speech and language milestones from his/her CI may not be due directly to the functioning of the cochlear implant itself but may reflect complex interactions among a number of contributing factors (Geers, Brenner & Davidson, 2003).

In our research program, we adopt the general working assumption that many profoundly deaf children who receive CIs, especially children who are performing poorly, may have other contributing neural, cognitive and affective sequelae resulting from a period of deafness and auditory deprivation combined with a language delay before implantation. The enormous variability observed in speech and language outcomes may not be due to hearing per se or to processes involved in the early sensory encoding of speech at the auditory periphery (Hawker, Ramirez-Inscoe, Bishop, Twomey, O’Donoghue & Moore, 2008). Evidence is now rapidly accumulating to suggest that other central cortical and subcortical neurobiological and neurocognitive processes contribute additional unique sources of variability to outcome and benefit that are not assessed by the traditional battery of speech and language measures.

Brain-Behavior Relations. Our approach to the problems of variability in outcome and benefit following cochlear implantation is motivated by several recent findings and new theoretical developments that suggest that deafness and hearing impairment in children cannot be viewed in isolation as a simple sensory impairment (see also Conrad, 1979; Luria, 1973; Myklebust, 1954, 1964; Myklebust & Brutten, 1953). The enormous variability in outcome and benefit reflects numerous complex neural and cognitive processes that depend heavily on functional connectivity of multiple brain areas working together as a complex integrated system (Luria, 1973). As W. Nauta (1964) pointed out more than forty years ago, “no part of the brain functions on its own, but only through the other parts of the brain with which it is connected" (p. 125). As described in the sections below, we believe this is a promising new direction to pursue in clinical research on individual differences in profoundly deaf children who use CIs.

Domain-General Cognitive Factors. Our recent work with preimplant visual-motor integration (VMI) tests that use only visual patterns and require reproduction and construction processes has found significant correlations with a range of conventional clinical speech and language outcome measures obtained from deaf children following implantation. Similarly, our recent findings on nonword repetition, talker discrimination and implicit learning of probabilistic sequential patterns presented in the sections below suggest that an important additional source of variance in speech and language outcomes in deaf children with CIs is associated with domain-general non-linguistic executive-organizational-integrative (EOI) processes that involve executive function, cognitive control and self-regulation (Blair & Razza-Peters, 2007; Figueras, Edwards & Langdon, 2008; Hauser, Lukomski & Hillman, 2008).

There is now good agreement among cognitive scientists that these so-called "control processes" rely critically on global system-wide executive attention processes that reflect organization-integration, coordination, functional connectivity and close interactions of multiple neural circuits and subsystems that are widely distributed across many areas of the brain (Sporns, 2003). Although these EOI processes overlap partially with elements of the traditional construct of global intelligence, these two broad domains of cognitive functioning can be distinguished in several ways.

First, global intelligence includes functions and abilities such as crystallized knowledge, reasoning, long-term memory, and concept formation, which are not generally thought to be a part of executive-organizational-integrative processes (Kaufman & Lichtenberger, 2006; Lezak, Howieson, Loring & Hannay, 2004). Second, executive-organizational-integrative processes are minimally dependent on the specific content of the information being processed; that is, they can be applied to almost any kind of neural or cognitive representation such as verbal, nonverbal, visual-spatial, sensory-motor (Hughes & Graham, 2002; Van der Sluis, deJong & van derLeij, 2007). Global intelligence includes a component of content in the form of explicit declarative knowledge, accumulated experience and acquired algorithms for problem solving. Third, recent neuroimaging studies have found differences in executive-organizational-integrative processing ability and its relationship to brain function, even in groups of subjects that are matched on measures of global intellectual ability (e.g., Mathews, Kronenberger, Wang, Lurito, Lowe & Dunn, 2005).

Executive-Organizational-Integrative Abilities in CI Outcomes. We hypothesize that EOI abilities are particularly important for speech-language development following CI because of strong reciprocal relations between the development of spoken language processing skills and the development of EOI abilities (Deary, Strand, Smith & Fernandes, 2007; Hohm, Jennen-Steinmetz, Schmidt & Laucht, 2007). Spoken language and verbal mediation processes provide the schemas and knowledge structures for symbolic representations that can be used for comprehension-integration (e.g., mental representation using language) and cognitive control, both of which are important EOI abilities (Bodrova & Leong, 2007; Diamond, Barnett, Thomas & Munro, 2007; Lamm, Zelazo & Lewis, 2006).

Additionally, early auditory experience promotes the ability to integrate temporal sequences into wholes (e.g., chunking auditory patterns into meaningful sounds and linguistic units) and to engage in fluent processing of temporal patterns. EOI processing also allows for the active control of selective attention, use of working memory, fluent speeded processing, and integration of multiple sources of information during spoken language processing. More efficient EOI processing therefore promotes better spoken language skills while better language provides the key building blocks for the development of EOI abilities through verbal mediation and feedback processes (Bodrova & Leong, 2007). Because hearing loss, even mild hearing loss, interferes with critical early spoken-language experiences, we suggest that development of key EOI skills may be at risk in deaf children. A CI restores some of the components of auditory experience to a "fragile" EOI system, which in turn, becomes a fundamental influence on the ability to use spoken language to build speech and language processing skills that are key outcomes following CI.

Preliminary research from our research center, taken together with the theoretical approach articulated above, suggests that four key EOI areas may be involved in speech-language outcome following CI: working memory, fluency-efficiency-speed, concentration-vigilance-inhibition, and organization-integration. These abilities allow spoken language to be processed rapidly (fluency-efficiency-speed) into meaningful symbolic units (organization-integration), stored (working memory), and actively assigned meaning (organization-integration) while the individual maintains a focus on the relevant stimulus information (concentration-vigilance) and resists distracting impulses (inhibition). The need to process enormous amounts of novel auditory sensory input in the development of speech-language skills following CI therefore draws heavily on these domain-general EOI areas. A child's ability to effectively integrate, coordinate and utilize these EOI abilities will impact on speech-language outcomes.

The hypothesis motivating our research program at Indiana is that many deaf children who use CIs may display delays or dysfunctions in several neurocognitive information processing domains in addition to their primary hearing loss and language delay. Some deaf children with CIs may not show "age-appropriate" scores on a variety of conventional neuropsychological tests that on the surface appear to have little, if anything, directly to do with domain-specific sensory aspects of hearing or speech perception and spoken language processing, but reflect instead domain-general processes. Variability in these basic elementary information processing skills may ultimately be responsible for some of the individual differences observed in audiological, speech and language outcome measures.

Nonword Repetition and Phonological Decomposition

To obtain additional knowledge about the underlying cognitive and linguistic processes that are responsible for the variation in speech and language outcomes following implantation and to broaden the information processing domains used in assessment, we carried out an unconventional nonword repetition study with a large group of deaf children to examine how they use sublexical phonological knowledge (Cleary, Dillon & Pisoni, 2002; Dillon, Burkholder, Cleary & Pisoni, 2004). When we first proposed using this novel experimental procedure, the CI clinicians in our center argued that deaf children would not be able to do a task like this and because they did not know any of the non-words and they could only do immediate repetition and reproduction tasks with words that they were familiar with and had in their mental lexicons. We explained that non-word repetition has been shown in numerous studies to be a valuable experimental methodology and research tool that could provide new fundamentally different information that was not available from any of the other standard clinical assessment instruments currently in use. Moreover, several studies have reported that non-word repetition scores were strongly correlated with vocabulary development and other language learning milestones in normal hearing children and other clinical populations (Gathercole & Baddeley, 1990; Gathercole, Hitch, Service & Martin, 1997).

In our first nonword repetition study, 88 pediatric CI users were asked to listen to recorded nonsense words that conformed to English phonology and phonotactics (e.g., "altupatory") and immediately repeat back what they heard over a loudspeaker to the examiner. Several measures of their performance on this task were obtained and then correlated with open-set word recognition scores from the Lexical

Neighborhood Test (LNT), Forward Digit Span, Speech Intelligibility, Speaking Rate, Word Attack, and Rhyme Errors. The Word Attack and Rhyme Errors were obtained from an isolated single word reading task that was collected as part of a larger research project carried out by Ann Geers and her colleagues (see Geers & Brenner, 2003).

As shown in Table I, the transcription scores for both consonants and vowels as well as the perceptual ratings of the nonwords obtained from a group of NH adults were all strongly correlated with the traditional clinical outcome measures we examined, suggesting that a common set of phonological representations and processing skills are used across a wide range of different language processing tasks, even single word reading tasks.

Table I

	Consonants (N=76)	**Vowels (N=76)**	**Accuracy Ratings (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 (Reading)	+.75***	+.72***	+.78***
Rhyme Errors (Reading)	-.63**	-.68**	-.54*

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) based on Cleary, Dillon and Pisoni (2002), Carter, Dillon and Pisoni (2002), Dillon, Cleary, Pisoni and Carter (2004) and Dillon, Burkholder, Cleary and Pisoni (2004).

Although nonword repetition appears at first glance to be a simple information processing task, it is actually a complex psycholinguistic process that requires the child to perform well on each of the individual component processes involving speech perception, phonological encoding and decomposition, verbal rehearsal and maintenance in working memory, retrieval and phonological reassembly, phonetic implementation and speech production. Moreover, the nonword repetition task like other imitation or reproduction tests requires additional organizational-integrative processes that link these individual component processes together to produce a unitary coordinated verbal response as output.

What unique and special properties does the nonword repetition task have that other conventional clinical speech and language tests lack? First, the stimuli used in nonword repetition tasks are novel sound patterns that children have not heard before. Thus, children must make use of robust adaptive behaviors and language processing strategies that draw on past linguistic experience in novel ways. Second, the nonword repetition task requires the child to consciously control and focus his/her attentional resources exclusively on the phonological sound properties of the stimulus patterns rather than the symbolic/linguistic attributes of the meanings of the words because there is no lexical entry in the mental lexicon for these particular stimulus patterns. Finally, the nonword repetition task, like other open-set spoken word recognition tests and reproduction tasks requires the subject to rapidly carry out phonological decomposition, reassembly of the structural description of the sound pattern and verbal rehearsal of a novel and unfamiliar phonological representation in working memory as well as implementation and reconstruction of a vocal articulatory-motor response linking perception and action.

Given these specific processing activities and the heavy demands on cognitive control and executive attention, it is not at all surprising that nonword repetition has proven to be very good at diagnosing a wide range of language disorders and delays which involve disturbances in rapid phonological processing of spoken language (Gathercole & Baddeley, 1990; Gathercole, Hitch, Service & Martin, 1997).

Executive Function and Organizational-Integration Processes

Inhibition Processes in Speech Perception. Our interest in executive function and cognitive control processes began almost by accident with a small-scale pilot study carried out by Miranda Cleary that was originally designed to assess the talker recognition skills of deaf children with CIs (Cleary & Pisoni, 2002). Using a same-different discrimination task, children heard two short meaningful English sentences in a row one after the other and were asked to determine if the sentences were produced by the same talker or different talkers. Half of the sentences in each set were produced by the same talker and half were produced by different talkers. Within each set, half of the sentences were linguistically identical and half were different. The results of this study are shown in Figure 1 for two groups of children, 8 and 9 year-old deaf children with cochlear implants, and a younger group of 3-5 year-old NH typically-developing children who served as a comparison group.

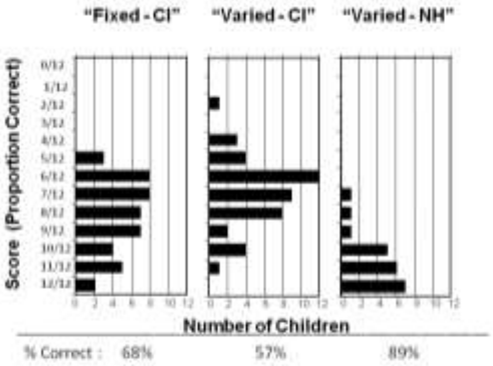

Figure 1. Distribution of "same-different" sentence discrimination scores for CI users and NH controls in fixed- and varied-sentence conditions based on data reported by Cleary & Pisoni (2002).

Although the normal hearing 3-5 year olds had little difficulty identifying whether two sentences were produced by different talkers in the varied-sentence condition, the deaf children with CI had considerably more difficulty in carrying out this task. If the linguistic content of the two utterances was the same, children with CIs performed the talker-discrimination task better than chance (67% correct). However, when the linguistic content of the two sentences differed in the varied sentence condition, the performance of deaf children with CIs did not differ from chance (58% correct) and was significantly worse than the group of younger normal hearing children (89% correct).

The talker discrimination findings obtained by Cleary and Pisoni (2002) are theoretically significant because they suggest that pediatric CI users have considerable difficulty inhibiting irrelevant linguistic information in this sentence processing task. When both sentences were linguistically the same in the fixed-sentence conditions, the child can simply judge whether the voice is the same in both conditions because there is no competing semantic information to affect their discrimination response. In the varied-sentence condition, however, the child must be able to consciously control his/her attention and actively ignore and inhibit the differences in sentence meaning, the more dominant response mode, in order to selectively focus attention on the sound structure to make a decision about the speaker's voice.

An examination of the errors produced in the varied-sentence condition showed that the CI users displayed a significant response bias to incorrectly respond "different" more often than "same" for these pairs of sentences. In contrast, the NH children showed no evidence of any response bias in this condition. The differences observed in talker discrimination performance across these two sentence conditions suggest that basic sensory-auditory discriminative capacities are not the primary factor that controls performance in the same-different sentence discrimination task. Rather, discrimination performance is influenced by differences among the two groups in their ability to actively use cognitive control strategies to encode, maintain, monitor and manipulate representations of the talker's voice in working memory and selectively attend to a specific component perceptual dimension of the speech signal. These findings were even more remarkable because the control group of NH children was three years younger than the group of deaf children with cochlear implants. This initial study was followed up with a more extensive investigation of talker discrimination skills of deaf children with CIs which revealed strong correlations between talker discrimination and a wide range of speech and language outcome measures (see Cleary, Pisoni & Kirk, 2005).

Neurocognitive Measures

To broaden the measures of outcome and benefit following implantation beyond the traditional endpoint speech and language assessments, we recently completed a new study that was designed to obtain additional neuropsychological measures of EF, CC and EOI processes from a group of 5-10 year-old deaf children with CIs. All but one of these children received their CIs before three years of age. A group of chronologically age-matched typically-developing NH children was also recruited to serve as controls. Both groups received a battery of neuropsychological tests designed to assess selected aspects of EF and CC including verbal and spatial working memory capacity, inhibition, processing speed, as well as fine motor control.

In addition to conventional performance measures of EF and CC obtained in the laboratory, we also obtained several additional measures using three parental report behavioral rating scales to assess EF and behavioral regulation, learning and executive attention and attentional control and self-regulation in everyday real-world settings. More details of this study are reported by Conway et al. (2008b). For now, however, we summarize of a subset of the findings here for three of the neurocognitive performance tests, Fingertip Tapping (FTT), Design Copying (DC) and Stroop Color-word Naming (Stroop) that revealed differences in EOI functioning between the two groups of children (see also Figueras et al., 2008; Hauser et al., 2008). In the next section, we report the results of the three behavior rating scales.

NEPSY Fingertip Tapping (FTT) and Design Copying (DC). Two of the NEPSY performance measures, the FTT and DC tests, revealed differences in performance between the CI and NH groups. The FTT subtest is part of the NEPSY sensory functions core domain and is designed to assess finger dexterity and motor speed. In the Repetitive Finger Tapping condition, the child is asked to make a circle with their thumb and index finger opening and closing it as fast as they can. In the Sequential Fingertip Tapping condition, the child taps the tips of his/her thumb to the index, middle, ring and pinky making a circle with each finger. Both tests are carried out with the preferred and nonpreferred hands. The DC subtest of the NEPSY is part of the visuo-spatial processing domain that is used to assess a child's non-verbal visuo-spatial skills such as body movement and hand-eye coordination. DC measures the child's ability to reproduce and construct visual patterns. The children were given eighteen geometric designs and were asked to copy each design using paper and pencil. DC is similar to the VMI that we used with younger children in our previous work that was successful in uncovering preimplant behavioral predictors of outcome (Horn et al., 2007). The results for both the FTT and DC tests indicated that deaf children with CIs performed more poorly than age-matched NH control children. Moreover, the mean scores for the CI group on the FFT were not only significantly lower than the NH children's scores but they were also atypical relative to the published normative data. These results revealed weaknesses and delays in sensory-motor and visual-spatial domains that are consistent with our hypothesis that domain-general organizational-integrative processes are at risk in deaf children with CIs.

Stroop Color Word Naming. Both groups of children were also administered the Stroop Color Word Test (SCWT), which consists of three subtests: a word reading subtest that requires reading a series of 100 alternating words (either red, green, or blue) aloud as quickly as possible, a color naming subtest that requires naming a series of 100 alternating colors (indicated by X's in the colors of red, green, or blue), and a color-word subtest that requires naming the color of ink used to print each of the words (red, green, or blue, when the word name and ink color are different). Because it is much easier to read words than name colors, the color-word subtest of the SCWT is considered to be an excellent measure of the ability to inhibit a more automatic dominant response (word reading) in favor of a more effortful color naming response. Automatic word reading interferes with color naming on the color-word subtest because the printed word and the color of the ink are different and compete with each other for attention and processing resources. Word reading causes interference with the more difficult controlled-processing task of color naming. However, this interference effect occurs only to the extent that word reading is more fluent and automatic than color naming.

Individuals with less fluent reading skills or delayed phonological processing skills often perform better on the color-word subtest because they experience less interference from the (normally automatic) word reading component of that subtest (Golden, Freshwater, & Golden, 2003). Increases in reading proficiency cause greater interference and, in turn, greater relative impairment in color-word subtest scores.

Results of the SCWT revealed similar performance speed on the color-word subtest for the two groups, although the CI group performed significantly more slowly on the word reading subtest. However, the two groups did not differ on the color naming subtest. The pattern of differences in word and color naming scores is consistent with less proficient phonological processing in the CI group and indicates that the word reading task was less automatized for the deaf children with CIs. The CI group showed less interference from the word reading component of the color-word task and would have been expected to do better on the color-word task compared to the NH group, a pattern that is consistent with findings that less proficient readers do better than more proficient readers on the color-word task (assuming that other cognitive abilities are matched between the groups). The failure to find differences between groups on the color-word subtest may also reflect greater resistance to interference in the NH group than in the CI group. The pattern of Stroop word reading subtest results observed with the CI group further suggests less robust automatized lexical representations of color words in memory as well as possible delays in verbal fluency and atypical attentional switching skills in reading isolated color words aloud.

BRIEF, LEAF and CHAOS Rating Scales of Executive Function. To obtain measures of executive function as they are realized in everyday real-world environments like home, school or preschool settings, outside the highly controlled conditions of the audiology clinic or research laboratory, we used a neuropsychological instrument called the BRIEF (Behavior Rating Inventory of Executive Function; Psychological Assessment Resources (PAR), Inc, 1996). Three different forms of the BRIEF are available commercially from PAR with appropriate norms. One form was developed for preschool children (BRIEF-P: 2.0- 5.11 years); another for school-age children (BRIEF: 5-18 years) and finally one was also developed for adults (BRIEF-A: 18-90 years). The BRIEF family of products was designed to assess executive functioning in everyday environments (Gioia, Isquith, Guy & Kenworthy, 2000).

The BRIEF consists of a rating-scale behavior inventory that is filled out by parents, teachers and/or daycare providers to assess a child's executive functions and self-regulation skills. It contains 8 clinical scales that measure specific aspects of executive function related to inhibition, shifting of attention, emotional control, working memory, planning and organization among others. Scores from these subscales are then combined to construct several aggregate indexes for BRI and MI. Each rating inventory also provides an overall GEC score.

The BRIEF has been shown in a number of recent studies to be useful in evaluating children with a wide spectrum of developmental and acquired neurocognitive conditions although it has not been used with deaf children who use cochlear implants. From our preliminary work so far, we believe that this instrument may provide new additional converging measures of executive function and behavior regulation that are associated with conventional speech and language measures of outcome and benefit in this clinical population. Some of these measures can be obtained preimplant and therefore may be useful as behavioral predictors of outcome and benefit after implantation. Others are obtained after implantation and have turned out to have clinical utility in the management and counseling of children with CIs, especially poorer-performing children.

The BRIEF parent-report rating inventory combined with clinical observations, parent interviews and speech perception, language and speech production assessments, have added an important new clinical component to our research and has generated numerous discussions with our colleagues in pediatric neuropsychology. Some parents of younger children often inquire whether their child’s behaviors are similar to other children of the same age. Parents of older children have reported that their child is having more difficulty socially or academically and can list concrete changes in behavior at home and performance in school. In both cases, parents are looking for normative benchmarks and suggestions because they either don’t know if their child’s behavior is typical for their age or they want to know how to address manifested problems.

The information on executive function and cognitive control provided by the BRIEF clinical scales provides a quantifiable platform for broadening our discussions with parents to include possible underlying causes of particular behaviors, the effects of those behaviors on everyday real-world activities, as well as, test performance. These discussions often lead to suggestions for intervention and aural rehabilitation. Discussions have included book recommendations, the role of parent training in effective behavior management and referrals to child behavior specialists in our Autism and ADHD clinic. The BRIEF has also been used in our center to track changes in executive function and cognitive control over time and document improvements from one assessment interval to the next.

Our initial analysis of scores obtained on the BRIEF from 30 NH 5-8 year-old children and 19 hearing-impaired 5-10 year-old children with CIs revealed elevated scores on several subscales. Figure 2a shows a summary of the BRIEF T-Scores for the GEC composite scale and the two aggregate scales, the MI and the BRI. The GEC, MI and BRI scores were all significantly higher elevated for deaf children with CIs than NH children although none of means for the CI group fell within the clinically significant range.

Figures 2b, 2c show the T-scores for the individual clinical scales of the BRIEF. Examination of the eight individual clinical subscales showed statistically significant differences in five of the BRIEF scales:

initiation (INT), working memory (WM), planning and organization (PO), shifting (SH) and emotional control (EC). No differences were observed in organization of materials (OM), monitoring (MNTR) or inhibition (INH). The BRIEF scores provide additional converging evidence from measures of everyday real-world behaviors that multiple processing systems are linked together in development and that disturbances resulting from deafness and language delay are not domain-specific and only narrowly restricted to hearing, audition and the processing auditory signals. The effects of deafness and language delay appear to be more widely distributed among many different neural systems and neurocognitive domains.

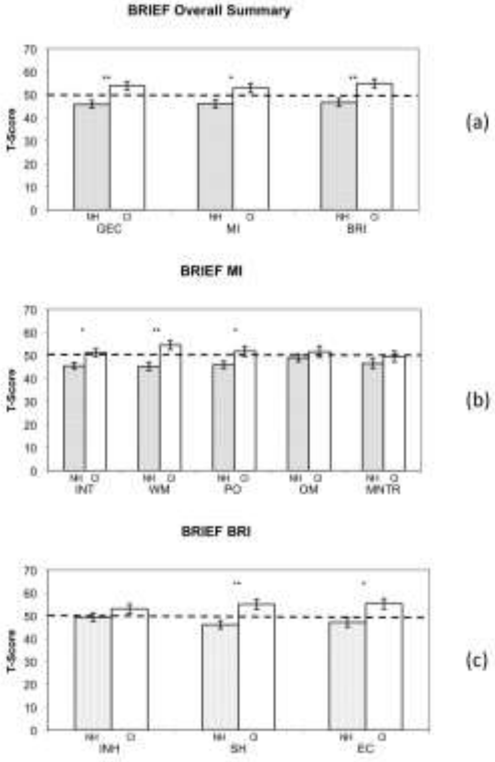

Figure 2. Mean T-Scores for NH children and deaf children with CIs obtained from the BRIEF parent-report behavioral rating inventory. Panel A shows the T-Scores for the Global Executive Composite (GEC), Meta Cognitive Index (MI) and the Behavior Regulation Index (BRI). Panel B shows the five individual MI clinical scales: Initiation (INT), Working Memory (WM), Planning and Organization (PO), Organization of Materials (OM) and Monitoring (MNTR). Panel C shows the three individual BRI clinical scales: Inhibition (INH), Shifting (SH) and Emotional Control (EC).

Two other parent- and teacher-report checklists have been developed at our ADHD clinic to evaluate executive functioning related to learning (Learning Executive and Attention Functioning scale (LEAF) and behavior problems (Conduct-Hyperactive-Attention Problem-Oppositional Scale (CHAOS) (Kronenberger & Dunn, 2008). Comparison of the CHAOS and LEAF scores between the CI and NH groups revealed elevated scores on most of the clinical subscales for the children with CIs. In particular, as shown in Figure 3a, 3b, statistically significant differences were observed on the learning, memory, attention, speed of processing, sequential processing, complex information processing, and novel problem solving subscales on the LEAF and on the attention problems and hyperactivity scales on the CHAOS shown in Figure 3c. No differences were observed for organization and reading on the LEAF or for the oppositional problems and conduct disorder subscales on the CHAOS.

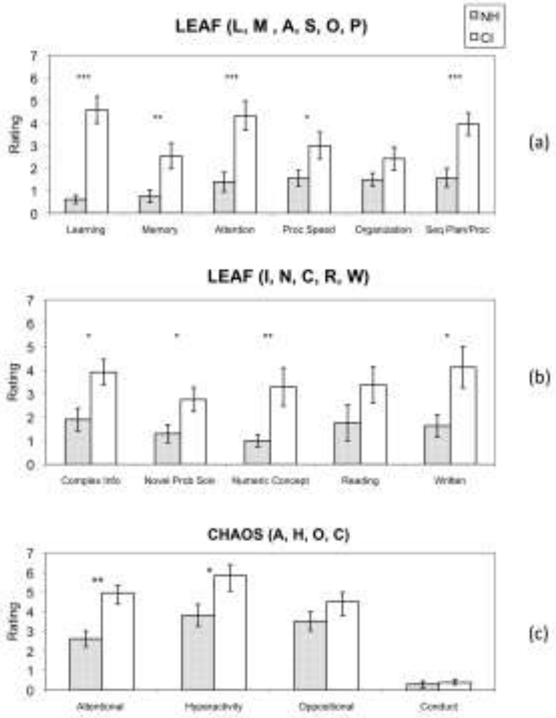

Figure 3. Mean ratings for the NH children and deaf children with CI's obtained from the LEAF parent-report behavioral rating inventory. Panel A shows the mean scores for: Learning, Memory, Attention, Processing Speed, Organization and Sequential Planning and Processing. Panel B shows the mean scores for: Complex Information Processing, Novel Problem Solving, Numerical Concepts, Reading and Writing. Panel C shows the mean ratings for the NH children and deaf children with CIs obtained from the CHAOS parent-report behavioral rating inventory for: Attention, Hyperactivity, Oppositional and Conduct Disorders.

These new findings suggest that a period of profound deafness and associated language delay before cochlear implantation not only affect basic domain-specific speech and language processes but also affect self-regulation and emotional control, processes not typically considered to be co-morbid with deafness and sensory deprivation. The scores on the BRIEF, LEAF and CHAOS rating scales provide additional converging evidence and support for the general hypothesis that multiple processing systems are linked together in development and that disturbances resulting from deafness and language delays are not domain-specific and restricted only to hearing, speech perception and processing spoken language. The disturbances appear to be more broadly distributed among many different brain systems that are used in language processing including other domains such as problem solving, writing and numerical cognition as well as emotional control, self-regulation and control of action in novel situations requiring adaptive behaviors

Implicit Learning of Sequential Patterns

Very little is currently known about how learning of complex sequential patterns contributes to language outcomes following cochlear implantation. At a fundamental level of analysis, all spoken language consists of a sequence of linguistic units (phonemes, syllables and words) built from a small inventory of elementary speech sounds organized in a linearly-ordered temporal sequence (Lashley, 1951). These units of spoken language do not occur randomly, but are highly regular and structured according to complex probabilistic relations that make human language predictable and learnable (Miller & Selfridge, 1950; Rubenstein, 1973). After acquiring knowledge about the probabilistic relations governing word order, an individual's knowledge of these sequential probabilities in language can enable a listener to reliably identify and predict the next word that will be spoken in a sentence (Elman, 1990; Kalikow, Stevens & Elliott, 1977; Miller & Selfridge, 1950).

Several researchers have argued recently that language development reflects the operation of fundamental learning processes related to acquiring knowledge of complex probabilistic patterns. Implicit or "statistical learning" is currently thought to be one of the basic learning mechanisms that is used in language acquisition (Altmann, 2002; Cleeremans, Destrebecqz, & Boyer, 1998; Saffran, Senghas, & Trueswell, 2001; Ullman, 2004). There are many published examples of infants (Saffran, Aslin, & Newport, 1996), children (Meulemans & Van der Linden, 1998), adults (Conway & Christiansen, 2005), neural networks (Elman, 1990), and even nonhumans (Hauser, Newport, & Aslin, 2000) demonstrating implicit learning capabilities.

These studies have demonstrated that humans, at least under typical (i.e., "normal") conditions of development, are equipped with the necessary raw learning capabilities to acquire the complex probabilistic structure found in language. Furthermore, recent findings from our research group have revealed a close empirical link between individual differences in implicit sequence learning and spoken language processing abilities (Conway, Baurenschmidt, Huang, & Pisoni, 2008a; Conway, Karpicke, & Pisoni, 2007).

In our initial studies, young healthy adults carried out a visual implicit sequence learning task and a sentence perception task that required listeners to recognize words under degraded listening conditions. The test sentences were taken from the Speech in Noise Test (SPIN) and varied on the predictability of the final word (Kalikow et al., 1977). Performance on the implicit sequence learning task was found to be significantly correlated with performance on the speech perception task – specifically, for the high predictability SPIN sentences that had a highly predictable final word. This result was observed even after controlling for common sources of variance associated with non-verbal intelligence, short-term memory, working memory, and attention and inhibition (see Conway et al., 2008a).

The findings obtained with adults suggest that general abilities related to implicit learning of sequential patterns are closely coupled with the ability to acquire and use information about the predictability of words occurring in the speech stream, knowledge that is critical for the successful acquisition of linguistic competence. The more knowledge that an individual acquires about the underlying sequential patterns of spoken language, the better one is able to use one's long-term knowledge of those patterns to perceive and understand novel spoken utterances, especially under highly degraded or challenging listening conditions. While these initial studies provided evidence for an important empirical link between implicit learning and language processing in NH adults, in order to better understand the development of implicit learning, it is necessary to investigate implicit sequence learning processes in both typically-developing and atypically-developing populations, specifically, profoundly deaf children who have been deprived of sound and the normal environmental conditions of development conducive/appropriate for language learning.

In a recent study, we investigated implicit sequence learning in a group of deaf children with CIs and a chronologically age-matched control group of NH typically-developing children to assess the effects that a period of auditory deprivation and delay in language may have on learning of complex visual sequential patterns (Conway et al., 2008c). Some evidence already exists that a period of auditory deprivation occurring early in development may have secondary cognitive and neural sequelae in addition to the obvious first-order hearing-related sensory effects (see Conrad, 1979; Luria, 1973; Myklebust & Brutten, 1953). Specifically, because sound is a physical signal distributed in time, lack of experience with sound may affect how well a child is able to encode, process, and learn sequential patterns and encode and store temporal information in memory (Fuster, 1995, 1997, 2001; Marschark, 2006; Rileigh & Odom, 1972; Todman & Seedhouse, 1994). Exposure to sound may also provide a kind of "auditory scaffolding" in which a child gains specific experiences and practice with learning and manipulating sequential patterns in the environment.

Based on our recent implicit visual sequence learning research with adults, we predicted that deaf children with CIs would show disturbances in visual implicit sequence learning because of their lack of experience with auditory temporal patterns early on in development. We also predicted that sequence learning abilities would be associated with several different measures of language development in both groups of children.

Two groups of 5-10 year old children participated in this study. One group consisted of 25 deaf children with CIs; the second group consisted of 27 age-matched typically-developing, NH children. All children carried out two behavioral tasks: an implicit visual sequence learning task and a sentence perception task. Several clinical measures of language outcome were available for the CI children from our larger longitudinal study. Scores on these tests were also obtained for the NH children. Our hypothesis was that if some aspects of language development draw on general learning abilities, then we should observe correlations between performance on the implicit visual sequence learning task and several different measures of spoken language processing. Measures of vocabulary knowledge and immediate memory span were also collected from all participants in this study in order to rule out obvious mediating variables that might be responsible for any observed correlations. The presence of correlations between the two tasks even after partialing out the common sources of variance associated with these other measures would provide support for the hypothesis that implicit learning is directly associated with spoken language development, rather than being mediated by a third contributing factor.

Visual Implicit Sequence Learning Task. Two artificial grammars (Grammars A and B) were used to generate the colored sequences used in the implicit learning task. These grammars specified the probability of a particular color occurring given the preceding color in sequence. Sequence presentation consisted of colored squares appearing one at a time, in one of four possible positions in a 2 x 2 matrix on a computer touch screen. The four elements (1-4) of each grammar were randomly mapped onto each of the four screen locations as well as four possible colors (red, blue, yellow, green). The assignment of stimulus element to position/color was randomly determined for each subject; however, for each subject, the mapping remained consistent across all trials. Grammar A was used to generate 16 unique sequences for the learning phase and 12 sequences for the test phase. Grammar B was used to generate 12 novel sequences for the test phase.

For the implicit learning task, the children were told that they would to see sequences of four colored squares displayed on the touch screen. The squares would flash on the screen in a pattern and their job was to remember the pattern of colors on the screen and reproduce each pattern at the end of each trial. The procedures for both the learning and test phases were identical and from the perspective of the subject, there was no indication of separate phases at all. The only difference between the two phases was which sequences were used. In the Learning Phase, the 16 learning sequences from Grammar A were presented first. After completing the reproduction task for all of the learning sequences, the experiment seamlessly transitioned to the Test Phase, which used the 12 novel sequences from Grammar A and 12 novel Grammar B test sequences.

The colored squares appeared one at a time, in one of four possible positions on the touch screen. The four colors (red, blue, yellow, green) of each grammar were randomly mapped onto each of the four screen location. The assignment of stimulus element to position/color was randomly determined for each subject; however, for each subject, the mapping remained consistent across all trials. The children were not told that there was an underlying grammar for any of the learning or test sequences or that there were two types of sequences in the Test Phase. The child just observed and then reproduced the visual sequences.

Eisenberg Sentence Perception Task. For this task, we used a set of English lexically-controlled sentences developed by Eisenberg, Martinez, Holowecky, and Pogorelsky (2002). The stimuli consisted of twenty lexically-easy (i.e., high word frequency, low neighborhood density) and twenty lexically-hard (i.e., low word frequency, high neighborhood density) meaningful English sentences. The sentences were presented through a loudspeaker at 65 dB SPL. The children were instructed to listen closely to each sentence and then repeat back what they heard to the examiner even if they were only able to perceive one word of the sentence. All of the test sentences were presented in random order to each child. Responses were recorded onto digital audio tape (DAT) and were later scored off-line based on number of keywords correctly repeated for each sentence. The sentences were played in the quiet without any degradation to the deaf children with CIs. For the NH children, the original sentences were spectrally degraded to simulate a cochlear implant with a four-channel sinewave vocoder (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; TigerSpeech.Com) to reduce their performance from ceiling levels.

In the implicit learning task, a sequence was scored correct if the participant reproduced each test sequence correctly in its entirety. Sequence span scores were then calculated using a weighted method in which the total number of correct test sequences at a given length was multiplied by the length and then scores for all lengths were added together (see Cleary, Pisoni, & Geers, 2001). We calculated separate sequence span scores for Grammar A and Grammar B test sequences for each subject.

For each subject we also calculated an implicit learning score (LRN), which was the difference in span scores between the learned grammar (Grammar A) and the novel grammar (Grammar B). The LRN score measures generalization indicating the extent that sequence memory spans improved for sequences that had been previously experienced during the initial Learning Phase. This score reflects how well memory spans improve for *novel* sequences that were constructed by the same grammar that subjects had previously experienced in the Learning Phase, relative to span scores for novel sequences created by the new grammar.

For the Eisenberg sentence perception task, percent keyword correct scores were calculated separately for easy and hard sentences. Each child received a forward and backward digit span score, reflecting the number of digit lists correctly repeated. Each child also received a standardized PPVT score based on how many pictures were correctly identified and their chronological age.

Group Differences in Implicit Learning. Figure 4a shows the average implicit learning (LRN) scores for both groups of children. For the NH children, the average implicit learning score (2.5) was significantly greater than 0, *t*(25)=2.24, *p*<.05, demonstrating that as a group the NH children showed better learning of test sequences with the same statistical structure as the sequences from the Learning Phase. On the other hand, the average implicit learning score for the CI children was -1.43, a value that was not statistically different from 0, *t*(22)=-.91, *p*=.372. We also conducted a univariate ANOVA with the factor of group (NH, CI), which revealed a statistically significant effect of group on the implicit learning scores, *F*(1, 47) = 4.3, *p* < .05. On average, the NH group showed greater implicit learning than the CI group, who in turn essentially showed no implicit learning on this task.

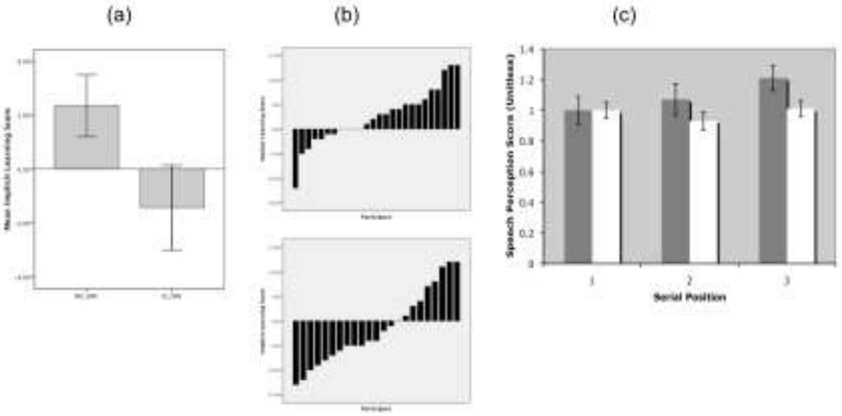

Figure 4. Panel A shows the average visual implicit learning scores for NH children (left) and deaf children with CIs (right). Panel B shows the implicit learning scores for individual children in the NH group (top) and CI group (bottom), ranked ordered from lowest to highest. Panel C shows the word recognition scores for NH (grey) and CI children (white) as a function of serial position within sentences for Word 1, Word 2, and Word 3. The ordinate shows a difference score that computed by dividing the score at each word position by the score at Word 1, for the NH and CI groups separately.

In addition to comparing group means, we also examined the distribution of individual scores for each of the two groups of children on the implicit learning task. Figure 4b shows the implicit learning scores for each individual participant in the NH group (top) and the CI group (bottom). Whereas 19 out of 26 (73%) of the NH children showed an implicit learning score of 0 or higher, only 9 out of 23 (39%) of the CI children showed a score above 0. Chi-square tests revealed that the proportion of learners to non-learners was significantly different for the NH children, $X^2(1) = 5.54$, *p*<.05, but not for the CI children, $X^2(1) = 1.08$, *p*=0.297. That is, more than half of the NH children showed an implicit learning effect whereas this was not the case with the CI children.

The present results demonstrate that deaf children with CIs show atypical visual implicit sequence learning compared to age-matched NH children. This result is consistent with the hypothesis that a period of deafness and language delay may cause secondary disturbances and/or delays in the development of visual sequencing skills. In addition, for the CI children, we computed a partial correlation between their implicit learning score and age at implantation, with chronological age partialed out. Implicit learning was negatively correlated with the age at which the child received their implant and positively correlated with the duration of implant use. That is, the longer the child was deprived of auditory stimulation, the lower the visual implicit learning scores; correspondingly, the longer the child had experience with sound via his/her implant, the higher the implicit learning scores. These correlations suggest that exposure to sound via a cochlear implant has secondary indirect effects on basic learning processes that are not directly associated with hearing, speech perception or language development per se; longer implant use appears to be associated with better ability to implicitly learn complex visual sequential patterns and acquire knowledge about the underlying abstract grammar that generated the patterns.

Implicit Learning and Sentence Perception. The observed individual differences in implicit learning were also correlated with performance on the Eisenberg sentence perception task which measured how well a child can perceive words in meaningful sentences, a language processing task that involves the use of both bottom-up sensory perceptual processes as well as top-down conceptual knowledge of language. Based on our earlier work with adults, we hypothesized that implicit sequence learning would be directly related to the use of top-down knowledge in speech perception and therefore we predicted a significant association between these two tasks. To assess this prediction, we calculated partial correlations between the implicit learning score and the two sentence perception scores (lexically-easy sentences and lexically-hard sentences), while controlling for the common variance associated with chronological age, forward digit span, backward digit span, PPVT, and for the CI children, age at implantation and articulation abilities as measured by scores obtained from the GFTA.

We found that implicit learning scores for deaf children with CIs were associated with their ability to effectively use sentence context to guide speech perception, as reflected by the sentence perception difference score for combined lexically-easy and –hard sentences. Implicit learning was significantly correlated with the sentence perception difference scores, specifically, for the lexically-hard sentences. These results suggest that better implicit learning abilities result in more robust knowledge of the sequential predictability of words in sentences, which leads in turn to better use of sentence context to aid speech perception, as reflected in the sentence perception difference score.

These results suggest that for the CI children, implicit learning is used to acquire information about word predictability in language, knowledge that can be brought to bear under conditions where sentence context can be used to help perceive the next word in an utterance. If this is the case, then we would expect that the CI children, who scored worse as a group overall on implicit learning, will also be impaired on their ability to make use of the preceding context of a sentence to help them perceive the next word.

Figure 4c shows the performance of correctly identifying the three target words in the sentence perception task, as a function of the position in the sentence (1^{st}, 2^{nd}, or 3^{rd}), for the NH and CI children. Sentence context can do very little to aid perception of the first target word in the sentence; however, context is useful to help perceive the second and third target words, but only if the child has sufficient top-down knowledge of word order regularities. Indeed, the NH children showed an improvement in speech perception for the second and third target words. Their performance on the last word was statistically greater than performance on the first word, $\underline{t}(25)=4.2$, $\underline{p}<.001$. In contrast, the CI children failed to show the same contextual facilitation. Their performance on the third word was no different than their performance on the first word in each sentence, $\underline{t}(22)=.106$, $\underline{p}=.92$. Unlike the NH children, the deaf children with CIs do not appear to be using sentence context predictably to help them perceive the final word in the sentence. Thus, one way in which weak implicit learning abilities may reveal themselves are in situations in which word predictability and sentence context can be used as a processing heuristic to guide spoken language perception.

Implicit Learning and Language Outcomes in Deaf Children with CIs. We also found that implicit learning was positively and significantly correlated with three subtests of the CELF-4: Concepts and Following Directions, Formulated Sentences and Recalling Sentences (Semel, Wiig, & Secord 2003. These subtests involve understanding and/or producing sentences of varying complexity, tasks in which knowledge of word order predictability – i.e., statistics of sequential probabilities in language – can be brought to bear to improve performance. Implicit learning was also positively and significantly correlated with receptive language on the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984).

The pattern of correlations obtained in this recent study suggests that implicit learning may be most strongly related to the ability to use knowledge of the sequential structure of language to better process, understand, and produce meaningful sentences, especially when sentence context can be brought to bear to aid processing. Importantly, this association does not appear to be mediated by chronological age, age of implantation, short-term or working memory, vocabulary knowledge, or the child's ability to produce speech. Moreover, these findings were modality-independent. The implicit sequence learning task used only visual patterns whereas the sentence perception task relied on an auditory-only presentation of spoken sentences

It is possible that experience with sound and auditory patterns via a CI, which are complex, serially-arrayed signals, provides a deaf child critical experience with perceiving and learning sequential patterns establishing strong links between speech perception and production. A period of deafness early in development deprives a child of experience in dealing with complex sequential auditory input, which affects their ability to encode and process sequential patterns in other sense modalities as well (Myklebust & Brutten, 1953). Once electrical hearing is introduced via a CI, a child begins for the first time to gain experience with auditory sequential input. The positive correlation between length of CI use and implicit learning scores obtained even when chronological age was partialed out suggests that early experience and interactions with sound via a CI improves a deaf child's ability to learn complex non-auditory visual sequential patterns. Thus, it is possible that given enough exposure and experiences with sound via a CI, a deaf child's implicit learning abilities will eventually improve to age-appropriate levels.

To explain these findings, we suggest that sound affects cognitive development by providing a perceptual and cognitive "scaffolding" of time and serial order, upon which temporal sequencing functions are based. From a neurobiological standpoint, it is known that lack of auditory stimulation early in development results in a decrease of myelination and fewer projections out of auditory cortex (Emmorey, Allen, Bruss, Schenker, & Damasio, 2003) – which may also include connectivity to the frontal lobe. Neural circuits in the frontal lobe, specifically the prefrontal cortex, are believed to play a critical role in learning, planning, and executing sequences of thoughts and actions (Fuster, 1995, 1997, 2001; Goldman-Rakic, 1988; Miller & Cohen, 2001). It is therefore possible that the lack of auditory input early on in development and the corresponding reduction of auditory-frontal connectivity fundamentally alters the neural organization of the frontal lobe and the extensive connections it has with other brain circuits (Wolff & Thatcher, 1990), impacting the development of sequencing functions regardless of input modality (Miller & Cohen, 2001).

Theoretical and Clinical Implications

Many of the deaf children with CIs tested in our studies also have comorbid disturbances and/or delays in several basic neurocognitive processes that subserve information processing systems used in spoken language processing, and these disturbances appear to be, at least in part, secondary to their profound hearing loss and delay in language development (Conrad, 1979; Rourke, 1989, 1995). A period of profound deafness and auditory deprivation during critical developmental periods before implantation affects neurocognitive development in a variety of ways. Differences resulting from both deafness and subsequent neural reorganization and plasticity of multiple brain systems may be responsible for the enormous variability observed in speech and language outcome measures following implantation. Without knowing what specific underlying neurobiological and neurocognitive factors are responsible for the individual differences in speech and language outcomes, it is difficult to recommend an appropriate and efficacious approach to habilitation and speech-language therapy after a child receives a cochlear implant. More importantly, the deaf children who are performing poorly with their CIs are not a homogeneous group and may differ in numerous ways from one another, reflecting dysfunction of multiple brain systems associated with congenital deafness and profound hearing loss. From a clinical perspective, it seems very unlikely that an individual child will be able to achieve optimal speech and language benefits from his/her CI without knowing why the child is having speech and language problems and which particular neurocognitive domains underlie these problems.

Some profoundly deaf children with CIs do extremely well on traditional audiological speech and language outcome measures while other children have much more difficulty. The enormous variability in outcome and benefit following cochlear implantation is a significant clinical problem in the field and it has not received adequate attention by research scientists in the past. Obtaining a better understanding of the neurocognitive basis of individual differences in outcomes will have direct implications for diagnosis, treatment and early identification of deaf children who may be at high risk for poor outcomes after implantation. New knowledge about the sources of variability in speech and language outcomes will also play an important role in intervention following implantation in terms of selecting specific methods for habilitation and treatment that are appropriate for an individual child. We have now identified two potential areas of neurocognitive functioning that may underlie variability in speech and language outcomes: EOI processes and implicit sequence learning abilities.

The bulk of clinical research on CIs has been intellectually isolated from the mainstream of current research and theory in neuroscience, cognitive psychology and developmental neuropsychology. As a consequence, the major clinical research issues have been narrowly focused on speech and language outcomes and efficacy of cochlear implantation as a medical treatment for profound hearing loss. Little basic or clinical research in the past has investigated the underlying neurobiological and neurocognitive bases of the individual differences and variability in the effectiveness of CIs. Moreover, few studies have attempted to identify reliable early neurocognitive predictors of outcome and benefit or systematically assessed the effectiveness of specific intervention and habilitation strategies after implantation. We believe these are important new areas of clinical research on CIs that draw heavily on basic research and theory representing the intersection of several closely related disciplines that deal with the relations between brain, behavior and development, memory and learning, attention, executive function and cognitive control.

References

Altmann, G.T.M. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences, 99,* 15250-15251.

Blair, C. & Razza-Peters, R. (2007). Relating effortful control, executive function, and false belief understanding to the emerging math and literacy ability in kindergarten. *Child Development, 78,* 647-663.

Bodrova, E. & Leong, D.J. (2007). *Tools of the mind*. Person-Merrill Prentice Hall: Columbus, OH.

Cicchetti, D., & Curtis, W. J. (2006). The developing brain and neural plasticity: Implications for normality, psychopathology, and resilience. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology (2nd ed.): Developmental neuroscience* (Vol. 2). New York: Wiley.

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. (2002). Talker discrimination by prelingually-deaf children with cochlear implants: Preliminary results. *Annals of Otology, Rhinology, & Laryngology Supplement-Proceedings of the 8th Symposium on Cochlear Implants in Children, 111,* 113-118.

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.

Cleary, M., Pisoni, D.B. & Kirk, K.I. (2005). 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, 48,* 204-223.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*, 406-416.

Conrad, R. (1997). *The deaf schoolchild*. London: Harper & Row, Ltd.

Conway, C.M., Bauernschmidt, A., Huang, S.S. & Pisoni, D.B. (2008a). Implicit statistical learning in language processing: Word predictability in the key. *Cognition*. (Under revision).

Conway, C.M. & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31*, 24-39.

Conway, C.M., Karpicke, J., & Pisoni, D.B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with normal-hearing adults. *Journal of Deaf Studies and Deaf Education, 12*, 317-334.

Conway, C.M., Karpicke, J., Anaya, E.M., Henning, S.C. & Pisoni, D.B. (2008b). Sequencing and executive function in hearing children and deaf children with cochlear implants.

Conway, C.M., Pisoni, D.B., Anaya, E.M., Karpicke, J., & Henning, S.C. (2008c). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science* (Under Revision).

Deary, I.J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13-21.

Diamond, A., Barnett, W.S., Thomas, J. & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387-1388.

Dillon, C.M., Burkholder, R.A., Cleary, M. & Pisoni, D.B. (2004). Nonword repetition by children with cochlear implants: Accuracy ratings from normal-hearing listeners. *Journal of Speech, Language and Hearing Research, 47,* 1103-1116.

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.

Eisenberg, L.S., Martinez, A.S., Holowecky, S.R. & Pogorelsky, S. (2002). Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. *Ear & Hearing, 23*, 450-462.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14,* 179-211.

Emmorey, K., Allen, J.S., Bruss, J., Schenker, N., & Damasio, H. (2003). A morphometric analysis of auditory brain regions in congentially deaf adults. *Proceedings of the National Academy of Sciences, 100,* 10049-10054.

Figueras, B., Edwards, L. & Langdon, D. (2008). Executive function and language in deaf children. *Journal of Deaf Studies and Deaf Education, 13,* 362-377.

Fuster, J. (2001). The prefrontal cortex—an update: Time is of the essence. *Neuron, 30,* 319-333.

Fuster, J. (1997). *The prefrontal cortex*. Philadelphia: Lippincott-Raven.

Fuster, J. (1995). Temporal processing. In J. Grafman, K.J. Holyoak, & F. Boller (Eds.), *Structure and* *functions of the human prefrontal cortex* (pp. 173-181). New York: New York Academy of Sciences.

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.

Geers, A. & Brenner, C. (2003). Background and educational characteristics of prelingually deaf children implanted for five years of age. *Ear & Hearing, 24,* 2S-14S.

Geers, A., Brenner, C., & Davidson, L. (2003). Factors associated with development of speech perception skills in children implanted by age five. *Ear & Hearing, 24,* 24S-35S.

Gioia, G.A., Isquith, P.K., Guy, S.C., & Kenworthy, L. (2000). *BRIEF™: Behavior Rating Inventory of Executive Function.*

Golden, C.J., Freshwater, S.M., Golden, Z. (2003). *Stroop Color and Word Test Children's Version for Ages 5-14.* Stoelting Company: Wood Dale, IL.

Goldman-Rakic, P.S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Reviews of Neuroscience, 11,* 137-156.

Hauser, M.D., Newport, E.L. & Aslin, R.N. (2000). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins, *Cognition, 75*, 1-12.

Hauser, P.C., Lukomski, J. & Hillman, T. (2008). Development of deaf and hard-of-hearing students' executive function. In M. Marschark & P.C. Hauser (Eds.), *Deaf cognition*. (pp. 268-308). Oxford: Oxford University Press.

Hawker, K., Ramirez-Inscoe, J., Bishop, D.V.M., Twomey, T., O'Donoghue, G.M. & Moore, D.R. (2008). Disproportionate language impairment in children using cochlear implants. *Ear & Hearing, 29,* 467-471.

Hohm, E., Jennen-Steinmetz, C. Schmidt, M.H., & Laucht, M. (2007). Language development at ten months. *European Child & Adolescent Psychiatry, 16,* 149-156.

Horn, D.L., Fagan, M.K., Dillon, C.M., Pisoni, D.B. & Miyamoto, R.T. (2007). Visual-motor integration skills of prelingually deaf children: Implications for pediatric cochlear implantation. *The Laryngoscope, 11*, 2017-2025.

Hughes, C. & Graham, A. (2002). Measuring executive functions in childhood: Problems and solution? *Child and Adolescent Mental Health, 7,* 131-142.

Kalikow, D.N., Stevens, K.N. & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Kaufman, A.S., Lichtenberger, E.O. (2006). *Assessing adolescent and adult intelligence, 3rd Ed.* New York: Wiley.

Kronenberger, W.G., & Dunn, D.W. (2008). *Development of a very brief, user-friendly measure of ADHD for busy clinical practices: The CHAOS Scale*. Poster presented at the 2008 National Conference on Child Health Psychology. Miami Beach, FL, April 11, 2008.

Lamm, C., Zelazo, P.D. & Lewis, M.D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia, 44*, 2139-2148.

Lashley, K.S. (1951). The problem of serial order in behavior. In L.A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-146). New York: Wiley.

Lezak, M.D., Howieson, D.B., Loring, D.W. & Hannay, H.J. (2004). *Neurological assessment*. New York: Oxford University Press.

Luria, A.R. (1973). *The working brain.* Penguin.

Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. *European Journal of Cognitive Psychology, 18*, 70-89.

Mathews, V.P., Kronenberger, W.G., Wang, Y., Lurito, J.T., Lowe, M.J., & Dunn, D.W. (2005). Media violence exposure and frontal lobe activation measured by fMRI in aggressive and non-aggressive adolescents. *Journal of Computer Assisted Tomography, 29*, 287-292.

Meulemans, T. & Van der Linden, M. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69,* 199-221.

Miller, E.K. & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews in Neuroscience, 24,* 167-202.

Miller, G.A. & Selfridge, J.A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176-185.

Myklebust, H.R. (1964). *The psychology of deafness*. New York: Grune & Stratton.

Myklebust, H.R. (1954). *Auditory disorders in children*. New York: Grune & Stratton.

Myklebust, H.R. & Brutten, M. (1953). A study of visual perception of deaf children. *Acta Otolaryngology, Suppl. 105,* p. 126.

Nauta, W.J.H. (1964). Discussion of 'Retardation and facilitation in learning by stimulation of frontal cortex in monkeys.' In J.M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior*. (pp. 125). New York: McGraw-Hill.

NIDCD (1988). *Cochlear implants*. NIH Consensus Statement, May 4, Vol. 7.

NIDCD (1995). *Cochlear implants in adults and children*. NIH Consensus Statement, May 15-17, 13, 1-30.

Psychological Assessment Resources, Inc. (1996). Lutz, FL 33549

Rileigh, K.K. & Odom, P.B. (1972). Perception of rhythm by subjects with normal and deficient hearing. *Developmental Psychology, 7,* 54-61.

Rourke, B. P. (1989). *Nonverbal learning disabilities.* New York: The Guilford Press.

Rourke, B. P. (1995). *Syndrome of nonverbal learning disabilities.* New York: The Guilford Press.

Rubenstein, H. (1973). Language and probability. In G.A. Miller (Ed.), *Communication, language, and meaning: Psychological perspectives* (pp. 185-195). New York: Basic Books, Inc.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Saffran, J.R., Senghas, A., & Trueswell, J.C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences, 98*, 12874-12875.

Semel, E., Wiig, E.H., & Secord, W.A. (2003). *Clinical evaluation of language fundamentals, fourth edition (CELF-4).* Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303-304.

Sparrow, S. Balla, D. & Cicchetti, D. (1984). *Vineland Adaptive Behavioral Scales*. Circle Pines, MN: American Guidance Service.

Sporns, O. (1998). Biological variability and brain function. In J. Cornwell (Ed.), *Consciousness and human identity* (pp. 38-56). Oxford: Oxford University Press.

Sporns, O. (2003). Network analysis, complexity, and brain function. *Complexity, 8*, 56-60.

Thelen, E. & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge: The MIT Press.

Todman, J. & Seedhouse, E. (1994). Visual-action code processing by deaf and hearing children. *Language & Cognitive Processes, 9*, 129-141.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231-270.

Ullman, M.T. & Pierpont, E.I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41,* 399-433.

Van der Sluis, S., de Jong, P.F. & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*, 427-449.

Wolff, A.B. & Thatcher, R.W. (1990). Cortical reorganization in deaf children. *Journal of_Clinical and Experimental Neuropsychology*, *12*, 209-221.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Implicit Sequence Learning in Hearing Children and Deaf Children with Cochlear Implants[1]

Christopher M. Conway[2], David B. Pisoni, Esperanza M. Anaya, Jennifer Karpicke[3], and Shirley C. Henning[4]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by DC00111 and DC00012. We wish to thank Linda Smith and Char Wozniak for their help with accessing the Kid Information Database. We also wish to thank Laurie Eisenberg for providing us with audio recordings for the sentence perception task, Luis Hernandez for invaluable technical assistance, and Lauren Grove for her help in manuscript preparation. Finally, we are especially indebted to all of the children and parents who took part in this experiment.

[2] Department of Psychology, Saint Louis University, St. Louis, MO 63103

[3] Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN 47907

[4] Indiana School of Medicine, Indianapolis, IN 46202

Implicit Sequence Learning in Hearing Children and Deaf Children with Cochlear Implants

Abstract. Understanding what cognitive mechanisms contribute to successful language acquisition and development has remained an important but elusive goal of the psychological sciences. Current theories suggest that the ability to learn structured sequential patterns – i.e. ,"implicit" or "statistical" learning -- may underlie language processing and development. In this paper, we investigate the role that implicit sequence learning plays in the development of spoken language skills in both normal-hearing children and deaf children with cochlear implants (CIs). Both groups of children completed a novel implicit sequence learning task, which measured learning through improvement to immediate serial recall for statistically-consistent visual sequences. The children were also assessed on several measures of language processing and development, including an online sentence perception task. The results demonstrated two key findings. First, the deaf children with CIs showed disturbances in their implicit visual sequence learning abilities relative to the normal-hearing children. Second, in both groups of children, implicit learning was significantly correlated with measures of spoken language processing and development. These findings suggest that a period of auditory deprivation has secondary effects related to general sequencing deficits, and that individual differences in sequence learning are associated with the development of spoken language under both typical and atypical conditions.

Introduction

For many children who are profoundly deaf, a cochlear implant (CI) provides the means to successfully develop spoken language skills, a possibility that would otherwise be impossible. Even so, it is well known that whereas many children are able to develop age-appropriate speech and language abilities, other children obtain little benefit other than the awareness of sound from their implant (American Speech-Language-Hearing Association, 2004). Some of this variation in outcome has been shown to be due to certain demographic factors, such as age at implantation and length of deafness (Kirk et al., 2002; Tomblin, Barker, & Hubbs, 2007). However, these demographic variables leave a large amount of variance unexplained. Very little is currently known about how intrinsic cognitive factors, especially fundamental abilities related to the learning of complex sequential patterns, contributes to language outcomes following implantation in this unique population.

At its most fundamental level, spoken language consists of a series of sounds (phonemes, syllables, words) occurring in a temporal stream (Lashley, 1951). These units of sound do not occur randomly or haphazardly, but are structured according to complex probabilistic relations that render language at least partially predictable (Rubenstein, 1973), given a sufficiently developed sequence learning mechanism. After sufficiently learning the probabilistic relations governing word order, one's knowledge of these sequential probabilities in language can enable a listener to better identify – and perhaps even implicitly predict – the next word that will be spoken (Elman, 1990; Kalikow, Stevens, & Elliott, 1977).

One promising and currently popular account is that fundamental learning abilities related to acquiring complex probabilistic patterns – i.e., implicit or statistical learning – are used in the service of language acquisition (Altmann, 2002; Cleeremans, Destrebecqz, & Boyer, 1998; Conway & Pisoni, in press; Saffran, Senghas, & Trueswell, 2001; Ullman, 2004). There are many published examples of infants (Saffran, Aslin, & Newport, 1996), children (Meulemans & Van der Linden, 1998), adults (Conway & Christiansen, 2005), neural networks (Elman, 1990), and even nonhumans (Hauser, Newport, & Aslin, 2000) demonstrating implicit learning capabilities. These "existence proofs" have proven beyond a doubt that the human (and possibly non-human) organism, at least under typical developmental conditions, is equipped with the necessary raw learning capabilities for acquiring the kind of complex, probabilistic structure found in language. Furthermore, recent work has revealed a direct empirical link between individual differences in implicit learning and spoken language processing abilities (Conway, Bauernschmidt, Huang, & Pisoni, submitted; Conway, Karpicke, & Pisoni, 2007). In these studies, healthy adults were engaged in a visual implicit sequence learning task and a speech perception task that required listeners to recognize words in degraded sentences varying on the predictability of the final word. Adults' performance on the implicit learning task was found to be significantly correlated with performance on the speech perception task – specifically, for sentences having a highly predictable final word – even after controlling for common sources of variance associated with non-verbal intelligence, short-term memory, working memory, and attention and inhibition (Conway et al., submitted).

These recent findings suggest that general abilities related to implicit learning of sequential patterns is closely coupled with the ability to learn about the predictability of words occurring in the speech stream, knowledge that is fundamentally important for successful spoken language competence. The more that an individual is sensitive to the underlying sequential patterns in spoken language, the better one is able to use one's long-term knowledge of those patterns to help perceive and understand spoken utterances, especially under degraded listening conditions. While these initial studies provided an important empirical link between implicit learning and language processing in normal-hearing adults, in order to better understand the role of implicit learning in development, it is of major theoretical interest to investigate such abilities in both typically-developing and atypically-developing populations.

Thus, the aims of this paper are twofold: to investigate the possible role that implicit learning has in spoken language development (in both typically-developing children and in deaf children with CIs); and, to assess the effects that a period of auditory deprivation may have on implicit learning for non-auditory sequential patterns. Examining deaf children with CIs represents a unique opportunity to study brain plasticity and neural reorganization following a period of auditory deprivation (Pisoni, Conway, Kronenberger, Horn, Karpicke, & Henning, 2008). In some sense, the research effort on deaf children with CIs can be thought of as the modern equivalent of the so-called "forbidden experiment" in the field of language development: it provides an ethically acceptable research opportunity to study the effects of the introduction of sound and spoken language on cognitive and linguistic development after a period of auditory deprivation.

There is in fact some indication that a period of auditory deprivation occurring early in development may have secondary cognitive and neural ramifications in addition to the obvious hearing-related effects (see Luria, 1973). Specifically, because sound by its very nature is a temporally-arrayed signal, a lack of experience with sound may affect how well one is able to encode, process, and learn serial patterns (Marschark, 2006; Rileigh & Odom, 1972; Todman & Seedhouse, 1994). Exposure to sound may provide a kind of "auditory scaffolding" in which a child gains vital experience and practice with learning and manipulating sequential patterns in the environment.

In this paper, we explore these issues by examining implicit visual sequence learning and language abilities in deaf children with CIs and an age-matched group of normal-hearing children. Our hypothesis is that sequence learning abilities will be associated with measures of language development in both populations. Furthermore, we also investigate whether deaf children with CIs will show disturbances in visual implicit sequence learning as a result of their relative lack of experience with auditory patterns early on in development.

Experiment

Two groups of children participated, deaf children with CIs, and an age-matched group of typically-developing, normal-hearing (NH) children. All children carried out two tasks: an implicit visual sequence learning task and an online speech perception task. Furthermore, we also collected several clinical measures of language outcome for the CI children. We reason that if language development is based in part on general and fundamental learning abilities, then it ought to be possible to observe empirical associations between performance on the implicit visual sequence learning task and measures of language processing and development. Several additional measures were also collected from all participants in order to rule out alternative mediating variables – such as vocabulary knowledge or immediate memory span -- responsible for any observed correlations. Observing a correlation between the two tasks even after partialing out the common sources of variance associated with these other measures would provide additional support for the conclusion that implicit learning is directly associated with spoken language development, rather than being mediated by a third underlying factor.

Method

Participants. Twenty-five prelingually, profoundly deaf children (aged 5-10 years; mean: 90.1 mos) who had received a cochlear implant by age 4 (mean: 21.2 mos) were recruited through the DeVault Otologic Research Laboratory at the Indiana University School of Medicine, Indianapolis. All the children had profound bilateral hearing loss (90-dB or greater) and had used their implant for a minimum of 3 years (mean: 69.0 mos). All subjects also were native speakers of English. Except for two children with bilateral implants and one child who had a hearing aid in the non-implanted ear, all children had a single implant. For the three children with bilateral hearing, testing was conducted with only one CI activated (the original implant). Although several of the children had been exposed to Signed Exact English, none of the children relied exclusively on sign or gesture, and all children were tested using oral-only procedures. Aside from hearing loss, there were no other known cognitive, motor, or sensory impairments. Data from two children were excluded because of inattention or refusal to participate on the experimental tasks, leaving a total of 23 children included in the final analyses. Table 1 summarizes the demographic characteristics of these 23 children. For their time and effort, the children's parents/caregivers received monetary compensation.

Twenty-seven typically developing, NH children (aged 5-9 years; mean: 87 mos) were recruited through Indiana University's "Kid Information Database" and through the Life Education and Resource Home Schooling Network of Bloomington, IN. All children were native speakers of English; parental reports indicated no history of a hearing loss, speech impairment, or cognitive or motor disorder. For their participation, children received a small toy and their parents received monetary compensation.

Apparatus. A Magic Touch® touch-sensitive monitor displayed visual stimuli for the implicit learning task and recorded participant responses. Auditory stimuli for the sentence perception and digit span tasks were presented through an Advent AV570 loudspeaker.

				Observed score range	
Variable		**M**	**SD**	**Minimum**	**Maximum**
CA		90.13	19.86	61.00	118.00
Age at Implantation		21.17	8.27	10.00	39.00
CI Use Length		68.95	19.36	36.00	98.00
Etiology (n)					
Unknown		17			
Genetic	3				
Ototoxicity		1			
Mondini dysplasia		2			

Table 1. CA, chronological age.

Stimulus Materials. For the visual implicit learning task, we used two artificial grammars to generate the stimuli (c.f., Jamieson & Mewhort, 2005). These grammars, depicted in Table 2, specify the probability of a particular element occurring given the preceding element. For each stimulus sequence, the starting element (1-4) was randomly determined and then the listed probabilities were used to determine each subsequent element, until the desired length was reached. Grammar A was used to generate 16 unique sequences for the learning phase (6 of length 2 and 5 each of lengths 3 and 4) and 12 sequences for the test phase (4 each of lengths 3-5). Grammar B was used to generate 12 sequences for the test phase as well (4 each of lengths 3-5). All learning and test phase sequences are listed in Appendix A.

	Grammar A (n+1)				**Grammar B (n+1)**			
Colors/locations (n)	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1	0.0	0.5	0.5	0.0	0.0	0.0	0.0	1.0
2	0.0	0.0	1.0	0.0	0.5	0.0	0.0	0.5
3	0.5	0.0	0.0	0.5	0.0	1.0	0.0	0.0
4	1.0	0.0	0.0	0.0	0.0	0.5	0.5	0.0

Table 2. Grammars show transition probabilities from position n of a sequence to position n+1 of a sequence for four colors labeled 1-4.

For the auditory sentence perception task, we used a set of English lexically-controlled sentences developed by Eisenberg, Martinez, Holowecky, & Pogorelsky (2002). Sentences consisted of twenty lexically-easy (i.e., high word frequency, low neighborhood density) and twenty lexically-hard (i.e., low word frequency, high neighborhood density) sentences. All sentences are listed in Appendix B. Audio recordings of the sentences were obtained from Laurie Eisenberg, and are the same as in Eisenberg et al. (2002). For the NH children only, the original sentences were spectrally degraded to four spectral channels using a sine wave vocoder (www.tigerspeech.com), which roughly simulates the listening conditions of a cochlear implant (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Procedure

The deaf children with CI's were tested by a trained Speech Language Pathologist at the Devault Otologic Research Laboratory, Department of Otolaryngology, Indiana University School of Medicine, Indianapolis. The NH children were tested in a sound-attenuated booth in the Speech Research Laboratory at Indiana University, Bloomington. For both groups of children, the study consisted of 10 tasks in a session lasting 60-90 minutes, with breaks provided as needed. However, data from only four of the tasks are reported in the present manuscript (implicit learning, speech perception, digit spans, and vocabulary test). Before beginning the experiment, all NH children received and passed a brief pure-tone audiometric screening assessment in both ears. Both groups of children were also given a brief color screening, which consisted of presenting four blocks to the children, each of a different color (blue, green, red, yellow), and asking them to point to each and name the color. This was done to ensure that the children could perceive and name each of the four colors used in the implicit learning task. All children passed this screening. Following the screenings, all children were given the experimental tasks in the following order: implicit learning, sentence perception, forward and backward digit spans, and the vocabulary test.

In addition, for the deaf children with CIs, we included several standardized clinical assessments to serve as additional measures of language outcome. The Vineland Adaptive Behavior Scales (VABS) is a standardized assessment of personal and social skills used in everyday living, assessed through caregiver reports (Sparrow, Balla & Cicchetti, 1984). All parents filled out the three communication subdomains (receptive, expressive, and written language) of the VABS at the time of testing. As part of the children's regular visits to the Department of Otolaryngology, 18 of the 23 children were assessed on three core subtests of the Clinical Evaluation of Language Fundamentals, 4th Edition (CELF-4), an assessment tool for diagnosing language disorders in children (Semel, Wiig, & Secord, 2003. These three subtests measure aspects of general language ability: Concepts and Following Directions (C&FD), Formulated Sentences (FS), and Recalling Sentences (RS). A brief description of these three subtests is provided in Table 3 (see Paslawski, 2005, for a review and description of all subtests). Finally, also as part of the children's regular visits, 20 out of 23 children were assessed on the Goldman-Fristoe Test of Articulation (GFTA), which measures spontaneous and imitative sound production (Goldman & Fristoe, 2000). The scores we report are raw number of articulation errors.

Subtest	Description
C&FD	Measures auditory comprehension and recall of utterances of increasing length and complexity
FS	Assesses morphology and pronoun use
RS	The child is given a word or words and must generate spoken sentences in reference to a picture cue

Table 3. C&FD, (Concepts and Following Directions), FS, (Formulated Sentences), RS, (Recalling Sentences).

For the visual implicit learning task, participants were given the following instructions:

> You are going to see some squares of different colors on this computer screen. The squares will flash on the screen in a pattern. Your job is to try to remember the pattern of colors that you see on the screen. After each pattern, you will see all four colors on the screen. You need to touch the colors in the same pattern that you just saw. For example, if you saw the pattern red-green-blue, you would touch the red square, then the green square, and then the blue square. Touch where it says 'continue' on the bottom of the screen when you're finished. Always use your (preferred) hand to touch the screen and rest your arm on this gel pad.

Unbeknownst to participants, the task actually consisted of two parts, a Learning Phase and a Test Phase. The procedures for both phases were identical and in fact from the perspective of the subject, there was no indication of separate phases at all. The only difference between the two phases was which sequences were used. In the Learning Phase, the 16 learning sequences were presented in three blocks: the 6 length-2 sequences first, then the 5 length-3 sequences, and finally the 5 length-4 sequences; within each block, sequences were presented in random order. After completing the sequence reproduction task for all of the learning sequences, the experiment seamlessly transitioned to the Test Phase, which used the 12 novel Grammar A and 12 novel Grammar B test sequences. Test sequences were presented in three blocks: the 8 length-3 sequences first, the 8 length-4 sequences next, and finally the 8 length-5 sequences; within each block, sequences were presented in random order.

Sequence presentation consisted of colored squares appearing one at a time, in one of four possible positions on the touchscreen (upper left, upper right, lower left, lower right). The four elements (1-4) of each grammar were randomly mapped onto each of the four screen locations as well as four possible colors (red, blue, yellow, green). The assignment of stimulus element to position/color was randomly determined for each subject; however, for each subject, the mapping remained consistent across all trials.

After a colored square appeared for 700 msec, the screen was blank for 500 msec, and then the next element of the sequence appeared. After the entire sequence had been presented, there was a 500 msec delay and then the four panels appeared on the touch screen that were the same-sized and same-colored as the four locations that were used to display each sequence. The subject's task was to watch a sequence presentation and then to reproduce the sequence they saw by pressing the appropriate buttons in the correct order as dictated by the sequence. When they were finished with their response, they were instructed to press a "Continue" button at the bottom of the screen, and then the next sequence was presented after a 3-sec delay. A schematic of the implicit learning task is shown in Figure 1.

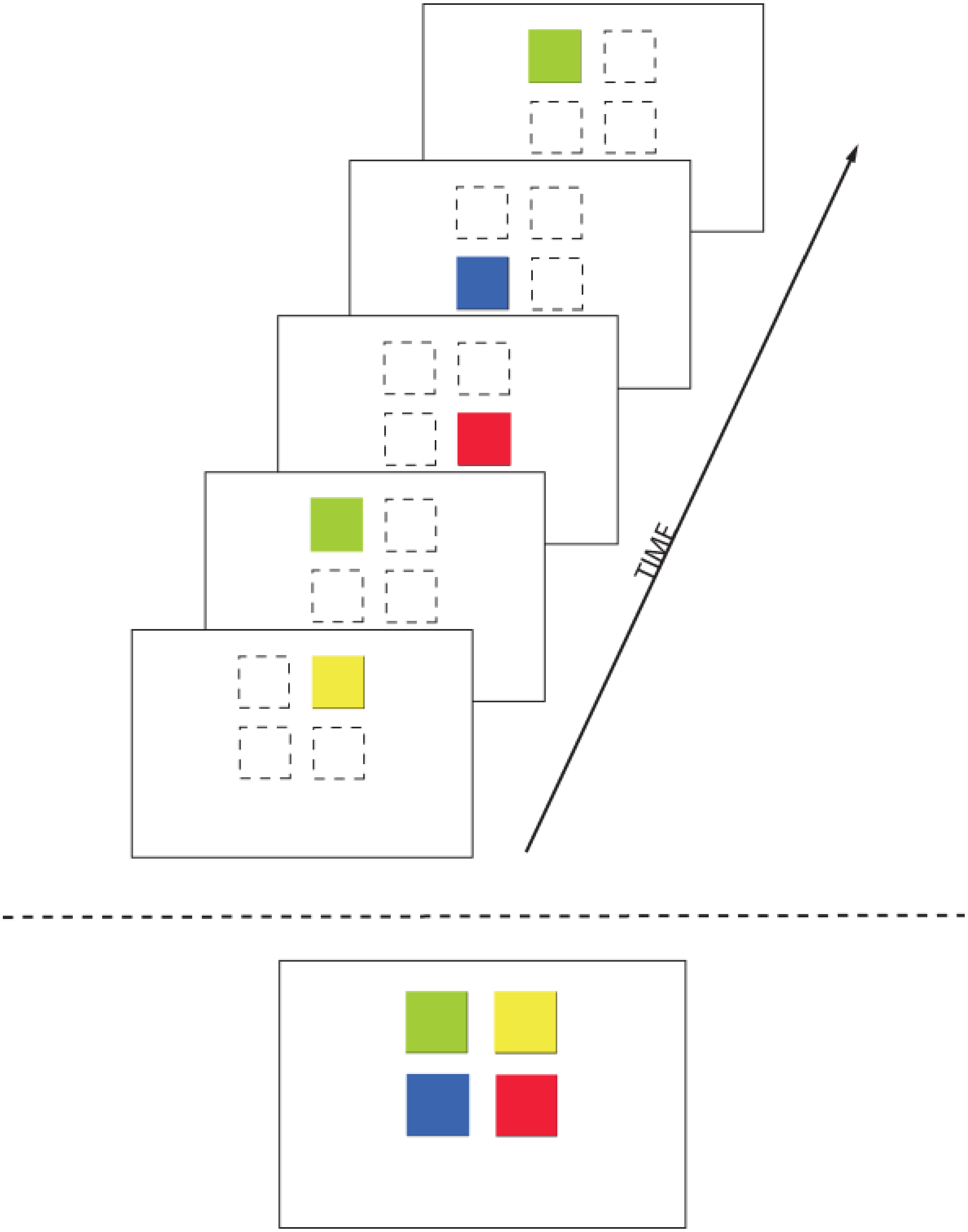

Figure 1. Depiction of the visual implicit learning task used in Experiments 1 and 2, similar to that used in previous work (Conway et al., 2007; Karpicke & Pisoni, 2004). Participants view a sequence of colored squares (700-msec duration, 500-msec ISI) appearing on the computer screen (top) and then, 500-msec after sequence presentation, they must attempt to reproduce the sequence by pressing the touch-panels in correct order (bottom). The next sequence occurs 3000-msec following their response.

Participants were not told that there was an underlying grammar for any of the learning or test sequences, nor that there were two types of sequences in the Test Phase. From the standpoint of the participant, the sequence task was solely one of observing and then reproducing a series of visual sequences. All test sequences (for both grammars) were scored off-line as either having been reproduced correctly in their entirety or not.

For the auditory sentence perception task, the 40 sentences described above were presented through a loudspeaker at 65 dB SPL. The children were instructed to listen closely to each sentence and then repeat back what they heard, even if they were only able to perceive one word of the sentence. Two practice sentences were presented before testing. Children were given feedback after they made their responses to the practice sentences, but received no feedback during testing. All 40 of the test sentences (20 'easy' and 20 'hard') were presented in random order to the children. The child's responses were recorded onto digital audio tape (DAT), and were later scored off-line based on number of keywords correctly repeated for each sentence. The sentences were played in the clear to the deaf children with CI's but were presented under perceptually-degraded conditions for the NH children.

For the forward and backward auditory digit spans, the procedure and materials followed the methods outlined in Wechsler (1991). In the forward digit span task, subjects were presented with lists of pre-recorded spoken digits with lengths (2-10) that became progressively longer. The subjects' task was to repeat each sequence aloud. In the backwards digit span task, subjects were also presented with lists of spoken digits with lengths that became progressively longer, but they were asked to repeat the sequence in reverse order. Digits were played through loudspeakers and the child's responses were recorded by a desk-mounted microphone. Subjects were scored on the longest sequence that they correctly recalled in each digit span task. Generally, the forward digit span task is thought to reflect the involvement of processes that maintain and store verbal items in short-term memory for a brief period of time, whereas the backward digit span task reflects the operation of controlled attention and higher-level executive processes that manipulate and process the verbal items held in immediate memory (Rosen & Engle, 1997).

The Peabody Picture Vocabulary Test (3rd edition) is a standard measure of vocabulary development (Dunn & Dunn, 1997). In this task, participants are shown four pictures on a single trial. They are prompted with a particular English word and then asked to pick the picture that most accurately depicts the word.

Results and Discussion

Data from any participant that scored more than two standard deviations from the mean on either the implicit learning or the sentence perception task was removed. This was done in order to reduce the undesirable effect that outliers might have on the correlation results. This resulted in data from only one NH participant being excluded from the final analyses, leaving a total of 26 NH children and 23 CI children included.

In the implicit learning task, a sequence was scored correct if the participant reproduced each test sequence correctly in its entirety. Span scores were calculated using a weighted method, in which the total number of correct test sequences at a given length was multiplied by the length, and then scores for all lengths added together (see Cleary, Pisoni, & Geers, 2001). We calculated separate span scores for Grammar A and Grammar B test sequences for each subject. Equations 1 and 2 show Grammar A (A_{span}) and Grammar B (B_{span}) span scores, respectively.

(1) $A_{span} = \sum(a_c * L)$

(2) $B_{span} = \sum(b_c * L)$

In Equation 1, a_c refers to the number of Grammar A test sequences correctly reproduced at a given length, and L refers to the length. For example, if a participant correctly reproduced 4 Grammar A test sequences of length 3, 3 of length 4, and 1 of length 5, then the A_{span} score would be computed as (4*3 + 3*4 + 1*5) = 29.

The B_{span} score is calculated in the same manner, using the number of Grammar B test sequences correctly reproduced at a given length, b_c.

For each subject we also calculated a learning score (Equation 3), which is the difference in span scores between the learned grammar (Grammar A) and the novel grammar (Grammar B). The LRN score measures the extent that sequence memory spans improved for sequences that had been previously experienced in the Learning Phase. Note that this score reflects how well memory spans improve for *novel* sequences that were constructed by the same grammar that they had been previously experienced in the Learning Phase, relative to span scores for sequences created by the new grammar.

(3) $LRN = A_{span} - B_{span}$

For the sentence perception task, a percent keyword correct score was calculated separately for easy and hard sentences. Each child received a forward and backward digit span score, reflecting the number of lists correctly repeated. Each child also received a standardized PPVT score based on how many pictures were correctly identified and their chronological age. A summary of the descriptive statistics for the CI and NH children are shown in Tables 4 and 5, respectively.

				Observed score range	
Measure	**n**	**M**	**SD**	**Minimum**	**Maximum**
ASpan	23	21.43	12.76	.00	48.00
BSpan	23	22.91	12.39	.00	44.00
LRN	23	-1.43	7.55	-13.00	12.00
LexEasy	23	.79	.20	.15	.97
LexHard	23	.77	.20	.22	.98
FWDigit	23	4.96	1.64	2.00	8.00
BWDigit	23	2.48	1.53	.00	5.00
PPVT	23	85.91	12.17	59.00	107.00
C/FD	18	6.33	4.06	1.00	13.00
FS	18	8.06	4.45	1.00	15.00
RS	18	5.50	3.70	1.00	13.00
Rec	23	15.00	1.20	11.00	18.00
Ex	23	14.43	3.00	10.00	20.00
Wr	23	14.83	2.99	10.00	20.00
GFTA	20	7.70	8.48	.00	23.00

Table 4. ASpan, Grammar A sequence span; BSpan, Grammar B sequence span; LRN, implicit learning score; LexEasy, % keywords correct for lexically easy sentences; LexHard, % keywords correct for lexically hard sentences; FWDigit, forward auditory digit span; BWDigit, backward auditory digit span; PPVT, Peabody Picture Vocabulary Test scaled score; C/FD, Concepts and Following Directions scaled score on the CELF-4 (mean score is 10, 1 S.D. is 3); FS, Formulated Sentences score on the CELF-4 (mean score is 10, 1 S.D. is 3); RS, Recalling Sentences score on the CELF-4 (mean score is 10, 1 S.D. is 3); Rec, Receptive language score on the VABS; Ex, Expressive language score on the VABS; Wr, Written language score on the VABS; GFTA, number of errors on the Goldman-Fristoe Test of Articulation.

			Observed score range	
Measure	**M**	**SD**	**Minimum**	**Maximum**
ASpan	26.96	10.06	7.00	48.00
BSpan	24.46	11.45	0.00	48.00
LRN	2.50	5.68	-12.00	13.00
LexEasy	0.42	0.16	0.10	0.75
LexHard	0.40	0.18	0.08	0.76
FWDigit	7.03	1.93	5.00	12.00
BWDigit	3.73	0.96	2.00	6.00
PPVT	114.30	12.66	90.00	139.00

Table 5. ASpan, Grammar A sequence span; BSpan, Grammar B sequence span; LRN, implicit learning score; LexEasy, % keywords correct for lexically easy sentences; LexHard, % keywords correct for lexically hard sentences; PPVT, Peaboy Picture Vocabulary Test scaled score; FWDigit, forward auditory digit span; BWDigit, backward auditory digit span.

Group Differences in Implicit Learning. Figure 2 shows the average implicit learning scores for both groups of children. For the NH children, the average implicit learning score (2.5) was significantly greater than 0, t(25)=2.24, p<.05, demonstrating that as a group the children showed better memory for test sequences with the same statistical/sequential structure as the ones from the Learning Phase. On the other hand, the average implicit learning score for the CI children was -1.43, a value that is statistically indistinguishable from 0, t(22)=-.91, p=.372. We also conducted a univariate ANOVA with the factor of group (NH, CI), which revealed a statistically significant effect of group on the implicit learning scores, F(1, 47) = 4.3, p < .05. Thus, on average, the NH group showed greater implicit learning than the CI group, who in turn essentially showed no learning on this task.

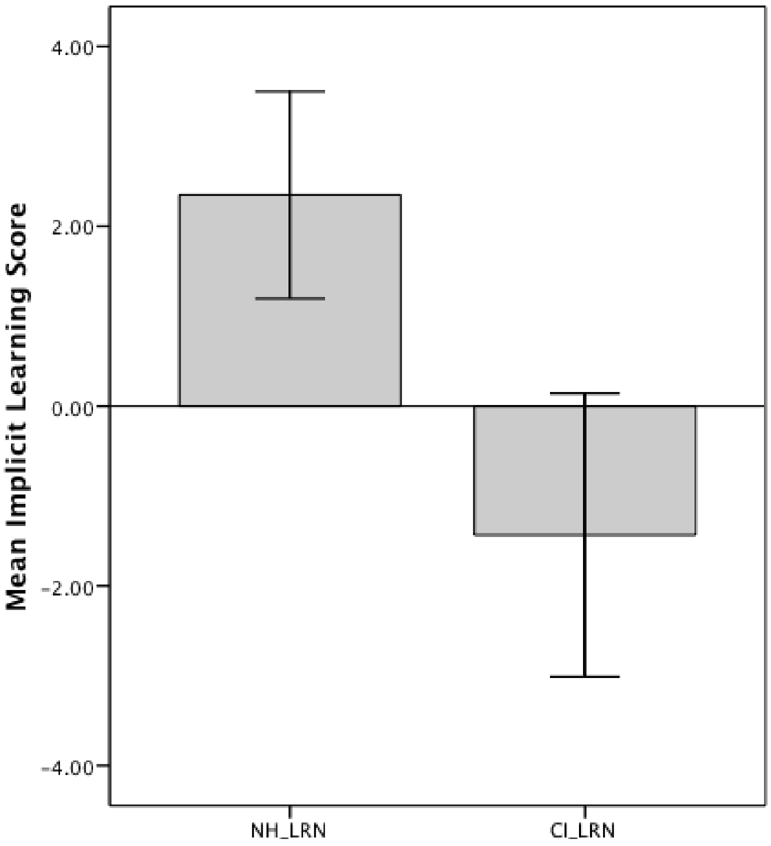

Figure 2. Average visual implicit learning scores for NH children (left) and deaf children with CIs (right). Error bars represent +/- 1 standard error.

In addition to examining group means, it is also important to compare the distribution of individual scores for each of the two groups of children on the implicit learning task. Figure 3 shows the implicit learning scores for each individual participant in the NH (top) and CI groups (bottom). Whereas 73.1% (19/26) of the NH children showed an implicit learning score of 0 or higher, only 39.1% (9/23) of the CI children showed such a score. Chi-square tests revealed that the proportion of learners to non-learners was significantly different for the NH children, $\underline{X}^2(1) = 5.54$, $\underline{p}<.05$, but not for the CI children, $\underline{X}^2(1) = 1.08$, $\underline{p}=0.297$. That is, more than half of the NH children showed a learning effect whereas this was not the case with the CI children.

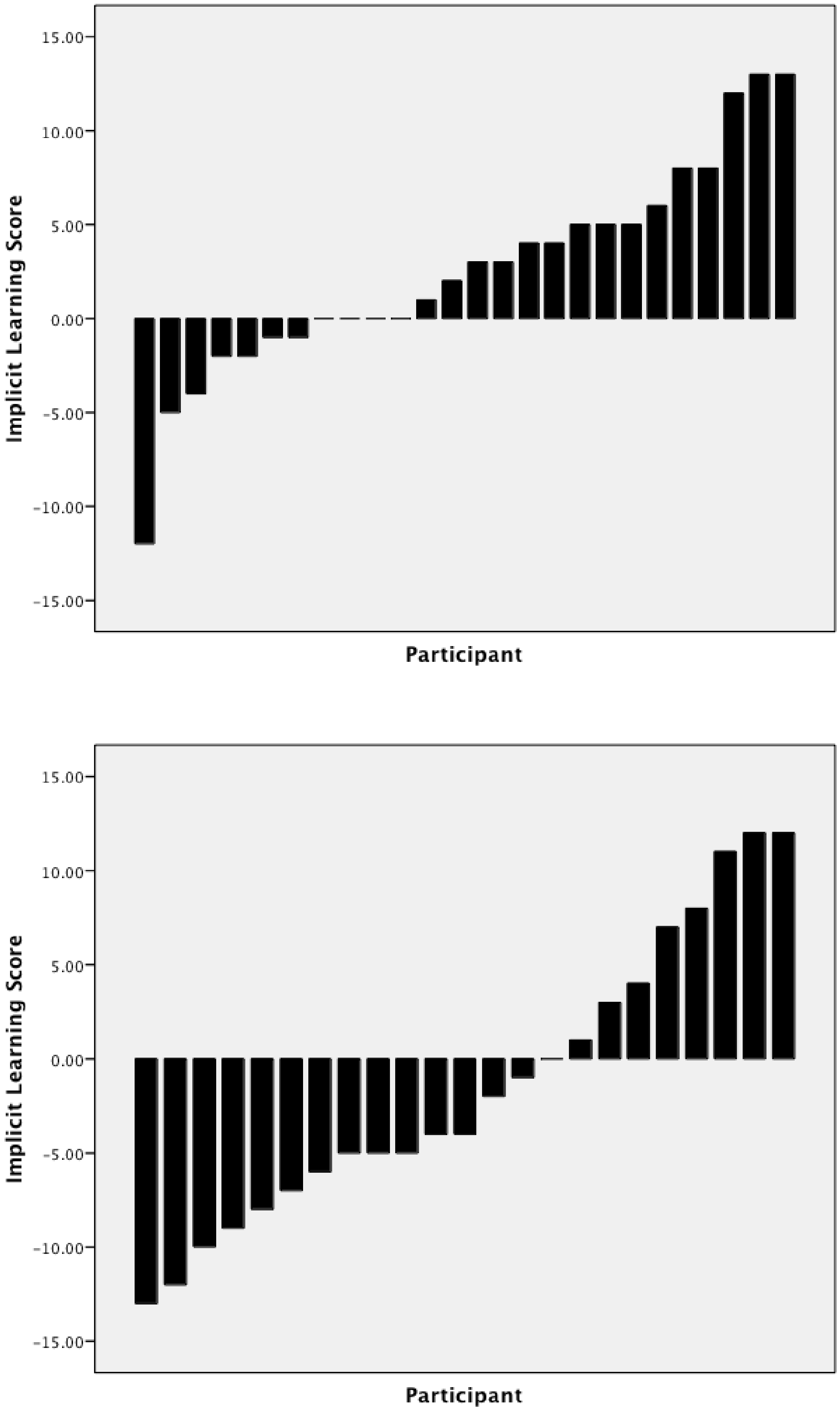

Figure 3. Implicit learning scores for each individual participant in the NH (top) and CI groups (bottom), ranked order from lowest to highest.

Consistent with the hypothesis that a period of deafness may cause secondary difficulties with sequencing skills, the present results suggest that deaf children with CI's show disturbances on visual implicit sequence learning. In addition, for the CI children, we computed a partial correlation between implicit learning and age at implantation, with chronological age partialed out. Implicit learning was negatively correlated with the age in which the child received their implant (r=-.292, p=.09, 1-tailed[5]) and positively correlated with the duration of implant use (r=.292, p=.09, 1-tailed). That is, the longer the child was deprived of auditory stimulation, the lower the visual implicit learning scores; correspondingly, the longer the child had experience with sound via the implant, the higher the implicit learning scores. These correlations, although non-significant, suggest that exposure to sound via a cochlear implant has secondary consequences not directly associated with hearing or language development per se; longer implant use is also associated with better ability to implicitly learn complex visual sequential patterns.

Implicit Learning and Sentence Perception. The question we next turn to is whether individual differences in implicit learning are associated with performance on the online sentence perception task. Performance on the sentence perception task assessed how well a child can perceive perceptually degraded words in sentences, which involves both bottom-up perceptual processes as well as top-down conceptual knowledge of language. We hypothesized that implicit learning is directly related to the use of top-down knowledge in speech perception and therefore we predicted a significant association between these two tasks. We calculated partial correlations between the implicit learning score with each of the two sentence perception scores, while controlling for the common variance associated with chronological age, forward digit span, backward digit span, PPVT, and for the CI children, age at implantation and articulation abilities.

The results are shown in Table 6 (Appendix C, Tables C-1 and C-2 display the full Pearson correlation matrix for all measures for the CI and NH children). For the NH children, implicit learning was positively and significantly correlated with performance on the lexically-hard sentences (+.36 < r's < +.42) and positively but non-significantly correlated with performance on the lexically-easy sentences (+.22 < r's < +.29). That is, children who showed the most learning on the implicit visual sequence learning task were the ones who scored best on the sentence perception task, especially for the sentences containing lexically hard words. This association does not appear to be mediated by forward or backward digit spans (measures of short-term and working memory, respectively) or vocabulary knowledge. The correlation with the implicit learning score was strongest for the lexically-hard sentences, containing words that have low frequencies and high neighborhood densities, suggesting that top-down linguistic knowledge (gained through implicit learning) is most heavily relied upon when processing degraded spoken sentences containing lexically-difficult words.

In contrast with the NH children, Table 6 reveals that for the CI children, implicit learning was not significantly correlated with the sentence perception task (r's < +.2). One possible reason may be that for the CI children, this task was not difficult enough. The average scores were quite high, with children correctly reproducing over 75% of all the target words, compared to roughly 40% for the NH children doing the task with spectrally degraded sentences. Rather than testing the deaf children's use of top-down knowledge to aid speech perception, this task instead likely tested their immediate serial recall abilities. As shown in Appendix C, Table C-2, the Pearson correlation between auditory digit spans and performance on the sentence perception task is very strong for both the lexically-easy (r=+.61, p<.01) and lexically-hard (r=+.60, p<.01) sentences. These correlations remain strong and significant even when

[5] One-tailed tests were used here and elsewhere (as noted) in instances where we hypothesized the specific direction in which a correlation would go (i.e., that implicit learning would be negatively correlated with age at implantation, and that it would be positively correlated with measures of language outcome).

controlling for chronological age (r’s>+.5, p’s<.01), age plus implicit learning (r’s>+.5, p’s<.05), and age plus vocabulary knowledge (r’s>+.45, p’s<.05).

	CI Children		NH Children	
Controlling for	**r (LexEasy)**	**r (LexHard)**	**r (LexEasy)**	**r (LexHard)**
CA	.188	.194	.286	.415*
CA + FWdigit	.033	.041	.251	.382*
CA + BWdigit	.075	.060	.324	.458*
CA + PPVT	.130	.126	.226	.362*
CA + AgeImp	.202	.181	---	---
CA + GFTA	.178	.186	---	---

Table 6. *p<.05, 1-tailed. LexEasy, lexically-easy sentences; LexHard, lexically-hard sentences; CA, chronological age; FWdigit, forward digit span; BWdigit, backward digit span; PPVT, Peabody Picture Vocabulary Test; AgeImp, age at cochlear implantation; GFTA, number of errors on the Goldman-Fristoe Test of Articulation.

The Use of Context to Aid Speech Perception. In order to obtain a more direct measure of the use of top-down knowledge to aid speech perception – which ought to tap into implicit learning abilities – we computed an additional score on the speech perception task for the CI children. A difference score was computed by subtracting the children’s performance on the first target word from their performance on the final target word, assessing how well the child used sentence context to guide speech perception (c.f., Bilger & Rabinowitz, 1979). Because the final target word is much more highly constrained by semantic and syntactic context compared to the first target word, a positive difference score reflects the beneficial effect of sentence context on speech perception.

We computed three difference scores, for lexically easy sentences (M=.043; S.D.=1.97), lexically-hard sentences (M=-.088; S.D=2.07), and combined lexically-easy and lexically-hard sentences (M=-.022; S,D,=1.27). As Figure 4 shows, deaf children’s implicit learning scores are associated with their ability to effectively use sentence context to guide speech perception, as reflected by the sentence perception difference score for combined lexically-easy and –hard sentences (r=+.36, p<.05, 1-tailed). Table 7 shows the partial correlations between implicit learning and all three difference scores, while controlling for the common sources of variance associated with chronological age, forward and backward digit spans, and vocabulary knowledge (PPVT). Implicit learning was significantly and positively correlated with the sentence perception difference scores, specifically for the lexically-hard sentences. These results suggest that better implicit learning abilities result in more robust knowledge of the sequential predictability of words in sentences, which leads to a better use of sentence context to aid speech perception, as reflected in the sentence perception difference score.

Controlling	LexEasy Diff Score	LexHard Diff Score	Both Diff Score
CA	.087	.375*	.371*
CA + FWdigit .	.030	.502**	.414*
CA + BWdigit .	-.050	.367*	.277
CA + PPVT	.079	.389*	.377*

Table 7. *p<.05, **p<.01, ***p<.001, 1-tailed. LexEasy, lexically-easy sentences; LexHard, lexically-hard sentences; CA, chronological age; AgeImp, age at cochlear implantation; FWdigit, forward digit span; BWdigit, backward digit span; PPVT, Peabody Picture Vocabulary Test.

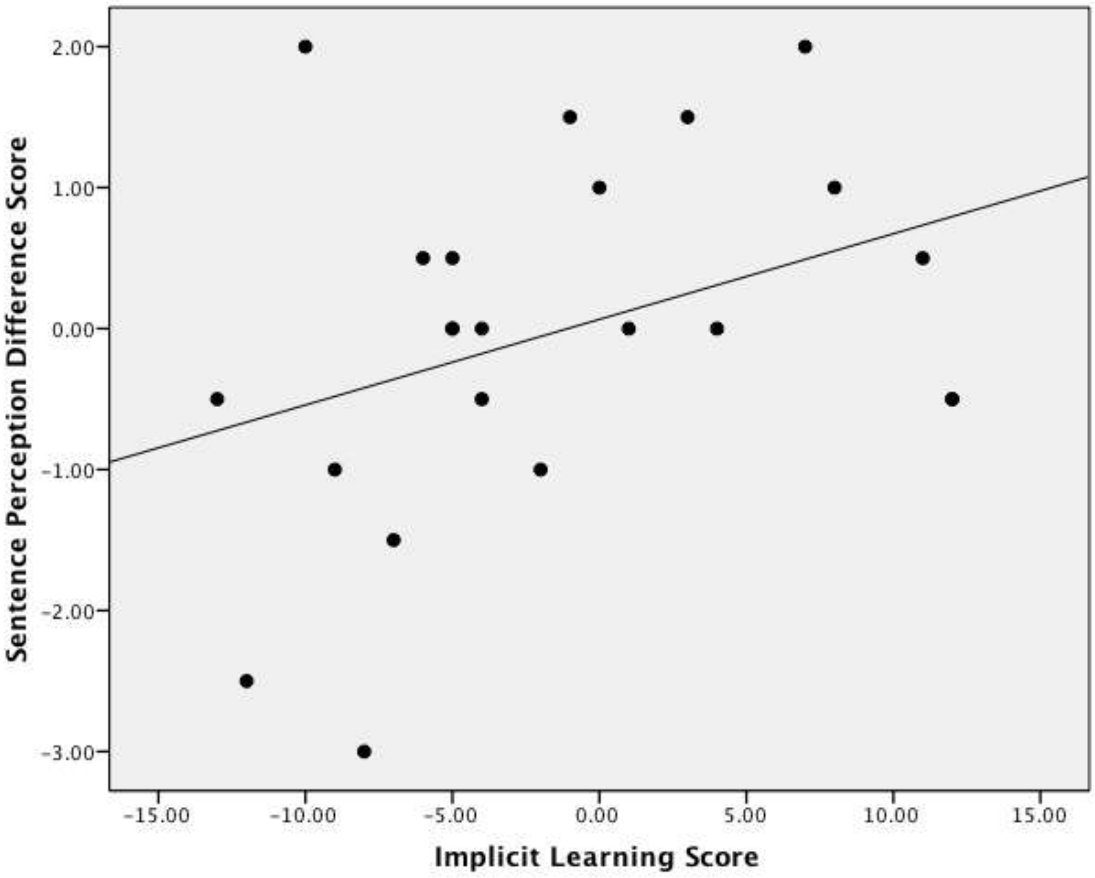

Figure 4. Scatter plot of implicit learning and the sentence perception difference score (for combined lexically-easy and –hard sentences) for deaf children with CIs.

These results suggest that for the CI children, implicit learning is used to learn about word predictability in language, knowledge of which can be brought to bear under conditions where sentence context can be used to help perceive the next word in an utterance. If this is the case, then we would expect that the CI children, who scored worse as a group on implicit learning, will also be impaired on their ability to make use of the preceding context of a sentence to help them perceive the next word. In support of this novel prediction is Figure 5, which shows the performance of correctly identifying the three target words in the sentence perception task, as a function of the position in the sentence (1st, 2nd, or 3rd), for the NH and CI children. Sentence context can do very little to aid perception of the first target word in the sentence; however, context can be very useful to help perceive the second and third target words, but only if one has sufficient top-down knowledge of word order regularities. Indeed, the NH children show a gradual improvement in speech perception for the second and third target words. Their performance on the last word is statistically greater than performance on the first word, t(25)=4.2, p<.001. In contrast, the CI children show no such contextual facilitation. Their performance on the third word is no different than their performance on the first word in each sentence, t(22)=.106, p=.92. Unlike the NH children, the deaf children with CIs do not appear to be using sentence context to help them perceive the final word in the sentence. Thus, one way in which poor implicit learning abilities may reveal themselves are in situations in which word predictability and sentence context can be used to guide spoken language processing.

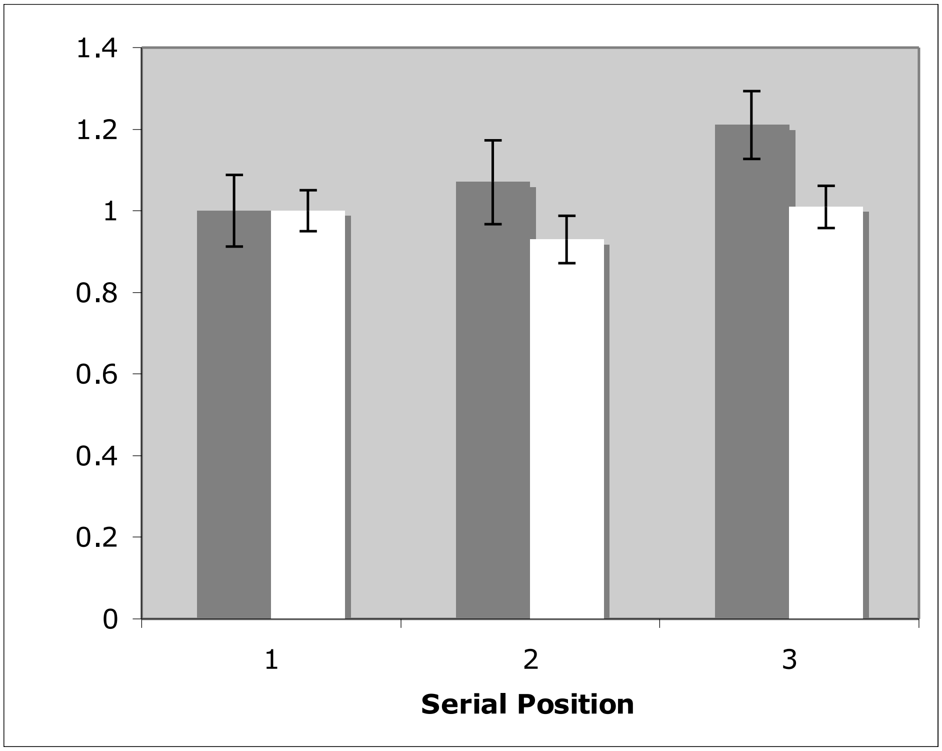

Figure 5. Speech perception accuracy and S.E. bars for NH (grey) and CI children (white), as a function of serial position (word 1, word 2, and word 3). The y-axis is a unitless measure, computed by dividing the score at each word position by the score at word 1, for the NH and CI groups separately.

Implicit Learning and Language Outcome in Deaf Children with CIs. Table 8 shows partial correlations between implicit learning and several standardized clinical outcome measures (the three scaled measures of language as measured by the CELF-4 and the three scaled measures of communication as assessed through parental responses to the VABS). Implicit learning was positively and significantly correlated with all subtests of the CELF-4: Concepts and Following Directions (+.36 < r's < +.57), Formulated Sentences (+.52 < r's < +.90), and Recalling Sentences (+.43 < r's < +.62). All three of the CELF-4 subtests involve understanding and/or producing sentences of varying complexity, tasks in which knowledge of word order predictability – i.e., statistics of sequential probabilities in language – can be brought to bear to improve performance. Implicit learning was also positively and significantly correlated with receptive language on the VABS (+.37 < r's < +.48), although the correlation was non-significant when the common variance associated with backward digit span, vocabulary knowledge, and articulation was controlled. Neither the expressive nor the written language scores on the VABS were significantly correlated with implicit learning.

Controlling	**C/FD**	**FS**	**RS**	**Rec**	**Exp**	**Wr**
CA	.494*	.616**	.514*	.379*	.252	.161
CA + AgeImp	.480*	.570*	.464*	..475*	.199	.200
CA + FWdigit	.428*	.569*	.447*	.371*	.162	.059
CA + BWdigit	.368	.527*	.434*	.288	.219	.071
CA + PPVT	.510*	.725***	.602**	.344	.185	.097
CA + GFTA	.567*	.895***	.616**	..375	.303	.150

Table 8. *p<.05, **p<.01, ***p<.001, 1-tailed. LexE, lexically-easy sentences; LexH, lexically-hard sentences; C/FD, Concepts and Following Directions score on the CELF-4; FS, Formulated Sentences score on the CELF-4; RS, Recalling Sentences score on the CELF-4; Rec, Receptive language score on the VABS; Ex, Expressive language score on the VABS; Wr, Written language score on the VABS.; CA, chronological age; AgeImp, age at implantation; FWdigit, forward digit span; BWdigit, backward digit span; PPVT, Peabody Picture Vocabulary Test; GFTA, Goldman-Fristoe Test of Articulation.

In summary, implicit learning abilities were found to be associated with several aspects of language outcome in deaf children who have received a cochlear implant and in NH children. The pattern of correlations suggests that implicit learning may be most strongly related to the ability to use knowledge of the sequential structure of language to better process, understand, and produce meaningful sentences, especially when sentence context can be brought to bear to aid processing. Importantly, this association does not appear to be mediated by chronological age, age of implantation, short-term or working memory, vocabulary knowledge, or the child's ability to adequately articulate speech. Moreover, these findings were found to be modality-independent. The implicit sequence learning task incorporated only visual patterns whereas the sentence perception task relied on an auditory-only presentation.

General Discussion

The goals of this paper were to determine 1) whether deaf children with CIs are impaired on their ability to learn complex visual sequential patterns; and 2) whether such a disturbance can account for the enormous range of variability in language outcome following cochlear implantation. The results showed that as a group, the CI children performed significantly worse on the visual sequence learning task compared to the age-matched group of NH children. Indeed, less than half of the deaf children with CIs showed learning on the task, compared to the roughly 75% of the hearing children who did display learning. Furthermore, in both the CI and NH children, implicit sequence learning was found to be significantly correlated with several measures of spoken language processing and development. We discuss both of these findings in turn.

The group differences in visual sequence learning are consistent with the hypothesis that a period of auditory deprivation may have major secondary effects on brain and cognition that are not specific to hearing or the processing of sound by the auditory modality. Sound is unique among sensory input in several important ways. Compared to vision and touch, sound appears to be more attention-demanding (Posner, Nissen, & Klein, 1976), especially early in development (Robinson & Sloutsky, 2004). Sound is also intrinsically a temporal and sequential signal, one in which time and serial order are of primary importance (Hirsh, 1967). Indeed, previous work in healthy adults suggests that auditory processing of time and serial order is superior to the other senses. Auditory advantages have been found in tasks involving temporal processing (Sherrick & Cholewiak, 1986), rhythm perception (Kolers & Brewster, 1985; Repp & Penel, 2002), immediate serial recall (Glenberg & Swanson, 1986; Penney, 1989), sequential pattern perception (Handel & Buffardi, 1969), and implicit sequence learning (Conway & Christiansen, 2005; Conway & Christiansen, in press). These findings suggest an intimate link between auditory cognition and the processing of temporal and sequential relations in the environment. In addition, previous work has suggested that the profoundly deaf (including those with and without CIs) show disturbances in (non-auditory) functions related to time and serial order, including: rhythm perception (Rileigh & Odom, 1972); attention to serially-presented stimuli (Horn, Davis, Pisoni, & Miyamoto, 2005; Knutson et al., 1991; Quittner, Smith, Osberger, Mitchell, & Katz, 1994); immediate serial recall (Marschark, 2006; Pisoni & Cleary, 2004; Todman & Seedhouse, 1994); motor sequencing (Horn, Pisoni, & Miyamoto, 2006); and aspects of executive function and cognitive control (Anaya, Conway, Pisoni, Geers, & Kronenberger, 2008; Hauser & Lukomski, in press; Pisoni et al., 2008). Furthermore, the introduction of sound via a cochlear implant appears to progressively improve certain sequencing abilities over time (Horn et al., 2005).

It is possible that experience with sound and auditory patterns, which are complex, serially-arrayed signals, provides a child vital experience with perceiving and learning sequential patterns. A period of deafness early in development deprives a child with the essential experience of dealing with complex sequential auditory input, which, it would appear, affects their ability to deal with sequential patterns in other sense modalities as well. Once hearing is introduced via the CI, a child begins for the first time to gain experience with auditory sequential input. The positive (though not statistically significant) correlation between length of CI use and implicit learning scores which we found – obtained even when chronological age was partialed out -- shows that experience with sound via a CI improves one’s ability to learn complex non-auditory sequential patterns. Thus, it is possible that given enough exposure to sound via a CI, a deaf child’s implicit learning abilities will eventually improve to age-appropriate levels.

To explain these findings, we suggest that sound impacts cognitive development by providing a perceptual and cognitive “scaffolding” of time and serial order, upon which sequencing functions are based. From a neurobiological standpoint, it is known that lack of auditory stimulation results in a decrease of myelination and fewer projections out of auditory cortex (Emmorey, Allen, Bruss, Schenker, & Damasio, 2003) – which may also include connectivity to the frontal lobe. The frontal lobe, and specifically the prefrontal cortex, is believed to play an essential role in learning, planning, and executing sequences of thoughts and actions (Fuster, 2001; 1995). It is therefore possible that the lack of auditory input early on in development, and corresponding reduction of auditory-frontal activity, fundamentally alters the neural organization of the frontal lobe and connections to other brain circuits (Wolff & Thatcher, 1990), impacting the development of sequencing functions regardless of input modality.

The second primary finding revealed associations between implicit sequence learning and spoken language development. Based on previous work with healthy adults (Conway et al., submitted; Conway et al., 2007), we hypothesized that implicit visual sequence learning abilities would be associated with spoken language processing and language outcomes in NH, typically-developing children as well as deaf children with CIs. In support of this hypothesis, we found that the NH children's implicit learning scores were positively and significantly correlated with their ability to perceive lexically-difficult sentences under degraded listening conditions. The group of deaf children with CIs also demonstrated a positive and statistically significant correlation between their learning scores and their ability to use sentence context to perceive the words in sentences, as well as with several clinical measures of language outcome. In both groups of children, these correlations remained significant even after partialing out the effects of chronological age, auditory digit spans, general vocabulary knowledge, and (in the case of the deaf children with CIs), age at implantation and articulatory abilities. These findings suggest a close coupling between the development of general (non-auditory) implicit sequence learning skills and the development of spoken language under both typical and atypical (language-delayed and auditory deprivation) conditions.

The current results demonstrating associations between implicit learning and language development extend recent findings (Conway et al., 2007; Conway & Pisoni, 2007: Conway et al., submitted) showing significant correlations between implicit learning and spoken language processing in healthy adults using experimental tasks very similar to those used here. The present developmental results suggest that general implicit learning abilities contribute to a child's ability to perceive and comprehend spoken language and are consistent with the hypothesis that implicit statistical learning is an important and foundational ability that underlies language acquisition and development. We should note in passing here that our present results cannot prove causality, but have rather established that there exists a close coupling between implicit learning and language development. There are at least two alternative explanations: implicit learning and spoken language processing may develop on a similar timescale but are independent of one another; or, differences in aspects of spoken language development may affect implicit sequence learning abilities, rather than the other way around. To help tease apart these alternative explanations, a longitudinal design could help determine if implicit learning abilities predict subsequent speech and language abilities assessed several years later (see Baddeley, Gathercole, & Papagno, 1998; Bernhardt, Kemp, & Werker, 2007; Gathercole & Baddeley, 1989; Newman, Bernstein Ratner, Jusczyk, Jusczyk, & Dow, 2006; Tsao, Liu, & Kuhl, 2004). Such a finding would provide converging support for the hypothesis that implicit learning plays a causal role in language development.

At a mechanistic level, what is the relation between implicit learning and language development? We suggest that implicit learning abilities allow the language learner to acquire knowledge about the predictability of the occurrence of patterns of sound sequences (i.e., words). Learning the predictability of words in spoken language is arguably crucial to understanding and representing both syntax and semantics (Kalikow et al., 1977; Rubenstein, 1973). A fine-grained long-term representation of the likelihood that any given word will be spoken next in an utterance can be used to help perceive speech under degraded listening conditions. Everyday speech communication is characterized by the use of context-based redundancy to facilitate real-time comprehension. Using top-down knowledge in this way is likely to be even more important when the words in the sentences themselves are infrequent and/or harder to perceive due to a large number of competing neighbors (i.e., lexically-hard sentences). Being able to use sentence context to constrain the number of possible alternatives to the ending of a sentence – and thus better perceive the ending – requires prior knowledge of word order statistics in language, which we argue is gained through implicit learning of the sequential auditory patterns in speech.

Aside from their theoretical importance, from a clinical standpoint, the findings with the CI children are important because they suggest that individual differences in basic implicit learning abilities may provide a principled explanation for why some deaf children with CIs achieve near-typical levels of speech and language outcomes whereas other children do not. Several recent studies have been devoted to understanding the nature of the enormous variation in language outcome in deaf children who receive a CI (e.g., Dawson, Busby, McKay, & Clark, 2002; Horn et al., 2005; Horn et al., 2006; Knutson, 2006; Pisoni & Cleary, 2004). The current results are clinically important because they may provide both the clinical prediction of audiological benefit from a CI and the formulation of new intervention programs that specifically target the development of implicit sequence learning skills in deaf children who are doing poorly with their CIs. In particular, interventions focused on the training of cognitive sequencing skills, executive functions, and cognitive control (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, in press; Klingberg et al., 2005) may provide benefits above and beyond the standard audiological-based treatment strategies.

In sum, we have shown a direct empirical link between implicit visual sequence learning and the development of spoken language in both typically-developing (normal-hearing) and atypically-developing children (deaf children with CIs). These findings suggest that basic cognitive learning abilities related to encoding sequential structure may be an important foundational aspect of language development, independent of the effects of differences in immediate serial recall (Baddeley et al., 1998). Our results also suggest that a period of auditory deprivation early in development may negatively impact implicit visual sequence learning abilities, which has profound implications for understanding variation in neurocognitive development and plasticity in both normal-hearing and deaf populations. Finally, the present findings suggest that it may be useful to investigate differences in implicit sequence learning as a contributing factor in other populations that show language delays or communication disorders, such as children with specific language impairment or autism.

References

Altmann, G. T. M. (2002). Statistical learning in infants. Proceedings of the National Academy of Sciences, 99, 15250-15251.

American Speech-Language-Hearing Association (2004). Cochlear implants [Technical report].Available from www.asha.org/policy.

Anaya, E.M., Conway, C.M., Pisoni, D.B., Geers, A., & Kronenberger, W. (2008). Effects of cochlear implantation on executive function: Some preliminary findings. Poster presented at the 10th International Conference on Cochlear Implants and other Implantable Auditory Technologies. San Diego, CA, April.

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

Bernhardt, B.M., Kemp, N., & Werker, J.F. (2007). Early word-object associations and later language development. First Language, 27, 315-328.

Bilger, R.C. & Rabinowitz, W.M. (1979). Relationships between high- and low-probability SPIN scores. The Journal of the Acoustical Society of America, 65(S1), S99.

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

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. Trends in Cognitive Sciences, 2, 406-416.

Conway, C.M., Bauernschmidt, A., Huang, S.S., & Pisoni, D.B. (submitted). Implicit statistical learning in language processing: Word predictability is the key. Cognition.

Conway, C.M. & Christiansen, M.H. (in press). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. European Journal of Cognitive Psychology.

Conway, C.M. & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. Journal of Experimental Psychology, 31, 24-39.

Conway, C.M., Karpicke, J., & Pisoni, D.B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. Journal of Deaf Studies and Deaf Education, 12, 317-334.

Conway, C.M. & Pisoni, D.B. (in press). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. Annals of the New York Academy of Sciences.

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, 3rd edition. Circle Pines, MN: American Guidance Service.

Eisenberg, L.S., Martinez, A.S., Holowecky, S.R., & Pogorelsky, S. (2002). Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. Ear and Hearing, 23(5), 450-462

Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.

Emmorey, K., Allen, J.S., Bruss, J., Schenker, N., & Damasio, H. (2003). A morphometric analysis of auditory brain regions in congentially deaf adults. Proceedings of the National Academy of Sciences, 100, 10049-10054.

Fuster, J. (2001). The prefrontal cortex-- an update: Time is of the essence. Neuron, 30, 319-333.

Fuster, J. (1995). Temporal processing. In J. Grafman, K.J. Holyoak, & F. Boller (Eds.), Structure and functions of the human prefrontal cortex (pp. 173-181). New York: New York Academy of Sciences.

Gathercole, S. E. and A. D. Baddeley (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.

Glenberg, A.M. & Swanson, N.G. (1986). A temporal distinctiveness theory of recency and modality effects. Journal of Experimental Psychology: Learning, Memory, & Cognition, 12, 3-15.

Goldman, R. & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation – 2nd Edition. Circle Pines, MN: American Guidance Service, Inc.

Handel, S., & Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. Quarterly Journal of Experimental Psychology, 21, 256-266.

Hauser, P.C. & Lukomski, J. (in press). Development of deaf and hard of hearing students' executive function. In M. Marschark & P. Hauser (Eds.), Deaf cognition: Foundations and outcomes. New York: Oxford University Press.

Hauser, M. D., Newport, E. L. & Aslin, R.N. (2000). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins, Cognition, 75,1-12.

Hirsh, I.J. (1967). Information processing in input channels for speech and language: The significance of serial order of stimuli. In F.L. Darley (Ed.), Brain mechanisms underlying speech and language (pp. 21-38). New York: Grune & Stratton.

Horn, D.L., Davis, R.A.O., Pisoni, D.B., & Miyamoto, R.T. (2005). Development of visual attention skills in prelingually deaf children who use cochlear implants. Ear & Hearing, 26, 389-408.

Horn, D.L., Pisoni, D.B., & Miyamoto, R.T. (2006). Divergence of fine and gross motor skills in prelingually deaf children: Implications for cochlear implantation. Laryngoscope, 116, 1500-1506.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., & Perrig, W.J. (in press). Improving fluid intelligence with training on working memory. Proceedings of the National Academy of Sciences.

Jamieson, R.K. & Mewhort, D.J.K. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 31, 9-23.

Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America, 61, 1337-1351.

Karpicke, J. D. and D. B. Pisoni (2004). Using immediate memory span to measure implicit learning. Memory & Cognition, 32(6), 956-964.

Kirk, K.I., Miyamoto, R.T., Lento, C.L., Ying, E., O'Neil, T., & Fears, B. (2002). Effects of age at implantation in young children. Annals of Otology, Rhinology, & Laryngology, 189, 69-73.

Klingberg, T., Fernell, E., Olesen, P.J., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C.G., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD – A randomized, controlled trial. Journal of the American Academy of Child & Adolescent Psychiatry, 44, 177-186.

Knutson, J.F. (2006). Psychological aspects of cochlear implantation. In H. Cooper & L. Craddock (Eds.), Cochlear implants: A practical guide, 2nd Ed. (pp. 151-178). London: Whurr Publishers.

Knutson, J.F., Hinrichs, J.V., Tyler, R.S., Gantz, B.J., Schartz, H.A., & Woodworth, G. (1991). Psychological predictors of audiological outcomes of multichannel cochlear implants. Annals of Otology, Rhinology, & Laryngology, 100, 817-822.

Kolers, P.A. & Brewster, J.M. (1985). Rhythms and responses. Journal of Experimental Psychology, 11, 150-167.

Lashley, K.S. (1951). The problem of serial order in behavior. In L.A. Jeffress (Ed.), Cerebral mechanisms in behavior (pp. 112-146). New York: Wiley.

Luria, A.R. (1973). The working brain: An introduction to neuropsychology. New York: Basic Books.

Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. European Journal of Cognitive Psychology, 18, 70-89.

Meulemans, T. & Van der Linden, M. (1998). Implicit sequence learning in children. Journal of Experimental Child Psychology, 69, 199-221.

Newman, R., Bernstein Ratner, N., Jusczyk, A.M., Jusczyk, P.W., & Dow, K.A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. Developmental Psychology, 42, 643-655.

Paslawski, T. (2005). The Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4): A review. Canadian Journal of School Psychology, 20, 129-134.

Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. Memory & Cognition, 17, 398-422.

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).

Pisoni, D.B., Conway, C.M., Kronenberger, W.G., Horn, D.L., Karpicke, J., & Henning, S. (2008). Efficacy and effectiveness of cochlear implants in deaf children. In M. Marschark & P. Hauser (Eds.), Deaf cognition: Foundations and outcomes, (pp. 52-101). New York: Oxford University Press.

Posner, M.I., Nissen, M.J., & Klein, R.M. (1976). Visual dominance: An information-processing accounts of its origins and significance. Psychological Review, 83, 157-171.

Quittner, A.L., Smith, L.B., Osberger, M.J., Mitchell, T.V., & Katz, D.B. (1994). The impact of audition on the development of visual attention. Psychological Science, 5, 347-353.

Repp, B.H. & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. Human Perception & Performance, 28, 1085-1099.

Rileigh, K.K. & Odom, P.B. (1972). Perception of rhythm by subjects with normal and deficient hearing. Developmental Psychology, 7, 54-61.

Robinson, C.W. & Sloutsky, V.M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387-1401.

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

Rubenstein, H. (1973). Language and probability. In G.A. Miller (Ed.), Communication, language, and meaning: Psychological perspectives (pp. 185-195). New York: Basic Books, Inc.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.

Saffran, J.R., Senghas, A., & Trueswell, J.C. (2001). The acquisition of language by children. Proceedings of the National Academy of Sciences, 98, 12874-12875.

Semel, E., Wiig, E.H., & Secord, W.A. (2003). Clinical evaluation of language fundamentals, fourth edition (CELF-4). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science, 270, 303-304.

Sherrick, C.E. & Cholewiak, R.W. (1986). Cutaneous sensitivity. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), Handbook of perception and human performance, Vol. I: Sensory processes and perception (pp. 1-12). New York: Wiley.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). Vineland Adaptive Behavioral Scales. Circle Pines, MN: American Guidance Service.

Todman, J. & Seedhouse, E. (1994). Visual-action code processing by deaf and hearing children. Language & Cognitive Processes, 9, 129-141.

Tomblin, J.B., Barker, B.A., & Hubbs, S. (2007). Developmental constraints on language development in children with cochlear implants. International Journal of Audiology, 46, 512-523.

Tsao, F.-M., Liu, H.-M., & Kuhl, P.K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. Child Development, 75, 1067-1084.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. Cognition, 92, 231-270.

Wechsler, D. (1991). Wechsler intelligence scale for children – Third edition. San Antonio, TX: The Psychological Corporation.

Wolff, A.B. & Thatcher, R.W. (1990). Cortical reorganization in deaf children. Journal of Clinical and Experimental Neuropsychology, 12, 209-221.

Appendix A

Learning and Test Sequences used in the Implicit Learning Task

Sequence Length	Learning Sequence	Test Sequence (A)	Test Sequence (B)
2	4-1		
2	3-1		
2	1-3		
2	2-3		
2	1-2		
2	3-4		
3	4-1-3	2-3-4	3-2-1
3	2-3-1	1-3-1	2-4-2
3	1-2-3	4-1-2	4-2-4
3	1-3-4	3-1-3	2-4-3
3	3-4-1		
4	1-2-3-4	1-3-1-3	3-2-4-2
4	3-1-2-3	3-4-1-2	1-4-2-4
4	1-2-3-1	4-1-2-3	4-2-1-4
4	4-1-3-1	3-1-3-4	2-1-4-3
4	2-3-1-3		
5		1-2-3-1-2	4-3-2-1-4
5		4-1-3-4-1	1-4-3-2-4
5		3-1-2-3-1	3-2-1-4-2
5		1-2-3-4-1	4-2-4-2-4

Appendix B

Sentences used in the Sentence Perception Task (from Eisenberg et al., 2002)

Lexically Easy	**Lexically Hard**
That kind of airplane is brown.	Tell him to sleep on his belly.
You can’t stand on your broken truck.	The bunny hid in my room.
The children cried at the farm.	She likes to share the butter.
I broke my finger at school.	His son played with the chickens.
My friend thinks her lipstick is cool.	Call if you ever find the toys.
Give the monkey some juice.	Grampa laughed at the goats.
I can wash the ducks myself.	Dad came to say hello.
I can draw a little snake.	The boys took turns locking the car.
Open the green one first.	Many kids can learn to sing.
The string can stay in my pocket.	She lost her mommy’s ring.
Please help her with the puzzle.	She knows where to leave the money.
Don’t scribble on the door.	The piggy moved the books.
I saw seven eggs in the street.	The gum is in the tiny box.
I just found the grey shoelace.	His tummy hurt for ten days.
I wonder who brought the food.	Start walking to your seat.
I know which space is black.	He taught us that funny trick.
It’s always fun to watch the fish.	The worm was stuck in the pool.
Let’s buy gas from that man.	I guess you were in the rain.
I hope the girl takes some milk.	The cups are in the pink bag.
The chair could break when I jump.	Both of the naughty cats are mine.

Appendix C

Full Correlation Matrices

Correlation Matrix for Deaf Children with CIs

Measure		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. CA	--	.268	.911***	.627***	.596**	.086	.317	.279	-.288	.340	
2. AgeCI	--	--	-.153	.139	.306	-.258	.099	.005	-.255	-.124	
3. CIUse		--	--	--	.584**	.480*	.198	.283	.284	-.187	.401
4. ASpan	--	--	--	--	.818***	.353	.222	.129	-.302	.262	
5. BSpan	--	--	--	--	--	-.249	.102	.002	-.404	.085	
6. LRN	**--**	--	--	--	--	--	--	.205	.210	.143	.300
7. LexEasy	--	--	--	--	--	--	--	.954***	.278	.608**	
8. LexHard	--	--	--	--	--	--	--	--	.369	.599**	
9. PPVT	**--**	--	--	--	--	--	--	--	--	--	.226
10. FWDigit	--	--	--	--	--	--	--	--	--	--	

Table C-1.

Measure	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1. CA	.596**	-.297	-.325	-.204	.080	.075	-.214	-.325
2. AgeCI	.136	-.200	-.448	-.357	.208	-.142	.021	.294
3. CIUse	.553**	-.196	-.100	-.025	-.007	.137	-.228	-.475*
4. ASpan	.512*	.223	.089	.118	.135	-.056	.034	-.079
5. BSpan	.301	-.065	-.269	-.195	-.065	-.186	-.018	-.023
6. LRN	.360	.444	.553*	.484*	.329	.208	.083	-.085
7. LexEasy	.430*	.542*	.713***	.660**	.253	.430*	-.115	-.385
8. LexHard	.451*	.535*	.684**	.650**	.245	.500*	-.167	-.461*
9. PPVT	-.129	.689**	.755***	.779***	.333	.687***	.466**	-.433
10. FWDigit	.588**	.283	.324	.384	.126	.341	.202	-.331
11. BWDigit	--	.234	.203	.167	.315	.122	.000	-.483*

12. C/FD	--	--	.855***	.882***	.292	.459	.473*	-.433
13. FS	--	--	--	.895***	.236	.583*	.443	-.583*
14. RS	--	--	--	--	.190	.585*	.411	-.519*
15. Rec	--	--	--	--	--	.473*	.238	-.295
16. Exp	--	--	--	--	--	--	.449*	-.635**
17. Wr	--	--	--	--	--	--	--	.162
18. GFTA	--	--	--	--	--	--	--	---

Table C-1. * p<.05, ** p<.01, *** p<.001 (two-tailed). CA, chronological age; ASpan, Grammar A sequence span; BSpan, Grammar B sequence span; LRN, implicit learning score; LexEasy, % keywords correct for lexically easy sentences; LexHard, % keywords correct for lexically hard sentences; PPVT, Peabody Picture Vocabulary Test scaled score; FWDigit, forward auditory digit span; BWDigit, backward auditory digit span; C/FD, Concepts and Following Directions score on the CELF-4; FS, Formulated Sentences score on the CELF-4; RS, Recalling Sentences score on the CELF-4; Rec, Receptive language score on the VABS; Ex, Expressive language score on the VABS; Wr, Written language score on the VABS; GFTA, number of errors on the Goldman-Fristoe Test of Articulation.

Correlation Matrix for Normal-Hearing Children

Measure	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. CA	--	.518**	.524**	-.137	.414*	.461*	.277	.449*	.270
2. ASpan	--	--	.868***	.023	.240	.147	.094	.363	.420*
3. BSpan	--	--	--	-.477*	.111	-.020	-.018	.271	.459*
4. LRN	--	--	--	--	.201	.302	.204	.097	-.179
5. LexEasy	--	--	--	--	--	.854***	.382	.408*	.280
6. LexHard	--	--	--	--	--	--	.414*	.599***	.289
7. PPVT	--	--	--	--	--	--	--	.306	.326
8. FWDigit	--	--	--	--	--	--	--	--	.502**
9. BWDigit	--	--	--	--	--	--	--	--	--

Table C-2. * p<.05, ** p<.01, *** p<.001 (two-tailed). CA, chronological age; ASpan, Grammar A sequence span; BSpan, Grammar B sequence span; LRN, implicit learning score; LexEasy, % keywords correct for lexically easy sentences; LexHard, % keywords correct for lexically hard sentences; PPVT, Peabody Picture Vocabulary Test scaled score; FWDigit, forward auditory digit span; BWDigit, backward auditory digit span.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

The Role of Implicit Learning in Spoken Language Processing: Word Predictability is the Key[1]

Christopher M. Conway[2], Althea Bauernschmidt, Sean S. Huang[3], and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by DC00111 and DC00012. We wish to thank Lauren Grove for her help in data collection and manuscript preparation, and Luis Hernandez for his technical assistance.

[2] Department of Psychology, Saint Louis University, St. Louis, MO 63103

[3] Indiana School of Medicine, Indianapolis, IN 46202

The Role of Implicit Learning in Spoken Language Processing: Word Predictability is the Key

Abstract. Fundamental learning abilities related to the implicit encoding of sequential structure have been postulated to underlie language acquisition and processing. However, there is very little direct evidence to date supporting such a causal link between implicit statistical learning and language. In three experiments using novel methods of assessing implicit learning and language abilities, we show that sensitivity to sequential structure -- as measured by improvements to immediate memory span for structurally-consistent input sequences -- is significantly correlated with the ability to use knowledge of word predictability to aid speech perception under degraded listening conditions. Importantly, the association remained even after controlling for participant performance on other cognitive tasks, including short-term and working memory, intelligence, attention and inhibition, and vocabulary knowledge. Thus, the evidence suggests that implicit learning abilities are essential for acquiring long-term knowledge of the sequential structure of language -- i.e., knowledge of word order predictability – and that individual differences on such abilities impact speech perception in everyday situations. These findings provide a new theoretical rationale linking basic learning phenomena to very specific aspects of spoken language processing in adults, and may furthermore indicate new fruitful directions for investigating both typical and atypical language development.

Introduction

Understanding the role that learning and memory abilities play in language acquisition and processing remains an important challenge in the cognitive sciences. Towards this end, a major advance has been in recognizing that language consists of complex, highly variable patterns occurring in sequence, and as such can be described in terms of statistical or distributional relations among language units (Redington & Chater, 1997). Due to the probabilistic nature of language, rarely is a spoken utterance perfectly predictable; most often, the next word in a sentence is only partially predictable based on the preceding context of the sentence (Rubenstein, 1973). Put another way, what a language speaker considers to be a "meaningful" sentence can be quantified in terms of how much the preceding context constrains or predicts the next spoken word (Miller & Selfridge, 1950). Due to the apparent importance of context and word predictability in language, sensitivity to such probabilistic relations among language units likely is crucial for successful language learning and understanding.

It is not surprising then, that it is now widely accepted that general abilities related to learning about complex structured patterns -- i.e., implicit statistical learning[4] – are important for language processing (Altmann, 2002; Conway & Christiansen, 2005; Conway & Pisoni, in press; Gupta & Dell, 1999; Kirkham, Slemmer, Richardson, & Johnson, 2007; Kuhl, 2004; Pothos, 2007; Reber, 1967; Saffran, 2003; Turk-Browne, Junge, & Scholl, 2005; Ullman, 2004). Implicit learning is thought to be important for word segmentation (Saffran, Aslin, & Newport, 1996), word learning (Graf Estes, Evans, Alibali, & Saffran, 2007), the learning of phonotactic (Chambers, Onishi, & Fisher, 2003) and orthographic (Pacton,

[4] We consider implicit learning and statistical learning to refer to the same underlying phenomenon: inducing structure from input following exposure to multiple exemplars (and see Perruchet & Pacton, 2006). For brevity, we use the term "implicit learning" throughout the remainder of this paper.

Perruchet, Fayol, & Cleeremans, 2001) regularities, aspects of speech production (Dell, Reed, Adams, & Meyer, 2000), and the acquisition of syntax (Gómez & Gerken, 2000; Ullman, 2004). What is more surprising, however, is that despite the voluminous work on implicit learning, few if any studies have demonstrated a direct causal link between implicit learning abilities and everyday language competence. Although there is some evidence suggesting that implicit learning is disturbed in certain language-impaired populations (e.g., Howard, Howard, Japikse, & Eden, 2006), other studies have revealed no such relationship between implicit learning and language processing, and the reason for the discrepancy is not entirely clear (for additional discussion, see Conway, Karpicke, & Pisoni, 2007).

We propose that if implicit learning supports language, then it ought to be possible to demonstrate an empirical association between individual differences in implicit learning abilities in healthy adults and some measure of language processing. However, a challenge lies in choosing language and implicit learning tasks that purportedly tap into the same underlying processes. Toward this end, we use Elman's (1990) now classic paper as a theoretical foundation, in which a connectionist model – a simple recurrent network (SRN) –was shown to represent sequential order implicitly in terms of the effect it had on processing. The SRN had a context layer that served to give it a memory for previous internal states. This memory, coupled with the network's learning algorithm, gave the SRN the ability to learn about structure in sequential input, enabling it to predict the next element in a sequence, based on the preceding context. Elman (1990) and many others since have used the SRN successfully to model both language learning and processing (Christiansen & Chater, 2001) and, interestingly enough, implicit learning (Cleeremans, 1993).

The crucial commonality between implicit (sequence) learning and language learning and processing may be the ability to encode and represent sequential input, using preceding context to implicitly predict upcoming units. To directly test this hypothesis, we explore whether individual differences in implicit learning abilities are related to how well one is able to use sentence context – i.e., word predictability -- to guide spoken language perception under degraded listening conditions.

Word Predictability in Spoken Language Perception

Previous work has shown that knowledge of the sequential probabilities in language can enable a listener to better identify – and perhaps even implicitly predict – the next word that will be spoken (Miller, Heise, & Lichten, 1951; Rubenstein, 1973; c.f., Bar, 2007). This use of top-down knowledge becomes especially apparent when the speech signal is perceptually degraded, which is the case in many real-world situations. When ambient noise degrades parts of a spoken utterance, the listener must rely on long-term knowledge of the sequential regularities in language to implicitly predict the next word that will be spoken based on the previous spoken words, thus improving speech perception and comprehension (Elliott, 1995; Kalikow, Stevens, & Elliott, 1977; Miller, et al., 1951; Pisoni, 1996).

For example, consider the following two sentences, which end with highly predictable and non-predictable endings, respectively:

(1) Her entry should win first <u>prize</u>.

(2) Mr. White swam with their <u>mugs</u>.

When these two sentences are presented to participants under degraded listening conditions, long-term knowledge of language structure can improve perception of the final word in sentence (1) but not in (2). We argue then, that performance on the first type of sentence ought to be more closely associated with implicit learning abilities, because it relies on one's knowledge of word predictability that accrued implicitly over many years of exposure to language. On the other hand, performance on the second type of sentence simply relates to how well one perceives speech in noise, where knowledge of word predictability is less useful. Bilger and Rabinowitz (1979) further suggested the use of a metric for how well any individual subject can make use of context and word predictability in spoken language. Their metric is computed by the difference between how well one perceives the final word in high-predictability sentences (sentences of type 1) minus how well they perceive the final word in low- or zero-predictability sentences (sentences of type 2). This difference score provides a means of assessing how well an individual can use word order predictability, based on the sentence context, to aid speech perception.

We propose that implicit learning abilities are used to implicitly encode the word order regularities of language, which, once learned, can be used to improve speech perception under degraded listening conditions. In the current study, we directly tested this hypothesis by assessing adult participants on both implicit learning and speech perception tasks. In the implicit learning tasks, learning was assessed by improvements to immediate memory span for statistically-consistent, structured sequences (Botvinick, 2005; Conway et al., 2007; Jamieson & Mewhort, 2005; Karpicke & Pisoni, 2004; Miller & Selfridge, 1950). This method for measuring implicit learning, based on improvement in the capacity of immediate memory, is arguably superior to that typically used because the dependent measure is indirect (Redington & Chater, 2002); that is, it does not require an explicit judgment from the participant. In the speech perception tasks, which also rely on an indirect processing measure, we use the difference score suggested by Bilger and Rabinowitz (1979), in which speech perception performance for highly predictable sentences and zero-predictability sentences under degraded listening conditions is measured (Elliott, 1995; Kalikow et al., 1977). Based on the preceding considerations, we predict that performance on the implicit learning task will be correlated with the difference score which reflects sensitivity to word predictability in spoken sentence perception.

Experiment 1

In the first experiment, participants engaged in an implicit learning task that indirectly assessed learning through improvements to immediate memory span for sequences containing redundant statistical structure (Conway et al., 2007; Karpicke & Pisoni, 2004; Miller & Selfridge, 1950). Participants also completed a speech perception task that used degraded sentences varying in the predictability of the final word. If implicit learning abilities are important for acquiring long-term knowledge of sequential probabilities of words in sentences, we should expect that performance on the learning task will be associated with the difference score metric derived from the speech perception task.

Method

Participants. Twenty-three undergraduate students (age 18-22 years old) at Indiana University received course credit for their participation. All subjects were native speakers of English and reported no history of a hearing loss, speech impairment, or other cognitive/perceptual/motor impairments at the time of testing.

Apparatus. For the implicit learning task, a *Magic Touch*® touch-sensitive monitor displayed visual sequences and recorded participant responses. For the sentence perception task, digital audio recordings were played through Beyer Dynamic DT-100 headphones.

Stimulus Materials. Fir the implicit learning task**, w**e used two artificial grammars to generate the stimuli (c.f., Jamieson & Mewhort, 2005). These grammars, depicted in Table 1, specify the probability of a particular element occurring given the preceding element. The grammar on the left was used to create constrained sequences whereas the control grammar on the right was used to create pseudorandom (unconstrained) sequences (all stimuli are listed in Appendix A). For each sequence, the starting element (1-4) was randomly determined and then the probabilities were used to determine each subsequent element, until a desired length was reached.

The constrained grammar was used to generate 48 unique sequences for the learning phase and 20 sequences for the test phase. The control grammar was used to generate twenty sequences for the test phase as well.

For the auditory-only sentence perception task, we used 50 English sentences that varied in terms of the final word's predictability (Kalikow et al., 1977): 25 high-predictability sentences with a final target word that is predictable given the preceding context of the sentence; 25 zero-predictability sentences with a final target word that is not predictable (see Appendix B). The two sets of sentences were balanced in terms of length and word frequency (for details, see Clopper & Pisoni, 2006). All 50 sentences were spoken by a male speaker and were acoustically degraded by processing them with a sinewave vocoder (www.tigerspeech.com) that reduced the signal to 6 spectral channels.

	Constrained grammar (n+1)				**Control grammar (n+1)**			
Colors/locations (n)	1	2	3	4	1	2	3	4
1	0.0	0.5	0.5	0.0	0.0	0.33	0.33	0.33
2	0.0	0.0	1.0	0.0	0.33	0.0	0.33	0.33
3	0.5	0.0	0.0	0.5	0.33	0.33	0.0	0.33
4	1.0	0.0	0.0	0.0	0.33	0.33	0.33	0.0

Table 1. Grammars show transition probabilities from position n of a sequence to position n+1 of a sequence for four colors labeled 1-4.

Procedure

All participants completed the implicit learning task first and the sentence perception task second. For the visual implicit learning task, input sequences consisted of colored squares (red, blue, yellow, green) appearing one at a time, in one of four possible quadrants on the screen (upper left, upper right, lower left, lower right). The task was to reproduce each sequence immediately following presentation by touching the colored squares displayed on the touch-sensitive monitor in the correct order. Figure 1 is a depiction of the task. No feedback was given. For each participant, the mapping of color to screen location was randomly determined, as was the mapping between the four sequence elements (1-4) to each of the four quadrants/colors; however, for each subject, the mapping remained consistent across all trials.

Unbeknownst to participants, the implicit learning task consisted of two parts, a learning phase and a test phase, which differed only in terms of the sequences used. In the learning phase, the 48 learning sequences were presented once each, in random order. After completing the learning phase, the experiment seamlessly transitioned into the test phase, which used the 20 novel constrained and 20 unconstrained test sequences, presented in random order, once each.

Following the implicit learning task, participants next were given the auditory-only sentence perception task. In this task, participants were told they would listen to spoken sentences that were distorted, making them difficult to perceive. Their task was to identify the last word in each sentence and write that word down on a sheet of paper provided to them. Sentences were presented using a self-paced format. The 50 sentences described above were presented in random order.

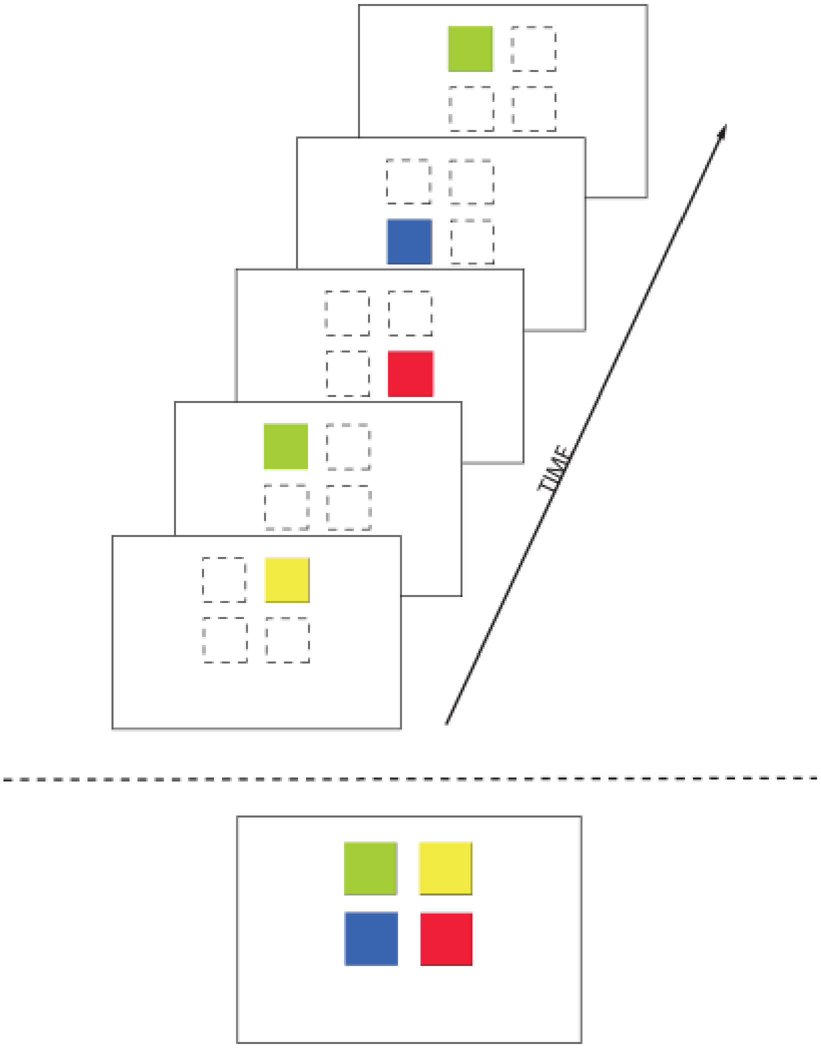

Figure 1. Depiction of the visual implicit learning task used in Experiments 1 and 3, similar to that used in previous work (Conway et al., 2007; Karpicke & Pisoni, 2004). Participants view a sequence of colored squares (700-msec duration, 500-msec ISI) appearing on the computer screen (top) and then, 2000-msec after sequence presentation, they must attempt to reproduce the sequence by pressing the touch-panels in correct order (bottom). The next sequence occurs 2000-msec following their response.

Results

Data from three participants were excluded from the final analyses because their performance on one or both tasks was greater than two standard deviations from the mean, leaving a total of twenty participants included. This was done in order to reduce the undesirable effect that outliers might have on the correlation results.

In the implicit learning task, a sequence was scored as correct if the participant reproduced each test sequence correctly in its entirety. Span scores for the “grammatical” (i.e., constrained) and “ungrammatical” (i.e., pseudorandom) test sequences were calculated using a weighted span method, in which the total number of correct test sequences at a given length was multiplied by the length, and then scores at all lengths added together. Equations 1 and 2 show the grammatical (G_{span}) and ungrammatical (U_{span}) span scores, respectively.

(1) $G_{span} = \sum(g_c * L)$

(2) $U_{span} = \sum(u_c * L)$

In Equation 1, g_c refers to the number of grammatical test sequences correctly reproduced at a given length, and L refers to the length. For example, if a participant correctly reproduced 4 grammatical test sequences of length 4, 4 of length 5, 3 of length 6, 2 of length 7, and 1 of length 8, then the G_{span} score would be computed as (4*4 + 4*5 + 3*6 + 2*7 + 1*8) = 76.

The U_{span} score is calculated in the same manner, using the number of ungrammatical test sequences correctly reproduced at a given length, u_c.

For each subject we also calculated a learning score (Equation 3), which is the difference between grammatical and ungrammatical span scores. The LRN score measures the extent that sequence memory spans improved for sequences that are constrained compared to pseudorandom sequences.

(3) $LRN = G_{span} - U_{span}$

For the sentence perception task, a sentence was scored correct if the participant wrote down the correct final word. For each participant, a word predictability difference score was calculated (Equation 4), which is the number of correctly identified target words in high-predictability sentences (HP_{corr}) minus the number of correctly identified target words in zero-predictability sentences (ZP_{corr}). This difference score reflects the participant’s ability to make use of context and word predictability to better perceive degraded speech.

(4) $PredDiff = HP_{corr} – ZP_{corr}$

A summary of the descriptive statistics is shown in Table 2. The average implicit learning score was significantly greater than 0 ($t(25) = 2.2$, $p < .05$), demonstrating that as a group, participants showed better memory for predictable test sequences compared to pseudorandom sequences. For the sentence perception task, participants’ correct identification of high-predictability and zero-predictability target words was 74.0% and 55.0%, respectively. The word predictability difference score was significantly greater than 0 ($t(19)=6.87$; $p<.001$), showing that participants were better on the task when context was present (i.e., when the final word in the sentence was highly predictable based on the preceding context).

Figure 2 shows a scatterplot of the implicit learning and word predictability difference scores. We computed a Pearson correlation between these two scores to assess the relation between implicit learning and knowledge of word predictability (the full correlation matrix for all dependent measures is presented in Appendix F-1). If implicit learning is associated with long-term knowledge of word order predictability in sentences, we would expect these two scores to be significantly positively correlated. This in fact was the case (r=.458, p<.05).

			Observed score range	
Measure	**M**	**SD**	**Minimum**	**Maximum**
GramSpan	90.00	11.96	67.00	109.00
UngramSpan	60.05	15.50	32.00	84.00
LRN	29.95	13.95	5.00	58.00
HighPred	18.50	2.82	14.00	23.00
ZeroPred	13.75	2.02	10.00	16.00
PredDiff	4.75	3.09	-2.00	10.00

Table 2. GramSpan, grammatical sequence span; UngramSpan, ungrammatical sequence span; LRN, implicit learning score; HighPred, number of high-predictability sentences correct; ZeroPred, number of zero-predictability sentences correct; PredDiff, word predictability difference score.

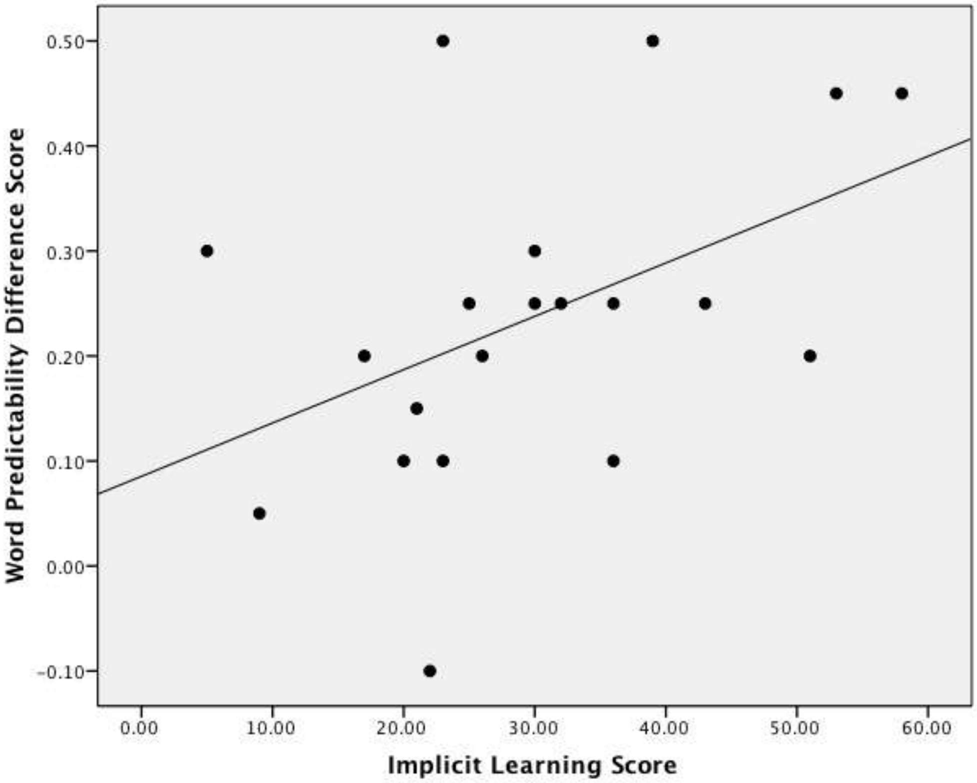

Figure 2. Scatterplot of data from Experiment 1. The x-axis displays the implicit learning scores; the y-axis displays the word predictability difference scores for the spoken sentence perception task. The best-fit line was drawn using SPSS 16.0.

Experiment 2

Experiment 1 demonstrated a statistically significant correlation between visual implicit learning and the use of knowledge of word order predictability in auditory-only speech perception. Experiment 2 served two purposes. First, it was designed to replicate the main finding of Experiment 1 but with a change in both the sensory modalities of the two experimental tasks (see Table 3) and the type of underlying structure used to generate the input sequences in the implicit learning task. If a significant correlation is still found between the two tasks even with these relatively dramatic changes, it would provide a convincing replication of the results of Experiment 1. Second, and perhaps more importantly, several additional measures were collected from participants in this study in order to determine whether there is a third mediating variable – such as general language abilities or intelligence -- responsible for the observed correlation. Observing a correlation between the two tasks even after partialing out the common sources of variance associated with these other measures would provide additional support for the conclusion that implicit learning is directly associated with knowledge of word order predictability in language, rather than being mediated by a third underlying factor.

	Experiments 1&3		**Experiment 2**	
Modality/Format	**IL**	**SP**	**IL**	**SP**
Input Modality	V (color/space)	A (words)	A (nonwords)	A/V (words)
Output Response	manual	written	spoken	written

Table 3. IL = implicit learning task; SP = sentence perception task; V = visual; A = auditory; A/V = audiovisual.

Method

Participants. Twenty-two undergraduate students (age 20-25 years old) at Indiana University received monetary compensation for their participation. All subjects were native speakers of English and reported no history of a hearing loss, speech impairment, or other cognitive/perceptual/motor impairment at the time of testing.

Apparatus. For the implicit learning task, Beyer Dynamic DT-100 headphones were used to present auditory sequences and a head-mounted microphone was used to record the participants' spoken responses. For the audiovisual spoken sentence perception task, the video display of the talker's face was presented on a Sony brand computer screen and the auditory signal was played through the headphones.

Stimulus Materials. For the auditory implicit learning task, an artificial grammar (Reber, 1967) was used to generate the stimuli (see Figure 3 and Appendix C): 18 sequences for the learning phase and 16 additional sequences for the test phase. Sixteen ungrammatical test sequences were also created by randomizing each grammatical test sequence and making sure that none of them were grammatical with respect to the grammar.

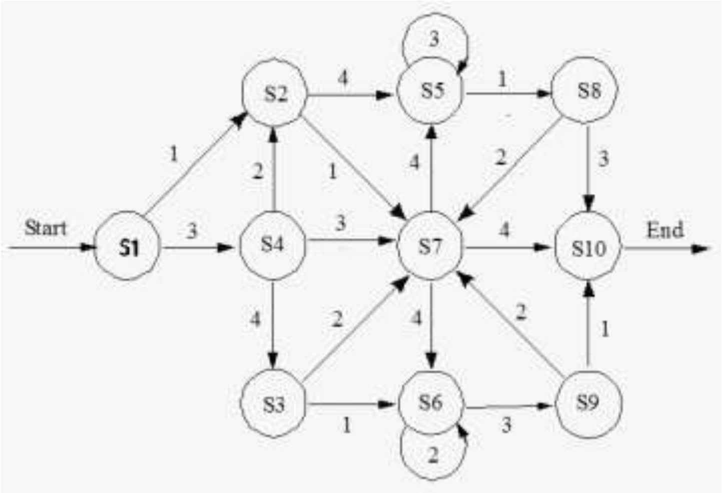

Figure 3. Artificial grammar used in Experiment 2. Each numeral was mapped onto one of four spoken nonwords.

Four spoken nonwords (“tiz”, “neb”, “dup”, and “lok”) were recorded from a 22 year-old female speaker and used to create four sound files, roughly 700 msec in duration. These nonwords were mapped onto each of the four elements (1-4) of the grammar, randomly determined for each participant. For example, if the mapping for a particular participant was 1 = “lok”; 2 = “neb”; 3 = “dup”; 4 = “tiz”, then the sequence 1-4-1-3 would be translated into the nonword sequence, “LOK-TIZ-LOK-DUP”.

For the audiovisual sentence perception task, we used 25 high-predictability and 25 zero-predictability sentences, listed in Appendix D. The same female speaker used for the implicit learning task was video recorded speaking all 50 sentences. The recordings were then converted into video clips using Final Cut Pro HD for the Macintosh. The audio portion of the clips were then processed digitally and degraded to 2 spectral channels.[5]

Procedure

The experimental tasks for this experiment took place in a sound-attenuated booth (Industrial Acoustics Company). All participants did the auditory implicit learning task first, the audiovisual sentence perception task second, and the language and intelligence assessments last.

For the auditory implicit learning task, the procedure was identical to the one used in Experiment 1 except that auditory sequences were presented instead of visual color patterns. Each nonword sequence was presented in the clear through the headphones at a level of 66-67 dB (SPL). The task was to verbally repeat each sequence immediately following presentation (see Figure 4). The timing parameters were identical to those used in Experiment 1. Sequence presentation was self-paced. Unbeknownst to participants, the task consisted of two parts, a learning phase and a test phase, which differed only in terms of the sequences used. In the learning phase, the 18 learning sequences were presented twice each,

[5] Pilot studies revealed that in the audiovisual speech perception task, only 2 spectral channels were needed to achieve performance comparable to the 6-channel audio-only speech perception task.

in random order. After completing the learning phase, the experiment seamlessly transitioned into the test phase, which used the 16 grammatical and 16 ungrammatical test sequences, presented in random order, once each.

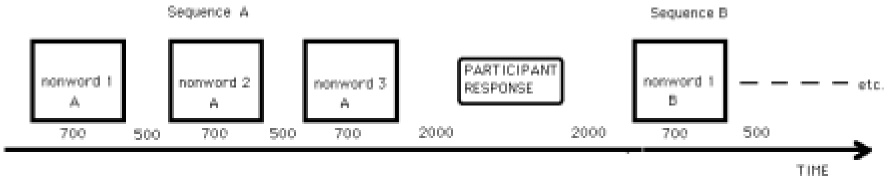

Figure 4. Depiction of the auditory implicit learning task used in Experiment 2. Participants listened to nonword sequences (700-msec duration, 500-msec ISI) through headphones and then, 2000-msec after sequence presentation, they must attempt to verbally reproduce the sequence by speaking into a microphone. The next sequence occurs 2000-msec following their response.

For the audiovisual sentence perception task, the procedure was identical to the auditory-only version of the task used in Experiment 1 except that participants watched a video of a person speaking the sentences and the auditory (but not visual) signal was distorted. Like Experiment 1, the participants' task was to identify the last word in each sentence and write it down on paper.

Following the two experimental tasks, participants were given the Reading/Vocabulary and Reading/Grammar subtests of the Test of Adolescent and Adult Language (TOAL-3; Hammill, Brown, Larsen, & Wiederholt, 1994). These subtests were used to assess receptive language abilities, specifically vocabulary and knowledge of grammar. In the Reading/Vocabulary subtest, participants read three stimulus words which all relate to a common concept. From four possible responses, the participant chooses the two words that are associated more closely with the three stimulus words. In the Reading/Grammar subtest, the participant reads five sentences that are meaningfully similar but syntactically different and then selects the two that most nearly have the same meaning. Participants completed a total of 30 Reading/Vocabulary items and 25 Reading/Grammar items and received a standardized, age-normed score for each subtest.

Participants next completed a brief (15-minute) online intelligence test (www.intelligencetest.com) consisting of 30 questions. Upon completion, the test provides a standardized, age-normed, score. The test is moderately correlated with other standardized tests of intelligence, including the Raven Progressive Matrices (r=0.42) and the Wechsler Scales (r=.32).

Results

As in Experiment 1, outliers on the implicit learning or sentence perception tasks (2 participants with scores > +/- 2 S.D.) were excluded from the final analyses, leaving a total of twenty participants included.

A summary of the descriptive statistics for the measures used in Experiment 2 are presented in Table 4. The implicit learning score was significantly greater than 0 (*t*(19) = 3.9, *p* < .01), demonstrating that participants had better memory for test sequences generated from the grammar compared to random sequences. For the sentence perception task, participants' correct identification of high-predictability and zero-predictability target words was 68.0% and 37.0%, respectively; the word predictability difference score was significantly greater than 0 (*t*(19)=9.2; *p*<.001).

			Observed score range	
Measure	**M**	**SD**	**Minimum**	**Maximum**
GramSpan	28.60	19.85	5.00	74.00
UngramSpan	22.60	19.45	0.00	75.00
LRN	6.00	6.79	-7.00	16.00
HighPred	0.64	0.22	0.21	0.96
ZeroPred	0.34	0.14	0.08	0.68
PredDiff	0.30	0.14	-0.20	0.52
TOAL-Vocab	13.15	2.03	9.00	15.00
TOAL-Grammar	12.85	1.39	10.00	135.00
IQ	117.60	13.01	93.00	135.00

Table 4. GramSpan, grammatical sequence span; UngramSpan, ungrammatical sequence span; LRN, implicit learning score; HighPred, number of high-predictability sentences correct; ZeroPred, number of zero-predictability sentences correct; PredDiff, word predictability difference score; TOAL-vocab, vocabulary as assessed by the TOAL-3 Reading/Vocabulary subscale; TOAL-grammar, knowledge of grammar as assessed by the TOAL-3 Reading/Grammar subscale; IQ, intelligence as assessed by www.intelligencetest.com.

Figure 5 shows a scatterplot of the implicit learning and word predictability difference scores. Like Experiment 1, the correlation between the implicit learning and word predictability scores was statistically significant (*r*=.423, *p*<.05, 1-tailed[6]). The full correlation matrix is shown in Appendix F-2. The strength of this correlation is strikingly similar to Experiment 1, suggesting that both sensory modality and the type of artificial grammar used have negligible effects on the nature of the association between implicit learning abilities and knowledge of word predictability.

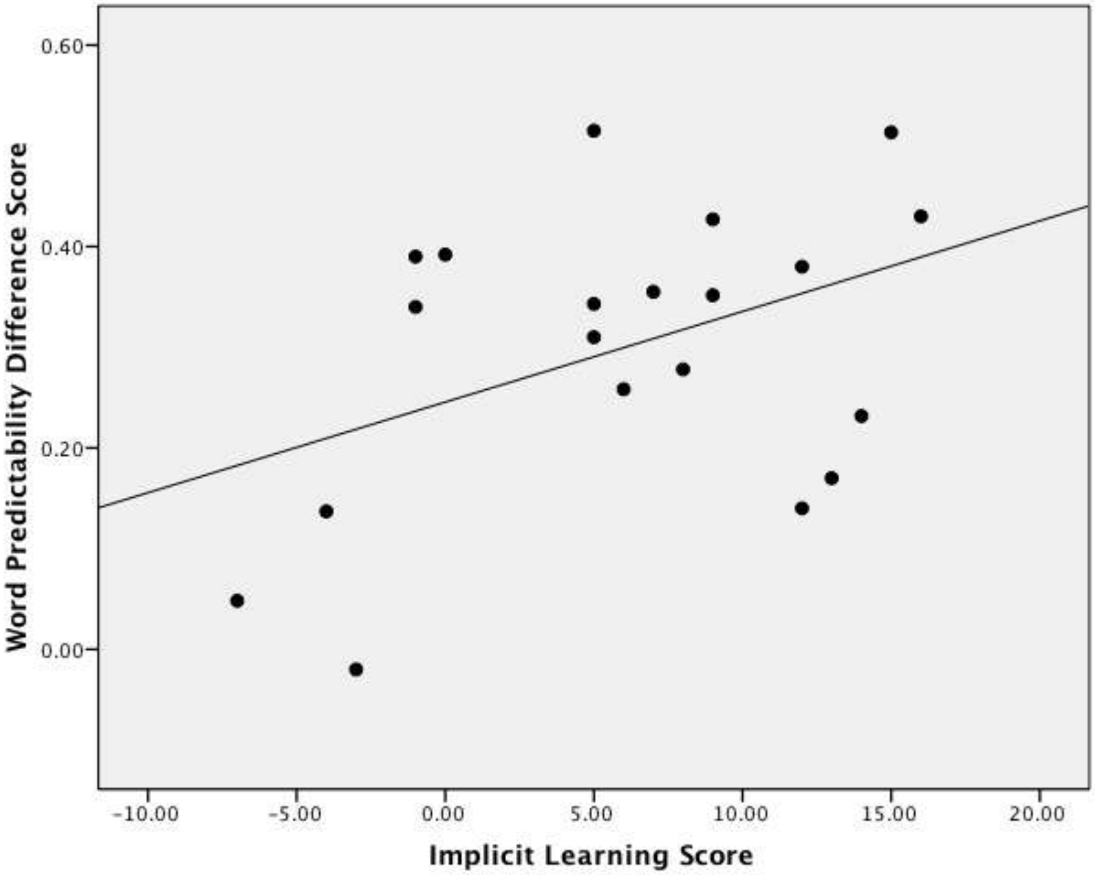

Figure 2. Scatterplot of data from Experiment 2. The x-axis displays the implicit learning scores; the y-axis displays the word predictability difference scores for the spoken sentence perception task. The best-fit line was drawn using SPSS 16.0.

We also computed correlations while partialing out the TOAL-3 and intelligence test scores. Partialing out the Reading/Vocabulary, Reading/Grammar, and intelligence scores did not reduce the strength of the correlations: *r*=.44, *p*<.05, *r*=.414, *p*<.05, and *r*=0.447, *p*<.05, respectively (all tests 1-tailed). Thus, it appears that the association between implicit learning and knowledge of word predictability is mediated neither by general linguistic knowledge nor global intelligence.

[6] One-tailed tests were used here and elsewhere (as noted) due to the results of Experiment 1 providing the specific hypothesis that implicit learning would be positively correlated with the word predictability difference score.

Finally, the implicit learning task provides a “built-in” measure of short-term memory for sequences in terms of each participant’s memory spans for the ungrammatical sequences in the test phase. Because the ungrammatical sequences contain no internal structure, this score presumably reflects short term, immediate recall. When controlling for ungrammatical sequence memory spans, the correlation between implicit learning and the word predictability difference score remains significant, and in fact is numerically stronger, *r*=.511, *p*<.05.

Experiment 3

Experiments 1 and 2 demonstrated that implicit learning abilities are associated with the ability to use knowledge of word order predictability to aid speech perception. This association remained strong even after controlling for the common variance associated with linguistic knowledge, general intelligence, and short-term sequence memory capacity. As a final replication and extension of these findings, we next include, in addition to a visual implicit learning and auditory-only sentence perception task, measures of immediate verbal recall and working memory (i.e., forward and backward digit spans), nonverbal intelligence as measured by the Raven Standard Progressive Matrices (Raven, Raven, & Court, 2000), and attention and inhibition as measured by the Stroop Color and Word Test (Golden & Freshwater, 2002).

Method

Participants. Twenty-seven undergraduate students (age 18-24 years old) at Indiana University received monetary compensation for their participation. All subjects were native speakers of English and reported no history of a hearing loss, speech impairment, or other cognitive/perceptual/motor impairment at the time of testing.

Apparatus. Experiment 3 used the same equipment as used in Experiment 1.

Stimulus Materials. The stimuli for the visual implicit learning task were identical to those used in Experiment 1. Like Experiment 1, the auditory-only sentence perception task incorporated two types of sentences: high-predictability and zero-predictability (18 of each, see Appendix E). Half of each sentence type were spoken by a female speaker and half by a male speaker. All sentences were degraded by reducing them to 6 spectral channels.

Procedure

All participants engaged in the implicit learning task first, the sentence perception task second, and the remaining assessments last.

The procedure for the visual learning task was identical to that used in Experiment 1. The procedure for the auditory-only sentence perception task also was identical to that used in Experiment 1, with the only exception being that there were 36 sentences total (half from Experiment 1 and half from Experiment 2), which were presented to participants in random order.

For the forward and backward digit spans, the procedure and materials followed that outlined in Wechsler (1991). In the forward digit span task, subjects were presented with lists of pre-recorded spoken digits with lengths (2-10) that became progressively longer. The subjects’ task was to repeat each sequence aloud. In the backwards digit span task, subjects were also presented with lists of spoken digits with lengths that became progressively longer, but they were asked to repeat the sequence in reverse order. Digits were played over headphones and recorded by a desk-mounted microphone. Subjects were scored on the longest sequence that they correctly recalled in each digit span task. Generally, the forward digit span task is thought to reflect the involvement of processes that maintain and store verbal items in short-term memory for a brief period of time, whereas the backward digit span task reflects the operation of controlled attention and higher-level executive processes that manipulate and process the verbal items held in memory (Rosen & Engle, 1997).

The Raven Standard Progressive matrices are a series of nonverbal reasoning tasks in which participants are asked to identify which of the given pictures will best complete the larger pattern in the matrix. The difficulty of the test item increases as the test goes on, so that each of the 5 subsets is progressively more difficult than the last. Subjects received either the odd half or the even half of a 60 item set taken from Raven et al. (2000). Responses were scored by total number of test items correct (out of 30).

For the Stroop Color and Word Test, we used the paper test created by Golden and Freshwater (2002). In this classic task, which measures the participant's ability to inhibit their tendency to read a *word* and not the *color*, participants are asked to read three lists aloud. The first list is a list of words 'red', 'green', and 'blue' in random order printed in black ink. The second list is a list of xxx's in red, blue, or green ink in random order. The last list is a random list of the words 'red', 'green' and 'blue' in ink color that is incongruent with the word. Each list is 100 items long. The responses are scored on how many items in the list were said aloud in 45 seconds. A standardized interference score is calculated as per Golden and Freshwater (2002), which represents how well a participant is able to selectively attend to the color of the word and inhibit the automatic reading response.

Results

As in Experiments 1 and 2, outliers on the implicit learning or sentence perception tasks (1 participant with a score > +/- 2 S.D.) were excluded from the final analyses, leaving a total of 26 participants.

A summary of the descriptive statistics for the measures used in Experiment 3 are presented in Table 5. The implicit learning score was significantly greater than 0 ($t(25) = 5.5$, $p < .001$). For the sentence perception task, participants' correct identification of HP and AN target words was 78.8% and 53.5%, respectively; the word predictability difference score was significantly greater than 0 ($t(25)=12.7$; $p<.001$).

			Observed score range	
Measure	**M**	**SD**	**Minimum**	**Maximum**
GramSpan	83.92	28.57	0.00	120.00
UngramSpan	67.08	30.83	0.00	123.00
LRN	16.73	15.57	-17.00	47.00
HighPred	0.78	0.10	0.44	0.89
ZeroPred	0.53	0.12	0.28	0.78
PredDiff	0.25	0.10	0.11	0.44
FWdigit	6.69	1.35	5.00	9.00
BWdigit	5.42	1.55	3.00	9.00
Raven	26.73	2.88	20.00	30.00
Stroop	54.11	9.16	31.00	72.00

Table 5. GramSpan, grammatical sequence span; UngramSpan, ungrammatical sequence span; LRN, implicit learning score; HighPred, number of high-predictability sentences correct; ZeroPred, number of zero-predictability sentences correct; PredDiff, word predictability difference score; FWdigit, forward digit span; BWdigit, backward digit span; Raven, non-verbal intelligence as measured by the Raven Progressive Matrices; Stroop, Stroop Interference Score.

Figure 6 shows a scatterplot of the implicit learning and word predictability difference scores. Like the previous two experiments, the correlation between these scores was positive and statistically significant (*r*=.348, *p*<.05, 1-tailed; see Appendix F-4 for full correlation matrix). However, the strength of the correlation in Experiment 3 was weaker than in the previous two experiments (which had *r*'s of .458 and .423). It may be that performing the speech perception task when there are two speakers (one male, one female), rather than one, involves additional cognitive processing resources not directly related to implicit learning, and thus slightly weakens the strength of the correlation.

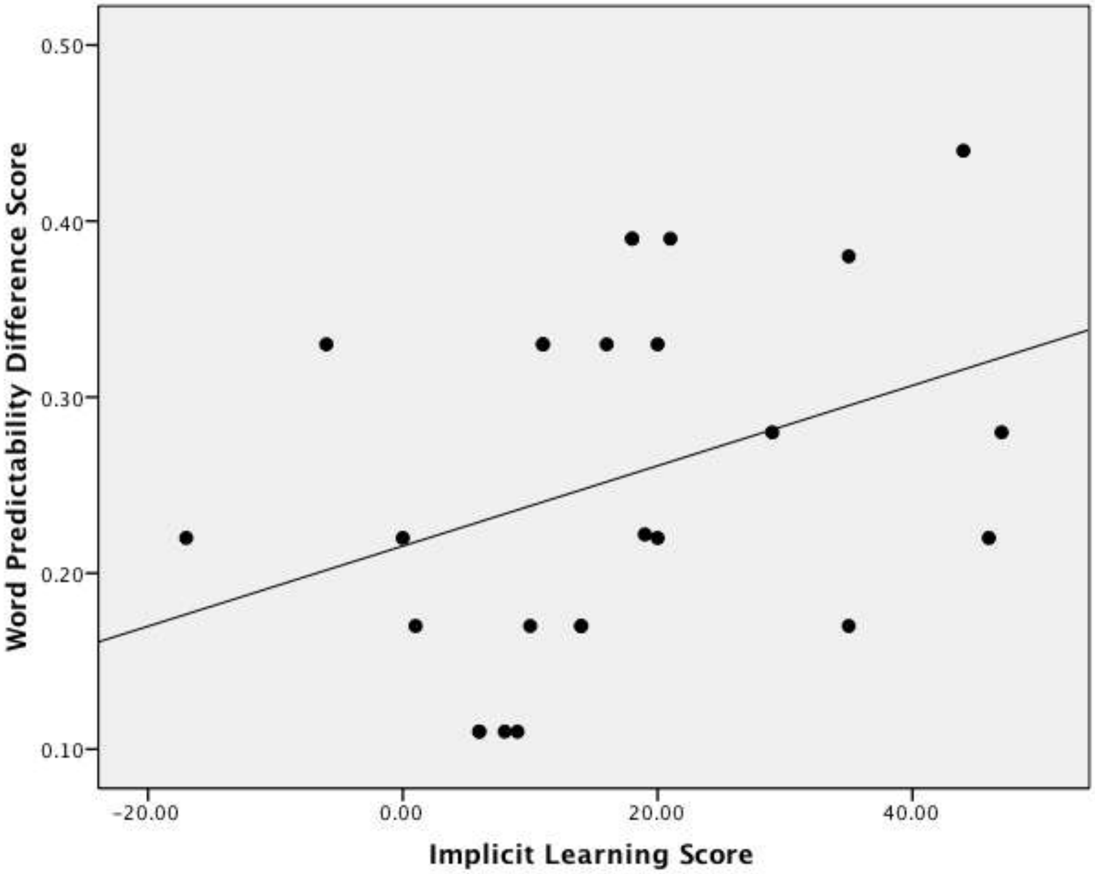

Figure 6. Scatterplot of data from Experiment 3. The x-axis displays the implicit learning scores; the y-axis displays the word predictability difference scores for the spoken sentence perception task. The best-fit line was drawn using SPSS 16.0.

Table 6 shows partial correlations between implicit learning and word predictability when controlling for each of the other measures individually. The correlations remained positive and statistically significant when controlling for the variance associated with the Stroop interference score and forward and backward digit spans. Controlling for nonverbal intelligence resulted in a slightly weaker correlation, though still positive and nearly significant (*r* = .301, *p*=.072, 1-tailed). Thus, neither verbal short-term and working memory nor attention and inhibition appear to mediate the association between implicit learning and knowledge of word predictability. However, it appears that some of the variance associated with nonverbal intelligence is shared with both implicit learning and knowledge of word predictability, but it is quite small.[7]

[7] As Table F-4 shows, nonverbal intelligence is only weakly (and negatively) correlated with both implicit learning (r=-.262, p=.2) and knowledge of word predictability (r=-.258, p=.2).

Controlling for	**r**
FWdigit	.344*
BWdigit	.346*
Stroop	.350*
Raven	.301

Table 6. *p<.05, 1-tailed. FWdigit, forward digit span; BWdigit, backward digit span; Stroop, Stroop Interference Score; Raven, non-verbal intelligence as measured by the Raven Progressive Matrices.

Finally, in order to provide converging evidence for the hypothesis that it is the ability to learn sequential patterns that is specifically associated with the ability to use knowledge of word predictability to guide speech perception, we conducted a principle component factor analysis to reduce the Experiment 3 data set. The factor analysis included all of the ten variables used in this experiment. The results of the analysis, run in SPSS 16.0, were restricted to factors with eigenvalues greater than 1. The resulting factor loadings for each of the variables on the three independent factors is shown in Table 7. The square of a factor loading represents the proportion of variance of that variable predicted by that factor.

		Factor	
Measure	**1**	**2**	**3**
GramSpan	**.741**	.251	.521
UngramSpan	**.733**	-.144	.567
LRN	-.044	**.706**	-.137
HighPred	**.650**	.355	-.424
ZeroPred	**.671**	-.422	-.448
PredDiff	-.173	**.834**	.128
FWdigit	**.741**	.304	-.109
BWdigit	**.695**	.248	-.203
Raven	**.445**	-.438	-.257
Stroop	.237	-.236	**.535**

Table . Boldface indicates the factor with the highest loading for each variable. GramSpan, grammatical sequence span; UngramSpan, ungrammatical sequence span; LRN, implicit learning score; HighPred, number of high-predictability sentences correct; ZeroPred, number of zero-predictability sentences correct; PredDiff, word predictability difference score; FWdigit, forward digit span; BWdigit, backward digit span; Raven, non-verbal intelligence as measured by the Raven Test of Progressive Matrices; Stroop, Stroop Interference Score.

As is apparent from Table 7, the distribution of factor loadings is consistent with the hypothesis that implicit learning specifically is associated with the ability to use knowledge of the structure of language to better predict and perceive words in sentences. The measures of sequence spans, speech perception, digit spans, and nonverbal intelligence, all load most strongly on Factor 1. Thus, this factor appears to relate to general memory and intelligence, abilities that may be important for deciphering and understanding degraded speech signals. Factor 3 represents inhibition and controlled attention as measured by the Stroop test. Factor 2, on the other hand, contains the implicit learning and word predictability difference scores, suggesting that this factor is associated with sensitivity to sequential statistical probabilities and the ability to make implicit predictions of the next item in a sequence.

General Discussion

The data from these three experiments show that performance on an implicit learning task is significantly correlated with performance on a spoken language measure that assesses sensitivity to word predictability in speech. The implicit learning tasks involved observing and reproducing visual color or auditory nonword sequences; a learning score was calculated for each individual by measuring the improvement to immediate serial recall for sequences with consistent statistical structure. The spoken language tasks involved perceiving degraded sentences that varied on the predictability of the final word; a difference score reflecting the use of sentence context (word predictability) to better perceive the final word was calculated by subtracting performance on zero-predictability sentences from performance on highly predictable sentences. A significant correlation between these two scores was found in all three experiments, even after controlling for sources of variance associated with intelligence[8] (Experiments 2 & 3), short-term and working memory (Experiment 3), attention and inhibition (Experiment 3), and knowledge of vocabulary and syntax (Experiment 2). We conclude that the common factor involved in both tasks -- and which mediated the observed correlations -- is sensitivity to the underlying statistical structure contained in sequential patterns, independent of general memory, intelligence, or linguistic abilities. We propose that superior implicit learning abilities result in more detailed and robust representations of the word order probabilities in spoken language. Having a more veridical representation of word predictability in turn can improve how well one can rely on top-down knowledge to help implicitly predict, and therefore perceive, the next word spoken in a sentence.

The role of top-down knowledge in influencing subsequent processing of input sequences has been mechanistically explored using a recurrent neural network model (Botvinick & Plaut, 2006). Unlike other models, Botvinick and Plaut's (2006) captures key findings in the domain of immediate serial recall while also simulating the role that background knowledge (previous learning) has on the processing of current sequences, in a manner analogous to our implicit learning task. We imagine that this model, too, could be used to capture the data on our speech perception task, showing that background knowledge of word predictability improves processing[9]. The model of Botvinick and Plaut (2006), then, appears to offer an explicit instantiation of the cognitive process that we have identified as being important: using previous knowledge to implicitly predict upcoming items in a sequence. In the terms of Botvinick and

[8] Controlling for performance on the Raven Progressive Matrices did result in a non-significant correlation; however, the factor analysis revealed that it did not load most heavily with the implicit learning and word predictability difference scores.

[9] Of course, the dependent variable in our task is word identification, not recall, but at an abstract level, the underlying mechanism – relying on background knowledge to improve subsequent processing – may be fundamentally the same.

Plaut (2006), this involves the "decoding" of imperfectly specified sequence representations through the use of long-term knowledge of sequential regularities.

An important finding from our experiments is that performance on the implicit learning task did not correlate simply with any task requiring verbalizing input (i.e., digit span tasks) or engaging linguistic knowledge (i.e., TOAL subtests). That is, implicit learning did not correlate with the measures of Reading/Vocabulary, Reading/Grammar, or digit spans. This suggests that despite the implicit learning task involving input patterns that are easy to verbalize[10], performance on the learning task is not simply due to general language abilities. Instead, implicit learning as assessed by this task is involved in language processing in a very specific way: acquiring knowledge about the predictability of items in a sequence.

Our sentence perception task involved materials that varied on word predictability, or semantic context (Kalikow, et al., 1977). There is also evidence that implicit learning is important for other aspects of language, such as syntax (Ullman, 2004) and phonotactics (Chambers, Onishi, & Fisher, 2003). From our perspective, both syntax and phonotactics can be described in terms of the predictability of items (words and phonemes, respectively) in a spoken sequence, and thus, these may be two additional aspects of language that depend upon implicit learning. However, given our present results, it may be the case that an association between implicit learning and syntax or phonotactics will best be revealed only when the language tasks rely on an implicit processing measure, not one that requires an explicit judgment (such as that used in the TOAL subtests). Toward this purpose, it should be possible to create an analogue of the present degraded speech perception task but to manipulate the underlying syntax or phonotactics of the sentences, rather than semantic context, while measuring improvements to speech perception.

In addition to exploring the role of implicit learning in other aspects of language processing, such as syntax and phonotactics, additional future work might fruitfully explore the role of implicit learning in language development specifically. For instance, a longitudinal design with children could help determine if implicit learning abilities predict subsequent speech and language abilities assessed several years later (see Bernhardt, Kemp, & Werker, 2007; Gathercole & Baddeley, 1989; Newman, Bernstein Ratner, Jusczyk, Jusczyk, & Dow, 2006; Tsao, Liu, & Kuhl, 2004). Such a finding would provide support for the hypothesis that implicit learning plays a causal role in language development. The current approach is also promising for exploring whether break-downs in implicit learning can help explain the underlying factors contributing to certain language and communication disorders, such as specific language impairment, dyslexia, and language delays associated with profound congenital hearing loss in children.

Implicit statistical learning has often been studied under the guiding assumption that it is important for language acquisition and processing (but see Casillas, in press). The present work bolsters the claim that general learning mechanisms are important for language, consistent with other recent evidence in neurophysiology (Christiansen, Conway, & Onnis, 2007; Friederici, Steinhauer, & Pfeifer, 2002) and clinical neuropsychology (Howard, Howard, Japikse, & Eden, 2006; Menghini, Hagberg, Caltagirone, Petrosini, & Vicari, 2006; Ullman, 2004; Vicari, Marotta, Menghini, Molinari, & Petrosini, 2003). However, to our knowledge, no other studies -- apart from our own preliminary findings (Conway

[10] Even the visual learning tasks in Experiments 1 and 3 involved sequences of colors that can be easily encoded into a verbal format. When using visual implicit learning stimuli that were not as easily verbalizable, Conway et al. (2007) did not find a significant correlation with speech perception for high-predictable sentences. This suggests that although the association between implicit learning and knowledge of word predictability does not appear to depend upon the sensory modality of the input, it may indeed depend upon whether the implicit learning task incorporates input that is easy to verbalize, i.e., encode into phonological and lexical representation in immediate memory. In turn, this suggests a possible dissociation between verbal and non-verbal forms of implicit learning. In fact, there is some reason to believe that implicit statistical learning may be at least partly mediated by a number of separate, specialized neurocognitive mechanisms (Conway & Christiansen, 2006). Goschke, Friederici, Kotz, and van Kampen, (2001) showed that Broca's aphasics perform normally on a spatiomotor implicit sequence learning task but are significantly impaired on one involving phonological sequences. We suggest that there may exist partially non-overlapping verbal and non-verbal implicit learning components, in a manner similar to Baddeley's (1986) theory of working memory.

et al., 2007) -- have uncovered an association between individual differences in implicit learning and spoken language abilities. One recent exception is Misyak and Christiansen (2007), who showed that adults' implicit learning performance was correlated with reading comprehension abilities. In fact, individual differences in implicit learning has been a topic infrequently explored (though see Feldman, Kerr, & Streissguth, 1995; Karpicke & Pisoni, 2004; Reber, Walkenfeld, & Hernstadt, 1991). Thus, the present results suggest that studying individual differences in implicit learning may in fact be a fruitful direction for future research, in the same way that it has been in other cognitive domains.

Finally, although not our primary aim, our data also showed that implicit learning task performance did not correlate with measures of short-term or working memory as assessed by the forward and backward digit spans (Experiment 3). Indeed, that implicit learning in this sequence reproduction task appears to be independent of verbal memory spans suggests that although serial order memory may be necessary in order to learn sequential patterns, it may not be sufficient. That is, the ability to encode and hold a series of items in immediate memory surely is necessary in order to learn about sequence structure; however, something else in addition – i.e., mechanisms involved in learning the underlying regularities – may be needed, as well. The exact relation between immediate memory capacity and implicit learning is an area in need of additional exploration (see Frensch & Miner, 1994; Karpicke & Pisoni, 2004). In fact, counter-intuitively, some research suggests that smaller memory capacities may actually be beneficial for learning complex input because it acts as a filter to reduce the complexity of the problem space, making it more manageable (Elman, 1993; Kareev, Lieberman, & Lev, 1997; Newport, 1990). Using the model of Botvinick and Plaut (2006) once more as a mechanistic framework, one could test the effect that larger or smaller sequence spans (see their footnote 9, p.213) may have on the model's ability to learn sequence structure (i.e., implicit learning).

In sum, we have presented empirical evidence showing that variation in implicit learning abilities in adulthood is directly related to sensitivity of word predictability in speech perception, specifically, sentence perception under degraded listening conditions. The correlation between the two tasks is striking given their apparent dissimilarity: one task involves using long-term knowledge of semantics and sentence context to guide speech perception, whereas the other task has to do with short-term learning and sensitivity to sequential patterns where there is no explicit semantic system. Everyday speech communication is characterized by the use of context-based redundancy to facilitate real-time comprehension; thus, these findings may be important for elucidating the underlying mechanisms involved in language processing and development, as well as for understanding and treating language and communication disorders.

References

Altmann, G. T. M. (2002). Statistical learning in infants. Proceedings of the National Academy of Sciences, 99, 15250-15251.

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

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. Trends in Cognitive Sciences, 11, 280-289.

Bernhardt, B.M., Kemp, N., & Werker, J.F. (2007). Early word-object associations and late language development. First Language, 27, 315-328.

Bilger, R. C. & Rabinowitz, W. M. (1979). Relationships between high- and low-probability SPIN scores. The Journal of the Acoustical Society of America, 65(S1), S99.

Botvinick, M. M. (2005). Effects of domain-specific knowledge on memory for serial order. Cognition, 97, 135-151.

Botvinick, M.M. & Plaut, D.C. (2006). Short-term memory for serial order: A recurrent neural network model. Psychological Review, 113, 201-233.

Casillas, G. (in press). The insufficiency of three types of learning to explain language acquisition. Lingua.

Chambers, K.E., Onishi, K.H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. Cognition, 87, B69-B77.

Christiansen, M.H. & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. Trends in Cognitive Sciences, 5, 92-98.

Christiansen, M.H., Conway, C.M., & Onnis, L. (2007). Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms. In D.S. McNamara & J.G. Trafton (Eds.), Proceedings of the 29th Annual Meeting of the Cognitive Science Society (pp. 173-178). Austin, TX: Cognitive Science Society.

Cleeremans, A. (1993). Mechanisms of implicit learning: Connectionist models of sequence learning. Cambridge, MA: MIT Press.

Clopper, C.G. & Pisoni, D.B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. Speech Communication, 48, 633-644.

Conway, C.M. & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. Journal of Experimental Psychology, 31, 24-39.

Conway, C.M. & Christiansen, M.H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. Psychological Science, 17, 905-912.

Conway, C.M., Karpicke, J., & Pisoni, D.B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings from normal-hearing adults. Journal of Deaf Studies and Deaf Education, 12, 317-334.

Conway, C.M. & Pisoni, D.B. (in press). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. Annals of the New York Academy of Sciences.

Dell, G.S., Reed, K.D., Adams, D.R., & Meyer, A.S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. Journal of Experimental Psychology: Learning, Memory, & Cognition, 26, 1355-1367.

Elliott, L. L. (1995). Verbal auditory closure and the speech perception in noise (SPIN) test. Journal of Speech, Language, and Hearing Research, 38, 1363-1376.

Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99.

Feldman, J., Kerr, B., & Streissguth, A.P. (1995). Correlational analyses of procedural and declarative learning performance. Intelligence, 20, 87-114.

Frensch, P. A. & Miner, C.S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. Memory and Cognition, 22(1), 95-110.

Friederici, A.D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. Proceedings of the National Academy of Sciences, 99, 529-534.

Gathercole, S. E. and A. D. Baddeley (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.

Golden, C.J. & Freshwater, S.M. (2002). The Stroop color and word test. Stoelting Co.: Wood Dale, IL.

Goschke, T., Friederici, A.D., Kotz, S.A., & van Kampen, A. (2001). Procedural learning in Broca's aphasia: Dissociation between the implicit acquisition of spatio-motor and phoneme sequences. Journal of Cognitive Neuroscience, 13, 370-388.

Graf Estes, K., Evans, J.L., Alibali, M.W., & Saffran, J.R. (2007). Can infants map meaning to newly segmented words? Psychological Science, 18, 254-260.

Grunow, H., Spaulding, T.J., Gómez, R.L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. Journal of Communication Disorders, 39, 158-170

Gupta, P. & Dell, G.S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), The emergence of language (pp. 447-481). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hammill, D.D., Brown, V.L., Larsen, S.C., & Wiederholt, J.L. (1994). Test of adolescent and adult language: Assessing linguistic aspects of listening, speaking, reading, and writing (3rd Edition). Austin, TX: Pro-Ed.

Howard, J.H., Jr., Howard, D.V., Japikse, K.C., & Eden, G.F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. Neuropsychologia, 44, 1131-1144.

Jamieson, R. K. and D. J. K. Mewhort (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 31, 9-23.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America, 61, 1337-1351.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. Journal of Experimental Psychology: General, 126, 278-287.

Karpicke, J. D. and D. B. Pisoni (2004). Using immediate memory span to measure implicit learning. Memory & Cognition, 32(6), 956-964.

Kirkham, N.Z., Slemmer, J.A., Richardson, D.C., & Johnson, S.P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. Child Development, 78, 1559-1571.

Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. Nature Reviews Neuroscience, 5, 831-843.

Menghini, D., Hagberg, G.E., Caltagirone, C., Petrosini, L., & Vicari, S. (2006). Implicit learning deficits in dyslexic adults: An fMRI study. Neuroimage, 33, 1218-1226.

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.

Miller, G. A. and J. A. Selfridge (1950). Verbal context and the recall of meaningful material. American Journal of Psychology, 63: 176-185.

Misyak, J.B. & Christiansen, M.H. (2007). Extending statistical learning farther and further: Long-distance dependencies and individual differences in statistical learning and language. In D.S. McNamara & J.G. Trafton (Eds.), Proceedings of the 29th Annual Meeting of the Cognitive Science Society (pp. 1307-1312). Austin, TX: Cognitive Science Society.

Newman, R., Bernstein Ratner, N., Jusczyk, A.M., Jusczyk, P.W., & Dow, K.A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. Developmental Psychology, 42, 643-655.

Newport, E.L. (1990). Maturational constraints on language learning. Cognitive Science, 14, 11-28.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. Journal of Experimental Psychology: General, 130, 401-426.

Perruchet, P. & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. Trends in Cognitive Sciences, 10, 233-238.

Pisoni, D. B. (1996). Word identification in noise. Language and Cognitive Processes, 11, 681-687.

Pothos, E.M. (2007). Theories of artificial grammar learning. Psychological Bulletin, 133, 227-244.

Raven, J., Raven, J.C., & Court, J.H. (2000). Standard progressive matrices. Harcourt Assessment: San Antonio, TX.

Redington, M. and N. Chater (1997). Probabilistic and distributional approaches to language acquisition. Trends in Cognitive Sciences, 1, 273-281.

Reber, A. S. (1967). Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior, 6, 855-863.

Reber, A.S., Walkenfeld, F.F., & Hernstadt, R. (1991). Explicit and implicit learning: Individual differences and IQ. Journal of Experimental Psychology, 17, 888-896.

Redington, M. and N. Chater (2002). Knowledge representation and transfer in artificial grammar learning (AGL). In R.M. French and A. Cleeremans (Eds.), Implicit learning and consciousness: An empirical, philosophical, and computational consensus in the making (pp. 121-143). Hove, East Sussex: Psychology Press**.**

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

Rubenstein, H. (1973). Language and probability. In G. A. Miller (Ed.), Communication, language, and meaning: Psychological perspectives (pp. 185-195). New York: Basic Books.

Saffran, J.R. (2003). Statistical language learning: Mechanisms and constraints. Current Directions in Psychological Science, 12, 110-114.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.

Tsao, F.-M., Liu, H.-M., & Kuhl, P.K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. Child Development, 75, 1067-1084

Turk-Browne, N. B., Junge, J. A., & Scholl, B.J. (2005). The automaticity of visual statistical learning. Journal of Experimental Psychology: General, 134, 522-564.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. Cognition, 92, 231-270.

Wechsler, D. (1991). Wechsler intelligence scale for children – Third edition. The Psychological Corporation: San Antonio, TX.

Vicari, , S., Marotta, , L., Menghini, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. Neuropsychologia, 41, 108-114.

Appendix A

Learning and Test Sequences used for Experiments 1 and 3 Visual ISL Task

Sequence Length	Learning Sequence	Test Sequence (C)	Test Sequence (UC)
3	4-1-2		
	1-3-4		
	2-3-4		
	3-4-1		
	4-1-3		
	1-3-1		
	1-2-3		
	2-3-1		
4	4-1-2-3	1-2-3-1	1-4-3-4
	3-4-1-3	1-3-4-1	2-1-2-4
	3-1-3-1	4-1-3-4	4-2-4-3
	2-3-1-2	4-1-3-1	1-4-1-2
	1-2-3-4		
	2-3-1-3		
	3-4-1-2		
	2-3-4-1		
5	2-3-1-3-4	4-1-3-4-1	4-1-2-3-2
	1-2-3-4-1	2-3-4-1-2	1-2-1-3-1
	3-4-1-3-1	3-4-1-2-3	1-4-2-3-1
	4-1-2-3-1	4-1-3-1-3	3-1-4-2-4
	2-3-4-1-3		
	3-4-1-3-4		
	4-1-2-3-4		
	2-3-1-2-3		
6	2-3-4-1-3-4	2-3-4-1-2-3	2-1-4-1-3-2
	3-1-3-4-1-3	3-4-1-2-3-1	4-1-3-4-3-2
	1-2-3-4-1-2	2-3-1-3-1-2	3-1-2-3-4-3

Sequence Length	Learning Sequence	Test Sequence (C)	Test Sequence (UC)
6 (cont'd)	2-3-1-2-3-1	3-1-3-1-3-1	2-1-2-1-3-4
	1-3-4-1-2-3		
	4-1-2-3-4-1		
	1-3-1-2-3-4		
	2-3-4-1-3-1		
7	4-1-3-4-1-3-1	3-4-1-3-4-1-3	2-3-1-2-3-1-3
	3-1-3-1-3-1-3	1-2-3-4-1-2-3	2-3-2-4-3-4-2
	1-2-3-4-1-3-1	3-1-3-1-2-3-4	4-3-4-2-3-2-4
	4-1-2-3-4-1-2	2-3-1-2-3-1-2	4-3-1-3-2-4-3
	3-4-1-3-1-2-3		
	2-3-1-3-1-3-4		
	1-3-1-3-4-1-2		
	3-1-3-4-1-2-3		
8	2-3-4-1-3-4-1-2	3-4-1-2-3-1-2-3	1-3-1-4-3-1-2-4
	1-3-1-3-4-1-2-3	2-3-1-2-3-1-3-1	4-3-4-2-3-4-2-4
	3-1-3-4-1-3-4-1	1-2-3-4-1-2-3-4	2-4-2-1-2-1-2-3
	3-4-1-3-4-1-3-4	3-4-1-2-3-4-1-2	2-3-2-3-1-4-2-4
	4-1-2-3-1-3-4-1		
	2-3-4-1-2-3-1-3		
	1-2-3-4-1-2-3-1		
	3-1-3-1-3-4-1-2		

Note: (C), constrained; (UC), unconstrained

Appendix B

Sentences used for Experiment 1 Auditory-Only Spoken Sentence Perception Task

High Predictability	Zero Predictability
Eve was made from Adam's rib.	The bread gave hockey loud aid.
Greet the heroes with loud cheers.	The problem hoped under the bay.
He rode off in a cloud of dust.	The cat is digging bread on its beak.
Her entry should win first prize.	The arm is riding on the beach.
Her hair was tied with a blue bow.	Miss Smith was worn by Adam's blade.
He's employed by a large firm.	The turn twisted the cards.
Instead of a fence, plant a hedge.	Jane ate in the glass for a clerk.
I've got a cold and a sore throat.	Nancy was poured by the cops.
Keep your broken arm in a sling.	Mr. White hit the debt.
Maple syrup was made from sap.	The first man heard a feast.
She cooked him a hearty meal.	The problems guessed their flock.
Spread some butter on your bread.	The coat is talking about six frogs.
The car drove off the steep cliff.	It was beaten around with glue.
They tracked the lion to his den.	The stories covered the glass hen.
Throw out all this useless junk.	The ship was interested in logs.
Wash the floor with a mop.	Face the cop through the notch.
The lion gave an angry roar.	The burglar was parked by an ox.
The super highway has six lanes.	For a bloodhound he had spoiled pie.
To store his wood, he built a shed.	Water the worker between the pole.
Unlock the door and turn the knob.	The chimpanzee on his checkers wore a scab.
We heard the ticking of the clock.	Miss Brown charged her wood of sheep.
Playing checkers can be fun.	Tom took the elbow after a splash.
That job was an easy task.	We rode off in our tent.
The bloodhound followed the trail.	The king shipped a metal toll.
He was scared out of his wits.	David knows long wheels.

Appendix C

Learning and Test Sequences used for Experiment 2 Auditory ISL Task

Sequence Length	Learning Sequence	Test Sequence (G)	Test Sequence (UG)
4	1-4-1-3		
	3-4-2-4		
5	1-1-4-1-3	1-1-4-3-1	3-1-4-1-3
	1-4-3-1-3	1-4-1-2-4	1-1-2-4-4
	3-2-4-1-3	3-3-4-1-3	4-3-1-3-3
	3-3-4-3-1	3-4-1-3-1	4-1-3-3-1
6	1-1-4-2-3-1	1-4-3-1-2-4	1-2-4-3-4-1
	3-2-1-4-1-3	3-2-4-1-2-4	4-2-2-4-1-3
	3-3-4-1-2-4	3-3-4-3-1-3	3-3-1-4-3-3
	3-4-1-2-3-1	3-4-2-4-1-3	3-4-1-4-3-2
7	1-1-4-3-3-1-3	1-1-4-2-2-3-1	1-2-4-3-1-2-1
	3-2-1-4-3-2-4	3-2-1-4-1-2-4	1-1-2-3-4-3-2
	3-3-4-3-3-1-3	3-3-4-2-2-3-1	1-2-3-3-2-3-4
	3-4-2-4-3-2-4	3-4-2-4-1-2-4	2-4-3-4-1-4-2
8	1-1-4-3-3-1-2-4	1-4-3-1-2-4-3-1	4-2-1-1-3-1-4-3
	3-2-1-4-3-3-1-3	3-3-4-2-2-3-2-4	4-4-3-3-2-2-3-2
	3-3-4-3-3-1-2-4	3-4-1-3-2-4-1-3	3-1-2-4-3-4-3-1
	3-4-2-4-2-3-2-4	3-4-2-4-3-3-1-3	3-1-4-2-3-3-4-3

Note: (G), grammatical; (UG), ungrammatical

Appendix D

Sentences used for Experiment 2 Audiovisual Spoken Sentence Perception Task

High Predictability	Zero Predictability
A bicycle has two wheels.	A duck talks like an ant.
Ann works in the bank as a clerk.	All the sailors were in bloom.
Banks kept their money in a vault.	For your football I prescribed a cake.
Bob was cut by the jackknife’s blade.	Lubricate the kitten to chew the draft.
Break the dry bread into crumbs.	Mr. White swam with their mugs.
Cut the meat into small chunks.	My teacher has a glad screen.
It was stuck together with glue.	Peter played and swabbed a white bruise.
Kill the bugs with this spray.	Ruth robbed the hay.
Paul hit the water with a splash.	She parked the camera with brief sheets.
Paul was arrested by the cops.	She tracked a bill in her cap.
Raise the flag up the pole.	Spread some soup from the pad.
Ruth had a necklace of glass beads.	Stir the class into strips.
The bird of peace is the dove.	Syrup is a fine sport.
The boat sailed across the bay.	The prickly bath knew about the track.
The bride wore a white gown.	The rosebush slept along the coast.
The cigarette smoke filled his lungs.	The sailor was followed from old wheat.
The cow gave birth to a calf.	The sand was filled with pine.
The nurse gave him first aid.	The stale game was spoken by steam.
The poor man was deeply in debt.	The steamship employed his crop.
The shepherds guarded their flock.	The steep man sat with the wax.
The wedding banquet was a feast.	The storm was trained by a dive.
The witness took a solemn oath.	The turn twisted the cards.
Tree trunks are covered with bark.	The useless knees escaped with the hive.
Watermelons have lots of seeds.	They milked a frightened entry of gin.
We swam at the beach at high tide.	This bear won’t drive in the lock.

Appendix E

Sentences used for Experiment 3 Auditory-Only Spoken Sentence Perception Task

High Predictability	**Zero Predictability**
He was scared out of his wits.	Mr. White hit the debt.
Greet the heroes with loud cheers.	The problem hoped under the bay.
He rode off in a cloud of dust.	Nancy was poured by the cops.
Her entry should win first prize.	Jane ate in the glass for a clerk
Her hair was tied with a blue bow.	The ship was interested in logs.
He’s employed by a large firm.	The turn twisted the cards.
Instead of a fence, plant a hedge.	.The stories covered the glass hen.
I’ve got a cold and a sore throat.	It was beaten around with glue.
Keep your broken arm in a sling.	The problems guessed their flock.
The baby slept in his crib.	Discuss the sailboat on the bend.
The candle flame melted the wax.	Face the cop through a notch.
The drowning man let out a yell.	The king shipped a metal toll.
The fruit was shipped in wooden crates.	The low woman was gladly in the calf.
The furniture was made of pine.	The old cloud broke his lungs.
The honey bees swarmed round the hive.	Water the worker between the poles.
The little girl cuddled her doll.	We scared a bomb of clever geese.
The lonely bird searched for its mate.	They milked a frightened entry of gin.
The railroad train ran off the track.	The rosebush slept along the coast.

Appendix F

Correlation Matrices for Experiments 1-3

Correlation Matrix for Experiment 1

Measure	1	2	3	4	5	6
1. GramSpan	--	.509*	.291	-.080	-.350	.157
2. UngramSpan		--	-.675**	-.350	-.042	-.292
3. LRN			--	.320	-.254	.458*
4. HighPred				--	.217	.770***
5. ZeroPred					--	-.456*
6. PredScore						--

Table F-1. * p<.05, ** p<.01, *** p<.001 (two-tailed)

Correlation Matrix for Experiment 2

Measure	1	2	3	4	5	6	7	8	9
1. GramSpan	--	.940***	.228	.652**	.452*	.528*	.039	.436	.323
2. UngramSpan		--	-.116	.550*	.434	.391	-.056	.407	.342
3. LRN			--	.329	.075	.422	.274	.106	-.036
4. HighPred				--	.756***	.746***	.207	.334	.319
5. ZeroPred					--	.129	.316	.367	.216
6. PredDiff						--	-.006	.132	.264
7. TOAL-Vocab							--	.494*	.255
8. TOAL-Grammar								--	.183
9. IQ									--

Table F-2. * p<.05, ** p<.01, *** p<.001 (two-tailed)

Correlation Matrix for Experiments 1 and 2 Combined

Measure	1	2	3	4	5	6
1. GramSpan	--	.759***	.248	.337*	.111	.348*
2. UngramSpan		--	-.412**	.119	.205	.099
3. LRN			--	.317*	-.108	.418**
4. HighPred				--	.488***	.697***
5. ZeroPred					--	-.124
6. PredDiff						--

Table F-3. * p<.05, ** p<.01, *** p<.001 (two-tailed)

Correlation Matrix for Experiment 4

Measure	1	2	3	4	5	6	7	8	9	10
1. GramSpan	--	.845***	.191	.343	.221	.067	.461*	.407*	.063	.274
2. UngramSpan		--	-.357	.212	.282	-.133	.380	.352	.200	.266
3. LRN			--	.196	-.130	.348	.118	.106	-.262	-.012
4. HighPred				--	.571**	.272	.515**	.395*	.225	-.058
5. ZeroPred					--	.129	.316	.367	.216	.088
6. DiffScore						--	.054	.040	-.258	-.151
7. FWdigit							--	.618***	.112	.087
8. BWdigiit								--	.322	-.074
9. Raven									--	.180
10. Stroop										--

Table F-4. * p<.05, ** p<.01, *** p<.001 (two-tailed)

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Combined Electric and Contralateral Acoustic Hearing for Speech Perception and Immediate Phonological Memory [1]

Marcia J. Hay-McCutcheon[2], Nathaniel R. Peterson[3] and David B. Pisoni[4]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by grants R03 DC008383 and T32 DC00012 from the National Institutes of Health and the National Institute of Deafness and Communication Disorders

[2] Department of Communicative Disorders, The University of Alabama, Tuscaloosa, Alabama

[3] Department of Otolaryngology-Head and Neck Surgery, Indiana University School Of Medicine, Indianapolis, Indiana

[4] Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana and Department of Otolaryngology-Head and Neck Surgery, Indiana University School Of Medicine, Indianapolis, Indiana

Combined Electric and Contralateral Acoustic Hearing for Speech Perception and Immediate Phonological Memory

Abstract. Purpose*:* This study examined the benefit of bimodal hearing (i.e., one cochlear implant and a contralateral hearing aid) for listening to words and sentences in quiet and in noise, the identification of vowels and immediate phonological memory.

Method: Two groups of twenty adult cochlear implant recipients, those who used bimodal hearing on a regular basis, and those who used one cochlear implant, were recruited for this study. They were asked to complete two speech perception tests, the Consonant Nucleus Consonant (CNC) test, and sentences in quiet and in noise from the Hearing In Noise Test. In addition, participants were asked to identify groups of vowels that were either acoustically similar or acoustically distinct. Immediate phonological memory skills were assessed by asking the participants to recall sequences of vowels.

Results: The outcomes suggested that the use of bimodal hearing did not provide sufficient acoustic cues to aid with the identification of words, sentences in quiet and in noise, isolated vowels, and immediate memory for sequences of vowels compared to the use of one cochlear implant. A larger degree of residual hearing prior to implantation was, however, associated with improved performance when identifying isolated words and listening to sentences in noise. The ability to recognize isolated vowels and the short-term memory for sequences of vowels were both positively correlated with speech perception skills.

Conclusions: The use of a contralateral hearing aid with a cochlear implant did not provide sufficiently greater benefit for identifying words, sentences in quiet and in noise, isolated vowels, and the immediate memory for those vowels, compared to the use of one cochlear implant alone. Vowel identification and immediate phonological memory, however, were moderately correlated.

Introduction

The limited spectral cues provided by a cochlear implant have been associated with poor speech understanding in noise (Fu, Shannon, & Wang, 1998; Turner, Gantz, Vidal, Behrens, & Henry, 2004) and difficulty identifying different talkers in a group (Vongphoe & Zeng, 2005). Both Fu et al. (1998) and Turner et al. (2004) used simulated speech (i.e., speech processed through a series of band-pass filters) with normal-hearing listeners and reported that low-frequency cues and fine spectral structure enhanced speech understanding abilities in noise. Also, Turner et al. (2004) found that for cochlear implant recipients, the additional spectral cues provided through increasing numbers of electrodes resulted in improved speech perception in noisy listening environments. Vongphoe & Zeng (2005) found that a greater number of talkers were correctly identified with speech tokens (i.e., spoken vowels) that were both amplitude and frequency modulated compared to just amplitude modulated. They also found that more spectral bands improved talker identification. These data suggest that spectral resolution is important for understanding speech in noise and identifying talkers.

Recent evidence from cochlear implant recipients has suggested that with bimodal hearing (i.e., one cochlear implant and a contralateral hearing aid) individuals can achieve improved speech understanding in noise. Using a group of adults who used bimodal hearing, Dunn, Tyler, & Witt (2005) reported that the identification of CUNY sentences in the presence of multi-talker speech babble was significantly improved when both their hearing aid and cochlear implant were used compared to the use of their cochlear implant alone. Seven out of eleven individuals showed this pattern. Additionally, Dorman, Gifford, Spahr, & McKarns (2008) found that the identification of single-syllable words and sentences in quiet and in noise was significantly improved when individuals used both their hearing aid and cochlear implant compared to when they used either device alone. Presumably, the added spectral information provided through a hearing aid and/or the use of bilateral hearing prostheses provided additional cues for these individuals which resulted in improved performance in background noise. One objective of this current study, therefore, was to try to replicate the findings of previous studies by using two groups of study participants, those who used one cochlear implant and those who used bimodal hearing.

Additionally, studies have suggested that improved recognition of vowels can occur with increased access to spectral cues and added low-frequency information. Nie, Barco, & Zeng (2006) demonstrated that as the number of electrodes increased from 4 to 12, significant improvement in vowel identification was observed for five Med-El cochlear implant recipients. Results from Chang & Fu (2006) also revealed that normal-hearing listeners identified more vowels when listening to sine-wave carrier vocoded speech in comparison to noise-band carrier simulated speech, suggesting that the added spectral cues from the sine-wave speech helped speech perception. Finally, Qin & Oxenham (2006) used a series of simulated speech processing schemes and demonstrated that increasing access to low-frequency information resulted in improved vowel identification in normal-hearing listeners. Taken together, these studies suggest that added spectral information is beneficial for vowel identification.

It has been commonly thought that consonants carry more information for sentence intelligibility than vowels (Freyman & Nerbonne, 1989; Preves, Fortune, Woodruff, & Newton, 1991). More recent evidence, however, has challenged this common notion. An early study performed by Cole, Yonghong, Mak, Fanty, & Bailey (1996) required individuals to listen to and transcribe TIMIT sentences in which either the consonants or the vowels were replaced with noise. Their findings indicated that the accurate recognition of the words in the sentences was more dependent on vowel information than on the consonant cues. These findings were recently replicated by Kewley-Port, Burkle, & Lee (2007) using young normal-hearing listeners and elderly hearing-impaired listeners. The young normal-hearing individuals identified significantly more words in sentences with the consonants replaced by noise (74%) compared to the vowels replaced by noise (34%). For the elderly hearing-impaired listeners, the words in sentences were recognized with 40% accuracy when the vowel information was maintained compared to 20% accuracy when the consonant information was maintained. The contribution of vowel cues, therefore, for sentence comprehension seems to be more important than previously thought.

Considering the contribution of spectral cues for vowel identification and the recent findings suggesting that vowel information is important for sentence perception, we also examined vowel identification in a group of individuals who use bimodal hearing and a group of adults who use unilateral cochlear implants. If the added spectral information provided through a hearing aid can improve speech perception, the recommended use of a hearing aid with a unilateral cochlear implant might be warranted.

It is also possible, however, that individuals who use a hearing aid in conjunction with a cochlear implant have a sufficient amount of aided residual hearing compared to individuals who do not use a contralateral hearing aid. Previous research has demonstrated that individuals with a greater degree of residual hearing prior to implantation perform better on speech perception tasks compared to individuals with less residual hearing (Francis, Yeagle, Bowditch, & Niparko, 2005; Gantz, Woodworth, Knutson, Abbas, & Tyler, 1993).

Researchers also have examined the role of low-frequency residual hearing prior to implantation on speech perception performance (Ching, Incerti, & Hill, 2004; Gifford, Dorman, McKarns, & Spahr, 2007). Ching, et al.(2004) and Gifford et al. (2007) suggested that the behavioral threshold at 250 Hz prior to implantation might be important for post-implantation performance but correlation analyses using these low-frequency thresholds measures of speech perception in quiet and noise revealed no significant findings. Considering the importance of low-frequency information for vowel identification, both traditional measures of residual hearing (i.e., pure-tone frequency threshold average at 500 Hz, 1000 Hz, and 2000 Hz) and a measure of low-frequency residual hearing (i.e., pure-tone frequency threshold average of 250 Hz, 500 Hz, and 1000 Hz) were obtained to determine the importance of residual hearing for both vowel identification and general speech perception skills.

For the first experiment of this study, it was hypothesized that profoundly deaf individuals who use cochlear implants would more successfully identify acoustically distinct vowels than acoustically similar vowels. In addition, it was expected that individuals who used bimodal hearing would more successfully identify acoustically similar vowels than individuals who used one cochlear implant without a hearing aid. The second experiment of this study examines how the acoustic characteristics of vowels might impact short-term phonological memory and, in turn, how proficiency with this cognitive task affects speech perception.

Experiment 1

Method

Participants Two groups of postlingually deaf adults who received a cochlear implant at the Department of Otolaryngology—Head and Neck Surgery at the Indiana University School of Medicine were recruited for this study. Twenty unilateral cochlear implant recipients and 20 cochlear implant recipients who were fitted with both a cochlear implant and a hearing aid in the opposite ear completed the study. Although previous studies have examined the skills of individuals who use bimodal hearing prostheses using a within-subject design (Armstrong, Pegg, James, & Blamey, 1997; Dorman, et al., 2008; Dunn, et al., 2005), we explored the potential benefits of bimodal hearing through the use of an between-subjects design. Using this methodological model it was possible to evaluate the potential benefits of the added spectral information via a hearing aid using “natural” listening conditions for the same set of participants.

The demographic information for the unilateral cochlear implant recipients and the bimodal hearing individuals is provided in Tables 1 and 2 respectively. The mean age at test was 59.3 years old (±13.0 years) for the unilateral cochlear implant group and 63.0 years old (±10.3 years) for the bimodal hearing group. All participants had at least 6 months of experience with their cochlear implant prior to testing (mean = 3.2 years, ± 2.6 SD), and the bimodal group had at least 3 years of experience with a hearing aid (mean = 33.6 years, ±15.0 SD) prior to implantation. All study participants wore their hearing devices for the majority of their waking hours. The degree of residual hearing prior to implantation was calculated using the pure-tone average (PTA) of the unaided audiometric thresholds at 500 Hz, 1000 Hz, and 2000 Hz for the better-hearing ear. A second PTA was calculated using the unaided audiometric thresholds at 250 Hz, 500 Hz, and 1000 Hz (i.e., Low-Frequency PTA) in order to account for the potential influence that residual hearing at 250 Hz might have on perceptual skills. All study participants used current speech processing strategies (i.e., SPEAK, ACE, CIS, MPS). Prior to testing, each hearing aid was analyzed using the Audioscan RM500 SL hearing aid analyzer to verify that each device was working according to manufacturer specifications. Throughout the testing session, hearing devices were set at everyday listening/volume levels.

Participant ID	Age at testing (yr)	Implant model	Duration of CI use (yr)	PTA (dB HL)	LF-PTA (dB HL)
AAH1	80.8	N24-CI24R (CS)	5.4	108	102
AAK1	52.5	N24-CI24R (CS)	3.9	112	117
AAL1	73.8	N24-CI24RE (CA)	0.7	62	33
AAN1	56.3	N24-CI24R (CS)	4.2	92	92
AAO1	81.8	ME-Combi 40+ H	3.0	77	80
AAP1	48.9	N22-CI22M	15.6	120	120
AAS1	61.1	ME-Combi 40+ H	3.5	107	97
AAT1	77.4	N24-CI24RE (CA)	3.8	107	82
AAV1	55.3	N24-CI24M	10.1	105	83
AAW1	66.9	N24-CI24RE (CA)	1.2	92	90
AAX1	30.4	N24-CI24RE (CA)	2.1	112	98
AAY1	49.7	N22-CI22M	17.4	110	98
ABA1	59.7	N24-CI24R (CS)	5.8	87	73
ABD1	59.7	CL-HiRes 90K	2.9	78	55
ABE1	48.9	N24-CI24RE (ST)	2.5	77	52
ABF1	48.4	ME-Combi 40+ H	4.4	80	80
ABI1	64.5	N24-CI24RE (CA)	2.1	110	108
ABL1	64.4	N24-CI24R (CS)	4.9	90	90
ABS1	43.7	N24-CI24RE (CA)	2.6	100	87
ABU1	61.7	CL-HiRes 90K	2.9	108	107
Mean	59.3		5.0	97	87
SD	13.0		4.4	16	22

Table 1. Demographics for unilateral cochlear implant recipients.

Participant ID	Age at testing (yr)	Implant model	Duration of CI use (yr)	PTA (dB HL)	LF-PTA (dB HL)	HA model	Duration of HA use (yr)
AAD1	72.7	ME-Combi 40+ H	2.4	87	62	Unitron digital SF	32
AAG1	54.7	N24-CI24R (CS)	4.9	105	83	Unitron US 80-PP	45
AAM1	58.0	ME-Combi 40+ H	2.7	100	82	Marcon	47
AAQ1	68.7	CL-HiFocus	7.1	95	83	Widex Senso	50
AAR1	74.2	N24-CI24M	9.2	98	77	Widex Q32	22
AAU1	57.6	N24-CI24RE (CA)	0.9	100	88	CN Resound 780 D	49
ABB1	73.5	ME-Combi 40+ HS	7.2	93	73	Oticon 39 PL	33
ABC1	70.4	N24-CI24RE (CA)	1.2	82	55	Phonak Perseo 311 DAZ	28
ABG1	64.8	ME-Combi 40+ H	3.1	93	93	Phonak Senso-Forte 331X	46
ABH1	52.4	N24-CI24R (CA)	3.2	97	98	Oticon 390 PL	3.7
ABJ1	70.4	N24-CI24RE (CA)	0.5	105	105	Oticon Digifocus II SP	20
ABK1	73.7	ME-Combi 40+ H	8.0	78	72	Starkey A575 Sequel AH	12
ABM1	69.6	N24-CI24RE (CA)	2.6	88	60	CN Resound	39
ABN1	57.7	CL-HiRes 90K	2.6	92	83	CN Resound 780 D	45
ABO1	62.7	N24-CI24RE (CA)	1.7	108	87	Siemens Infinity Pro SP	32
ABQ1	44.8	N24-CI24RE (CA)	0.6	83	73	Oticon Digifocus II	40
ABR1	70.7	CL-HiRes 90K	2.3	113	108	AVR XP 675	62
ABT1	71.0	CL-HiRes 90K	1.2	107	97	Phonak Supero 413 AZ	31
ABV1	54.8	N24-CI24RE (CA)	2.1	80	82	Phonak Perseo 311 DAZ	11
ABW1	38.5	N24-CI24RE (CA)	0.6	95	97	Rexton Energy	25
Mean	63.0		3.2	95	83		33.6
SD	10.3		2.6	10	14		15.0

Table 2. Demographic information for adults who used bimodal hearing

Additionally, for control purposes a group of 20 adults with normal hearing participated in this study. The data from the normal-hearing group provided information about the identification of vowels that was compared to the vowel identification data obtained from the individuals who used cochlear implants. Audiometric thresholds at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz were ≤ 20 dB HL bilaterally for this group of control study participants. The mean age at testing was 34.2 years old (± 8.6 years) with a range of 23.7 to 49.3 years old.

Stimuli and Procedures Word and sentence recognition tests were administered to all study participants. One 50-word list of the Consonant-Nucleus-Consonant (CNC) word recognition test (Peterson & Lehiste, 1962) and the Hearing in Noise Test (HINT) sentence recognition test (Nilsson, Soli, & Sullivan, 1994) were presented to the cochlear implant recipients in a sound-treated booth. The CNC words were presented at 65 dB SPL, and the HINT sentences were presented at 65 dB SPL, and in two background noise conditions (i.e., +5 dB and +10 dB signal-to-noise ratios). Two HINT lists of 10 sentences were presented in each listening condition. The standard procedure for the HINT test (i.e., determining a sentence reception threshold in noise) was not used. Rather, the sentence lists from this test were used to assess speech understanding in quiet and in the two noise conditions. The CNC word test was administered first, followed by the HINT sentence tests. All study participants were instructed to repeat the stimuli they heard and guessing was encouraged. For all tests, a percentage correct score was obtained as the dependent measure.

Vowel tokens were created as described in Cleary (1996). Eight vowels, four of them acoustically dissimilar or comprising a "far" vowel space (i.e., [i], [u], [a], and [æ]), and four of them acoustically similar or in a "near" vowel space (i.e., [ ], [ʊ], [Λ], and [ɛ]) were used. . Each vowel was edited from the natural speech of one male speaker producing the vowels in isolation. The recordings were made in a single-walled IAC sound booth using a Shure (SM98) microphone. A 16-bit analog-to-digital conversion of the speech was conducted using Tucker-Davis Technologies (TDT) System II equipment with a sampling rate of 22.05 kHz and an anti-aliasing filter of 10.4 kHz. The TDT system was controlled using a customized speech acquisition Program (Dedina, 1987; Hernandez, 1995). For each vowel token, 300 ms of the sample was selected from the center of the waveform. The first and last 50 ms of each sample was ramped to mimic the natural onset and offset of speech. The amplitude of each .wav file was then normalized to use 80% of the available bit space using the SoundEdit™ 16 Version 2 (1995) software program by Macromedia. Spectral analyses of the vowels were conducted using Waves+ digital signal processing software (Entropics Corporation). Table 3 displays the formant frequencies, obtained at approximately 145 ms after onset, for each of the vowel tokens.

Vowel (as in)	F0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
[i] "heed"	130	263	2206	3071
[u] "who'd"	138	287	771	2290
[æ] "had"	124	634	1767	2613
[a] "hawd"	127	604	924	2662
[Λ] "hud"	125	604	1257	2670
[ɛ] "head"	124	609	1655	2458
[ ] "hid"	131	364	1922	2586
[ʊ] "hood"	134	438	1101	2470

Table 3. Fundamental and formant frequencies for vowel stimuli (adapted from Cleary 1996).

The results from the spectral analyses were used to place the vowels into two groups, those that were acoustically dissimilar (i.e., "far vowels") and those that were acoustically similar (i.e., "near vowels") in the vowel space. Specifically, the formants were converted to Bark values using the following equation in radians from Zwicker & Terhardt (1980):

$$B = 13 \arctan (0.76f) + 3.5 \arctan (f/7.5)^2 \qquad (1)$$

where f = frequency in kHz. The differences in the Bark values of F1 and F0, and F3 and F2 for each of the vowels are plotted on Figure 1 adapted from Cleary (1996). The vowels were then categorized into two groups based on their acoustic similarity. From the figure it can be observed that the vowels [i], [u], [a], and [æ] comprise the outer quadrilateral or "far vowels," and [ ], [ʊ], [Λ], and [ɛ] comprise the inner quadrilateral or "near vowels."

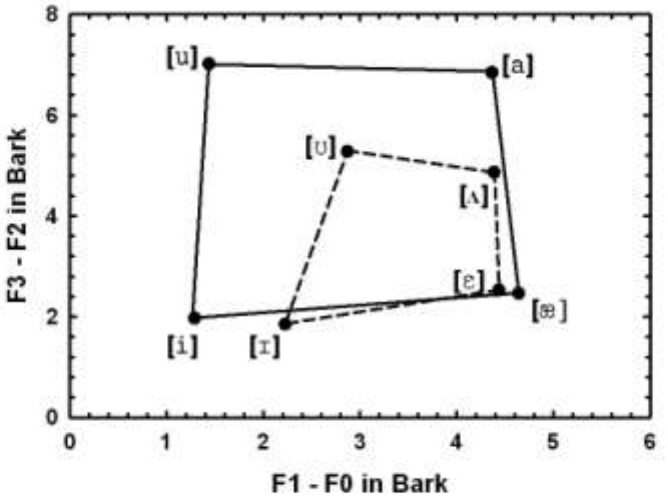

Figure 1. The differences in the Bark values of F1 and F0, and F3 and F2 for each of the vowels are plotted [adapted from Cleary (1996)]. Two groups of vowels are displayed based on their acoustic similarity. The vowels [i], [u], [a], and [æ] comprise the outer quadrilateral or far vowels, and [ ], [ʊ], [Λ], and [ɛ] comprise the inner quadrilateral or near vowels.

Both the study participants who were hearing-impaired and the participants who had normal hearing completed the vowel identification and vowel span tasks in quiet. Normal hearing adults were used as a control group. We were interested, therefore, in how normal-hearing individuals process these cues in a "natural" listening environment, and therefore, both tests were conducted in quiet. The stimuli were presented in a computerized four-choice format using a Dell 1707 FPT monitor and a MacMini 1.42 GHz Power PC G4 computer. The software programming tool, PsyScript 5.5d3, was used to generate all of the vowel tasks. The stimuli were presented to the listener at 65 dB SPL via two Advent AV570 speakers each placed at 45º azimuth.

Two vowel identification tasks, one using the four far vowels and one using the four near vowels, were administered. Each task consisted of two phases, a familiarization phase and a testing phase. In the first phase, participants were familiarized with the vowel sounds for each of the two groups of four vowels by selecting a word presented on the computer monitor in hVd format that corresponded to a particular vowel (i.e., "heed," "who'd," "hawd," "had" for the far vowels and "hid," "hood," "head," "hud" for the near vowels) and listening to the stimuli. Participants were allowed to select each vowel twice before proceeding to the testing phase. During the testing phase, the participants selected the hVd word that corresponded to the presented vowel token. A crosshair was presented on the computer screen prior to each vowel presentation to prepare the listener for the upcoming stimulus presentation. Each vowel was presented four times during the testing phase. No feedback was provided to the listeners. A percentage correct score was calculated from the raw score and used in the statistical analyses.

Data Analyses ANOVA analyses were conducted to determine differences in performance in the tasks between the CI-only and CI plus HA experimental groups. These analyses were conducted using the SPSS 16.0 statistical software. Additionally, ANCOVA models were used to examine the relationship between device (i.e., HA plus CI or CI-only), PTA, and low-frequency (LF) PTA on the outcomes from the word and sentence tasks, and the vowel identification task. Statistical mixed-effect models were also used to analyze the results from the HINT sentence test. The three different presentation levels (i.e., quiet, +10 dB SNR, and +5 dB SNR) were combined into one outcome measure using a condition variable of 3 levels. The mixed-effect model was used to account for the repeated measures from the same study participant. The ANCOVA and repeated measure mixed-effect model were conducted using the SAS for Windows (version 9.1) statistical analysis program. Finally, Pearson Correlation Coefficients were obtained using the SPSS software to determine significant relationships between the dependent and independent variables.

Results

Figure 2 displays box plots showing the results from the CNC word test for the cochlear implant recipients. Percent correct is displayed as a function of the stimulation mode (i.e., CI-only or CI+HA). For these figures and all subsequent figures, the horizontal edges of each box represent the 25th and 75th percentiles and the solid line within the box represents the median. The whiskers represent the 10th and 90th percentiles and the solid circles show the suspected outliers. The means were not significantly different between these two groups of study participants as determined by a one-way ANOVA. However, ANCOVA analyses revealed that the PTA [F(1, 37) = 9.74, p= 0.004] and the LF_PTA [F(1,37) = 4.69, p = 0.04] prior to implantation had a significant effect on the identification of isolated single-syllable words.

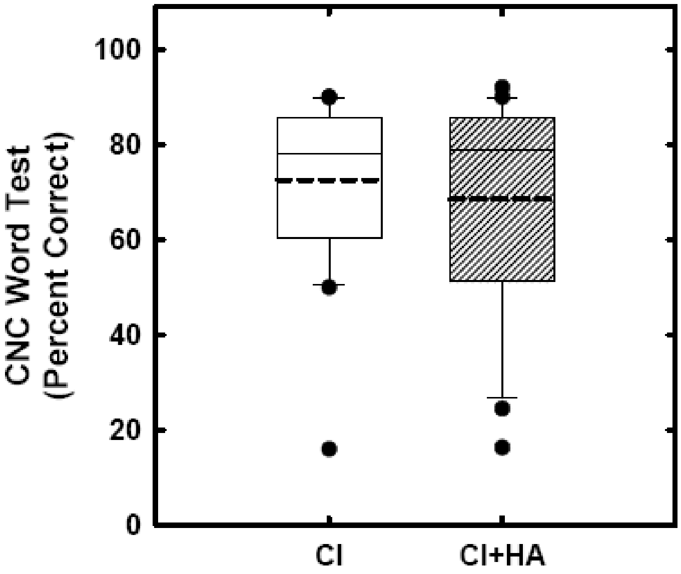

Figure 2. Box plots displaying the results from the CNC word test for the cochlear implant recipients. The percent correct is displayed as a function of the stimulation mode (i.e., CI-only or CI+HA). For these figures, and all subsequent figures, the horizontal edges of each box represent the 25th and 75th percentiles and the solid line within the box represents the median. The whiskers represent the 10th and 90th percentiles and the solid circles show the suspected outliers.

The results from the HINT sentence test are displayed using box plots presented in Figure 3. The percent correct scores are shown for the different listening conditions (i.e., quiet, +10 dB SNR, +5 dB SNR). Although the means for the HINT in quiet, HINT with +10 dB SNR, and HINT with +5 dB SNR listening conditions were similar for the two experimental groups, the medians were generally higher for the participants who used bimodal hearing (i.e., 94.8, 87.6, 82.4) in comparison to the results obtained for the unilateral cochlear implant recipients (i.e., 90.9, 83.7, 69.5). ANCOVA analyses revealed that the PTA, but not the LF_PTA or the type of device, influenced the identification of HINT sentences in quiet [F(1,37) = 5.39, p = 0.03] and in noise using a +10 dB SNR [F(1, 37) = 7.44, p = 0.01].

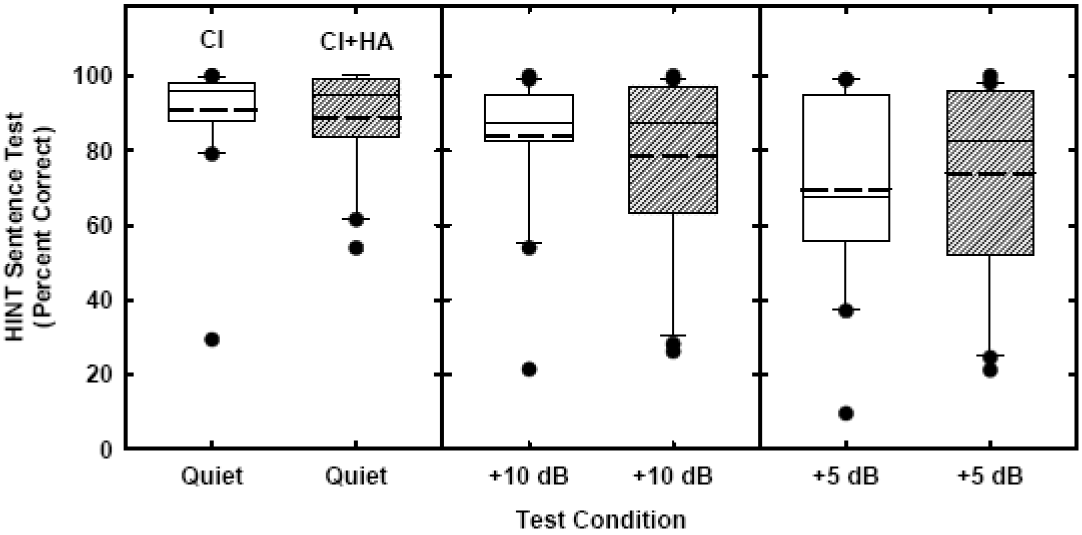

Figure 3. The results from the HINT sentence test are displayed using box plots. The percent correct scores are shown for the different listening conditions (i.e., quiet, +10 dB SNR, +5 dB SNR). The horizontal edges of each box represent the 25th and 75th percentiles and the solid line within the box represents the median. The whiskers represent the 10th and 90th percentiles and the solid circles show the suspected outliers.

In order to further explore these findings, mixed-effect models were performed. Specifically, the results from the three listening conditions were combined into one outcome measure with a condition variable of three levels (i.e., Quiet, +10 dB SNR, and +5 dB SNR). The mixed-effect model was used to account for the repeated measures from the same study participant. Two-way interactions (i.e., condition variable with the type of device, and the condition variable with the PTA) were not significant. With this model, the PTA [F(1,78) = 5.68, p<0.02] and presentation condition [F(2,78) = 29.31, p<0.0001] were both significantly related to the performance on this sentence test. A pair-wise Bonferroni post hoc analysis indicated that the performance in the quiet listening condition was significantly higher than the performance in the +5 dB SNR condition (t-value$_{78}$ = -7.65, p<0.0001), and the +10 dB SNR condition (t-value$_{78}$ = -3.66, p = 0.001). In addition, performance in the +10 dB SNR condition was significantly better than performance in the +5 dB SNR listening condition (t-value$_{78}$ = 4.00, p = 0.0004).

The vowel identification results are presented in Figures 4. The percent correct identification is shown as a function of the test group (i.e., CI, CI+HA, NH – normal hearing). As expected, recognition of the near vowels was much poorer compared to the far vowels for both groups of cochlear implant recipients. A one-way ANOVA revealed significant differences in vowel identification of the far and near vowels between these two groups [F(1,78) = 35.6, p<0.0001]. The ANCOVA procedure indicated that the effect of the type of device, the LF_PTA and the PTA were not significant.

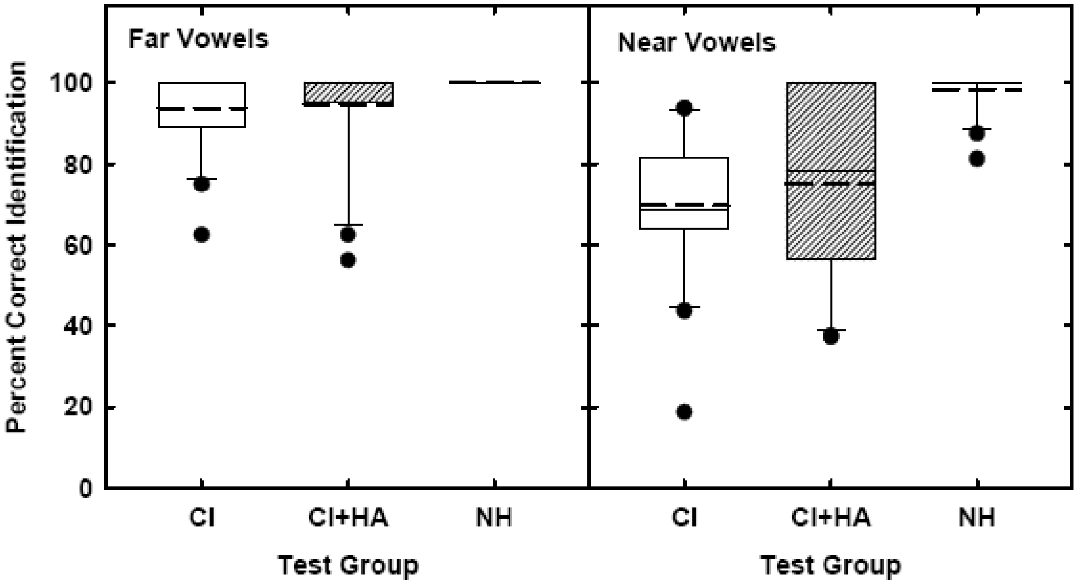

Figure 4. The vowel identification results are presented. The percent correct identification is shown as a function of the test group (i.e., CI, CI+HA, NH). Again, the horizontal edges of each box represent the 25th and 75th percentiles and the solid line within the box represents the median. The whiskers represent the 10th and 90th percentiles and the solid circles show the suspected outliers.

Although the identification of acoustically similar vowels for the participants with normal-hearing was slightly poorer than the identification of acoustically dissimilar vowels the difference was not found to be significant. Differences in performance between the adults with normal hearing and the cochlear implant recipients, however, were statistically significant. A two-way ANOVA was performed using the listening mode (i.e., NH, CI, CI+HA) and acoustic vowel space (i.e., far, near) as factors. The results revealed a significant difference in listening mode [F(2, 114) = 18.8, p<0.0001], vowel space [F(1,114) = 36.9, p<0.0001], and a significant interaction between these variables [F(2,114) = 7.7, p<0.001]. Post hoc Bonferroni pairwise testing indicated that the adults with normal hearing performed significantly better than the adults who used one cochlear implant (p<0.0001) as well as the adults who used a cochlear implant and a hearing aid (p<0.0001). It is possible, therefore, that the differences in performance that exist between the adults with normal hearing and the adults who use a cochlear implant can be partially attributed to differences in hearing acuity.

Discussion

This first study examined the effect that acoustic information provided through a contralateral hearing aid might have on the identification of spoken words and sentences for individuals who use cochlear implants. Although previous studies have examined the benefit of contralateral hearing aid use with a cochlear implant through the examination of within-subject analyses, the intention of this study was to examine the benefit of hearing aid use using a between-subject design. Typically, the findings from the within-subject design studies have suggested that the use of a hearing aid is beneficial for individuals who use bimodal hearing (Ching, et al., 2004; Dorman, et al., 2008; Dunn, et al., 2005). The findings from the current study suggest that the use of a contralateral hearing aid does not provide sufficient additional spectral information to help with the identification of isolated words, isolated vowels, and sentences presented in quiet and in noise. The results from this study, therefore, suggest that greater access to spectral cues should be provided to profoundly deaf adults through advances in cochlear implant technology rather than through the use of a contralateral hearing aid.

The findings did reveal, however, that acoustic similarity did affect vowel identification for individuals who use cochlear implants. Vowels that were acoustically similar were more poorly identified than acoustically dissimilar vowels. Neither the degree of residual hearing prior to implantation nor the use of bimodal hearing significantly affected these outcomes.

The use of bimodal hearing also did not help when identifying words and sentences presented in quiet or in the presence of background noise. For the CNC words, the degree of residual hearing had a greater impact on word identification than did the use of bimodal hearing. Additionally, for the HINT sentences, a significant effect with the degree of background noise was observed. The results for the quiet condition were significantly higher than the results for the +10 dB SNR condition, which were in turn significantly higher than the scores for the +5 dB SNR listening condition. Although the use of bimodal hearing did not have a significant effect on the performance, the degree of residual hearing prior to implantation had a significant impact on the outcomes. Greater residual hearing was associated with higher speech recognition scores.

Experiment 2

To further explore the influence that the acoustic similarity of speech signals had on speech perception performance, short-term phonological memory tasks were completed in this second experiment. Immediate memory skills have been previously examined with groups of words, consonants, and vowels that contain acoustically similar tokens. Research has found that sequences of words, vowels, and consonants that are acoustically similar are recalled more poorly than items that are acoustically discriminable (Cleary, 1996; Conrad & Hull, 1964; Wickelgren, 1965). Additionally, work conducted by Baddeley (1968) and Drewnowski (1980) found that in serial recall tasks, individuals encode and store acoustically similar items in short-term memory, but do not retrieve these items as efficiently compared to acoustically dissimilar items. The effect that immediate phonological memory has on speech perception skills within normal-hearing listeners and profoundly deaf individuals who use cochlear implants is unknown. The objective of this study, therefore, was to determine the effect that added spectral cues provided via a hearing aid might have on immediate phonological memory skills, and also to evaluate the impact that these skills have on speech perception.

We expected that individuals who had access to additional spectral cues via a hearing aid would be able to store, retrieve and recall longer sequences of vowels compared to individuals who did not use a hearing aid. We also predicted that the ability to retrieve previously stored phonemes would be related to overall speech perception performance. If the storage and retrieval of information does not effectively take place, the ability to discriminate among and identify presented speech tokens should be reduced.

Method

Participants The same three groups of study participants (i.e, 20 CI-only adults, 20 CI+HA adults, 20 NH adults) that completed experiment 1 also completed experiment 2. The demographic information for both groups of cochlear implant recipients is displayed in Tables 1 and 2.

Stimuli and Procedures The experimental procedure for the vowel sequence task consisted of four different phases: 1) "Familiarization," 2) "Learn," 3) "Practice," and 4) "Sequence." During the Familiarization phase, participants heard an isolated vowel and saw the corresponding word in hVd context on the computer screen flash. During the learning stage the participants were asked to select the flashing word on computer monitor which corresponded to the presented vowel. In the "Practice" condition, the participants selected the word on the screen that matched the vowel they heard. If they selected the incorrect corresponding word, the correct word flashed (i.e., feedback was provided). Finally, during the "sequence" condition the listeners were instructed to reproduce the sequence of presented sounds with no feedback by selecting the appropriate hVd token on the computer screen. The stimuli were presented at a rate of one per second. Prior to each sequence presentation a crosshair was displayed on the screen to signal the next vowel sequence presentation. An adaptive procedure was used for the sequence. Specifically, when two sequences of vowels were correctly reproduced, the next sequence in the test was increased by one vowel. The task ended after two incorrect vowel sequence reproductions. This task was completed twice, one for each of the two vowel groups. The length of the longest vowel sequence that was correctly reproduced twice was used for data analysis.

Data Analysis The data were analyzed using similar procedures as those completed with the data obtained in Experiment 1. Specifically, ANOVA and ANCOVA procedures were conducted to determine the impact of bimodal hearing on immediate phonological memory. In addition, Pearson Correlation Coefficients were obtained using the speech perception data from Experiment 1 and the short-term memory data from this second experiment.

Results

Figure 5 shows the results obtained from the sequence memory tasks. For each test group, the longest sequence of vowels that was correctly recalled at least twice is presented. As shown in this figure, the sequences of acoustically similar vowels (i.e., near vowels) were shorter than sequences of acoustically dissimilar vowels (i.e., far vowels) for all groups. That is, fewer near vowels than far vowels were recalled when presented in a sequence format. The mean length of the span for the far and near vowels was 3.7 and 1.8 for the CI-only study participants, respectively, and 3.8 and 2.3 for the CI + HA participants, respectively. When analyzing the combined results, a one-way ANOVA revealed a significant difference in the ability to successfully recall acoustically dissimilar vowels compared to acoustically similar vowels [$F(1, 78) = 34.4$, $p<0.0001$]. The ANCOVA analyses revealed that the device-type, the LF_PTA and the PTA did not have a significant effect on the sequence span.

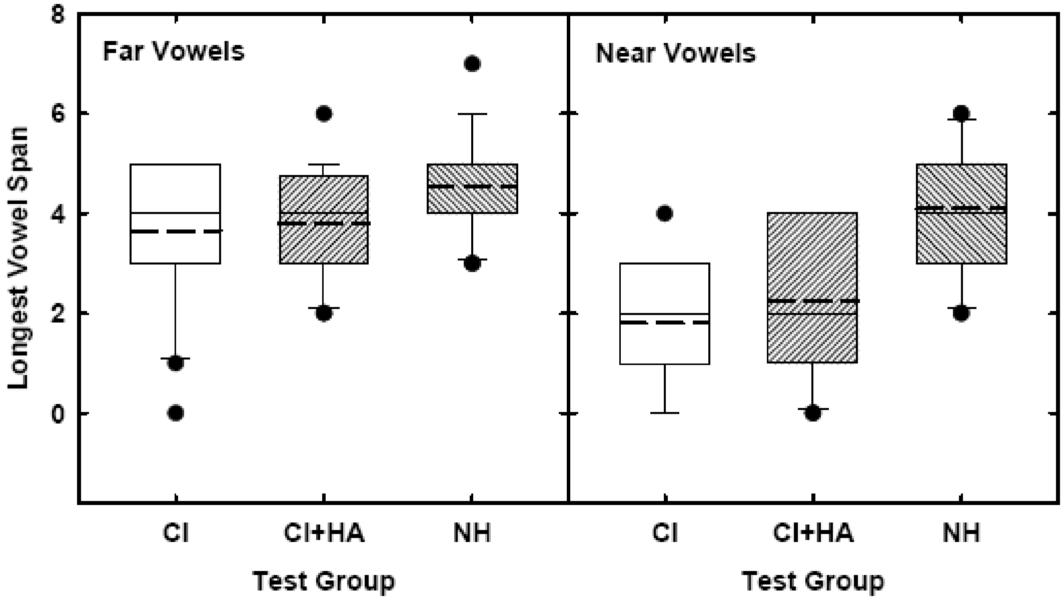

Figure 5. Results obtained from the immediate phonological memory task are displayed in this figure. For each test group, the longest span of vowels that was correctly recalled at least twice is presented. The horizontal edges of each box represent the 25th and 75th percentiles and the solid line within the box represents the median. The whiskers represent the 10th and 90th percentiles and the solid circles show the suspected outliers.

The data from the participants with normal-hearing revealed that the mean vowel span length for the far vowels was 4.6 and 4.1 for the near vowels. A one-way ANOVA indicated that this difference in performance was not statistically significant. However, these results were significantly better than the results obtained by study participants who used one cochlear implant and by those who used bimodal hearing. A two-way ANOVA with listening mode (i.e., normal hearing, CI-only, CI+HA) and vowel acoustic similarity (i.e., far, near) indicated significant findings for the listening mode [F(2,114) = 18.8, p<0.0001] and vowel acoustic similarity [F(1,114) = 32.1, p<0.0001]. A significant interaction also was found [F(2,114) = 3.53, p = 0.03]. Post hoc Bonferroni pairwise analyses revealed that the normal hearing listeners recalled significantly longer sequences of near and far vowels than both the participants with unilateral cochlear implants (p<0.0001), and than those using a cochlear implant and a hearing aid (p<0.0001). These findings also suggest that the differences in hearing acuity can at least partially explain the differences in immediate memory between listeners with normal hearing and those who use cochlear implants.

Another goal of this study was to examine how the ability to store and retrieve phonemes was related to overall speech perception performance for individuals who used cochlear implants. Table 4 displays the Pearson correlation analyses for the mean speech perception data (i.e., HINT sentences and CNC words) and the mean vowel identification and vowel span data for all 40 study participants. The analyses indicated that the vowel identification and the immediate memory tasks were moderately correlated with the speech perception tasks. All of the correlation analyses were found to be significant at the 0.05 level or better with the exception of the correlation between the far vowel span and the CNC word results (italized).

	Identification		Vowel Span	
CI-only and CI+HA (N = 40)	Far Vowels	Near Vowels	Far Vowels	Near Vowels
CNC word	0.65 (p<0.0001)	0.42 (p = 0.007)	***0.29*** ***(p = 0.07)***	0.54 (p<0.0001)
HINT quiet	0.67 (p<0.0001)	0.42 (p = 0.007)	0.39 (p = 0.01)	0.42 (p = 0.007)
HINT +10	0.64 (p<0.0001)	0.40 (p = 0.01)	0.33 (p = 0.04)	0.44 (p = 0.004)
HINT +5	0.62 (p<0.0001)	0.40 (p = 0.01)	0.40 (p = 0.01)	0.46 (p = 0.003)

Table 4. Pearson Correlation analyses for speech perception and vowels tasks for both experimental groups. Note: CNC word is Consonant Nucleus Consonant isolated word identification task; HINT is Hearing in Noise Test sentences only presented in quiet, +10 dB SNR, and +5 dB SNR; Acoustically dissimilar vowels (Far Vowels) and acoustically similar vowels (Near Vowels); Vowel Span: retrieval of sequences of vowels.

Table 5 displays the correlation analyses for each experimental group (i.e., the CI-only and CI+HA groups). These analyses indicated that the correlations for the unilateral cochlear implant group were stronger than the correlations for the bimodal group. This finding might be partially attributed to the fact that the medians for the data from the HINT sentence tests in noise were generally higher for the listeners who used bimodal hearing than the unilateral cochlear implant recipients. If a large percentage of the bimodal group were performing closer to ceiling in comparison to the performance from the CI-only group, obtaining meaningful correlations with the vowel tasks would, therefore, be challenging. If the ceiling effect was reduced, a stronger relationship between the speech perception tasks and the vowel tasks might have been obtained for the study participants who used bimodal hearing.

	Identification		Vowel Span	
CI-only (N = 20)	Far Vowels	Near Vowels	Far Vowels	Near Vowels
CNC word	0.60 p = 0.005	0.62 p = 0.003	0.50 p = 0.03	0.70 p = 0.001
HINT quiet	0.56 p = 0.01	0.61 p = 0.004	0.50 p = 0.03	0.50 p = 0.03
HINT +10	0.78 p<0.0001	0.55 p = 0.01	0.63 p = 0.003	0.51 p = 0.02
HINT +5	0.65 p = 0.002	0.51 p = 0.02	0.58 p = 0.007	0.58 p = 0.008
CI + HA (N = 20)				
CNC word	0.69 p = 0.001	0.32 p = 0.17	0.09 p = 0.71	0.48 p = 0.34
HINT quiet	0.80 p<0.0001	0.29 p = 0.22	0.24 p = 0.32	0.40 p = 0.08
HINT +10	0.56 p = 0.01	0.33 p = 0.16	0.06 p = 0.81	0.45 p = 0.05
HINT +5	0.60 p = 0.005	0.32 p = 0.17	0.16 p = 0.51	0.35 p = 0.13

Table 5. Pearson Correlation results for speech perception and vowel tasks separated by experimental group.

Discussion

For all study participants, longer sequences of vowels were obtained for the acoustically distinct vowels than for the acoustically similar vowels. The added spectral information provided through a hearing aid was not sufficient for the bimodal listeners to significantly improve the encoding of the near and far vowels compared to listeners who used one cochlear implant. Individuals who use both a cochlear implant and contralateral hearing aid, therefore, are not able to store and retrieve sequences of vowels with more proficiency than individuals who use one cochlear implant.

The results from Experiment 2 suggested that the ability to perceive isolated words and connected speech is dependent on the ability to recognize individual components of speech (i.e., vowels) and the processing of these vowels for storage and retrieval in immediate memory. When examining all study participants, the scores from the vowel tasks (i.e., vowel identification and vowel span tasks) were found to moderately correlate with the recognition of isolated words and words in sentences. Separate correlation analyses revealed that the vowel identification and the vowel span tasks were significantly correlated with the speech perception tasks for the unilateral cochlear implant recipients but generally not for the adults who used bimodal hearing. This finding can be partially attributed to the ceiling effects for the HINT sentences in the bimodal group of study participants.

General Discussion

Bimodal Hearing and Listening in Noise

Previous research has suggested that the use of bimodal hearing can be beneficial for identifying speech in noise (Dorman, et al., 2008; Dunn, et al., 2005). This earlier work examined one group of adults who used both a cochlear implant and a contralateral hearing aid and revealed that performance in noise was much improved when both the cochlear implant and hearing aid were used compared to the use of each device alone. The present study examined two groups of study participants – those who used bimodal hearing and those who used unilateral cochlear implants – and, although the median performance with the HINT sentence test for the adults who used bimodal hearing was generally higher in comparison to the performance for individuals who used unilateral cochlear implants, no significant differences between the groups were observed. The differences in outcomes across studies could be due to differences in methodology. The study participants in both the Dorman, et al. (2008) and Dunn, et al. (2005) studies were asked to identify speech in quiet and noisy listening environments when using either their hearing aid or cochlear implant only. The participants, therefore, might have been at a disadvantage when identifying speech in these conditions which could have artificially affected the outcomes. The participants in the current study identified speech using their everyday listening devices, and therefore, the findings from the current study might be more representative of listening abilities in noise.

Degree of Residual Hearing and Speech Perception

Although this study failed to demonstrate that the use of bimodal hearing significantly improves listening in noise, the identification of vowels, or immediate phonological memory for sequences of vowels, the findings did reveal that the degree of residual hearing prior to implantation did have a significant impact on the ability to perceive speech in noise. Previous work by Francis, et al. (2005) demonstrated that after 12 and 24 months of cochlear implant use, the ability to recognize HINT sentences in noise was significantly influenced by the PTA (i.e., pure-tone frequency threshold average of 500 Hz, 1000 Hz, and 2000 Hz) in the better hearing non-implanted ear. In addition, Armstrong, et al.(1997) hypothesized that the study participants who had better speech understanding skills in noise did so because they had more residual hearing than participants who did not do as well in noise.

Additionally, the results from this study failed to show that residual hearing at low to mid frequencies was beneficial for listening in noise. The behavioral threshold average of 250 Hz, 500 Hz, and 1000 Hz in the better hearing ear prior to implantation was calculated for each participant and used as a predictor in the ANCOVA analyses. This LF_PTA was not a significant variable in predicting how well the study participants would perform in noise. Ching, et al. (2004) and Gifford, et al. (2007) reported similar outcomes in their studies. Although the statistical analyses performed in those studies were slightly different from the current study (those studies used correlation analyses and this study used analysis of variance) all of the studies have reported that the degree of residual hearing in the low-frequencies did not aid speech understanding in noise.

Acoustic Similarity and Speech Understanding in Adults who use Cochlear Implants

When identifying sequences of isolated vowels, the results from this study suggest that their acoustic similarity influences how successfully they are recognized by cochlear implant recipients. Specifically, a group of isolated vowels that are acoustically discriminable will be more successfully identified than a group of isolated vowels that are acoustically similar. In addition, the ability to store and recall a sequence of vowels from immediate memory was also found to be highly dependent upon the acoustic distinctness of the presented vowels. Longer sequences of acoustically distinct vowels (i.e., far vowels) were recalled in comparison to sequences of acoustically similar vowels (i.e., near vowels). These findings were in agreement with previous results suggesting that acoustic similarity affects the ability to encode, maintain, and retrieve sequences of vowels from short-term memory (Baddeley, 1968; Cleary, 1996; Drewnowski, 1980).

As opposed to previous studies that examined the effect of acoustic similarity on the recall of vowels sequences in individuals with normal hearing, this current study examined a group of adults who were hearing-impaired and used either one cochlear implant or a cochlear implant and a contralateral hearing aid, in addition to a group of adults with normal hearing. The performance on the vowel identification tasks and the immediate phonological memory tasks for the adults who used cochlear implants, either a CI-only or a CI+HA, was significantly poorer than the performance by the adults with normal hearing. This finding suggests that individuals who have a profound hearing loss that has been partially corrected through a cochlear implant are still at a disadvantage, compared to adults with normal hearing, when processing (i.e., storing, retrieving, and recalling) sequences of isolated vowels that are acoustically similar. The lack of auditory access to all of the spectral aspects of speech affected not only the identification of vowels, but also the retrieval of sequences of vowels in immediate memory.

The results of this study also suggested that the ability to successfully perceive and process groups of acoustically similar and dissimilar vowels is related to overall speech perception skills. Moderate correlations were obtained with the identification of vowels and word understanding, and with immediate memory and speech recognition. These relationships were robust when combining the data from the two groups of study participants, and were much stronger for the adults with unilateral cochlear implants compared to the adults who used bimodal hearing. These findings might be attributed to the greater power associated with the data from larger numbers of study participants. Additionally, the lack of a normal distribution of the scores obtained on the HINT sentence tests in noise by the bimodal group of listeners could have partially attributed to the absence of significant correlations with the data from the vowel tasks.

For post-lingual adults who use cochlear implants, therefore, it is important that the successful processing of vowels occurs in order for speech to be successfully perceived. These findings also support the data from Cole, et al. (1996) and Kewley-Port, et al. (2007) who reported that sentences were much more difficult to identify when the vowels in the sentences, as opposed to the consonants, were replaced with noise. Although this study did not replicate the methodology of Cole, et al. (1996) and Kewley-Port, et al. (2007), the data from all three studies confirm the important role of vowels for speech perception.

Conclusion

The results of this study suggest that the use of bimodal hearing did not provide a significant advantage for the identification of words and sentences in quiet and in noise, the identification of groups of acoustically similar and dissimilar vowels, and immediate phonological memory in comparison to the use of one implant alone. The degree of residual hearing prior to implantation, however, was determined to be in important factor for the identification of words and sentences. Additionally, moderate correlations were found with the vowel identification and speech perception skills, and also with immediate memory and speech perception abilities.

Acknowledgements

Funding for this study was provided by grants from the NIH/NIDCD (R03 DC008383 to the first author and T32 DC00012 to the third author) and by the Psi Iota Xi Philanthropic Organization. We would also like to thank Christian Rosado who helped with data collection.

References

Armstrong, M., Pegg, P., James, C., & Blamey, P. J. (1997). Speech perception in noise with implant and hearing aid. *American Journal of Otology, 18*, S140-141.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology, 20*, 249-264.

Chang, Y. P., & Fu, Q. J. (2006). Effects of talker variability on vowel recognition in cochlear implants. *Journal of Speech, Language, and Hearing Research, 49*, 1331-1341.

Ching, T. Y. C., Incerti, P., & Hill, M. (2004). Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear & Hearing, 25*, 9-21.

Cleary, M. (1996). *Measures of phonological memory span for sounds differing in discriminability: Some preliminary findings. Research on Spoken Language Processing Report No. 21*. Bloomington, IN: Indiana University.

Cole, R. A., Yonghong, Y., Mak, B., Fanty, M., & Bailey, T. (1996). *The contribution of consonants versus vowels to word recognition in fluent speech.* Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, GA.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429-432.

Dedina, M. J. (1987). *SAP: A speech acquisition program for the SRL-VAX* (No. 13 Research on Speech Perception Progress Report). Bloomington, IN: Speech Research Laboratory, Indiana University.

Dorman, M. F., Gifford, R. H., Spahr, A. J., & McKarns, S. A. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiology & Neurotology, 13*, 105-112.

Drewnowski, A. (1980). Memory functions for vowels and consonants: A reinterpretation of acoustic similarity effects. *Journal of Verbal Learning and Verbal Behavior, 19*, 176-193.

Dunn, C. C., Tyler, R. S., & Witt, S. A. (2005). Benefit of wearing a hearing aid on the unimplanted ear in adult users of a cochlear implant. *Journal of Speech, Language, and Hearing Research, 48*, 668-680.

Francis, H. W., Yeagle, J. D., Bowditch, S., & Niparko, J. K. (2005). Cochlear implant outcome is not influenced by the choice of ear. *Ear & Hearing, 26 (Suppl)*, 7S-16S.

Freyman, R. L., & Nerbonne, G. P. (1989). The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants. *Journal of Speech and Hearing Research, 32*, 524-535.

Fu, Q. J., Shannon, R. V., & Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *Journal of the Acoustical Society of America, 104*, 3586-3596.

Gantz, B. J., Woodworth, G. G., Knutson, J. F., Abbas, P. J., & Tyler, R. S. (1993). Multivariate predictors of audiological success with multichannel cochlear implants. *Annals of Otology Rhinology and Laryngology, 102*(12), 909-916.

Gifford, R. H., Dorman, M. F., McKarns, S. A., & Spahr, A. J. (2007). Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. *Journal of Speech, Language, and Hearing Research, 50*, 835-843.

Hernandez, L. R. (1995). *Current computer facilities in the Speech Research Laboratory* (No. 20 Research on Spoken Language Processing Progress Report). Bloomington, IN: Speech Research Laboratory, Indiana University.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America, 122*, 2365-2375.

Nie, K., Barco, A., & Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear & Hearing, 27*, 208-217.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*, 1085-1099.

Peterson, G. E., & Lehiste, I. (1962). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders, 27*, 62-65.

Preves, D. A., Fortune, T. W., Woodruff, B., & Newton, J. (1991). Strategies for enhancing the consonant to vowel intensity ratio with In the Ear hearing aids. *Ear & Hearing, 12*, 139S-153S.

Qin, M. K., & Oxenham, A. J. (2006). Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. *Journal of the Acoustical Society of America, 119*, 2417-2426.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., & Henry, B. A. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *Journal of the Acoustical Society of America, 115*, 1729-1735.

Vongphoe, M., & Zeng, F. G. (2005). Speaker recognition with temporal cues in acoustic and electric hearing. *Journal of the Acoustical Society of America, 118*, 1055-1061.

Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *The American Journal of Psychology, 78*, 567-574.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America, 68*, 1523-1525.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Rule Reliability in Natural and Artificial Grammar: The Case of Velar Palatalization[1]

Vsevolod Kapatsinski

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] Many thanks to the NIH for financial support through grants DC00111 and DC00012. I would also like to thank Bill Croft, Ken de Jong, Bruce Hayes, Janet Pierrehumbert, and David Pisoni for helpful feedback.

Rule Reliability in Natural and Artificial Grammar: The Case of Velar Palatalization

Abstract: Russian velar palatalization changes velars into alveopalatals before certain suffixes, including the stem extension –i and the diminutive suffixes –ok and –ek/ik. While velar palatalization always applies before the relevant suffixes in the established lexicon, as depicted by dictionaries, it often fails with nonce loanwords before –i and –ik but not before –ok or –ek. A model of rule induction and weighting (the Rule-based Learner, developed by Albright and Hayes 2003) is trained on the established lexicon of Russian, in which velar palatalization is exceptionless, and tested on new borrowings. Despite the fact that velar palatalization is exceptionless in the training set for every suffix, it is correctly predicted to often fail with novel words before –i and –ik but not before –ek or –ok based on information in the lexicon. This success can be traced to the model's weighting of competing rules according to their reliability. Reliability-driven competition between rules is shown to predict that a morphophonological rule will fail if the triggering suffix comes to attach to inputs that are not eligible to undergo the rule. This prediction is confirmed in an artificial grammar learning experiment. A method for distinguishing between source- and product-oriented mental grammars is developed and product-oriented generalizations are shown to be unable to account for the data from the examined artificial grammar learning paradigm (Bybee & Newman 1995). The influence of the learning paradigm on the shape of the learned grammar is discussed. Finally, the winning model (the Rule-based Learner) is shown to succeed only if the suffix and the stem shape are chosen simultaneously, as opposed to the suffix being chosen first and then triggering or failing to trigger a stem change and if the choice between competing rules is stochastic.

Introduction

One puzzling phenomenon in language change is the loss of productivity by morphophonemic alternations. Why would an alternation start accumulating exceptions and stop being extended to new words entering the language despite starting out with no exceptions and an abundance of examples supporting it in the lexicon? A particularly interesting historical development happens when an alternation has no exceptions in the lexicon but is not extended fully to new words entering the language. Thus, the alternation loses productivity while not gaining exceptions.[2] In the present paper I will show that a particular alternation, velar palatalization (k→tʃ, g→ʒ), has lost productivity before some Russian suffixes but not others. Velar palatalization has no exceptions in the Russian lexicon, as depicted by dictionaries, before the verbal stem extension –i and the diminutive suffixes –ok and –ek/ik. However, with recent loanwords found in the discourse of Russian-speaking Internet users, velar palatalization fails about 50% of the time before –i and –ik while remaining fully productive before –ek and –ok. Thus Russian provides an opportunity to examine the factors influencing productivity by examining what makes velar palatalization lose productivity before –i and –ik but not before –ek or –ok.

The Rule-Based Learner (RBL, Albright and Hayes 2003) is a computational model that induces rules from a lexicon and weights them relative to each other. The RBL proposes that the productivity of a linguistic rule is determined by its estimated reliability relative to other rules that can apply to the same

[2] Later stages of productivity loss have been documented in wug tests by Zimmer (1969) and Zuraw (2000).

input forms.[3] For instance, suppose we have a language where the velar consonant [k] at the end of a singular stem changes into the alveopalatal [tʃ] when followed by a plural suffix [i]. This singular-plural mapping can be stated as the rule k#→tʃi#. The reliability of a rule is defined by the number of words to which the rule applies divided by the number of words to which it could apply. In the case of k#→tʃi#, this is the number of singular-plural pairs in which a final [k] in the singular corresponds to [tʃi] in the plural divided by the number of singulars that end in [k]. If all singulars that end in [k] correspond to plurals that end in [tʃi], the rule is completely reliable. The more reliable a rule is, the more productive it is expected to be. The present paper is intended to show that reliability-driven competition between rules predicts that a morphophonological rule will lose productivity if the triggering suffix comes to attach mostly to inputs that cannot undergo the rule. This prediction is confirmed both by Russian loanword adaptation data, where –i and –ik are the suffixes that tend not to attach to velar-final inputs whereas –ok and –ek are favored by velar-final inputs, and by data from artificial grammar learning. In both paradigms, the degree of productivity of a palatalizing rule like k#→tʃi is explained by its reliability relative to the more general rule like C#→Ci, which competes with k#→tʃi for velar-final inputs and stipulates that the consonant remains unchanged.

While the RBL is used to model the data in the present paper, it is used only as a representative example of models that embody the hypothesis of statistically resolved competition between source-oriented generalizations. Source-oriented generalizations specify a mapping between a specific category of inputs to a specific category of outputs, e.g., k→tʃi, rather than simply describing what the outputs should be like (Bybee 2001:126-9). This assumption is also made in Analogical Modeling of Language (Skousen 1989), which relies on rule type frequency instead of reliability,[4] and the account of productivity of English velar softening developed in Pierrehumbert (2006). Product-oriented generalizations specify only the shape of the output, e.g., 'plurals end in /tʃi/', and are derived via generalization over outputs. A novel product-oriented account that could account for the data is presented in section 3.4. The product-oriented account is argued to require reliability-driven competition between *conditional* product-oriented generalizations, such as 'if the plural ends in –i, the preceding consonant must be /tʃ/', and paradigm uniformity constraints.

Loanword Adaptation in Russian

The pattern in the lexicon

If one looks at a dictionary of modern Russian, velar palatalization appears to involve several exceptionless morphophonological rules, which can be stated simply as "velars become alveopalatals before the derivational suffixes X" where the relevant derivational suffixes either begin with a front vowel or used to begin with a front vowel historically. For the purposes of this article, we will be concentrating on Russian verbs with the highly productive stem extension –i, and the diminutive suffixes for masculine nouns, -ik/ek and -ok, which obligatorily trigger velar palatalization in the lexicon, as depicted by dictionaries (e.g., Levikova, 2003; Sheveleva et al., 1974).

[3] Throughout this paper, by 'input' and 'output' I mean the input to the rule and the output of the rule, which could both be either underlying or surface forms. A rule here is an input-output mapping, in which both the input and the output are classical categories.

[4] While Analogical Modeling of Language is not usually described as rule-based, (supra)context-outcome pairings are input-output mappings in which the input and the output are classical categories.

Example (1) shows that in Russian verbs are derived from consonant-final nouns by adding the stem extension (in this case /i/) followed by verbal inflection (in this case the infinitival marker t^j). As shown in (1), velars at the ends of noun roots change into alveopalatals when a verb is derived from the root. This does not happen with all stem extensions, as evidenced by the existence of Russian verbs like n^{j}ux+a+t^j, plak+a+t^j, and stalk+iva+t^j, but it always happens with the stem extension -i.

(1)

k → tʃi	
klok	klotʃ+i+t^j
durak	duratʃ+i+t^j
polk	poltʃ+i+t^j
jamʃtʃik	jamʃtʃitʃ+i+t^j
g → ʒɨ	
flag	flaʒ+i+t^j
dolg	dolʒ+i+t^j
x → ʃɨ	
grex	greʃ+i+t^j

The mappings between velar consonants and the corresponding alveopalatals are constant across Russian. Thus, if velars change into alveopalatals in some context, /k/ always becomes [tʃ], /g/ becomes [ʒ], and /x/ becomes /ʃ/. The Russian phone inventory does not contain [dʒ]. The phone [ɨ] cannot follow velars or [tʃ] while the phone [i] cannot follow [ʃ] or [ʒ]. Whether [i] and [ɨ] are allophones of /i/ and chosen during a separate allophone selection stage or separate stem extensions does not influence the qualitative results presented here. The reported graphs are based on a model that treats the choice between [i] and [ɨ] as happening after the morphophonological competition modeled.

In the Russian lexicon, -a is favored over -i by velar-final roots while -i is favored elsewhere. The distribution in the diminutive system is quite different. Only masculine diminutive suffixes will be considered for the purposes of this paper, because the loaned English nouns end in a consonant, consequently being adopted into the masculine gender. There are three highly productive masculine diminutive suffix morphs, -ik, -ek, and -ok. The morphs –ek and –ik are in complementary distribution in the established lexicon and thus can be considered allomorphs of a single morpheme. The suffixes that trigger palatalization in the lexicon, -ok and -ek, are heavily favored by velar-final nouns, with -ek attaching only to velar-final bases. The suffix -ik, on the other hand, does not attach to velar-final bases, thus one could argue that the Russian lexicon provides no evidence in whether –ik would trigger or fail to trigger velar-palatalization if it were to be attached to a velar-final base, although I will argue that the lexicon does in fact provide the relevant information and Russian speakers use this information in loanword adaptation.[5] Examples are shown in (2).

[5] Since -ek and -ik are unstressed, they have the same phonetic realization, the choice between them may be part of orthography. However, the answer to the question of whether the choice is made in orthography or in phonology is not relevant to the modeling of output stem shape as long as the choice of the allomorph follows the decision on whether to change the stem.

(2)

lug	→	luˈʒok		
luk	→	luˈtʃok		
lutʃ	→	ˈlutʃik		
ˈfartuk	→	ˈfartutʃek		
kaˈbluk	→	kabluˈtʃok		
tʃeloˈvek	→	tʃeloˈvetʃek		
rog	→	roˈʒok		
noʒ	→	ˈnoʒɨk	→	ˈnoʒɨtʃek
ʃag	→	ʃaˈʒok	→	ʃaˈʒotʃek
tʃas	→	tʃaˈsok		
tʃas	→	ˈtʃasik		

Methods

Data collection. When an English verb is borrowed into Russian, it must be assigned a stem extension. In order to get a sample of such borrowings, I took all verbs found in the British National Corpus retrieved by searching for "*x.[vvi]" in the online interface provided by Mark Davies (http://corpus.byu.edu/bnc/) where 'x' is any letter. The resulting verbs were transliterated into Cyrillic.

For each verb, possible Russian infinitival forms were derived. For instance, if the English verb is *lock*, some possible Russian infinitives are /lotʃit^j/, /lokit^j/, /lokat^j/, /lokovat^j/ and /lokirovat^j/. Verbs for which an established Russian form already existed (e.g., format > /formatirovat^j/) were excluded. Existence was determined by the occurrence in either the Reverse Dictionary of Russian (Sheveleva, 1974), Big Dictionary of Youth Slang (Levikova, 2003), or the present author's memory. This yielded 472 different verbs. For 56 of them, the final consonant of the English form was a velar, for 99 it was a labial, and for 317 it was a coronal. In the case of the nouns, all possible English monosyllables ending in /k/ or /g/ were created and transliterated into Russian manually. Then possible diminutive forms were created from them and submitted to Google. An additional sample of non-velar-final nouns was then created by matching the distribution of final consonant types in terms of manner and voicing and preceding vowels in the sample of velar-final nouns.

The frequencies of the possible infinitives and nominative diminutives on the web were determined by clicking through the pages of results returned by Google to eliminate identical tokens and to allow Google to 'eliminate similar pages', which increases speaker diversity by eliminating results that come from the same server, e.g., different pages from the same bulletin board. In addition, clicking through is necessary when one of the possible forms has a homonym.

Finally, to have a reasonably reliable estimate of the likelihood of failure of velar palatalization before –i for each verb, velar-final verbs and nouns that had 10 or fewer tokens containing the palatalizing suffixes were excluded from the sample. This yielded 36 velar-final verbs and 19 velar-final nouns that could undergo velar palatalization and had a reasonably large number of tokens containing the relevant suffixes.

Modeling

Introducing the Rule-Based Learner. The Rule-based Learner (Albright and Hayes 2003) is a computational model of rule induction and weighting. The model starts with a set of morphologically related word pairs as in (3).

(3)

mot	motat^j
tʃmok	tʃmokat^j
drug	druʒit^j
krug	kruʒit^j
golos	golosovat^j

For each word pair, the model creates a word-specific rule as in (4).

(4)

[]→a/mot_
[]→a/tʃmok_
g→ʒi/dru_
g→ʒi/kru_
[]→ova/_golos_

Then, rules that involve the same change are combined. Contexts in which the same change, e.g., []→i, happens are compared by matching segments starting from the location of the change. If segments match, they are retained in the specification of the context for the change and the pair of segments further away from the change is compared. When this comparison process reaches the nearest pair of segments that do not match, the phonological features they share are extracted and retained in the specification of the context. Segments that are further away from the location of the change than the closest pair of non-matching segments are not compared and are replaced by a free variable in the specification of context.

For instance, the rules in (5) are combined into the rule in (6). Since the change involves the end of the stem, comparison starts from the end. The last segments in the context are both /u/, so they are retained and preceding segments are compared. Since the preceding segments are both /r/, they are retained as well and comparison proceeds to the preceding segment. These segments do not match but they are the closest pair of segments to the change that doesn't match, so the matching features are retained in the rule.

(5)

g→ʒi/dru_
g→ʒi/kru_

(6)

g→ʒi/[+cons;-cont;-son;-Labial]ru_

The resulting more general rules are then compared to each other and even more general rules derived if the same change occurs in multiple contexts, eventually resulting in quite general rules, such as []→i/C_. However, all rules are retained in the grammar. Instead of removing non-maximally-general rules from the grammar, the RBL weights each rule by its reliability. Reliability is defined as the number of words to which the rule applies divided by the total number of words to which it *could* apply. For instance, the reliability of the rule in (6) is the number of words of the form in (7) that are derived from words with the shape in (8) divided by the total number of words with the shape in (8) in the lexicon.

(7)
[+cons;-cont;-son;-Labial]ruʒi

(8)
[+cons;-cont;-son;-Labial]rug

A reliable rule is more likely to apply to a novel word than a less reliable rule. For instance, if the rule in (9) is more reliable than the rule in (10), and these are the only rules that can apply to the novel verb /dig/, the verb should be more likely to be borrowed as /diʒi/ than as /diga/.

(9)
Vg→Vʒi

(10)
Vg→Vga

The set of rules extracted from the lexicon, i.e., the grammar, is used only on novel words entering the lexicon. Existing morphologically complex words are stored in memory and retrieved from the lexicon as wholes rather than being generated by the rules of the grammar. Storage and retrieval of morphologically complex words is essential for a rule to be able to lose productivity while not gaining exceptions. If existing words were generated by the grammar, they would not continue to obey a rule as the rule loses productivity.

For the purposes of the present paper, this model has four essential features: 1) the model generalizes over input-output mappings, as opposed to just outputs (Bybee 2001:126-129, Pierrehumbert 2006), 2) input-output mappings compete for inputs, 3) the outcome of this competition is driven by reliability, and 4) morphologically complex words are retrieved from the lexicon in production.

Training the model. The model of the stem extension process was presented with the set of stem-verb pairings found in the Reverse Dictionary of Russian (Sheveleva 1974) and/or the Big Dictionary of Youth Slang (Levikova 2003). The Reverse Dictionary contains 125,000 words extracted from the four major dictionaries of Russian that existed in 1965 (Sheveleva et al. 1974: 7). The Slang Dictionary is much smaller, containing 10,000 words. The main results presented below held regardless of whether the Reverse Dictionary, the Slang Dictionary, or both were used. Only the results based on the full training set will be presented. Only stems that occurred independently as separate words were included. No stem extensions were excluded from the training set. Thus, aside from verbs featuring the highly productive –i and –a, verbs having –ova, -irova, -izirova, and –e were also included. The full training set consisted of 2,396 verb-stem pairs, of which 286 stems had final /k/ and 85 had final /g/. There were 22 examples of g→ʒi and 62 examples of k→tʃi. The model of diminutive formation was trained on a set of 1,154 diminutive nouns extracted from the Reverse Dictionary of Russian. All diminutive nouns whose base ends in a consonant were extracted regardless of the diminutive suffix used. The Slang Dictionary contains only a very small number of diminutives and thus was not used.

The learner models competition between input-output mappings. Therefore it is crucial to define what is meant by the input and the output. For the present paper, we are interested in modeling competition between input-output mappings in which some mappings require velar palatalization. The input form for these mappings may or may not have the stem extension already specified. If it does, rules specifying that a velar changes into an alveopalatal compete with rules that retain the consonant in the context of a stem extension that always triggers the change in the lexicon. If not, rules specifying that a velar changes into an alveopalatal also specify the stem extension. Thus a rule like k→tʃi would compete with k→ka as well as C→Ci.

In addition, the output of the competition can be either a phonetic form, specifying the allophone of /i/ used or a phonemic form, which does not include this specification. Both of these possibilities were examined in modeling but the choice between phonetic and phonemic outputs did not influence the qualitative results. In the case of the diminutive suffixes -ek and –ik, which can be considered allomorphs (or even orthographic variants), it also did not make a difference whether the choice between –ek and –ik followed the stage in which the decision on whether to palatalize the stem was made.

Testing the model. The model is presented with the set of English verbs found to be borrowed into Russian in the corpus study. To estimate the probability of a given verb undergoing velar palatalization given that a particular suffix is chosen, we can divide the reliability of the most reliable rule that requires palatalization by the sum of its reliability and the reliability of the rule that does not require palatalization but still attaches the same suffix. For instance, suppose the verb is /dig/ and the model has extracted the rules in (11) with reliability estimates shown in parentheses. The only rules that can apply to /dig/ are (a), (d), (e), (h), (i), and (j). Of these, the only rules that require velar palatalization are rules (h) and (i). Rule (h) is more reliable than rule (i), so it would get to apply. Its reliability is .272. The rule that attaches –i without palatalizing the stem-final /g/ is rule (j). Its reliability is .232. Therefore, the predicted probability that the final consonant of /dig/ will be palatalized, given that –i is selected as the stem extension, is .272 / (.272 + .232) = 54% (cf. Albright and Hayes, 2003:128).

(11)

a. []→a/{i;l}g_ (.723)
b. []→a/Cag_ (.718)
c. []→a/{l;r}eg_ (.718)
d. []→a/{i;l;n;r}g_ (.670)
e. []→a/[velar]_ (.641)
f. g→ʒi/$V_{[+back;-high]}$_ (.475)
g. g→ʒi/$V_{[-high]}$_ (.350)
h. g→ʒi/V_ (.272)
i. g→ʒi/[+voice]_ (.195)
j. []→i/$C_{[+voiced]}$_ (.232)

Results

Figure 1 shows that most velar-final verbs are highly unlikely to take –i while most labial-final and coronal-final verbs are very likely to take –i. Thus, the stem extension that triggers a stem change in the lexicon is disfavored by the stems that can undergo the change.

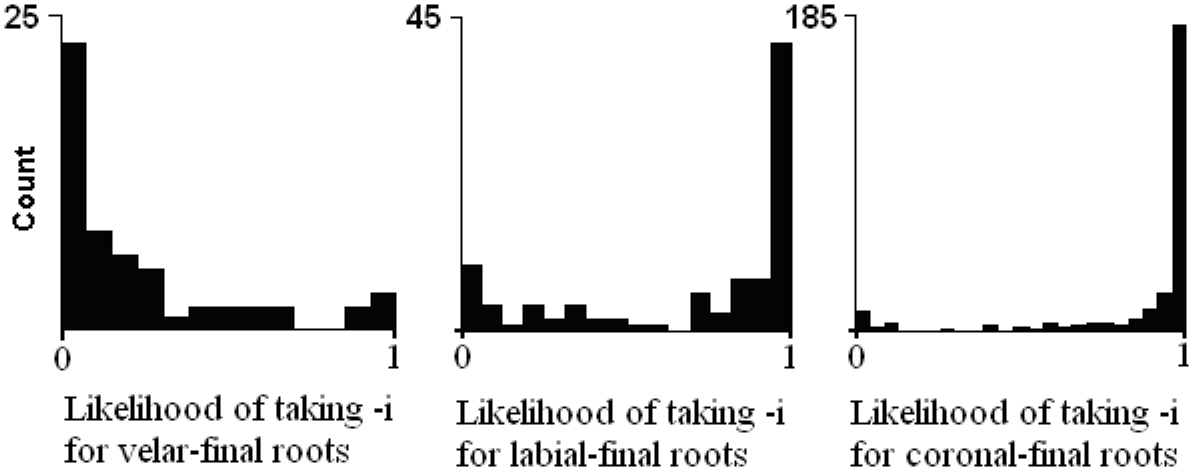

Figure 1. Histograms showing the probabilities of attaching –i to roots ending in consonants with various places of articulation.

Since the population distribution is skewed and bimodal, there is no monotonic transformation that will restore normality, which makes standard statistical tests inapplicable, which means that bootstrapping should be done. For this test, I treated the labial-final roots and coronal-final roots as the null population and generated 2,000 samples of 56 verbs from this population, calculating mean rate of taking -i in each sample. The mean rate of taking –i in the sample of velar-final stems (33%) falls very far outside the distribution of 2,000 samples of 56 verbs from the null population, thus *p*<.0005 (1/2,000). All versions of the model are able to predict that –i is less productive with velar-final stems than with coronal-final and labial-final stems.

Figure 2 shows just the velar-final stems that take –i as the stem extension. These are the only stems that undergo velar palatalization in the data, suggesting that the speakers are using a source-oriented generalization mapping velars onto alveopalatals, rather than a purely product-oriented generalization requiring alveopalatals before –i (Pierrehumbert 2006). A product-oriented generalization specifies only the shape of the output, thus imposing no restrictions on what changes can be done to the input to produce the output (for examples of such product-oriented behavior, see Bybee 2001:126-129).

The white bars show the observed likelihood of failure of velar palatalization before –i in various contexts while the dark bars show probabilities of velar palatalization failure predicted by the model. Figure 2 shows that velar palatalization is more likely to fail with /g/ than with /k/ (*t*(26)=4.803, *p*<.0005), and when the verb ends in a consonant cluster (left pairs of bars in each box) as opposed to a VC sequence (*t*(22)=3.415, *p*=.003). There is also a trend for the rule to fail more often after front vowels than after back vowels but it is not statistically significant. In other words, speakers tend to retain the velar if it is /g/ and if it is preceded by a consonant. They tend to replace the velar with an alveopalatal if it is a /k/ preceded by a vowel, especially if the vowel is back.

Despite the fact that the model is trained on a lexicon in which velar palatalization is exceptionless, the model predicts that velar palatalization will not be exceptionless with the borrowed verbs. Mean rate of failure of velar palatalization varies between 43% and 62% depending on parameter settings and approximates the actual mean rate of failure of velar palatalization in the data (56%).

While the mean predicted rate of failure for velar palatalization is similar to the observed rate of failure, the model's predictions are less variable than the data. In order to make them comparable, failure rates predicted by the model were rescaled to have the same standard deviation as the observed failure rates.[6] The qualitative results shown in Figure 2 hold for all versions of the model that assume that the stem extension and the stem shape are chosen simultaneously. These versions of the model correctly predict that velar palatalization is more likely to fail when the stem ends in a consonant cluster than when it ends in a single consonant, that penultimate front vowels disfavor palatalization compared to back vowels, and that /k/ is more likely to be palatalized than /g/ (however, all versions of the model underestimate the difference between /k/ and /g/).[7] If the stem extension is chosen first with the decision on whether to change the stem consigned to a subsequent decision stage, the predicted rate of failure of velar palatalization is not significantly affected but the effect of penultimate segment identity disappears.

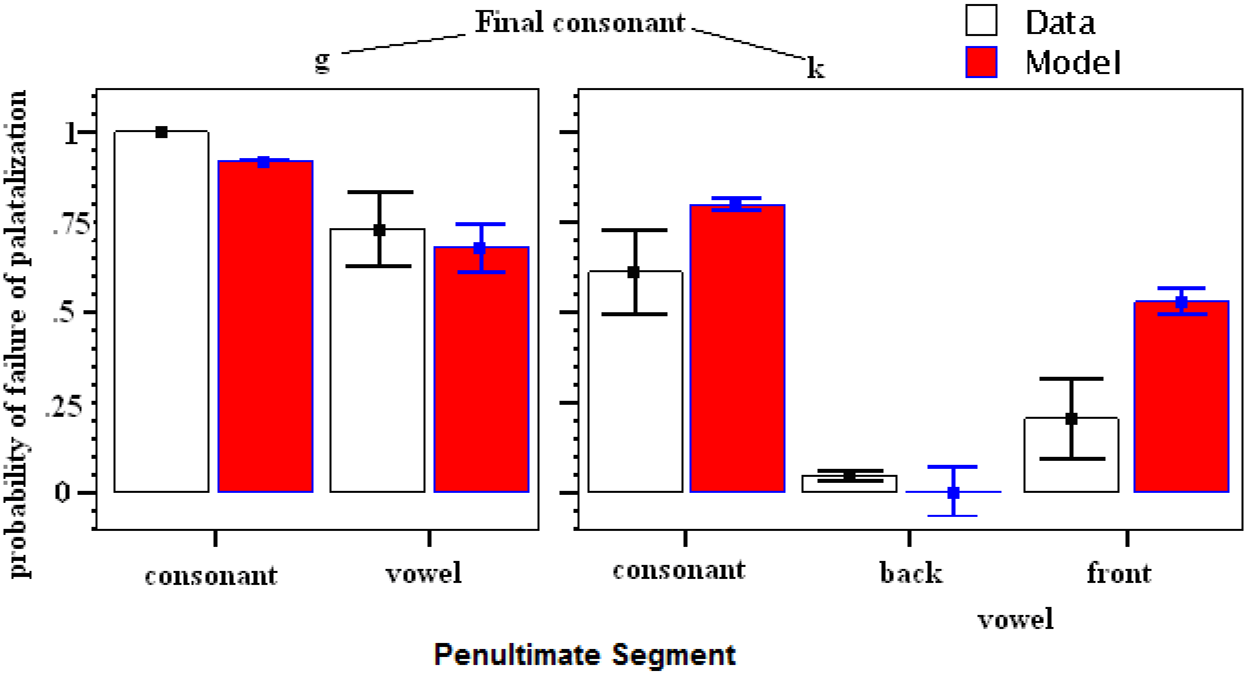

Figure 2. Observed (white bars) vs. predicted (grey bars) probabilities of failure of velar palatalization before the stem extension –i depending on segmental content of the stem. (Back and front vowels are not distinguished before [g] because there are relatively few roots ending in [g]).

Observed and predicted rates of failure of velar palatalization in front of diminutive suffixes are shown in Figure 3. As with the stem extensions, velars are the only consonants that change into alveopalatals, suggesting a source-oriented generalization. The rate of failure of velar palatalization is significantly higher before the suffix -ik (mean rate of failure = 40%) than before the suffix -ok (mean rate of failure = 0%), according to the paired-samples Wilcoxon signed ranks test (Z(16)=3.516, *p*<.0005). Failure of palatalization (which only happens before -ik) is more likely with /g/ (67%) than with /k/ (29%), *t*(15)=2.496, *p*=.025. The likelihood of using -ik is lower after /k/ than after /g/ (*t*(17)=5.729, *p*<.0005) and is higher after non-velars than after velars (*t*(45)=12.461, *p*<.0005). Thus, the

[6] This is why one of the error bars goes negative.

[7] This problem is exacerbated when impugnment is used. While versions of the model without impugnment are able to predict that /k/ is more likely to be palatalized than /g/, versions with impugnment incorrectly predict the opposite result except for stems with a penultimate back vowel.

suffixes –i and –ik tend to attach to non-velar-final inputs and often fail to trigger velar palatalization. The suffixes –ek and –ok tend to attach to velar-final inputs and are strong triggers of velar palatalization. Furthermore, in both the domain of verbal stem extensions and nominal diminutives, the productivity of k→tʃ is greater than the productivity of g→ʒ.

The model successfully learns that –ik is disfavored by velars and that palatalization is likely to fail only if –ik is chosen as the suffix, although the rate of failure of velar palatalization before -ik is overestimated. It predicts that –ek should be more productive with bases ending in /k/ than with bases ending in /g/, a numerical trend in the data. It fails to predict that /k/ is more likely to undergo palatalization and less likely to be followed by –ik than /g/. These predictions are parameter-independent, holding for all versions of the model.

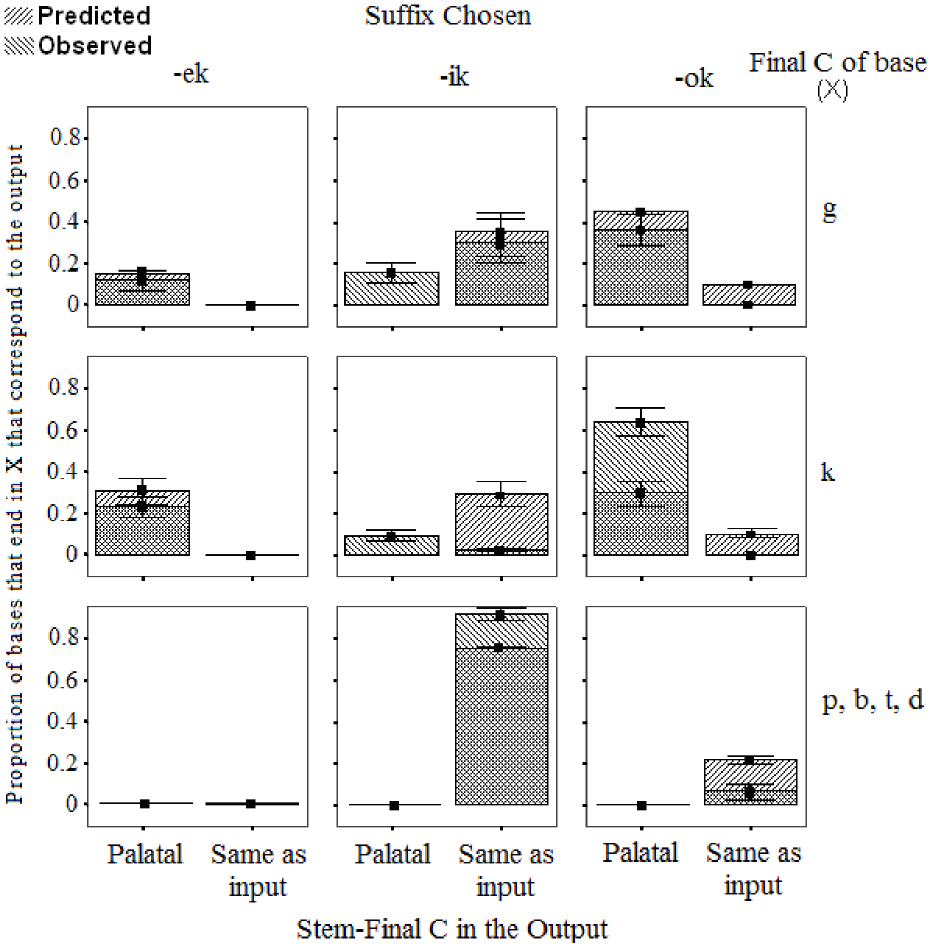

Figure 3. Relative likelihoods of various base-diminutive mappings for velar-final and non-velar-final bases of diminutive nouns (likelihoods sum to one across each row of panels) in the data (lines going from upper left to lower right) and in the model (lines going from lower left to upper right). The overlap between observed and predicted probabilities is shown by the intersection of the lines.

Explaining successes and failures of the model

In the present study, the RBL is used as only an example of a general class of models that postulate that input-output mappings are involved in a competition that is resolved by the mappings' relative reliability. Therefore it is important to determine the extent to which the successes and failures of the RBL are due to its reliance on this assumption.

In order to explain why the model performs the way it does let us examine the rules that it abstracts from the lexicon and uses when a velar-final verb is presented. The full list of applicable rules for [g]-final verbs is shown in (11) above. For both [k]-final and [g]-final verbs, there is only one rule that favors adding –i and leaving the final consonant of the stem unchanged. For /g/-final roots, this is the rule $C_{[+voiced]} \rightarrow C_{[+voiced]}i$ and for /k/-final roots this is the rule C→Ci. Thus, in order for the more specific rules requiring /k/ to change into /tʃ/ or /g/ to change into /ʒ/ to fail, they must lose to an extremely general rule. For this outcome to be likely, 1) a very general rule must be extracted from the lexicon, 2) it should be quite reliable relative to the less general rules requiring stem changes, and 3) it must compete with those rules.

In the Russian lexicon used to train the model, coronal-final and labial-final stems tend to take –i while velar-final stems tend to take -a. Since most stems in the lexicon end up taking –i, the model extracts a very general rule C→Ci and assigns it a moderate reliability. On the other hand, the fact that velar-final stems favor –a drives down the reliabilities of rules that add other stem extensions to velar-final stems. This includes the rules that add -i and change the root-final consonant. As a result, these rules will sometimes lose the competition for application to the more general rule C→Ci. Thus, the model predicts that velar palatalization will often fail before an affix if and only if the affix is more productive after non-velars than after velars. This holds for the stem extension –i and the diminutive suffix –ik but not for the diminutive suffixes –ek and –ok. Therefore, the model correctly predicts that velar palatalization should fail often before –i and -ik and rarely before –ek and –ok. This prediction follows directly from the hypothesis that input-output mappings compete with the outcome determined by reliability.

The model systematically fails to capture the difference in rate of palatalization between /k/ and /g/, which is observed in both stem extension and diminutive formation. In both cases, the rate of palatalization is underestimated for /k/. Palatalization of /k/ to [tʃ] is much more phonetically natural than palatalization of /g/ to [ʒ]. Bhat (1974:41) notes that velar stops generally become affricates or remain stops as a result of palatalization and if a language palatalizes voiced velars, it also palatalizes voiceless velars but not necessarily vice versa, which suggests that the g→ʒ change is typologically marked. Hock (1991:73-77) proposes that palatalization arises when a fronted velar stop develops a fricative release, suggesting that the velar stop is more similar to an alveo-palatal affricate than to an alveopalatal fricative. In addition, the voiceless velar stop [k] is more acoustically similar to [tʃ] than [g] is to [dʒ] in terms of peak spectral frequency and duration of aperiodic noise, leading listeners to misperceive [ki] as [tʃi] much more often than they misperceive [gi] as [dʒi] (Guion, 1998). Thus, [g] and [ʒ] can be argued to be more perceptually and articulatorily distinct than [k] and [tʃ] and the g→ʒ alternation can be argued to be less phonetically natural than the k→tʃ alternation. Phonetic naturalness has been argued to influence learnability of an input-output mapping when the reliability of the mapping is controlled (Finley and Badecker, to appear; Wilson 2006). The [k]/[g] asymmetry observed in Russian may be another case of this phenomenon. If the palatalization rule for [g] is more difficult to learn than the rule for [k] and the diminutive suffixes –ok and -ek do not permit a velar to precede it without a loss of naturalness, the speaker is driven to choose –ik as the diminutive suffix after [g] more often than after [k], accounting for the relatively high productivity of –ik following [g]. Phonetic naturalness alone cannot account for the data because velar palatalization is much more likely before –ok than before –ik, despite the fact that [o] is a less natural trigger of palatalization than [i].

Another shortcoming of the model is that it overpredicts the rate of velar palatalization before the suffix –ik, especially when –ik attaches to a /k/-final noun. This prediction follows from the fact that –ik never attaches to velar-final inputs in the native lexicon and thus is predicted not to trigger velar palatalization. There are at least two possible explanations for why it should still sometimes trigger velar palatalization. First, the alveopalatal stem-final consonant may be used as a diminutive marker in its own right, especially when the consonant is /tʃ/. This hypothesis is supported by the fact that some labial-final bases take -tʃik rather than -ik as the diminutive marker, e.g., sup 'soup' → suptʃik. Secondly, -ik and –ek are usually phonetically identical due to being unstressed (cf. Shvedova et al., 1980: 27-28). Despite being phonetically identical to -ik, –ek is a much stronger trigger of velar palatalization, thus the two suffixes must constitute different choices in phonology. However, it is possible that some instances of –ik in the (written) data can be cases in which the speaker chose –ek (which triggered velar palatalization) and misspelled it as the more frequent –ik.

The affix and the stem shape are chosen simultaneously. Perhaps, the most interesting parameter in the RBL is the sequence of stages assumed in modeling morphophonological processing. Interestingly, the penultimate segment effect on palatalization rate for stem changes is only obtained if a particular assumption is made about the sequence of processing stages, allowing us to distinguish between the two models in (12). Each stage in (12) is modeled by a separate Rule-Based Learner trained on the relevant input-output mappings.

(12)

Two-stage Model:

Stage I:

Choose the suffix based on the borrowed base:
[] → suffix / Base_

Stage II:

Modify the base to fit the suffix:
/Base/ → [Base] /_suffix

One-stage Model:

Stage I:

Choose the suffix based on the borrowed base and modify the base to fit the suffix:
/Base/ → [Base] + suffix

The effects of the penultimate segment shown in Figure 2 (and only those effects) are not predicted if we assume that the stem extension (-i vs. –a) is chosen first, followed by the decision on whether to change the stem (the two-stage model). Let us now examine why this is the case.

In the one-stage model, the palatalizing rules that are applicable to a given stem differ in their reliability, with some rules being more likely to outcompete the general non-palatalizing rule than others. For instance, the stem /overlok/ is likely to undergo velar palatalization because the most reliable palatalizing rule that can apply to it (k→tʃi/[+cons;+son]o_) is very reliable (.805) and can easily outcompete the applicable general rule (C→Ci) with its .2 reliability. By contrast, the most reliable palatalizing rule that can apply to the stem /drink/ ([+son]k→[+son]tʃi) has a reliability of only .125, which means that it is likely to lose to the more general rule C→Ci whose reliability is .2, resulting in failure of palatalization.

Suppose instead that the suffix has already been chosen and it is –i. The model now needs to decide whether to palatalize the stem. Interestingly, although the rules changing k→tʃ and g→ʒ are exceptionless and thus have a reliability value of 1, they can still sometimes lose to the more general rule "do nothing" because the reliability of "do nothing" is also quite high (86%). This is because most stems in the lexicon take –i and remain the same after the addition of -i.

However, with the stem change choice following affix choice raw reliability predicts no effect of penultimate segment identity. In this model, the reliabilities of all stem-changing rules are at 1, regardless of penultimate segment identity because velar palatalization never fails before –i in the lexicon on which the model is trained. Therefore, the model can capture segmental context effects only if they correspond to differences in rule type frequency (i.e., the number of word pairs supporting the rule), which in this case they do not. Thus, the effect of the penultimate segment is accounted for by the model only if the stem change and the affix are chosen during a single decision stage in which the palatalizing rules compete with rules adding other stem extensions, such as –a (the one-stage model).

Artificial Grammar Learning

Introduction

The data from Russian strongly suggest that the productivity of velar palatalization is connected to whether the palatalizing affix is used mostly with inputs that can undergo velar palatalization or with inputs that cannot. However, the data are correlational in nature, so the direction of causation is uncertain. It is possible that the low productivity of velar palatalization before –ik and –i, whatever its cause, makes speakers of Russian avoid using –i and –ik with velar-final inputs. If changing an input consonant in front of a certain suffix is difficult while keeping the consonant unchanged results in a suboptimal output, the best course of action may be to avoid using the suffix altogether.[8] Furthermore, the dictionary is not a perfect model of the Russian lexicon as it exists in the mind of a Russian speaker. Therefore, the data on whose basis velar palatalization is acquired by the model are different from the data on whose basis velar palatalization is acquired by Russian speakers.

A way to address both of these issues is provided by artificial grammar learning. By training the subjects and the model on the same language featuring velar palatalization, we can maximize the similarities between their relevant learning experiences. Furthermore, by varying the lexical distribution of the palatalizing suffix and keeping all other aspects of the competing rules constant, we can determine whether the distribution of the palatalizing suffix can influence the productivity of palatalization.

Native English speakers were randomly assigned to two groups exposed to two different artificial languages. Both groups were presented with an artificial language featuring two plural suffixes, -a and –i, and an exceptionless rule that palatalized velars before –i, turning /k/ into [tʃ] and /g/ into [dʒ]. Since the same input-output mappings are used in languages presented to both groups, phonetic naturalness is controlled. In both languages, velar-final singulars always corresponded to plurals ending in –tʃi or -dʒi. Both subject groups were presented with 30 singular-plural pairs in which the singular ended in a velar. Therefore, the palatalizing rule has the same type (and token) frequency in both languages. The difference between the two languages was that in Language 1 –i was not very productive with non-velar-final singulars, being used in only 25% of the cases with –a being used 75% of the time. In Language 2, the rates were reversed: -a was used 25% of the time with non-velar-final bases while –i was used 75% of the time. In both cases, 40 non-velar-final bases were used. Just like velar-final bases, the non-velar-final

[8] For instance, Thomason (1976) proposes that speakers may avoid an affix if it triggers opaque rules.

bases ended in oral stops (/p/, /b/, /t/, and /d/). Non-velar consonants did not change when a suffix was added.

The only rules that are applicable for novel velar-final bases and are extracted by the RBL upon exposure to the two languages are presented in Table 1. As the table shows, the two languages differ only in the reliabilities of the rules that do not require velar palatalization. The rule attaching –i without changing the preceding consonant is much more reliable in Language 2 than in Language 1. Therefore, velar palatalization is predicted to fail before –i in Language 2 more often than in Language 1. Importantly, in both Language 1 and Language 2, the most reliable rules that can apply to a velar-final input are palatalizing. Thus if subjects always used the most reliable applicable rule, there would be no difference between the two languages. Thus, the model predicts a difference between Language 1 and Language 2 only if the choice between rules is probabilistic.

	Language 1	Language 2
k→tʃi/V_	0.85	
g→dʒi/V_		
[ ]→i/C_	0.18	0.57
[ ]→a/C_	0.57	0.18

Table 1. Rules and the corresponding reliability values extracted by the Rule-Based Learner when it is exposed to the same artificial languages presented to human participants.

Methods

The experiment consisted of a training stage and a testing stage. In the training stage, subjects repeated singular-plural pairs presented to them auditorily over headphones. There were 62 word pairs, with each word pair presented twice to each subject during training. The aural presentations of the words were accompanied by visual presentations of the referents on a computer screen. The participants' productions were used to assess whether the training stimuli were perceived correctly. If a participant made perception errors on more than 5% of singular-plural pairs, s/he was excluded from the experiment (N=2). In the testing stage, subjects were presented with novel singular forms (not presented during training) and asked to orally produce the plural. There were 20 velar-final singulars, and 34 non-velar-final singulars presented during testing. Participants who used –i fewer than 10 times with velar-final singulars (N=4) were excluded from the present analyses. The analyses below are based on thirty participants, half of whom were exposed to each language.

Results and Discussion

First, it is important to note that velars become alveopalatals much more often than other consonants do in both languages (the rate of labials changing into alveopalatals is 0% for both languages; for coronals, it is 2% for Language 1 and 8% for participants exposed to Language 2, a non-significant difference). The difference between rate of coronal palatalization and the rate of velar palatalization is significant for both Language 1 (Wilcoxon signed ranks test, Z(14)=3.05, *p*<.01) and Language 2 (Z(14)=2.63, *p*<.01). Thus, there is evidence that subjects really have acquired input-output mappings specifying that velars change into alveopalatals, rather than simply learning that –i should be preceded by an alveopalatal.

The participants were able to discover the distribution of –i and –a in the lexicon. Figure 4 shows that participants exposed to Language 1 used –i after alveolars and labials 30% of the time while participants exposed to Language 2 used –i 67% of the time (*t*(28)=4.4, *p*<.001). Thus the training was successful in making –i more productive after non-velars in Language 2 than in Language 1. The proportions of –i use by the subjects in the two groups are similar to proportions in the data to which they were exposed: 25% for Language 1 and 75% for Language 2.

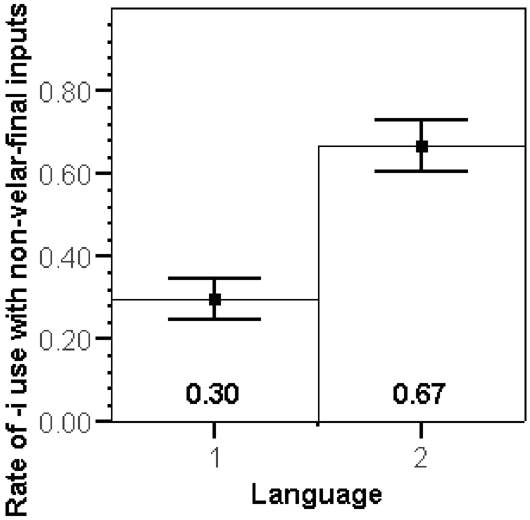

Figure 4. Subjects exposed to Language 2 are more likely to use –i to form the plural than subjects exposed to Language 1.

More interestingly, Figure 5 shows that participants exposed to Language 1, the language predicted to favor velar palatalization by virtue of disfavoring the use of –i with non-velar-final singulars, palatalized the velar before -i 67% of the time, while participants exposed to Language 2 palatalized the velar before –i only 38% of the time (*t*(28)=2.316, *p*<.05). Thus, the predictions of rule reliability are confirmed: even if the rules changing velars into alveopalatals before –i are exceptionless in the language, the more productive –i is with non-velar-final bases, the more likely velar palatalization is to fail.

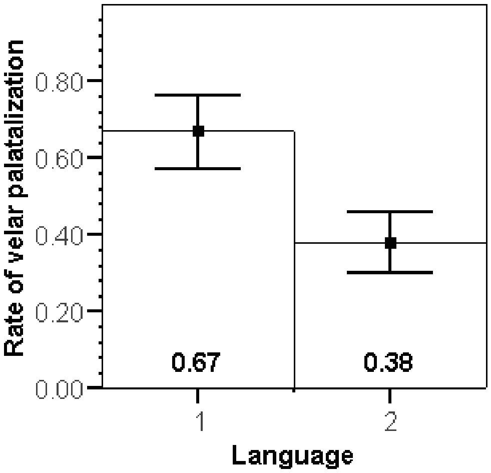

Figure 5. Subjects exposed to Language 2 are less likely to palatalize the velar before –i than subjects exposed to Language 1.

Like speakers of Russian, subjects exposed to the artificial languages do not simply match the rate of velar palatalization to which they are exposed (100% for all subjects, regardless of whether they were exposed to Language 1 or Language 2). Rather, learners appear to be sensitive to the reliability of the 'just add –i' rule relative to the palatalizing rule. Figure 6 shows that there is a strong and significant negative correlation (*r*=-.68, *p*<.001) between how much a subject uses –i with non-velar-final inputs and how likely s/he is to palatalize a velar before –i.

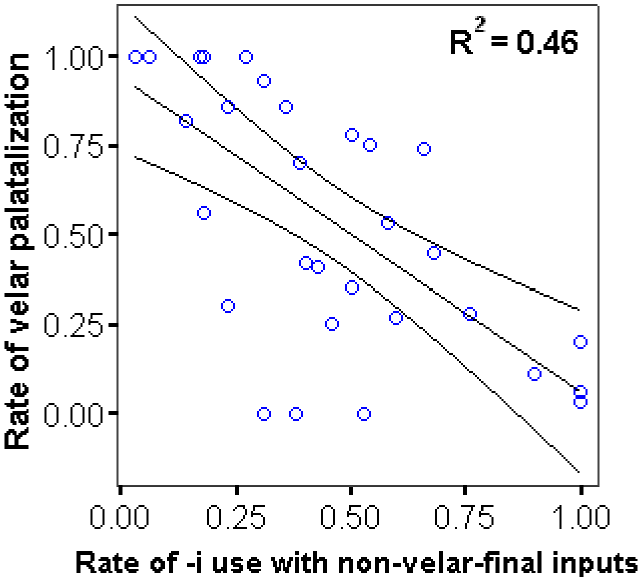

Figure 6. Subjects for whom –i is productive with inputs that cannot undergo velar palatalization are the subjects for whom velar palatalization is unproductive. Curves show the 95% confidence region for the regression line.

Once the correlation between rate of velar palatalization exhibited by a subject and his/her rate of –i use with non-velar-final inputs (Figure 4) is taken into account, the difference between subject groups no longer contributes towards explaining between-subject differences in velar palatalization productivity (according to an ANCOVA with rate of velar palatalization as the independent variable, the rate of –i use with non-velar-final inputs as a covariate, and Language as a fixed factor, rate of –i use is significant, F(1,27)=14.23, *p*<.001, while Language is not, F(1,27)=.082, *p*>.5). Thus, the difference in productivity of –i with non-velar-final inputs, which is the independent variable predicted to influence productivity of velar palatalization by the RBL, accounts for all differences in productivity of velar palatalization between subjects that can be attributed to the artificial language they are exposed to.

Source-oriented vs. Product-oriented Generalization

The present paper has examined the predictions of a rule-based model. We will now examine in more detail whether the same predictions can be derived from a product-oriented and/or constraint-based model. The simplest product-oriented model is one in which the possible generalizations have the form 'outputs must (not) have X' (Bybee, 2001). Thus, in the case of our two artificial languages, the relevant palatalizing schemas would have the form 'plurals end in {tʃ;dʒ}i (in context X)'. While attractively simple, this account fails to predict a difference between the two artificial languages: since there is the same number of examples of velar palatalization in both languages, the palatalizing schema has the same type frequency in both languages and is expected to be equally productive.

	Language 1	Language 2
‘end in -tʃi/dʒi’	30	

Table 2. Palatalizing product-oriented generalizations and the number of plural noun types supporting them in the two languages.

Things get worse if we assume that the learner develops a preference against /ki/, which increases whenever a learner expect to but do not in fact hear it. /Ci/ plurals are more common in Language 2 than in Language 1 while /kV/ plurals don’t occur in either language, thus the learner generalizing over plurals would expect (and fail) to hear /ki/ more often in Language 2 than in Language 1, incorrectly predicting that velar palatalization should be more productive in Language 2 than in Language 1.

	Language 1	Language 2
*ki/gi	22.5	7.5

Table 3. A negative palatalizing product-oriented generalization and the number of times the ungrammatical output is expected but not observed in the two languages.

What seems to be required is a *conditional* product-oriented palatalizing schema of the form ‘if the plural ends in –i, the preceding consonant must be {tʃ;dʒ} (in context X)’. The reliability of this generalization (given as the number of plurals that end in {tʃ;dʒ}i divided by the number of plurals that end in –i) differs between the two languages, since the denominator is much greater in Language 2 than in Language 1, thus palatalization is correctly predicted to fail more often in Language 2 than in Language 1. Equivalently this notion may be formalized as the learner attempting to simultaneously satisfy ‘plurals must end in –{tʃ;dʒ}i’ and ‘plurals must end in –Ci’. The support for the second generalization is greater in Language 2 than in Language 1, thus it will be satisfied more often. The support for the first generalization is the same across the two languages, thus it would be satisfied equally often. Thus the proportion of times a plural ending –i features velar palatalization is expected to be lower in Language 2 than in Language 1.[9]

[9] A second problem with the product-oriented account is the lack of restrictions on inputs that can give rise to outputs ending in {tʃ ;dʒ }i (Pierrehumbert 2006). A possible solution, albeit one relying on generalization over word pairs, is that the palatalizing product-oriented generalization is in competition with some version of paradigm uniformity constraints (see Downing, Hall, and Raffelsiefen, 2005, for possible formalizations), stipulating that the singular and the plural have the same value on the place feature of a stem-final consonant, such as Ident-[velar].

	Language 1	Language 2
plurals that end in -tʃi/dʒi	30	
plurals that end in -Ci	38	54
'if the plural ends in –i, the preceding consonant must be {tʃ;dʒ}'	30/38 = .79	30/54 = .56

Table 4: A conditional palatalizing product-oriented generalization and the number of times the ungrammatical output is expected but not observed in the two languages.

A crucial difference between the product-oriented accounts and the source-oriented account is the treatment of singular-plural mappings in which the singular ends in {tʃ;dʒ} while the plural ends in {tʃ;dʒ}i. Under the product-oriented account, these mappings exemplify the palatalizing generalizations 'plurals must end in –{tʃ;dʒ}i' or 'if the plural ends in –i, the preceding consonant must be {tʃ;dʒ}'. Thus, their addition to the training set should increase the productivity of velar palatalization. Under the source-oriented account, these singular-plural pairings exemplify the rule 'just add –i', which militates against velar palatalization. Thus, their addition should reduce the productivity of velar palatalization.

A new group of 68 adult native English speakers was presented with one of the four languages in Table 5. Languages I and II are the same languages as the ones presented to the original group of subjects. Languages III and IV differ from Language 1 and Language 2 respectively in having 20 additional singular-plural pairs in which a singular ending in {tʃ;dʒ} corresponded to a plural ending in {tʃ;dʒ}i.

	Language 1	Language 2	Language 2I	Language 1V
{k;g} → {tʃ;dʒ}i	30			
{t;d;p;b} → {t;d;p;b}i	8	24	8	24
{t;d;p;b} → {t;d;p;b}a	24	8	24	8
{tʃ;dʒ} → {tʃ;dʒ}i	0		20	

Table 5. The four languages presented to learners in Experiment II

Figure 7 shows that the addition of {tʃ;dʒ} examples reduced the rate of velar palatalization. When the presence of {tʃ;dʒ}→{tʃ;dʒ}i examples and the probability of attaching –i to alveolars and labials are entered into an ANCOVA, both are significant (F(1,41) = 23.15, p<.001 for rate of attaching to alveolars and labials; F(1,41) = 6.06, p=.01, for presence of {tʃ;dʒ}→{tʃ;dʒ}i examples). Thus velar palatalization rate is reduced if –i often attaches to non-velars, which may be labials and coronals.

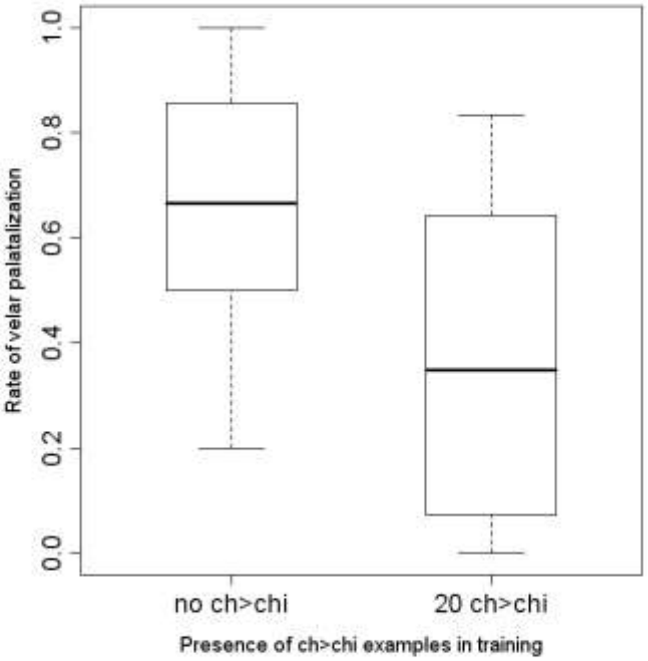

Figure 7. The addition of singular-plural pairs exemplifying {tʃ;dʒ} → {tʃ;dʒ}i reduces the productivity of velar palatalization.

This result directly contradicts the hypothesis that the learners are extracting product-oriented generalizations, where the generalization responsible for velar palatalization is 'plurals must end in {tʃ;dʒ}i'. The examples whose addition reduces the productivity of velar palatalization in the present paradigm exemplify the product-oriented generalization that is supposed to favor velar palatalization. By contrast, the results are expected under the hypothesis that the learners are using source-oriented generalizations. The examples in which an alveopalatal is mapped onto an alveopalatal followed by –i are examples of the rule 'just add –i' (C→Ci), which disfavors velar palatalization. Thus, at least in the present training paradigm, the generalizations that learners extract from the lexicon are source-oriented.

Conclusion

The hypothesis that rules compete for inputs with the outcome of this competition determined by differences in reliability or type frequency between the competing rules predicts that a morphophonemic rule will lose productivity if the triggering affix comes to be used increasingly with inputs that cannot undergo the rule due to not being in the class of inputs to which the rule can apply. This hypothesis is supported by loanword adaptation data in Russian as well as experimental data from artificial grammar learning.

The present data place three restrictions on the rule-based account. First, the affix and the 'triggered' stem change are actually chosen at the same time, rather than the affix being chosen first and then triggering or failing to trigger a stem change. The generality of this finding remains a matter for future research. It is possible that it is a hallmark of phonological processes triggered by specific suffixes or a property of phonological processes occurring in derived environments more generally.

Second, as predicted by Albright & Hayes (2003), the choice between rules must be probabilistic in nature, rather than the subjects always applying the most reliable applicable rule. An important caveat is that the present artificial grammar learning experiments examine generalization in adults. Hudson Kam & Newport (2005) have shown that children exposed to a probabilistic artificial grammar are more prone to regularize the variation, choosing to apply the most productive rule 100% of the time, than adults are. Thus, it would be important to see if the prediction is upheld if the experiment is repeated with children.

Finally, existing morphologically complex words are stored in memory and retrieved for production (cf. Bybee 1985, 2001; Halle 1973; Vennemann 1974), which allows for accurate retrieval of all forms of an existing word accompanied by probabilistic extension of the rules for creating those forms to new words. This feature of natural language is not replicated in the experimental paradigm, where accuracy of plural form production for singulars presented in training is the same as for novel singulars that are first introduced at test. In future work, it is important to modify the experimental paradigm so that learners are asked and enabled to learn individual plural forms.

This modification of the paradigm may make subjects more prone to product-oriented generalization (Bybee, 1985, 2001), which is disconfirmed for the present experimental paradigm but not for the corpus data. In order to determine whether source-oriented generalization in the present artificial grammar learning paradigm is due to characteristics of the training task, future experiments will expose learners with either a singular form or a plural form one at a time, rather than being exposed to singular-plural pairs, and reduce the number of singular-plural pairs to allow for memorization of individual plural forms.

Finally, there is an enormous amount of individual variability in how successful learners are at matching the statistics of the input. This high variability suggests that another fruitful avenue for future research would be correlating the subjects' generalization behavior in artificial grammar learning with their linguistic experience and success at natural language learning and acquisition. Artificial grammar learning is a precisely controlled experimental paradigm that seems ideally suited for examining the intriguing possibility that poor language learners have different generalization patterns than good language learners.

References

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition, 90,* 119–161.

Bhat, D. N. S. (1974). A general study of palatalization. *Working Papers on Language Universals, 14,* 17-58.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Downing, L. J., Hall, T. A., & Raffelsiefen, R. eds. (2005). *Paradigms in phonological theory*. Oxford: Oxford University Press.

Finley, S., & Badecker, W. (To appear). Towards a substantively biased theory of learning. *Proceedings of the 33rd Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society.

Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica, 55,* 18-52.

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry, 4,* 3-16.

Hock, H. (1991). *Principles of historical linguistics*. New York: Mouton de Gruyter.

Hudson Kam, C. L., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1,* 151-195.

Levikova, S. I. (2003). *Bol'shoj slovar' molodezhnogo slenga*. Moscow: Fair-Press.

Pierrehumbert, J. B. (2006). The statistical basis of an unnatural alternation. In L. Goldstein, D.H. Whalen, & C. Best (eds), *Laboratory Phonology VIII, Varieties of Phonological Competence* (pp. 81-107). Berlin: Mouton de Gruyter.

Sheveleva, M. S. (1974). *Obratnyj slovar' russkogo jazyka.* Moscow: Sovetskaja Enciklopedija.

Shvedova, N. Ju., Arutjunova, N. D., Bondarko, A. V., Ivanov, V. V., Lopatin, V. V., Uluxanov, I. S., & Filin, F. P. (1980). *Russkaja grammatika I*. Moscow: Nauka.

Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer.

Thomason, S. G. (1976). What else happens to opaque rules? *Language, 52,* 370-381.

Vennemann, T. (1974). Words and syllables in natural generative phonology. In A. Bruck, R. A. Fox, & M. W. La Galy (eds.), *Papers from the Parasession on Natural Phonology* (pp. 346-374). Chicago: Chicago Linguistic Society.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science, 30,* 945-82.

Zimmer, K. E. (1969). Psychological correlates of some Turkish morpheme structure conditions. *Language, 45,* 309-321.

Zuraw, K. (2000). Patterned exceptions in phonology. Unpublished Doctoral Dissertation, UCLA.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Some Links between Executive Function and Spoken Language Processing: Preliminary Findings using Self-Ordered Pointing and Missing Scan Tasks[1]

Esperanza M. Anaya, Christopher M. Conway[2] and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This study was supported by NIH Research Grant No. DC-00111 and Training Grant No. DC-00012. I wish to thank Dr. Robert M. Joseph for providing us with stimuli for the self-ordered pointing task, Luis Hernandez for his invaluable technical assistance, and Lauren Grove for her help with data collection.

[2] Department of Psychology, Saint Louis University, St. Louis, MO 63103.

Some Links between Executive Function and Spoken Language Processing: Preliminary Findings using Self-Ordered Pointing and Missing Scan Tasks

Abstract. Early auditory deprivation and a delay in language acquisition may result in disturbances to the normal development of executive function behaviors in congenitally deaf children with cochlear implants. The purpose of this study was to examine the relationship between executive function and spoken language processing using two new executive function measures, the Self-Ordered Pointing (SOP) and Missing Word task (MWT). Other cognitive and language measures were also collected in this study. Twenty adult subjects completed the experiment. Results showed that subjects performed better in the verbal condition of SOP in comparison to the non-verbal condition. Subjects also showed a trend for improvement over block presentations. These results replicate earlier studies that used SOP. Additional analyses revealed associations of executive functions and language measures. Specifically, correlations were found between the non-verbal condition of SOP and receptive vocabulary as well as non-verbal SOP and backward Digit Span. Few results were found with MWT. Subjects showed a trend for a decrease in performance as set size increased. A weak relationship was observed for MWT and forward Digit Span. The results contribute to our understanding of the relationship between executive functions and spoken language processes. Future directions for this research include the use of these new behavioral measures with normal hearing children and congenitally deaf children with cochlear implants.

Introduction

Enormous variability in audiologic outcomes exists for congenitally deaf children who receive cochlear implants. Some children seem to flourish after implantation and resemble normal hearing children while others gain little to no benefit from their implant. These low-performing children continue to show little benefit even though they have a fully functioning device. The cause of these differences in performance is still unknown. Several recent findings from our lab suggest that these individual differences in speech and language outcomes cannot be attributed to a dysfunction in the auditory domain alone. Other areas of cognition may be affected by the period of auditory deprivation and subsequent delay in language development that these congenitally deaf children experience. Identifying the underlying factors of speech perception can help researchers better understand the cognitive deficits found in deaf children with cochlear implants. Being able to account for the large variance observed following implantation can help in the development of new treatment and intervention programs for deaf children with cochlear implants. By identifying the implant user’s cognitive impairments, unique learning programs can be created that address these specific deficits. Early intervention and treatment may help reduce some of the variance in implant success that is seen later on. However, more research needs to be conducted into the differences in implant benefit before these intervention and/or prevention programs can be developed.

Concerns with the Medical Assessment of Cochlear Implantation

The medical field commonly assesses cochlear implant benefit with traditional “endproduct” speech and language measures (Miyamoto, Robbins, Myres, Pope, & Punch, 1986; Robbins, Osberger, Miyamoto, Kienle, & Myres, 1985). These measures were created by audiologists and speech language pathologists to measure outcomes after cochlear implantation according to FDA guidelines (Pisoni, Conway, Kronenberger, Henning, & Anaya, 2009). Many clinicians working in this field have focused on the performance of the device itself or the hearing problems this population faces. In other words, these medical practitioners may see a lack of implant benefit as stemming from a malfunctioning device or a structural malformation of the ear and neglect the notion of the problem lying elsewhere. It is very likely that several areas of cognition may have been affected by the sensory deprivation and language delay these children experienced. This is an unconventional way of approaching research into this hearing impaired population. Very little research has investigated the relationship between cognition and language in cochlear implant users (see however Pisoni, 2000). However, researchers are now beginning to address the issue of large individual differences for speech outcomes. The 1988 and 1995 National Institutes of Health (NIH) consensus statement on cochlear implants called for more research to investigate this significant clinical problem, "Research must be attempted to explain the wide variation in performance across individual implant users." (p. 12). The NIH went on to state that "Investigations of the role of higher level cognitive processes in cochlear implant performance are needed." (p. 12).

Early Auditory Structures

In order to address the major issues in language outcomes for implant users, researchers need to understand the nature of the auditory deprivation this population experiences. The auditory system begins to develop in the early embryonic stages. As early as the fourth week of gestation, some auditory structures are identifiable. Research has also shown that the fetus to be capable of hearing auditory stimuli (Sohmer & Freeman, 1995). A study reported by Gagnon (1989) showed fetal heart rates increased as their mothers' stomach was exposed to vibroacoustic stimulation. Another study revealed that normal hearing babies to have an increased fetal heart rate while still in the womb when exposed to vibroacoustic stimulation, whereas babies with hearing loss at birth did not show a change in their fetal heart rate when exposed to the same stimulation (Birnholz & Benacerraf, 1983). Additional research has shown that fetuses are able to hear their mothers' voice (Querleu, Renard, & Crepin, 1981). Consequently, babies are more accustomed to and prefer their mothers' voice after they are born (Hammond, 1970; DeCasper & Fifer, 1980). In applying these research findings to congenitally deaf children, it would appear that auditory deprivation is much greater than just the time period from birth through implantation. For this population, sensory deprivation begins prior to birth. Length of deprivation can then be understood as beginning in the womb and extending until implantation.

Important Demographic Factors Concerning Implantation

Several factors have been shown to influence later performance on speech and language measures. Length of deprivation or duration of deafness before implantation is one factor that has been studied. Children who have shorter periods of auditory deprivation have been found to perform better on a wide range of speech and language measures than children who experience longer periods of deprivation (Kirk, Pisoni, & Miyamoto, 2000; Miyamoto, Osberger, Todd, Robbins, Stroer, Zimmerman-Phillps, & Carney, 1994). Early linguistic experience also affects later performance measures of speech and language (Geers, Spehar, & Seday, 2002; Kirk, Miyamoto, Ying, Perdew, & Zuganelis, 2002; Sarant, Blamey, Dowell, Clark, & Gibson, 2001). Implant children in auditory-oral schools use only spoken language to communicate. Total communication schools encourage children to use spoken language as well as some manual coded form of English to communicate. Research done by Kirk et al. (2000) has shown that deaf children in oral-only schools perform better than deaf children in total communication schools on speech measures as well as other language measures which assess areas outside the auditory domain.

One of the best demographic factors that predict implant benefit has been the age at which a child is implanted (Harrison, Panesar, El-Hakin, Abdolell, Mount, & Papsin, 2001; Kileny, Zwolan, & Ashbaugh, 2001; Miyamoto et al., 1994; Sharma, Dorman, & Spahr, 2002). Children who are implanted at an early age show greater benefit with their implant in comparison to those children implanted at later ages. Some infants have been implanted as early as 6 months old; however, most are implanted sometime after the age of 12 months, the FDA approved age of implantation. Children who are implanted early, prior to the age of three years old, generally perform better on speech and language measures than children implanted after this age (Miyamoto, Kirk, Svirsky, & Sehgal, 1999). However, early implantation is not a guarantee for implant success. Large variability in speech and language outcomes still exists in children implanted early. It would appear using language measures alone to assess cochlear implant benefit does not tell us the full story of how the child is performing with his/her implant. Recent studies have shown that children with cochlear implants often show deficits in several areas in addition to the auditory domain. The identification of cognitive factors that aid in language development may help in better predicting cochlear implant benefit and identify children who may be at high risk for poor outcomes after implantation.

Executive Function

Executive function is an umbrella term used to describe higher order cognitive abilities that control and guide goal-oriented behaviors (Bernstein & Waber, 2007; Hauser, Lukomski, & Hillman; 2008; Penington, 1997; Visu-Petra, Benga, & Miclea, 2007). This term encompasses functions such as attention, inhibition, emotional control, and working memory. These processes are involved in solving novel and complex problems and are thought to be mediated by language (Luria, 1973; Vygotsky, 1934). In his book *The Working Brain* (1973), Luria states that lesions to the frontal lobe can affect speech's regulatory function. An inability to regulate our internal thoughts can lead to deficits in cognitive control, such as the failure to attend to important stimuli and inhibit distracting stimuli.

Executive functions are thought to be mediated by brain circuits in the prefrontal cortex (Miller & Cohen, 2001; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005). Research has shown that damage to the prefrontal cortex can cause deficits in self-monitoring, planning, inhibition, working memory, and the ability to shift attention from one task to the next (Chase, Clark, Sahakian, Bullmore, & Robbins, 2008; Gouveia, Brucki, Malheiros, & Bueno, 2007; Luria, 1966; Petrides & Milner, 1982; Stuss & Benson, 1984). The infamous story of Phineas Gage provides an excellent example of deficits seen in cognition as a result of frontal lobe damage. In 1848, Phineas Gage who was a railroad foreman at the time was injured at work when an explosion sent a tamping iron through his cheek, left eye, and frontal lobe. Prior to the accident, Gage was said to be a well-mannered, hard working individual. Yet after the accident, however, his demeanor dramatically changed. Gage was profane and lacked an ability to inhibit his thoughts and actions. He also was unable to plan ahead and organize himself. Consequently, he struggled in his personal and work life after he suffered damage to his prefrontal cortex.

The frontal lobes have also been shown to be the last areas of the brain to mature. The prefrontal cortex continues to develop through early adulthood (Golden, 1981; Hughes & Graham, 2002). Consequently, the development of executive functions has been shown to have a protracted period of development. Executive functions begin to manifest in infancy (Anderson, 2002; Anderson, 2002). However, aspects of the development of these cognitive abilities, such as working memory and the ability to inhibit competing input signals, parallel the development of the frontal lobe and do not become fully developed until late adolescence.

A study recently carried out in our lab assessed the executive functions of children and adolescents. Our study examined three groups of profoundly deaf children with cochlear implants

(Anaya, Conway, Pisoni, Geers, & Kronenberger, 2008). The first group consisted of fourteen children between the ages of 2 and 5; the second group was composed of eighteen 6 to 10 year old children; and the third group had twenty-six adolescences between the ages of 15 and 18. The Behavior Rating Inventory of Executive Function (BRIEF) was administered to all three groups. BRIEF is a parental questionnaire and rating scale that was developed to assess executive function in children and adolescents in real-world environments (Gioia, Andrews, & Isquith, 1996; Gioia, Isquith, Guy, & Kenworthy, 1996). The questionnaire consists of nine subscales and several composite scores that rate executive functions. Higher scores are indicative of behavioral problems (Mahone et al., 2002; Mares, McLuckie, Schwartz, & Saini, 2007).

Results of this first study showed that all three age groups of deaf children had significantly elevated average BRIEF scores relative to age norms for normal hearing children. Children in Group 1 (2-5 year olds) showed elevation in one subscale measuring inhibition. The children in Group 2 (6-10 year olds) had five elevated scales including the subscales of shift, emotional control, and working memory as well as two composite scales. Finally, the adolescences in Group 3 (15-18 year olds) showed elevated scores across nine different clinical scales including the subscales shift, emotional control, working memory, organizational, initiate, and monitoring as well as all three composite scales.

These findings suggest that disturbances in executive function become more evident with increased age. This may have occurred because executive functions have a protracted period of development. The absence of elevated scores across additional subdomains in the youngest group could be explained by the continuing development of executive functions. The findings from our initial study using the BRIEF suggest that some deaf children with cochlear implants may not only experience hearing loss and delay in language development but they may also have comorbid disturbances and/or delays in executive function and cognitive control. The disturbances observed in executive function in some of these deaf children could reflect more basic neurobiological changes that occur in the frontal lobe as a result of a period of sensory deprivation prior to implantation (Wolff & Thatcher, 1990).

Working Memory

Other recent studies from our laboratory have also shown that working memory is impaired in children with cochlear implants (Pisoni & Geers, 2000; Cleary, Pisoni, & Kirk, 2000). In one study, Pisoni and Geers examined 43 cochlear implant children and found significant associations between verbal digit span, a measure of verbal working memory, and several other speech perception and language production measures. The deficit found in verbal working memory may be due to a delay or disorder of the phonological loop and basic verbal rehearsal processes. The phonological loop is a part of Baddeley and Hitch's (1974) working memory model and is used for verbal rehearsal of information in short-term memory. Children with language disorders or delayed language learners, such as congenitally deaf children with cochlear implants, are thought to be poor users of the phonological loop (Gathercole & Baddeley, 1990). The problem cochlear implant children have with short-term memory may be the result of their delay and lack of experience with verbally rehearsing auditory stimuli (Pisoni & Cleary, 2003).

Another study in our lab done by Pisoni and Cleary (2003) further examined the working memory deficit observed in this population. A total of 176 cochlear implant children between the ages of 8 and 9 years old were recruited for the study. An additional 45 normal hearing aged matched children served as a comparison group. All subjects were tested with the WISC III forward and backward digit span. Results showed that hearing impaired children had significantly shorter forward and backward digit spans in comparison to the normal hearing age-matched group. The authors concluded that "the hearing-impaired children may have been unable to utilize strategies already mastered by normal-hearing children to improve immediate recall under simple forward recall conditions" (p. 110). Pisoni and Cleary also found significant correlations between digit span and verbal rehearsal speed. Subjects who spoke more quickly showed longer digit spans. Additional research from our lab has provided further converging support for the claim that congenitally deaf children with cochlear implants have disturbances/deficits in their verbal working memory (Burkholder & Pisoni, 2006; Dillon, Burkholder, Cleary, & Pisoni, 2004; Fagan, Pisoni, Horn, & Dillon, 2007).

Burkholder & Pisoni (2003) examined speech timing and working memory spans in normal hearing and congenitally deaf children with cochlear implants. They compared 36 normal hearing children and 37 deaf children between the ages of 8 and 9 on speech and working memory measures. The hearing impaired group had longer sentence durations, larger pauses between recall of items, and shorter digit spans when compared to normal hearing children. A relation was also found between articulation rates and digit span for both groups. The authors concluded that slower subvocal rehearsal and scanning processes may contribute to the shorter immediate memory spans seen in the hearing impaired group.

Testing Executive Function

There are many experimental tests that have been developed to assess executive function and cognitive control. Two popular assessment measures are the STROOP test (Stroop, 1935) and the Wisconsin Card Sorting Task (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). Both measures have been useful in the past for assessing executive functions in normal developing and clinical populations. However, new measures of executive function need to be developed to better understand the different aspects that make up these complex cognitive processes. Some of these new measures can be used to explore the possible deficits in executive function and cognitive control in congenitally deaf children with cochlear implants.

This study used two new measures of executive function. The first assessment was the self-ordered pointing test (SOP). This procedure was originally created by Petrides and Milner (1982) to evaluate the cognitive deficits of patients with frontal lobe lesions. Another group of patients with temporal lobe lesions was also used as a comparison group. In the original SOP task, both groups of brain-damaged patients were shown pictures of common objects on a piece of paper. Subjects were told to pick an object. The experimenter then showed the subject another piece of paper with the same objects as before but with the pictures rearranged in a different spatial location than the previous one. The patient was told to select another object that was different from the one previously chosen. This continued until the patient selected all of the different items.

Subjects completed each SOP task using varying set sizes. There were also two conditions, a “Verbal” and “Non-verbal” condition. Each condition contained two tasks. Patients completed two verbal and two non-verbal tasks. In the verbal condition, the images were easy to name. In the nonverbal condition the images were pictures of abstract objects that were difficult to name. The self-ordered pointing task is considered to be challenging because it requires the subject to encode, store, and monitor stimuli that were presented and to keep in working memory the objects they already selected on previous trials.

The Petrides and Milner study revealed that patients with frontal lobe lesions showed a significant deficit in performing the task while patients with temporal lobe lesions were not impaired in this task. The self-ordered pointing task has also been used with normal developing and clinical populations. Joseph et al. (2005) used the SOP task with high functioning autistic children as well as normal developing children. Their results showed that autistic children performed equally well when compared to the control group for the non-verbal condition. However, the autistic group showed significant impairments in the verbal condition when compared to the normal developing children.

Other research on SOP has been carried out by Cragg & Nation (2007) using typically developing children between the ages of 5 and 11. Young adults were also tested. Cragg and Nation found that children performed worse than the adults for both the verbal and non-verbal conditions. Results from the children also revealed that the older age group had longer memory spans than the younger age groups. Craig and Nation concluded that the self-ordered pointing task was a sensitive measure of executive functioning and could be useful with several other clinical populations.

The second new measure of executive function used in this study was the Missing Word Task (MWT). This task assesses verbal working memory. The Missing Word task used in this experiment was a modified version of Herman Buschke's "Missing Scan" task (see also Yntema & Trask, 1963). In the original MWT task, sequences of numbers were visually presented to subjects (Buschke, 1968). Participants first saw a sequence of numbers. They then saw a second list which contained all but one of the numbers from the initial list. The subjects were required to report the missing number. For our study, the MWT task was modified in several ways. The task was first changed from a visual to an auditory task. This modified version also used words for stimuli instead of numbers which were in the original task.

This study explored the relationship between executive function and spoken language processing using two relatively new measures. The research was designed to assess the sensitivity and effectiveness of the self-ordered pointing and missing word tasks. In addition to these new measures, subjects were tested on several other executive function and language measures. The present report is the initial part of a two-part study. The experimental study described here used normal hearing adults. The data gathered in this experiment will be used in a second study involving normal-hearing and hearing-impaired children, specifically, a population of congenitally deaf children with cochlear implants at the Indiana University School of Medicine. Our hypothesis was that these two new measures of executive function would correlate with speech and language measures and several other conventional measures of executive function and cognitive control, such as digit span and scores from the BRIEF.

Method

Participants

Twenty adult students (age 18-32 years old) from Indiana University were recruited for this experiment. All subjects received monetary compensation for their participation. Inclusion in this study required subjects to be free of any cognitive, hearing, or speech impairments. Participants were also required to be native speakers of American English. Data were collected in one testing session that took place in a sound-attenuated booth room in the Speech Research Lab in Bloomington, Indiana. Subjects were asked to carry out several tasks which assessed their cognitive and language skills.

Materials

Apparatus. A "Magic Touch"® touch-sensitive monitor was used to display images for the implicit sequence learning and Self-Ordered Pointing tasks. Computer tasks were run on a Power Mac G4. All auditory stimuli were presented over an Advent AV570 loud speaker at a comfortable listening level of 65 dB SPL.

Self-Ordered Pointing Task. Subjects were shown a set of visual images on a touch screen monitor and were required to select one alternative by touching the screen. Once an image had been selected, a new display was presented in which the pictures were rearranged on the screen and the subject was required to select another image that was different from the previously selected one. This process continued until all the images had been selected.

Each image was a GIF file and measured at 100 x 100 pixels on the touch screen monitor. Pictures pseudorandomly fell on a 4 x 5 grid. When images were rearranged on the grid, they were not allowed to fall in the same location as the previous screen. The SOP consisted of a verbal and a nonverbal condition. The verbal condition consisted of black and white line drawings of objects that were easy to name. These images were taken from the International Picture Naming Project, University of California, San Diego (Szelkely et al., 2004). A total of 40 pictures were selected. The names of the objects were all high frequency, high density words and frequency and density were checked using the Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984). The non-verbal condition consisted of black and white abstract images that were difficult to verbalize. Some of these images were kindly provided by Dr. Robert M. Joseph from Boston University School of Medicine (Joseph et al., 2005). The remaining images were created using Microsoft Paint 2001. Examples of the verbal and non-verbal stimuli are shown in Figures 1 and 2.

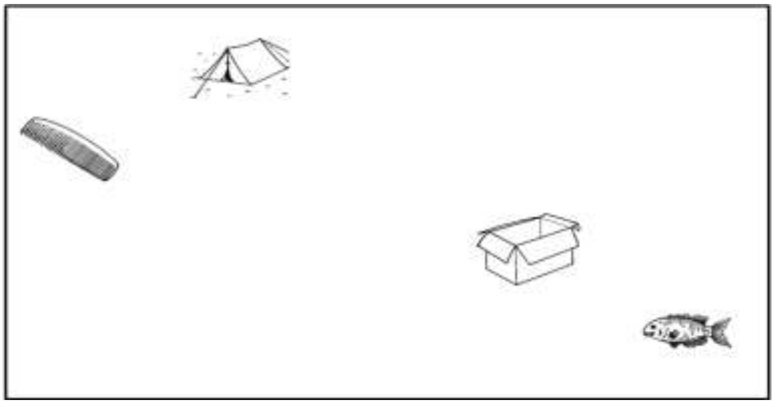

Figure 1. Example of stimuli in the verbal condition for the self-ordered pointing task.

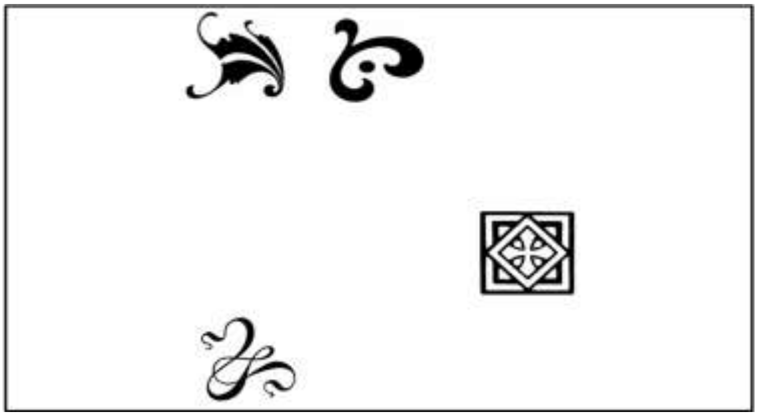

Figure 2. Example of stimuli in the non-verbal condition for the self-ordered pointing task.

Set sizes of 4, 6, 8, 10, and 12 objects were presented three times to each subject. There were three blocks of trials. Each block consisted of all five set sizes. Unique images were presented within each block with no repeated presentations. Images were repeated across blocks. No feedback was given to subjects. Participants completed the task in a fixed order. The non-verbal condition always preceded the verbal condition. This was done to discourage the use of verbal rehearsal in the non-verbal condition.

All subjects completed a practice session at a set size of four. There were no time restrictions. Participants normally completed the task within 15 minutes.

Performance was assessed by identifying the longest correct span for each set size condition. The longest span for each block was identified and then an average of the three blocks was taken to form a single score for a particular condition. Comparisons were made between conditions for the highest correct span attained. Within conditions, performance was also assessed by looking at performance across blocks.

Missing Word Task. The missing word task was similar to the Missing Scan test originally developed by Buschke (1968). In this task, subjects hear a list of words. They were then immediately presented with another list of words which consisted of all but one of the words from the initial list. Subjects were then required to recall the missing word. Subjects had to remember the item that was presented in the initial list but missing from the second.

A total of 54 words were selected for this task. Stimuli were presented at a rate of one second per item. There was a one second pause between the presentation of the first and second list. Stimuli were randomly presented in the initial list but pseudorandomly assigned in the second. Words were not allowed to be presented in the same order as they were in the first list. Words were taken from the Modified Rhyme Test (House, Williams, Hecker, Kryter, 1965). Auditory recordings were made by a female speaker. All stimuli were high frequency, high density CVC concrete nouns (i.e. bat, cave, pig, king, lip, meat) and were checked against the Hoosier Mental Lexicon. Words that were used in the SOP were excluded from this task.

Subjects were presented with set sizes of 3, 4, 5, and 6 items. Each set size was presented three times. There were three blocks of set sizes. Each block contained each set size. Stimuli were unique within set sizes. Prior to the testing portion of the task, subjects completed two practice trials. Subjects were instructed to say and write down the missing word during practice trials but were only required to write their response during the testing phase. Subjects were required to say the missing word aloud during the practice trial in order to insure that they understood the instructions.

Scoring for Missing Word was similar to that of SOP. The longest span for each block was identified and then an average of the three blocks was taken to form a single overall score. Performance was also assessed by looking at the longest correct span across blocks.

Digit Span. The WISC III auditory forward and backward digit span (Wechsler, 1991) was administered in order to measure subject's information processing capacity of immediate memory. Lists of spoken digits were played through a loudspeaker. Auditory recordings of the lists were made with a female talker. Digits were presented at a rate of one second per item. List lengths started at two items per list and increased by one item to the longest list which had a length of ten digits. Lists increased in length only if the subject was successfully able to repeat back at least one of the lists at a given length. The task ended when subjects could not repeat back both of the lists at the same length. The forward digit span task required subjects to simply repeat back what they heard. In the backward digit span task, subjects were told to repeat the list of digits in the reverse order in which the items were presented. Performance was assessed by recording the longest correct span for both the forward and backward conditions.

Implicit Sequence Learning. This task was administered in order to assess participants' procedural memory, specifically, their ability to implicitly learn sequences of items. Subjects were exposed to visual sequences that were generated according to a finite-state "grammar" which specified the order of the sequence elements (Karpicke & Pisoni, 2004). The stimuli consisted of sequences of colored squares. Subjects were required to observe the entire sequence first and then reproduce it by pressing the appropriate location on the touch screen in the order that was shown. Subjects observed each visual sequence and then recorded their response by pressing a touch screen monitor.

The sequence learning task consisted of two phases, a Learning Phase and a Test Phase. Each phase consisted of thirty-two sequences. The sequences were divided into four blocks with each block consisting of two exposures to set sizes 5, 6, 7, and 8. During the Learning Phase, subjects were exposed to one grammatical sequence whereas in the Testing Phase subjects saw test sequences generated by two different grammars. Half of these test sequences followed the grammar used in the Learning Phase. The other grammar used to create the color sequences was an unfamiliar grammar that was not used during the initial training phase. Presentation of the trained grammar and novel grammar sequences were also randomized during the testing phase.

Subjects' implicit sequence learning ability was assessed by examining changes in their memory span for sequences for the trained grammatical versus the novel grammatical stimuli. A learning score was calculated by subtracting subjects' novel grammar sequence score from their trained grammar sequence score.

BRIEF. The Behavioral Rating Inventory of Executive Function-A (Ruth, Isquith, & Gioia, 1996) is a self-report rating scale that was developed to assess executive function in real-world environments. The BRIEF-A can be administered to adults between the ages of 18-90. The inventory consists of 75 questions. Subjects rate the questions as applying to them rarely, sometimes, or often. The questions cover several different domains of executive functioning. The BRIEF is composed of nine clinical subscales and three composite scores. The subscales measure inhibition, shifting, emotional control, self-monitoring, initiation, working memory, planning and organization, task monitoring, and organization of materials. Higher scores are indicative of problematic behavior.

Language Assessment

Spoken Sentence Perception Task. To obtain a measure of sentence perception, subjects were asked to listen to degraded sentences that were difficult to perceive. Participants were required write down the last word of each test sentence.

A total of seventy-five speech intelligibility in noise (SPIN) sentences were presented through a loudspeaker (Kalikow et al., 1977). The SPIN test was designed to assess speech perception and spoken word recognition in predictable or unpredictable sentence contexts. The SPIN sentences are meaningful English sentences that vary in the predictability of the last word in the sentence. The stimuli consisted of high predictability (HP), low predictability (LP), and anomalous sentences (AN). Examples of these sentences are presented in Table 1. An HP sentence contained semantic context that aided in the identification of the final word whereas LP sentences had little semantic content. AN sentences contained semantically meaningless content and provided no aid in the identification of the target word. Subjects heard twenty-five sentences from each set of sentences. Auditory stimuli were created using a male speaker. The sentences were designed to simulate the auditory experience of a CI user and were first degraded down to six spectral channels and then further degraded using a sinewave vocoder (http://www.tigerspeech.com). Sentences were presented in random order to each subject. A written response was counted as correct if it matched the final target word in each sentence.

High Predictability	Hold the baby on your lap.
Low Predictability	Mr. Brown can’t discuss the slot.
Anom	For a bloodhound he had spoiled pie.

Table 1. Examples of SPIN sentences.

Vocabulary Test. The Peabody Picture Vocabulary Test (3rd edition) was administered to all subjects in order to assess age-appropriate vocabulary levels. Participants were shown four pictures and were asked to select the image that best depicts the word that was spoken by the examiner.

The PPVT is a standard measure of vocabulary for ages 2.5- 90 (Dunn & Dunn, 1997). Stimulus words are broken up into numbered sections. Each section contained twelve words. Words within each section became more difficult as the test progressed. A baseline vocabulary level was established when a subject scored one error or less in a section. The testing session ceased when a subject made eight or more errors in a section. The subject’s vocabulary level was assessed by taking the number of the highest word correctly answered and subtracting the number of errors made.

Nonword Repetition Task. The nonword repetition task was administered in order to assess subjects’ phonological awareness. In the task, subjects were told they would hear an unfamiliar word and their task was to repeat back what they heard. Two nonwords were presented to each subject prior to the start of testing to serve as practice trials.

Participants heard a total of 44 nonwords. The nonword stimuli were originally constructed by Edwards et al. (2004). Auditory recording of these stimuli were later created by Dillon (2005). Half of the stimuli were composed of phonotactic sequences that occur with a low frequency in English; whereas, the remaining stimuli had phonotactic sequences that occur with a high frequency in English. In addition, words at each frequency contained different syllable lengths. Half of the stimuli within a frequency were either 2 or 3 syllables in length. In total, subjects heard eleven 2-syllable nonwords and eleven 3-syllable nonwords for both the low and high frequencies categories.

Subjects’ responses were scored by three different experimenters. A response was scored as correct if all three judges agreed that the subject was able to accurately repeat the nonword. Judges scored subjects’ responses separately.

Results

Executive Function and Verbal Working Memory

Descriptive statistics for executive function and verbal working memory measures are listed in Table 2.

Measure	M	SD	Minimum	Maximum
SOP Verbal	10.66	1.36	7.33	12
SOP Non-verbal	9.63	1.95	5.33	12
Missing Word	5.34	.65	4	6
Digit Forward	7.7	1.41	5	10
Digit Back	6.15	1.34	3	9
Grammar A	78.8	20.92	19	104
Grammar B	58.05	21.73	17	99
Learning	20.75	21.61	-16	53

Table 2. Summary of Descriptive Statistics for the Executive Function Measures (N=20): Mean score, standard deviation, minimum value, and maximum value. (Note. SOP Nonverbal, Self-ordered pointing Nonverbal condition highest correct span; SOP Verbal, Self-ordered pointing Verbal condition highest correct span; Digit Forward, Digit Span Forward; Digit Back, Digit Span Backwards; Grammar A, Implicit Sequence Learning Task Grammar A; Grammar B, Implicit Sequence Learning Task Grammar B; Learning, Implicit Sequence Learning Task Learning.)

Self-Ordered Pointing. Initial analyses of SOP were carried out in order to assess subjects' performance across the different conditions and blocks. Performance was assessed by recording the longest correct span for each subject. The longest span was the average correct span from the three block presentations. The mean span for the verbal condition was 10.66 with a standard deviation of 1.36; the mean span for the non-verbal condition was 9.63 with a standard deviation of 1.95. A two-way analysis of variance was conducted with condition and block as the independent factors and subjects' responses as the dependent variable. Main effects of condition, F(1, 114)= 6.857, p= .01, and block, F(2, 114)= 7.47, p= .001, were found. No interaction was found.

Overall subjects had a longer SOP span for the verbal condition (M= 10.66, SD= 1.36) in comparison to the non-verbal condition (M= 9.63, SD= 1.95). A paired samples t-test revealed a significant difference between the conditions, t(19)= 3.24, p= .004. This difference is most likely due to the ease at which images in the verbal condition can be scanned in working memory in comparison to the non-verbal conditions. In further comparing the verbal and the non-verbal conditions, follow up t-tests showed a significant difference between the two conditions for Block 1, t(19)= 2.67, p< .015, but not for

Block 2, t(19)= 1.36, p= .189, or Block 3, t(19)= 1.37, p= .186. This result may have been due to subjects performing close to ceiling for the verbal condition. There was little room for improvement across blocks for this condition. Figure 3 displays subjects' responses for the verbal and non-verbal conditions across all three blocks.

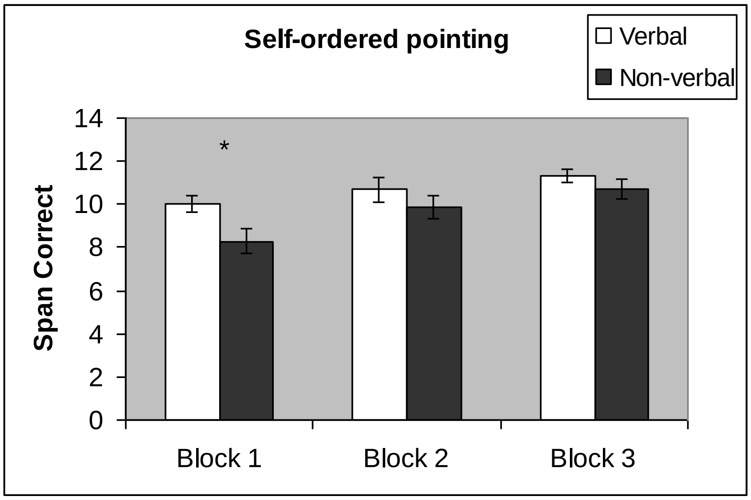

Figure 3. Correct span scores for the self-ordered pointing task.

To further assess SOP performance, analyses were done examining each condition separately. A one-way analysis of variance test was conducted for the non-verbal condition with block as the independent variable and the subjects' score as the dependent variable. A main effect was found across blocks, F(2, 57) = 5.317, p= .008. A set of t-tests showed significant differences between Blocks 1 and 2, t(19)= -2.89, p= .009, and Blocks 1 and 3, t(19)= -4.188, p< .000. No difference was found for Blocks 2 and 3 for the non-verbal condition. For the non-verbal SOP condition, there was a trend for subjects to perform better across block presentations. Another analysis of variance test was run in order to assess subjects' performance in the verbal condition. No main effect of block was found in the verbal condition, F(2, 57)= 2.272, p= .112. Although for the verbal SOP condition, there was a trend for subjects to show better performance across blocks but this trend was not as strong as that found in the non-verbal condition.

Additional analyses were done comparing the SOP task and language outcome measures. All prior analyses of SOP were done looking at the different set-size conditions and blocks within the self-ordered pointing task. Regression analyses were conducted to assess how much subjects' performance on the SOP task predicted their speech and language abilities. Linear regression analyses revealed that higher vocabulary scores on PPVT were associated with executive function scores for the non-verbal condition, t(18)= 3.41, p= 0.003. Figure 4 displays this result. No association was shown between PPVT and the verbal condition of self-ordered pointing task, t(18)= 1.35, p= .194. It would appear a better vocabulary aided subjects' performance more in the non-verbal condition where objects are difficult to name than in the verbal condition where objects were easy to name. Subjects with a better vocabulary most likely used some form of verbal coding to encode the abstract images for the self-ordered pointing task.

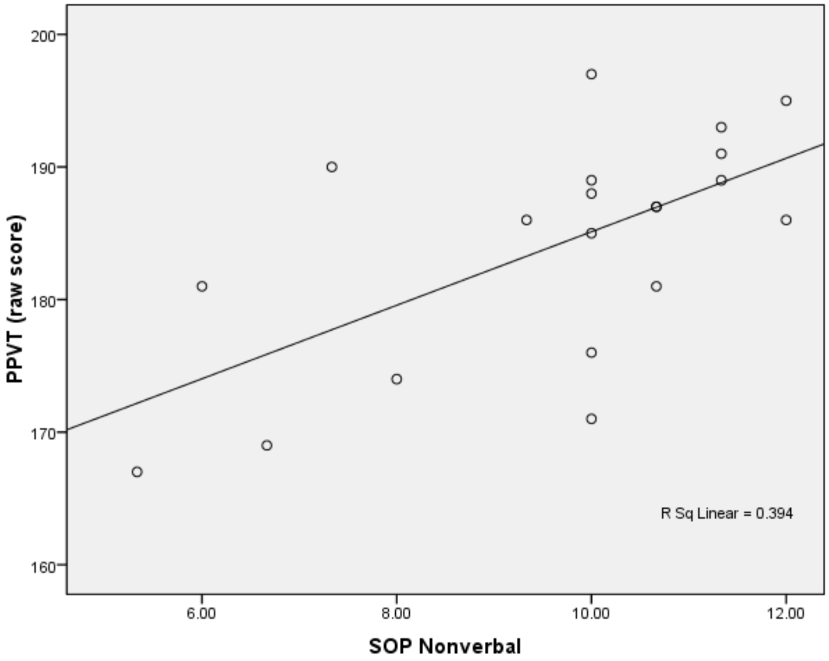

Figure 4. Scatter plot of PPVT and SOP Non-verbal data. The x-axis displays the self- ordered pointing scores for the non-verbal condition; the y-axis displays the subjects' PPVT scores.

Missing Word Task. Initial analyses for the missing word task assessed subjects' performance at different set sizes and across block presentations. Performance was assessed by examining whether the subject was able to recall the missing word at each set size presentation. A two-way ANOVA was run with set size and block as the independent variables and subjects' performance as the dependent variable. A main effect of set size was found, F(3, 228)= 9.76, p< .000. There was a trend for subjects to perform worse as set size increased. No main effect of block was found, F(2, 228)= .000, p= 1.00. No interaction effect was found between set size and block, F(6, 228)= .24, p= .963.

Performance was also assessed by looking at the longest correct span across blocks. Additional analyses between the missing word task and other measures from the study were done using this performance assessment. Subjects had a mean missing span score of 5.34 (SD= 1).

Digit Span. Immediate memory span capacity was assessed by identifying the longest correct span for both the forward and backward conditions. The mean for forward span was 7.7; the mean for backward span was 6.15. A paired samples t-test revealed a significant difference between the forward and backward spans t(19)= 5.431, p< .000. In this task, it is common for subjects to perform worse in the backward condition because it requires the use of more complex cognitive processes than the forward condition. The backward condition is a working memory task where stimulus items must be maintained and manipulated while the forward condition is a short term memory task which requires passive storage and retrieval of items (Baddeley, 1986).

Sequence Learning. A paired sample t-test was conducted and showed a significant difference between subjects' scores for grammar A sequences and grammar B sequences, t(19)= 4.29, p< .000. Differences between the two grammars are shown in Figure 5. Overall, subjects performed better in the trained grammar than the novel grammar. A single sample t-test revealed that the implicit learning score was significantly greater than zero, t(19)= 4.29, p< .000. Figure 6 shows individual subject sequence scores. With the exception of four, all subjects had higher grammar A scores in comparison to grammar B. The majority of subjects were able to implicitly learn the underlying grammar.

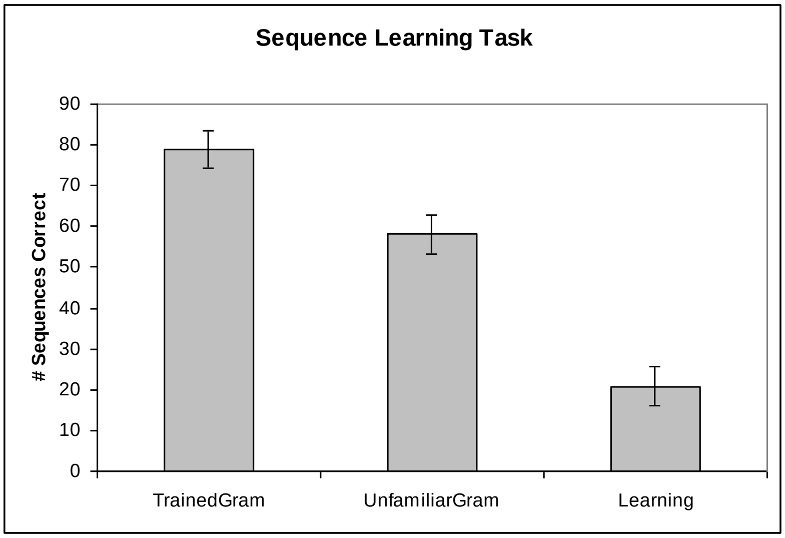

Figure 5. Performance scores for the implicit sequence learning task. (Note. TrainedGram, Number of correctly replicated sequences that were from the trained grammar; UnfamiliarGram, Number of correctly replicated sequences that were from an untrained grammar; Learning, Difference scores for subjects' performance on the trained and untrained grammar sequences.)

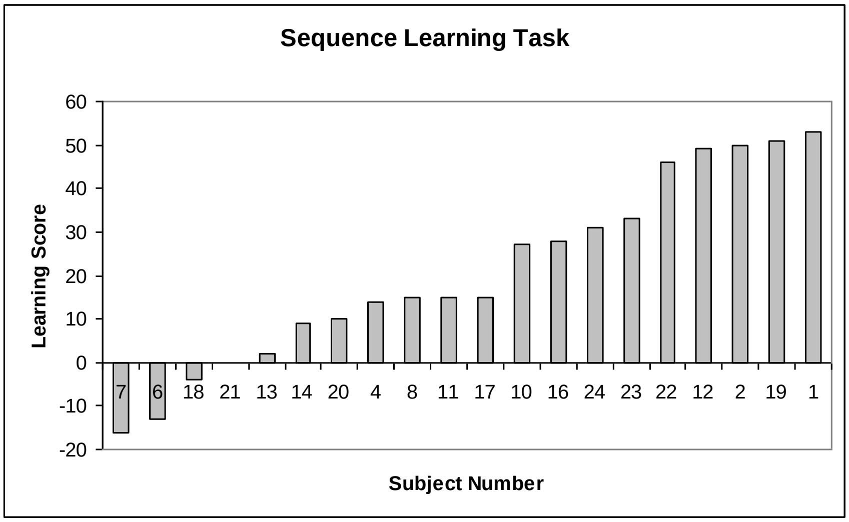

Figure 6. Bar graph of subjects learning scores for the sequence learning task. The x-axis displays each individual subject number; the y-axis displays the learning score or the difference between the grammar A score and grammar B score.

BRIEF. A t-test revealed that subjects were elevated above the age normed mean of 50 for four subscales and one composite scale. Subjects were significantly elevated on the scales for working memory, t(19)= 2.52, p= .021, planning and organization, t(19)= 2.54, p= .02, task management, t(19)= 2.37, p= .028, and metacognition t(19)= 3.31, p= .004. Subjects scores were elevated but did not surpass the clinical range which consists of scores greater than 65. Figure 7, 8, and 9 show subjects' results for the subscales and composite scales.

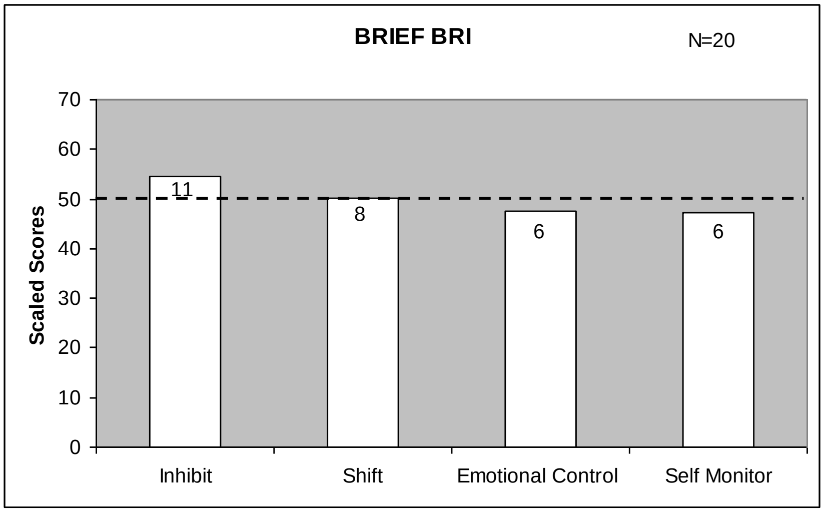

Figure 7. Bar graph of subjects' BRIEF BRI subscale scores. (Note. Numbers above or within the bars indicates the number of subjects who were elevated above the norm mean of 50.)

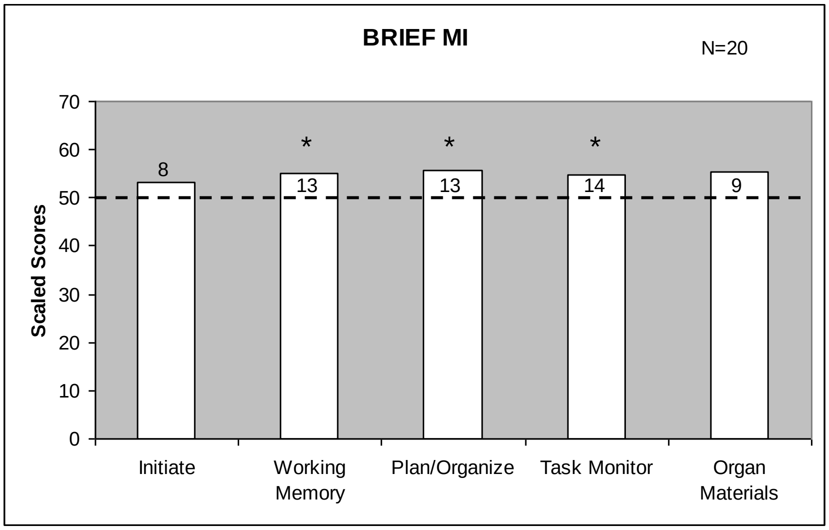

Figure 8. Bar graph of subjects' BRIEF MI subscale scores. (Note. Numbers above or within the bars indicates the number of subjects who were elevated above the norm mean of 50.)

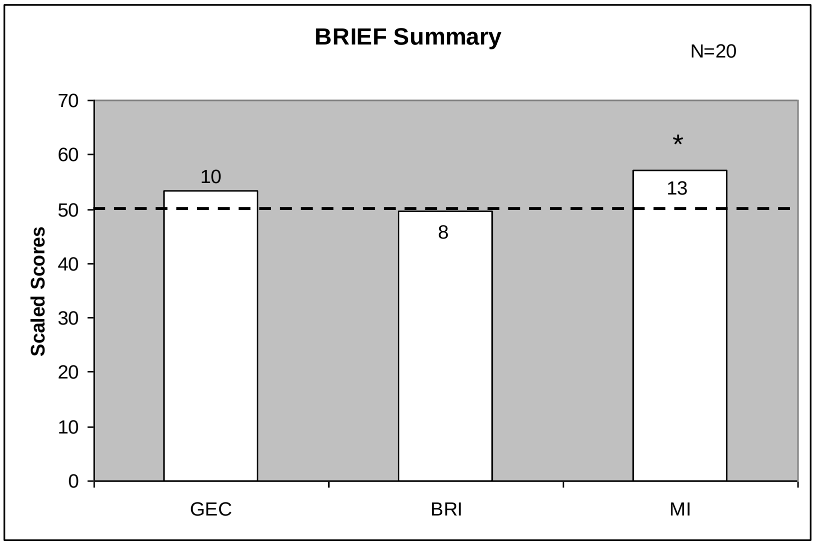

Figure 9. Bar graph of subjects' BRIEF composite scores. (Note. Numbers above or within the bars indicates the number of subjects who were elevated above the norm mean of 50.)

Spoken Language Processing Measures

Descriptive statistics for these measures are summarized in Table 3.

Measure	M	SD	Minimum	Maximum
SPIN HP	18.7	3.23	13	24
SPIN LP	13.05	3.45	5	19
SPIN AN	14.2	3.27	7	18
HP-LP	5.65	2.81	-1	10
HP-AN	4.5	2.32	-1	8
LP-AN	-1.15	1.89	-4	3
PPVT	112.5	10.94	94	137
Nonword	38.55	2.67	32	43

Table 3. Summary of Descriptive Statistics for the Language Measures (N=20): Mean score, standard deviation, minimum value, and maximum value. (Note. SPIN HP, number of high predictability sentences correct; SPIN LP, number of low predictability sentences correct; SPIN AN, number of anomalous sentences correct; HP-LP, SPIN

difference between HP and LP categories; HP-AN, SPIN difference between HP and AN categories; LP-AN, SPIN difference between LP and AN categories; PPVT, Peabody Picture Vocabulary Test assesses receptive vocabulary; Nonword, number of correctly repeated nonwords.)

SPIN. Initial analyses for this task were done comparing the different conditions as well as difference scores between conditions. For the sentence perception task, subjects on average correctly identified 80% of words from the high probability sentences, 54% low probability, and 60% of anomalous sentences. A t-test showed significant differences between means for the high and low conditions t(19)= 8.97 p< .000, low and anomalous conditions t(19)= -2.71 p< .014, and the high and anomalous conditions t(19)= 8.64 p< .000. Overall, subjects performed better on the high probability sentences in comparison to the low probability sentences and the anomalous sentences.

Performance was also assessed by computing difference scores between the HP-LP, HP-AN, and LP-AN conditions. Taking the difference between sentence conditions allows for analyses of individual improvement in performance. Additional analyses were then done comparing these difference scores to other speech and language measures. Regression analyses were conducted in order to examine how much subjects' performance on the SOP task predicted their speech perception abilities. Linear regression analyses were done comparing the SPIN difference scores to the self-ordered pointing task. SPIN difference scores were used as the dependent variable and self-ordered pointing performance in the non-verbal condition was used as the independent variable. No association was found between HP-LP and the nonverbal SOP performance t(18)= 1.79, p= .089. A relationship was found between the HP-AN difference score and SOP non-verbal performance, t(18)= 2.38, p= .028. Better recall of abstract objects in the SOP task indicates a greater improvement in performance between the HP and AN conditions. No association was also found for the LP-AN difference score t(18)= .09, p= .926. Analyses comparing SOP verbal scores and the different SPIN conditions revealed no significant results.

Linear regression analyses were also done comparing performance on SPIN and the implicit sequence learning task. Regression analyses were again conducted to assess how much subjects' performance on the sequence learning task predicted their speech perception abilities. SPIN performance was used as the dependent variable and the learning score for the sequence task was used as the independent variable. No relationship was found between SPIN HP and the learning scores, t(18)= -1.57, p= .132. However, higher sequence learning scores were associated with poor performance on SPIN LP, t(18)= -2.22, p= .039 and on SPIN AN words, t(18)= -2.24, p= .038. In other words, subjects who had higher learning scores for the sequence task had lower scores on the speech perception task for low and anomalous words. This was an unexpected finding; we expected that a higher learning score would indicate better speech perception performance on the SPIN task.

PPVT and Nonword Repetition. The mean score for PPVT was 112.5 with a standard deviation of 10.94. Subjects were significantly above the age norm mean of 100 for the PPVT vocabulary test t(19)= 5.11, p<.000. For the Nonword Repetition task, responses were scored as correct if all three judges agreed that the subject accurately repeated the nonword. Bivariate correlation analyses were done to compare judges' scores and to assess inter-rater reliability. Weak associations were shown between judges 1 and 2, r= .381, p< .000; judges 1 and 3 r= .569, p< .000; and judges 2 and 3, r= .336, p< .000. Overall judges agreed on which subject responses were correct. A paired sample t-test showed no difference between 2 syllable and 3 syllable nonwords, t(19)= -.364, p= .72. No significant difference was also found between high frequency and low frequency nonwords, t(19)= -.343, p= .735. After examing these findings, all nonwords regardless of frequency or syllable size were collapsed together to form one category.

In order to assess the relations between executive function/cognitive control measures and language tasks, bivariate correlations were also computed among the measures. Table 4 shows the correlations for the different tasks. To account for the many comparisons used in this study, a Bonferroni correction was conducted with an alpha level set at .003. As expected, performance on the SOP verbal condition was moderately correlated with performance in the non-verbal condition, r= .683, p< .000. Subjects with high SOP memory spans in one condition also had high spans in the other. An association was also found between SOP non-verbal condition and the Digit Span backward condition, r= .648, p= .001. This indicates that subjects who had longer memory spans in the SOP non-verbal condition also had longer backward memory spans. Also as expected, a few significant correlations were found within an individual test. Performance on SPIN high probability (HP) and low probability (LP) were correlated, r= .647 p= .002, as was performance on SPIN HP and SPIN anomalous, (AN) r= .743 p< .000 and SPIN LP and SPIN AN, r= .841 p< .000. Table 4 shows additional correlations found between measures, although many of these correlations failed to meet the Bonferroni alpha level.

Measure	SOPVerb	SOPNonV	MissingW	DigitF	DigitB	GramA	GramB	SPIN *HP*	SPIN *LP*	SPIN *AN*
1. SOP Verb	--	--	--	--	--	--	--	--	--	--
2. SOP NonV	.683***	--	--	--	--	--	--	--	--	--
3. Missing W	.031	.164	--	--	--	--	--	--	--	--
4. Digit F	.287	.162	.468*	--	--	--	--	--	--	--
5. Digit B	.266	.648***	.278	.575***	--	--	--	--	--	--
6. GramA	.216	.150	.083	-.023	.139	--	--	--	--	--
7. GramB	.181	.049	-.011	.313	.225	.487*	--	--	--	--
8. SPIN *HP*	.382	.103	.242	.243	.421	.108	.450*	--	--	--
9. SPIN *LP*	.039	-.185	.153	.303	.269	.002	.463*	.647***	--	--
10. SPIN *AN*	.028	-.219	.117	.172	.124	-.157	.313	.743***	.841***	--
11. HP-LP	.346	.39	.089	-.093	.153	.121	-.052	.353	-.485*	-.181
12. HP-AN	.451*	.49*	.171	.095	.41	.37	.185	.343	-.285	-.373
13. PPVT	.303	.627**	.106	.312	.518**	-.054	.244	.537*	.031	.109
14. NonRep	.243	.067	-.111	.449*	.371	-.081	.223	.258	.219	.252

Table 4. Correlation Matrix for Executive Function and Language Measures ((Note. * p<.05, ** p<.01, *** p<.001 (two-tailed). SOP Verb, Self-ordered pointing highest correct span; SOP NonV, Self-ordered pointing Non-verbal highest correct span; Missing W, Missing Word task highest correct span; Digit F, Digit Span Forward highest correct span; Digit B, Digit Span Backwards highest correct span; SPIN *HP*, SPIN High Probability # of key words correct; SPIN *LP*, SPIN Low Probability # of key words correct; SPIN *AN*, SPIN Anomalous # of key words correct; HP-LP, SPIN difference between High Probability score and Low Probability score; HP-AN, SPIN difference between High Probability score and Low Probability score; PPVT, Peabody Picture Vocabulary Test

scaled score; NonRep, # of nonwords correctly repeated.))

Measure	HP-LP	HP-AN	PPVT	NonRep
1. SOP Verb	--	--	--	--
2. SOP NonV	--	--	--	--
3. Missing W	--	--	--	--
4. Digit F	--	--	--	--
5. Digit B	--	--	--	--
6. GramA	--	--	--	--
7. GramB	--	--	--	--
8. SPIN *HP*	--	--	--	--
9. SPIN *LP*	--	--	--	--
10. SPIN *AN*	--	--	--	--
11. HP-LP	--	--	--	--
12. HP-AN	.742***	--	--	--
13. PPVT	.578**	.592**	--	--
14. NonRep	.027	.004	.475*	--

Table 4 continued ((Note. * p<.05, ** p<.01, *** p<.001 (two-tailed).)

Discussion

This study examined the relations between executive function and several language outcome measures in a group of young adults. Specifically, our goal was to further explore the relationship between working memory and spoken language processing. This was done through the use of two new measures of executive function, the self-ordered pointing task (SOP) and the missing word task (MWT). The purpose of this study was to assess the effectiveness and feasibility of the self-ordered pointing and missing word tasks as measures of executive function and cognitive control in young children.

We replicated previous research involving SOP (Petrides & Milner, 1982; Cragg & Nation, 2007). Cragg and Nation showed that young adults have longer working memory spans for the verbal condition in comparison to the non-verbal condition. In our study, subjects also performed better in the verbal condition compared to the non-verbal condition. The difference in performance in the two conditions is most likely due to the ease of verbal labeling in one condition versus the other. In the verbal condition, objects are easy to name; whereas, images in the non-verbal condition are abstract and difficult to verbalize. As with the other studies done using the SOP, this study showed a trend for subjects to improve in performance across successive blocks.

An unexpected result occurred in the analyses of the missing word task (MWT) and SOP. We expected to find an association between the verbal condition of SOP and the MWT task because both are measures of verbal working memory. However, no association was found. This may have been due to the ceiling effects that were obtained in the verbal condition of SOP. The ceiling effects could have minimized any significant findings that may exist between SOP and MWT. The lack of association between these two tasks may have also occurred because the SOP task is more cognitively demanding than the MWT. SOP draws upon different aspects of executive functioning such as attention, inhibition, working memory, and the ability to organize materials, whereas, MWT is a less demanding working memory task. This explanation receives some support from the correlations we observed. The non-verbal condition of SOP was highly correlated with backward Digit Span. The other finding was the association, although a weak one, observed between MWT and forward Digit Span. SOP and backward digit span are difficult tasks which draw upon higher cognitive processes, such as cognitive control abilities required to maintain and monitor the contents of working memory. Moreover, MWT and the forward Digit Span require fewer and different processes. In light of these results, it seems unlikely for SOP to correlate with MWT.

We also replicated some of the earlier findings of Conway, Karpicke, and Pisoni (2007) who showed a relationship between implicit sequence learning and spoken language processing. Subjects who had higher sequence learning spans also performed better in the speech perception task for high and low probability words. Overall, subjects were also able to recall more words from high probability sentences than low or anomalous sentences. However, we found an unexpected result that does not support Conway, Karpicke, and Pisoni. An association was observed between the learning score for the sequence task and subjects' performance on the SPIN speech perception task. Subjects who had higher learning scores showed lower scores for low predictability and anomalous words. No correlations were found for the difference scores between the high and low predictability speech perception conditions and sequence learning scores. It is unclear why this result occurred. We expected that higher learning scores would be associated with better speech perception skills.

We also found a relation between performance on the non-verbal part of the SOP task and subjects' vocabulary level. Subjects who had higher receptive vocabulary scores on the PPVT also performed better on the non-verbal portion of SOP. This was an unexpected finding because we had anticipated an association between the SOP verbal condition and vocabulary scores. We expected that subjects would have more difficulty in assigning names to the non-verbal images. However, it appears that subjects were able to use some form of verbal coding to remember the non-verbal objects. Better vocabulary scores aided subjects more in the non-verbal condition than in the verbal condition where the names of images were basic elementary level words (i.e. girl, boat, and star).

The present study had several limitations. One goal of this research was to further examine the relationship between executive functions and spoken language processing abilities. Only weak associations were found among the cognitive and language measures obtained in the study. As stated earlier, a significant correlation was found between SOP and vocabulary level. Nevertheless, many of the findings were only close to or marginally significant. Perhaps the use of more subjects would have helped to elucidate the relationship between executive functions and spoken language processing.

Another limitation lies in the tests chosen to measure executive functioning in this study. The SOP task only measures certain aspects of executive function as does MWT, Digit Span, and BRIEF. It has been argued that none of these measures are able to assess the full range of executive functions. The ability for any single test to successfully accomplish this enormous feat is questionable considering the many different areas that encompass executive function. It may be more advantageous to approach research on executive function and cognitive control with multiple measures that assess different areas of executive function rather than rely on only a single measure.

It is also doubtful that in everyday real-world situations all aspects of executive function work independently of one another. Tests such as STROOP and the Wisconsin Card Sorting Task, which measure inhibition and the ability to shift and focus attention, work in this way in that each measure assesses one specific area of executive function. Executive function should be measured with the systems working together to accomplish a goal. Although tests such as SOP, do not measure the full range of executive functioning, the usefulness of the task in assessing selective aspects of executive function and cognitive control should not be overshadowed by what some researchers may see as a limitation of its ability to measure executive functioning. The self-ordered pointing task requires the subject to encode objects in working memory, perform a response, and monitor his or her responses that were already made. This is a challenging task which measures several different aspects of executive functioning. The missing word task is almost as equally demanding as SOP because subjects must retain a series of auditory stimuli and recall a single item from amongst the list. SOP and MWT are set apart from other executive function tasks in that they test various aspects of executive function and require the subject to retain, maintain, manipulate, and process information over time.

As stated earlier, this study is the initial part of a larger two-phase research project. The measures used in this study will also be used to test groups of normal hearing and hearing impaired children. Specifically, we plan to look at congenitally deaf children with cochlear implants. It is our hope that the new measures developed in this study will provide more detailed insights into the cognitive deficits seen in implanted children. This initial study done with adults provides additional support showing links between executive function and spoken language processing. However, more research needs to be done with the clinical population of pediatric implant users in order to understand why some users gain little to no benefit from their device. In future research, we hope to gain a better understanding of the relation between executive function and spoken language processing in implanted children. The ultimate goal of our research program is to identify the underlying neurocognitive factors that predict speech and language outcomes in deaf children with cochlear implants. Identifying these factors can help researchers better understand the large individual differences observed in deaf children following cochlear implantation.

References

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8,* 71-82.

Anderson, V. (2002). Executive function in children: Introduction. *Child Neuropsychology, 8,* 69-70.

Anaya, E.M., Conway, C.M., Pisoni, D.B., Geers, A., & Kronenberger, W. (2008). *Effects of cochlear implantation on executive function: Some preliminary findings.* Poster Presentation at the 10th International Conference on Cochlear Implants and other Implantable Auditory Technologies. San Diego, CA, April.

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

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *Recent Advances in Learning and Motivation, Volume 8* (pp. 47-89). New York: Academic Press.

Bernstein, J.H., & Waber, D.P. (2007). Executive capacities from a developmental perspective. In L. Meltzer (Ed.), *Executive Function in Education*. Guilford Press.

Birnholz, J.C., & Benacerraf, B.R. (1983). The development of human fetal hearing. *Science, 222,* 516-518.

Burkholder, R.A., & Pisoni, D.B. (2006). Working memory capacity, verbal rehearsal speed, and scanning in deaf children with cochlear implants. In P.E. Spencer & M. Marschark, *Advances in the spoken language development of deaf and hard-of-hearing children. Perspectives on deafness.* Oxford University Press.

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

Buschke, H. (1968). Relative vulnerability of item-information in short-term memory for the missing scan. *Journal of Verbal Learning and Verbal Behavior, 7,* 1043-1048.

Chase, H.W., Clark, L., Sahakian, B.J., Bullmore, E.T., & Robbins, T.W. (2008). Dissociable roles of prefrontal subregions in self-ordered working memory performance. *Neuropsychologia, 46,* 2650-2661.

Cleary, M., Pisoni, D.B., & Kirk, K.I. (2000). Working memory spans as predictors of spoken word recognition and receptive vocabulary in children with cochlear implants. *The Volta Review, 102,* 259-280.

Cochlear implants in adults and children. NIH Consensus Statement Online 1995 May 15-17; 13(2): 1-30.

Cragg, L., & Nation, K. (2007). Self-ordered pointing as a test of working memory in typically developing children. *Memory, 15,* 526-535

DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208,* 1174-1176.

Dillon, C.M. (2005). Phonological processing skills and the development of reading in deaf children who use cochlear implants. *Research on Spoken Language Processing Technical Report, 14,* 1-84.

Dillon, C.M., Burkholder, R.A., Cleary, M., & Pisoni, D.B. (2004). Nonword repletion by children with cochlear implants: Accuracy ratings from normal-hearing listeners. *Journal of Speech, Language, and Hearing Research, 47,* 1103-1116.

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

Edwards, J., Beckman, M.E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probabilities effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47,* 421-436.

Fagan, M.K., Pisoni, D.B., Horn, D.L., & Dillon, C.M. (2007). Neuropsychological correlates of vocabulary, reading, and working memory in deaf children with cochlear implants. *Journal of Deaf Studies and Deaf Education, 12,* 461-471.

Gagnon, R. (1989). Stimulation of human fetuses with sound and vibration. *Semin Perinatol, 13,* 393-402.

Gathercole, S.E., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection. *Journal of Memory and Language, 29,* 336-360.

Geers, A., Spehar, B., & Sedey, A., (2002). Use of speech by children from total communication programs who wear cochlear implants. *American Journal of Speech-Language Pathology, 11,* 50-58.

Gioia, G., Andrews, K., Isquith, P., (1996). *Behavioral Rating Inventory of Executive Function-Preschool Version Professional Manual*. Psychological Assessment Resources Inc.

Gioia, G., Isquith, P., Guy, S., Kenworthy, L. (1996). *Behavioral Rating Inventory of Executive Function Professional Manual*. Psychological Assessment Resources Inc.

Golden, C.J. (1981). The lubria-nebraska children's battery: Theory and formulation. In G.W. Hynd & G.E. Obrzut (Eds.), *Neuropsychological assessment and the school-aged child.* New York: Grune & Stratton.

Gouveia, P.A., Brucki, S.M., Malheiros, S.M., & Bueno, O.F. (2007). Disorders in planning and strategy application in the frontal lobe lesion patients. *Brain and Cognition, 63,* 240-246.

Hammond, J. (1970). Hearing and response in the newborn. *Developmental Medicine & Child Neurology, 12,* 3-5.

Harrison, R.V., Panesar, J., Ef-Hakim, H., Abdolell, M., Mount, R.J., & Papsin, B. (2001). The effects of age of cochlear implantation on speech perception outcome in prelingually hearing impaired children. *Scandinavian Audiology, 30,* 73-78.

Hauser, P.C., Lukomski, J., & Hillman, T. (2008). Development of deaf and hard-of –hearing students' executive function. In M. Marshark & P.C. Hauser, *Deaf Cognition Foundations and Outcomes.* Oxford University Press.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting TestManual: Revised and Expanded*. Odessa, FL: Psychological Assessment Resources.

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.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions. *Child and Adolescent Mental Health, 7,* 131-142.

Joseph, R.M., Steele, S.D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: Failure to use verbal mediation in the service of working memory?. *Neuropsychologia*, 43(10), 1400-1411.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Karpicke, J.D., & Pisoni, D.B. (2004). Using immediate memory span to measure implicit learning. *Memory and Cognition, 32,* 956-964.

Kileny, P.R., Zwolan, T.A., & Ashbaugh, C. (2001). The influence of age at implantation on performance with a cochlear implant in children. *Otology & Neurotology, 22,* 42-46.

Kirk, K.I., Miyamoto, R.T., Ying, E.A., Perdew, A.E., & Zuganelis, H.Z. (2002). Cochlear implantation in young children: Effects of age at implantation and communication mode. *The Volta Review, 102,* 127-144.

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* (pp. 252-254). New York: Thieme.

Luria, A.R. (1973). *The Working Brain an Introduction to Neuropsychology*. Basic Books.

Luria A.R. (1966). *Higher cortical functions in man*. Basic Books, New York.

Mahone, E.M., Cirino, P.T., Cutting, L.E., Cerrone, P.M., Hagelthorn, K.M., Hiemenz, J.R., Singer, H.S., & Denckla, M.B. (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology, 17,* 643-662.

Mares, D., McLuckie, A., Schwartz, M., & Saini, M. (2007). Executive function impairments in children with attention-deficit hyperactivity disorder: Do they differ between school and home environments. *Canadian Journal of Psychiatry, 52,* 527-534.

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews Neuroscience, 24,* 167-202.

Miyamoto, R.T., Kirk, K.I., Svirsky, M., & Sehgal, S.T. (1999). Communication skills in pediatric cochlear implant recipients. *Acta oto-laryngologica, 119,* 219-224.

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

Miyamoto, R.T., Robbins, A.M., Myres, W.A., Pope, M.L., & Punch, J.L. (1986). Long-term intracochlear implantation in man. Otolaryngology Head and Neck Surgery, 95, 63-70.

Nusbaum, H.C., Pisoni, D.B, & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 200,000 words. *Research on Speech Perception Progress Report No. 10 Indiana University*.

Pennington B. (1997). Dimensions of executive functions in normal and abnormal development. In N.A. Krasnegor, G.R. Lyon, & P.S. Goldman-Rakic (Eds.), *Development of the prefrontal cortex* (pp. 265-281). Baltimore: Brookes.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, 20(3), 249-262.

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. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24,* 106-120.

Pisoni, D.B., Conway, C. M., Kronenberger, W., Henning, S., & Anaya, E.M. (2009). Executive function and cognitive control in deaf children with cochlear implants. In *Oxford Handbook of Deaf Studies, Language, and Education*. Oxford University Press.

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, 109,* 92-93.

Querleu, D., Renard, X., & Crepin, G. (1981). Auditory perception and fetal reaction to sound stimulation [author's translation]. *Journal de gynecologie, obstetrique et biologie de la reproduction, 10,* 307-314.

Robbins, A.M., Osberger, M.J., Miyamoto, R.T., Kienle, M.L., & Myres, W.A. (1985). Speech-tracking performance in single-channel cochlear implant subjects. *Journal of Speech and Hearing Research, 28,* 565-578.

Rueda, M.R., Rothbart, M.K., McCandliss, B.D., Saccomanno, L., & Posner, M.I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences, 102,* 14931-14936.

Ruth, R.M., Isquith, P., & Gioia, G., (1996). *Behavioral Rating Inventory of Executive Function Professional Manual*. Psychological Assessment Resources Inc.

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

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.

Sohmer, H., & Freeman, S. (1995) Functional development of auditory sensitivity in the fetus and neonate. *Journal of basic and clinical physiology and pharmacology, 6,* 95-108.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Stuss, D.T., & Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95,* 3–28.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A new online resource for online psycholinguistic studies. *Journal of Memory and Language, 51*, 247.

Visu-Petra, L., Benga, O., & Miclea, M. (2007). General developmental trends in EF. *Cognition, Brain, Behavior, 11,* 585-608.

Vygotsky, L.S. (1962). *Thought and Language* (E. Hanfmann & G. Vakar, Trans.). The M.I.T. Press, (Original work published 1934).

Wechsler, D. (1991). *Wechsler intelligence scale for children- Third edition.* The Psychological Corporation: San Antonio, TX.

Wolff, A., & Thatcher, R.W. (1990). Cortical reorganization in deaf children. *Journal of clinical and experimental neuropsychology, 12*(2), 209-221.

Yntema, D.B., & Trask, F.P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, *2,* 65-74.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Perception of Temporal Asynchrony in Audiovisual Phonological Fusion[1]

Melissa Troyer[2], Jeremy Loebach[3] and David B Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] The research described in this report was supported by NIH-NIDCD Research Grant R01DC00111 to Indiana University. The authors would like to specially thank Luis Hernandez, Darla Sallee, Thomas Busey, Dale Sengelaub, Althea Bauernschmidt, and all members of the Speech Research Lab for their helpful comments and support of this project.

[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

[3] Department of Psychology, Macalester College.

Perception of Temporal Asynchrony in Audiovisual Phonological Fusion

Abstract. Audiovisual phonological fusion (AVPF) is a recently discovered perceptual phenomenon in speech perception in which visual information (i.e., *back*) combines with auditory information (i.e., *lack*) to create a fused percept (i.e., *black*) (Radicke, 2007). The current study investigated the effects of temporal asynchrony on perception of AVPF. Subjects were presented with stimuli that differed in the amount of temporal offset ranging from 300 ms of auditory lead to 500 ms of visual lead and were asked to perform two tasks. In the fusion task, subjects were asked to report what they thought the speaker said. In the asynchrony judgment task, subjects were asked to determine whether the auditory and visual portions occurred at the same time ("in sync") or at different times ("out of sync."). The stimuli presented in both tasks were the same, but the ordering of the tasks was manipulated to test whether completing either task first would affect performance on the other. We found that (1) AVPF was moderately robust to temporal asynchrony; (2) synchrony judgments were robust for AVPF stimuli; and (3) the ordering of the tasks can modulate performance, at least for the asynchrony judgment task. Implications for current theories of audiovisual integration are discussed.

Introduction

Integration (or fusion) of complex stimuli is a fundamental part of perception (see Calvert, Spence, & Stein, 2004). All organisms exist in complex environments, and the coherent perception of events in the world depends on the interaction between the sensory information we are exposed to and our prior knowledge about how the world is structured. Perception involves comprehending the sensory information we are presented with. Speech perception is no exception. We are constantly integrating information from multiple modalities in order to understand spoken words. Although auditory information is weighted much more heavily for normal-hearing listeners than visual information in speech perception, the addition of visual information can enhance auditory information and contribute to higher intelligibility of speech. In a seminal study of multimodal speech perception, Sumby and Pollack (1954) demonstrated that visual information enhances the perception of speech presented in noise. In their experiment, subjects were seated around a table and asked to transcribe the speech of a speaker who was either facing them (i.e., the speaker's face was visible to the subjects) or was facing away from them (i.e., the subjects received no visual information from the speaker). Although the speech was presented in real-time, it was embedded in background noise and presented through a set of headphones at one of seven signal-to-noise ratios (SNRs). Speech presented without noise was consistently perceived accurately; however, subjects who received the visual information in addition to the auditory information performed better than subjects who only received auditory information across all SNRs. In particular, visual information improved response accuracies more for the less favorable SNRs where the noise was louder than the speech. Sumby and Pollack's study was the first of many to show that modalities other than audition can improve perception of speech.

Neural correlates of audiovisual speech perception have also indicated that the integration of visual cues is important for speech perception. In a recent electroencephalography study, Van Wassenhove and colleagues (2007b) found that for congruous audiovisual speech stimuli, the presence of a visual cue facilitated auditory processing, particularly when the cue was very salient, as for the bilabial stop /p/. That is, when visual as well as auditory cues were present, the peak-latency and amplitude of the N1 and P2 components, which are two ERP components sensitive to auditory stimuli, were significantly lower than for auditory speech alone. Imaging studies have also demonstrated that activation of visual and auditory cortices to audiovisual stimuli exceeds the response to stimuli presented in a single modality alone (Calvert, Brammer, Bullmore, Campbell, Iversen, and David, 1999). Further, imaging studies have shown that cortical areas known to be involved in auditory processing are activated during a visual lip-reading task (Sams, Aulanko, Hamalainene, Hari, Rounasmaa, Lu, and Simola, 1991; Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen, and David, 1997). In addition, fMRI studies have shown that distinct neural areas for cross-modal speech perception exist and differ from general sensory integration areas (Miller & D'Esposito, 2005). These neurophysiological studies indicate that the integration of visual and auditory information is important for the processing of speech.

In addition to improving speech perception in noise, visual cues can also alter the perception of spoken words. Previous research has demonstrated that when asked to identify syllables that were presented with conflicting auditory and visual information, subjects often responded with sounds containing a combination of the information from the two modalities (McGurk & MacDonald, 1976; MacDonald & McGurk, 1978). For instance, when subjects were presented with a visual velar consonant-initial syllable (e.g., /ga/) and an auditory bilabial-initial syllable (e.g., /ba/), subjects reported hearing an alveolar consonant that was in between the two (e.g., /da/) 64% of the time (MacDonald & McGurk, 1978). In this example, subjects reported hearing a stop consonant that has a place of articulation in between that of the velar stop /ga/ and the bilabial stop /ba/. Another way of understanding the observed response pattern is to consider that a syllable like /ba/ has a salient visual cue; however, a syllable beginning with a velar (such as /ga/) has no such salient visual cue because it is produced farther back in the mouth. If individuals hear /ba/ paired with a conflicting visual stimulus like /ga/, their perception of a bilabial /b/ may be reduced since they do not see the salient visual gesture of a bilabial, and the most proximal sound (by place of articulation), /d/, is heard instead.

This type of fusion, known as *psychoacoustic fusion*, has also been found to occur in the auditory domain when conflicting synthetic speech sounds (such as /b/ and /g/) are presented to each ear but only one fused sound (such as /d/) is perceived (Cutting, 1976). Psychoacoustic fusion differs from *phonological fusion*, another type of unimodal fusion first reported in the auditory domain (Cutting, 1975; 1976). In psychoacoustic fusion, two distinct events lead to an integrated percept of one event (e.g., /b/ and /g/ fuse to form the percept /d/, which is not completely specified by either of the input signals). In phonological fusion, however, two distinct events lead to a combined percept which integrates both events in a different manner; instead of two speech sounds fusing to one speech sound, the two speech sounds are maintained and both are perceived in a constrained temporal order.

Cutting (1975; 1976) studied auditory phonological fusion using a dichotic listening task in which subjects were presented with a different speech sound in each ear. For instance, when presented with *back* in one ear and *lack* in the other, subjects reported perceiving the fused word *black*. Cutting and Day (1975) found that the fusion of information across the two ears was consistent with the phonotactics of English. For instance, when presented with the word *pay* in one ear and *lay* in another, subjects report hearing *play* but never perceive the ungrammatical consonant cluster, such as /*lp/, that would result in the response *lpay*. Overall, when stop consonants (like /p/) were presented in one ear and liquids (like /l/) in the other, subjects fused the two into a percept of a stop-liquid cluster 57% percent of the time (Cutting & Day, 1975).

More recently, phonological fusion has been demonstrated to occur in response to multimodal speech stimuli. Radicke (2007) discovered a new form of audiovisual fusion, Audiovisual Phonological Fusion (AVPF), which occurs when subjects are presented with a bilabial visual stimulus (e.g., *back*) and an auditory liquid (e.g., *lack*) and report a percept that fuses the two into a permissible consonant cluster (e.g., *black*). Unlike unimodal phonological fusion, AVPF only occurs in response to stimuli containing visually presented bilabial stop consonants (/b/ and /p/) paired with auditory liquids (/l/ and /r/) (Radicke, 2007). In particular, fused responses included a voiced bilabial stop consonant /b/ much more often than a voiceless bilabial stop consonant /p/ (Levi, Radicke, Loebach, and Pisoni, 2008). Since there is no voicing information present in the visual bilabial, it is not surprising that individuals report hearing a fusion response containing a /b/ even when the visual stimulus contained the voiceless stop /p/. In addition, Radicke (2007) found that the proportion of fusions to visual stimuli which specified a bilabial and auditory stimuli which specified a liquid were much higher for the liquid /l/ (48% fused) than for the liquid /r/ (2% fused). Radicke suggested that individuals may be less likely to fuse the conflicting visual and auditory stimuli when the auditory stimulus is /r/ because visually, /r/ contains some lip rounding information (Radicke, 2007). Therefore, there is less conflict between the bilabial stop consonant (/b/ or /p/) and the auditory liquid, which contains bilabial visual cues.

Research therefore suggests that a certain amount of conflict between auditory and visual stimuli may result in multi-sensory integration leading to a perceptual fusion. In the McGurk effect, the information specified by the auditory and visual streams is incompatible, and subjects resolve the conflict by integrating the two streams into a combined percept that is in between (McGurk & MacDonald, 1978). In AVPF, the visual information from the /b/ or /p/ adds more information than was specified in the auditory stimulus (/l/ or /r/ alone), and conflict is resolved by integrating the information into a percept that contains information from both streams (Radicke 2007). One possible reason that individuals resolve conflict by perceiving one coherent event is that they may be strongly biased to believe that different environmental stimuli are likely to result from one single, unified event; this has been referred to as the *unity assumption* (Welch & Warren, 1980).

Several studies have endeavored to experimentally test the unity assumption. In a temporal order judgment task, Vatakis and Spence (2007) presented subjects with audiovisual speech stimuli in which the gender of the speaker of the auditory token was either the same as or different from the gender of the speaker of the visual token. In addition, the auditory and visual components of the speech stimuli differed in timing up 300 ms of visual or auditory lead. Subjects were told to report whether they thought the visual stream or the audio stream occurred first. When the gender of the video and audio portions of the stimulus differed, subjects were more accurate at reporting whether the visual or auditory signal came first. These findings suggest that when the unity assumption is already violated (i.e., when the gender of the auditory and visual streams differs), subjects are much better at detecting other differences in a multimodal stimulus display. Furthermore, the McGurk effect is reduced when a familiar face is paired with an unfamiliar voice (Walker, Bruce, and O'Malley, 1995). In this instance, there may also be a violation of the unity assumption; since the listener knows that the auditory and visual stimuli were not produced by the same person, she may be less likely to fuse the conflicting stimuli into one coherent event.

The McGurk effect has also been demonstrated when there is yet another element of conflict in addition to the incongruent information from the auditory and visual modalities. When subjects were presented with McGurk stimuli (such as visual /ga/ and auditory /ba/) in which the gender of the speaker of the auditory token differed from the gender of the visual token, subjects still fused the multiple conflicts and reported hearing the fused response, *da* (Green et al., 1991). In addition, the McGurk effect is robust to some temporal asynchrony, generally around the same levels at which subjects begin to judge congruous audiovisual items as asynchronous (Massaro & Cohen, 1993; Munhall, Gribble, Sacco, and Ward, 1996). These findings may be explained under the unity assumption, suggesting that individuals try to integrate conflicting stimuli (at least to some extent) in order to perceive a single coherent event (Vatakis & Spence, 2007).

There is an important distinction between studies of the McGurk effect (i.e., psychoacoustic fusion) and phonological fusion, which was the focus of the current study: they differ in the extent to which the fusion is constrained by timing. In psychoacoustic fusion, two different events occur simultaneously and are perceptually combined into a new simultaneous event. In phonological fusion, however, the two different events occur at the same time but must be perceived as occurring at slightly different times in order for the percept of two distinct sounds to occur. Studies of audiovisual speech in which the auditory and visual portions only differ in *timing* (rather than conflicting in quality, as in the McGurk effect) have revealed that a slight asynchrony in the alignment of auditory and visual information does not adversely affect perception and is not detected until up to a certain threshold (Conrey & Pisoni, 2006; Dixon & Spitz, 1980).

Conrey and Pisoni (2006) demonstrated that individuals judge items as being synchronous at up to 300 milliseconds of audiovisual asynchrony. However, several variables were shown to contribute to the exact amount of asynchrony that could be tolerated. In particular, certain words were shown to demonstrate less tolerance to audiovisual asynchrony than other words. For example, the word *back* was robust to only 100 ms of auditory lead to 200 ms of visual lead, but *theme* was robust to 167 ms of visual lead to 300 ms of visual lead. In addition, a general tendency was observed for greater tolerance to asynchrony in visual-lead videos than in audio-lead videos. A priori, this finding is not surprising because while it would be strange to perceive a sound before seeing any motion in the speaker's face, we often see a person's face moving before we perceive the sound they produce. Individuals are therefore able to tolerate a certain amount of audiovisual asynchrony in the presentation of auditory and visual events, integrating the two streams into a single synchronous event. However, in phonological fusion, the opposite must also be true. Subjects are able to perceive two synchronous events as occurring at slightly disparate times. Discovering how much temporal asynchrony can be tolerated in the perception of audiovisual phonological fusion is therefore an important question in understanding multimodal speech perception.

The present study sought to investigate the relationship between audiovisual asynchrony and audiovisual conflict in items that have been shown to result in audiovisual phonological fusion. Subjects were presented with audiovisual stimuli that could conflict both temporally (i.e., stimuli began with different levels of visual or auditory lead) and qualitatively (i.e., the auditory stimulus contained a different speech segment from the visual stimulus). The experimental stimuli used were modified versions of a subset of the materials developed by Radicke (2007) and consisted of a visual bilabial onset (/b/ or /p/) and an auditory liquid onset (/l/ or /r/). The experiment consisted of two tasks. In the asynchrony detection task, subjects were asked to determine whether the auditory and visual portions of the stimulus were presented at the same time (synchronous) or at different times (asynchronous). In the perceptual fusion task, subjects were asked to identify the word or syllables that they perceived by typing their response on a computer keyboard.

One goal of the study was to determine whether AVPF is robust (i.e., perceptually invariant) even when the auditory and visual stimuli portions of a stimulus do not occur at the same time. Another goal was to determine whether subjects showed similar performance in synchrony judgments to experimental AVPF stimuli as to control stimuli (where the visual and auditory tokens were the same word). It was expected that subjects would demonstrate greater tolerance for visual lead than for auditory lead in the perceptual task. Not only is there more tolerance for visual lead in asynchrony tasks (Conrey & Pisoni, 2006), but the phonotactics of English also predict that for the stimuli used in the current experiment, the visual component would always precede the auditory component. For instance, given the visual stimulus *back* and the auditory stimulus *lack*, the fused target is *black* and not **lback*. The onset cluster /lb/ does not occur in English and is therefore highly unlikely (or perhaps nearly impossible) for an English listener to perceive with the given stimuli. The general tolerance for visual lead combined with prior information about the phonotactics of English lead to the specific hypothesis that there would be more visual lead tolerance to stimuli in both the asynchrony detection and perceptual tasks.

The presentation order of these two tasks has also varied between subjects, and it was predicted that if subjects were aware that stimuli might be presented at varying levels of asynchrony (i.e., if the asynchrony detection task was presented first), they might also be less likely to fuse the stimuli. If this reduction of perceptual fusion occurred, it could be argued that higher-order cognitive (or "top-down") mechanisms might be interfering with the otherwise robust fusion of conflicting audiovisual stimuli. If, however, the presentation order of the two tasks did not affect performance (i.e., fusion rates) on the perceptual task, it is more likely that fast, stimulus-driven ("bottom-up") mechanisms may account for the phonological fusion of conflicting audiovisual stimuli. In addition, it was hypothesized that task ordering might also have an effect on the asynchrony detection task. Specifically, if subjects performed the perceptual fusion task first, it was expected that they might be more likely to report items as synchronous in the asynchrony detection task due to the processes used for integration in the perceptual fusion task.

Finally, it was expected that the two different tasks (speech perception versus asynchrony detection) might occur at different thresholds of asynchrony and that this difference could also be an indicator of the level of processing at which the two different types of judgments occur. In the asynchrony detection task, subjects were informed that the stimuli they saw might be conflicting; however, in the perceptual fusion task, subjects were not informed that the auditory and visual constituents of the stimuli might not result from the same unified event.

It is possible that the threshold for asynchrony detection might be lower than the threshold for the occurrence of phonological fusion in the perception task. This would be the predicted result if phonological fusion resulted more from prior perceptual biases and cognitive factors (such as the unity assumption) and less from stimulus-driven, lower-level perceptual factors. If there is no difference between the two thresholds, this might be an indicator that being explicitly aware of the possibility of discord between stimuli does not modulate phonological fusion and that phonological fusion is a robust effect which is not influenced by cognitive factors such as prior experience and perceptual bias based on the phonotactics of English.

Methods

Participants

Participants consisted of 46 undergraduate students at Indiana University between the ages of 18 and 23 (*M* = 19.57, *SD* = 1.20). All participants were native speakers of American English and reported no history of speech or hearing disorders at the time of testing. Subjects were compensated with partial course credit in an introductory psychology course. Due to computer error, the data for 10 subjects was not recorded, and their data will not be reported here.

Stimuli

The stimulus materials used in this experiment were short video clips of a male speaker with a neutral American dialect. All original video clips were taken from the database created by Radicke (2007) and were modified by altering the relative onset of the audio and video tracks using Final Cut Pro software as described below.

The control stimuli consisted of video recordings of six monosyllabic English words (*back*, *lack*, *rack*, *pay*, *lay*, and *ray*). For the control stimuli, the video and audio tracks were taken from the same token of the word and modified by altering the relative onset of the audio and video tracks in the following manner. For all stimuli, the auditory track was held constant and the video track was moved forwards or backwards in time to create stimuli which differed in timing at 25 different steps (300 ms audio lead to 500 ms visual lead, including stimuli in which the auditory and visual portions actually matched)[4]. This was done in order to maintain the same overall length of videos within a stimulus set so that subjects would not be biased to respond based on differences in the length of the stimulus. Since the rate of sampling was higher for the auditory stimuli than for the visual stimuli, the auditory track was held constant so that some ambient noise could be heard throughout the "still" audio frames in order to reach the appropriate overall stimulus length. To achieve this, the first frame of the audio track was copied and pasted for 15 frames for a total of 500 ms before the "live" audio track began. The visual track was then moved to the appropriate level of asynchrony (up to 15 frames before or 9 frames after the onset of the live auditory track, where each frame is 33 ms). In the cases in which the audio portion led the video portion, the first frame of the video track was copied for the appropriate number of frames backwards so that the audio and video tracks appeared to begin at the same time, though the "live" tracks still differed in onset. The final frames of these stimuli were made to end at the same time by copying and pasting the final frame of the audio track for the appropriate number of frames. These procedures ensured that there was always information in the auditory and visual signals. After such manipulation, there were a total of six words at 25 different asynchrony intervals each (15 visual-lead intervals + 9 auditory-lead intervals + 1 interval without asynchrony), making a combined total of 150 control stimuli. All manipulations to the videos were done using Final Cut Pro. Mention the Software used for editing videos. I imagine you used final cut pro.

The experimental stimuli consisted of the same six monosyllabic English words *back*, *lack*, *rack*, *pay*, *lay*, and *ray* that were combined so that the auditory track always consisted of a liquid-initial word (beginning with *l*- or *r*-) and the visual track always consisted of a bilabial stop-initial word (beginning with *b*- or *p*-). The words were matched so that the audio and video tracks differed only in the first consonant of the word, but not in its rhyme: back-lack, back-rack, pay-lay, and pay-ray. Other than the difference in initial consonant, the experimental stimuli were constructed in the same way as the control stimuli at the same 25 intervals of audiovisual asynchrony (of 33 ms steps between 300 ms of auditory lead and 500 ms of visual lead). There were a total of four pairs at each of 25 different asynchrony levels, leading to 100 experimental stimuli. Added to the 150 tokens of control stimuli, a total of 250 videos were used in the experiment. The same stimuli were used in both experimental tasks, which are outlined below.

Apparatus

Subjects were seated at Macintosh computers with 17" CRT monitors in an experimental room containing multiple testing stations. Auditory stimuli were presented over Beyer Dynamic DT-100 headphones. Subjects were tested in groups of one to four participants at a time. The volume of the videos was held constant at approximately 65 dBV SPL for all participants. PsyScript was used for experiment presentation.

Procedure

Subjects participated in two experimental tasks: a perceptual fusion task and a forced-choice audiovisual asynchrony detection task. The ordering of the two tasks was counterbalanced over subjects. Half of the participants completed the perceptual task first while the other half of the participants completed the asynchrony detection task first. Subjects were given instructions both verbally and via text presented on the computer screen for the each task they were to perform.

[4] These asynchrony levels were chosen based on previous work by Conrey & Pisoni (2006).

In the perceptual fusion task, subjects were told that they would be asked to watch videos of a speaker and to type what they thought the speaker said. Subjects were told that it was important to both watch and listen to the videos carefully and that they should take special care not to make typographical errors when entering their responses. The stimuli (both experimental and control videos) were presented in a random order which differed for each subject. Immediately following the presentation of a video, a dialogue box appeared on the screen in order for the subject to record his or her response.

In the audiovisual asynchrony detection task, subjects were told that they would be watching videos of a speaker saying English words and that they would be asked to respond whether the audio and video tracks of the word were "in sync" (i.e., synchronous, meaning that the audio and video events happened at the same time) or "out of sync" (i.e., asynchronous, meaning that the audio and video events occurred at different times). The subjects were told to respond by pressing the left button (labeled "in sync") on a response box if they thought the audio and video tracks were synchronous and by pressing the right button on a response box (labeled "out of sync") if they thought the audio and video tracks were asynchronous. Subjects were instructed to respond as quickly and as accurately as possible. Subjects were also told that they might find this part of the experiment difficult but that it was very important to pay attention and monitor the timing of the audio and video tracks. Subject responses were recorded at any point beginning at the onset of the videos so that subjects could respond as soon as they wished.

After subjects had completed both tasks, they were debriefed. The experimenter also asked the subjects whether they had any questions about the experiment and offered to explain the purpose of the experiment if they wished to discuss it any further.

Scoring

For the perceptual fusion task, open-set responses were collected and scored semi-automatically in Microsoft Excel. Responses were hand-checked and corrected for misspellings. Responses for experimental stimuli were scored as falling in one of the following categories: visual stimulus, auditory stimulus, fused (visual + auditory) target, or other. For the purposes of this report, however, subject responses for the perceptual fusion task were treated either as being a fused response or not. A fused response contained a bilabial stop consonant in all cases followed by the appropriate liquid consonant specified by the auditory stimulus (/l/ or /r/). The bilabial stop consonant in the response did not have to match the visual stop consonant in voicing[5].

For the asynchrony detection task, closed-set responses (a button press indicating that stimuli were "in sync" or "out of sync") were collected. These responses were assigned a score of 1 (for "in sync") or 0 (for "out of sync"). In addition, reaction times in milliseconds from onset of the stimulus were collected.

Data analysis

For the perceptual fusion task, the proportion of responses fused was analyzed as a function of asynchrony level and subject group (perceptual fusion task first vs. asynchrony detection task first). Differences between the proportions fused of all experimental stimulus sets were also analyzed for each asynchrony level. In addition, individual differences in overall proportion fused across subjects were examined.

[5] Generally, all fused responses in AVPF have been found to begin with a voiced consonant, regardless of the voicing of the visual stop consonant (see Radicke, 2007). Individuals are unable to distinguish visual /p/ and visual /b/, and it is thought that these two phonemes, in addition to the phoneme /m/, constitute a "viseme" class (see Lidestam & Beskow (2006) and Berger (1972) for a description of such viseme classes, or classes of indistinguishable visual phonemes).

For the asynchrony detection task, the proportion reported "in sync" was also analyzed as a function of asynchrony level and subject group. Differences between the proportion reported as "in sync" for all stimulus sets were also analyzed for each asynchrony level, and individual differences in overall proportion responded "in sync" across subjects were also examined. Differences in reaction time for "in sync" responses were also analyzed for each asynchrony level; however, these data are not reported here.

Results

Perceptual Fusion Task

A wide range of variability was observed across experimental stimulus sets (*back* + *lack*, *back* + *rack*, *pay* + *lay*, *pay* + *ray*) in the perceptual fusion task. The overall fusion rate across subjects and stimulus sets was .050 (*SD* = .219). The overall proportion fused for each stimulus set is shown in Figure 1. A one-way ANOVA with stimulus set as a factor was conducted on the data. The ANOVA revealed an overall effect of stimulus set on fusion rate *F*(3, 3596) = 80.223, *p* < .001. Post-hoc Bonferroni comparisons revealed greater fusion for the *back* + *lack* (*M* = .134, *SD* = .341) stimulus set than for any other set (for *back* + *rack*, *M* = .002, *SD* = .047; for *pay* + *lay*, *M* = .063, *SD* = .244; for *pay* + *ray*, *M* = .001, *SD* = .033). For all sets, *back* + *lack* yielded more fusions (*p* < .001 for each comparison). In addition, a higher fusion rate was obtained for the *pay* + *lay* set than for either *back* + *rack* or *pay* + *ray* (for both comparisons, *p* = .01). There was no difference between the *pay* + *ray* and *back* + *rack* stimulus sets, which very rarely fused. Participants thus reported the most fusion for the two sets which contained an /l/-initial auditory signal.

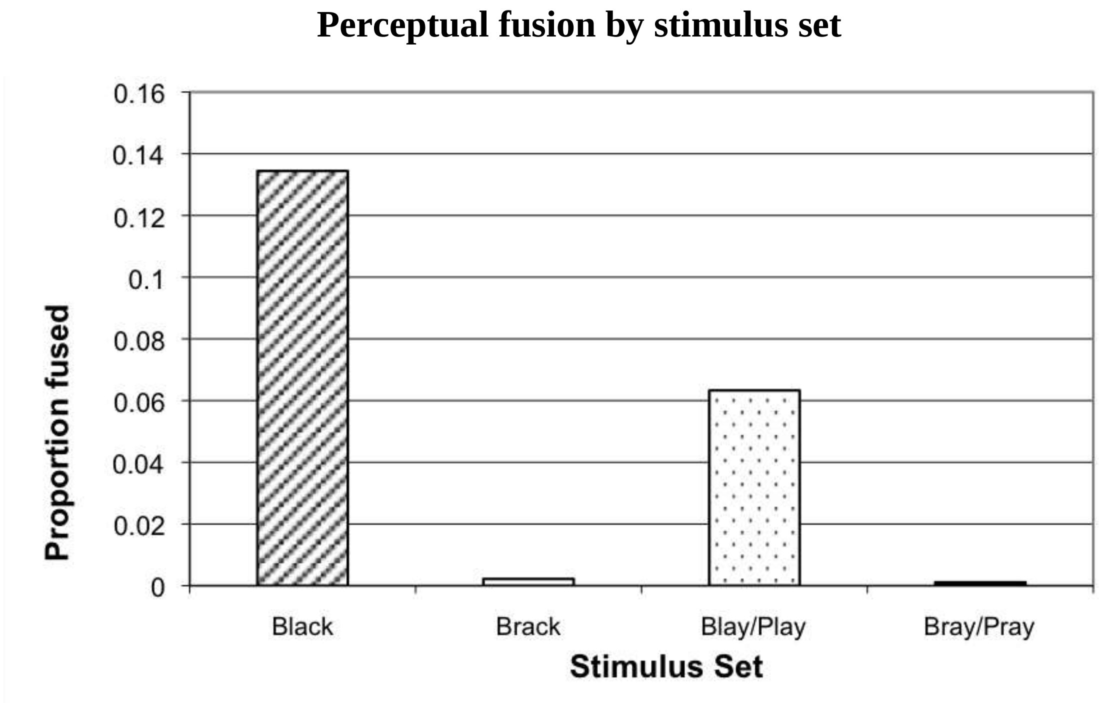

Figure 1. Proportion of fused items varies depending on the stimulus set. Participants reported the most fusion for the two sets which contained an /l/-initial auditory signal (*p* < .001). In particular, auditory *back* and visual *lack* were fused the most of any of the four sets (*p* < .001).

Fusion rates tended to be highest at shorter asynchrony levels than at longer asynchrony levels. In addition, fusion rates were higher for visual-lead stimuli than for auditory-lead stimuli at matched levels of asynchrony. The rate of fusion for each asynchrony level is shown in Figure 2. The response distribution was skewed towards the visual lead for proportion fused. The maximum proportion of fusions observed at any level of asynchrony was .12 at 100 ms visual lead.

One-sample t-tests comparing the fusion rate at each level of asynchrony against zero were performed. A summary of the results for each asynchrony level is given in Table 1. Overall, fusion occurred most often at shorter asynchronies. A one-way ANOVA with asynchrony level as a factor was also conducted on the data. A significant main effect of asynchrony level was found (*F*(24, 3575 = 2.537, *p* < .001). Using a threshold of 50% of the maximum proportion of fusion, we found that the boundaries for robust fusion of AVPF stimuli were located at 66 ms visual lead and 200 ms visual lead. An independent samples t-test was conducted to compare fusion of stimuli which fell between these boundaries to fusion of stimuli which fell outside the boundaries. Fusion of stimuli which fell between the boundaries of 66 ms auditory lead and 200 ms visual lead was significantly greater than fusion of stimuli which fell on either side of this boundary (*t*(3598) = 6.20, *p* < .001). An independent samples t-test was then conducted to test whether visual-lead stimuli were fused more often than auditory-lead stimuli. The t-test revealed no significant difference in fusion for visual-lead versus auditory-lead stimuli (*t*(3598) = .374, *p* = .709). Interestingly, the boundary for fusion for visual lead (at 200 ms) was nearly three times as long as the auditory lead boundary (at 66 ms). In addition, the peak of fusion occurred at a visual lead asynchrony of 100 ms rather than occurring closer to 0 ms of asynchrony.

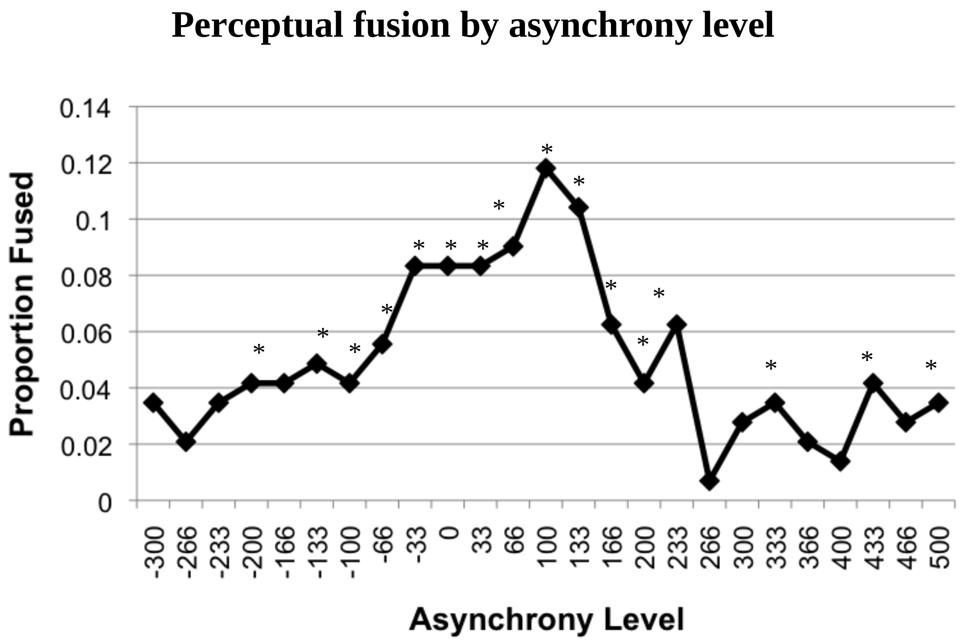

Figure 2. The proportion of experimental stimuli fused varies by asynchrony level. There was significant fusion at 233 ms auditory lead, from 166 ms of auditory lead to 233 ms visual lead, at 333 ms of visual, and from 433 to 466 ms of visual lead. Overall, fusion was more often significant at lower levels of asynchrony. An asterisk (*) above an asynchrony level indicates significant difference from 0 (*p* < .05).

Asynchrony Level	Mean	Standard Deviation	*t*-statistic	*p*-value
A300	0.035	0.184	3.924	**0.000**
A266	0.021	0.143	3.179	**0.003**
A233	0.035	0.184	2.907	**0.006**
A200	0.042	0.201	2.376	**0.023**
A166	0.042	0.201	2.497	**0.017**
A133	0.049	0.216	2.092	**0.044**
A100	0.042	0.201	1.972	0.057
A66	0.056	0.230	2.092	**0.044**
A33	0.083	0.277	1.784	0.083
0	0.083	0.277	1.673	0.103
V33	0.083	0.277	3.000	**0.005**
V66	0.090	0.288	3.494	**0.001**
V100	0.118	0.324	4.183	**0.000**
V133	0.104	0.307	3.993	**0.000**
V166	0.063	0.243	2.376	**0.023**
V200	0.042	0.201	2.376	**0.023**
V233	0.063	0.243	2.907	**0.006**
V266	0.007	0.083	1.000	0.324
V300	0.028	0.165	1.784	0.083
V333	0.035	0.184	2.092	**0.044**
V366	0.021	0.143	1.784	0.083
V400	0.014	0.117	1.435	0.160
V433	0.042	0.201	2.376	**0.023**
V466	0.028	0.165	2.092	**0.044**
V500	0.035	0.184	1.963	0.058

Table 1. Statistics for perceptual fusion by asynchrony level.

Asynchrony Detection Task

Performance in the asynchrony detection task also revealed wide variability in performance that was dependent on stimulus type. In particular, the stimulus sets which ended in the rhyme *-ack* were reported "in sync" more often than those which ended in the diphthong *-ay*. The proportion of synchrony judgments for each stimulus set is shown in Figure 3. A one-way ANOVA with stimulus type as a factor was conducted on the data. There was an overall effect of stimulus type on proportion reported "in sync" (*F*(3, 3596) = 70.789, *p* < .001). Post-hoc Bonferroni comparisons revealed that *back* + *lack* (*M* = .320, *SD* = .467) elicited significantly more "in sync" responses than the *pay* + *lay* (*M* = .134, *SD* = .341) and *pay* + *ray* (*M* = .141, *SD* = .348) stimulus sets, *p* < .001 for each comparison. In addition, *back* + *rack* (*M* = .354, *SD* = .479) elicited significantly more "in sync" responses than either of the *pay* + *lay* or *pay* + ray sets. There was no significant difference between the *pay* + *lay* and *pay* + *ray* sets. Overall, the results indicate that the stimulus sets which ended in a consonant (/k/) elicited more "in sync" responses than did stimulus sets which ended in a diphthong (/e^{i}/). Figure 4 shows the proportion of fused responses for each stimulus set across all asynchrony levels.

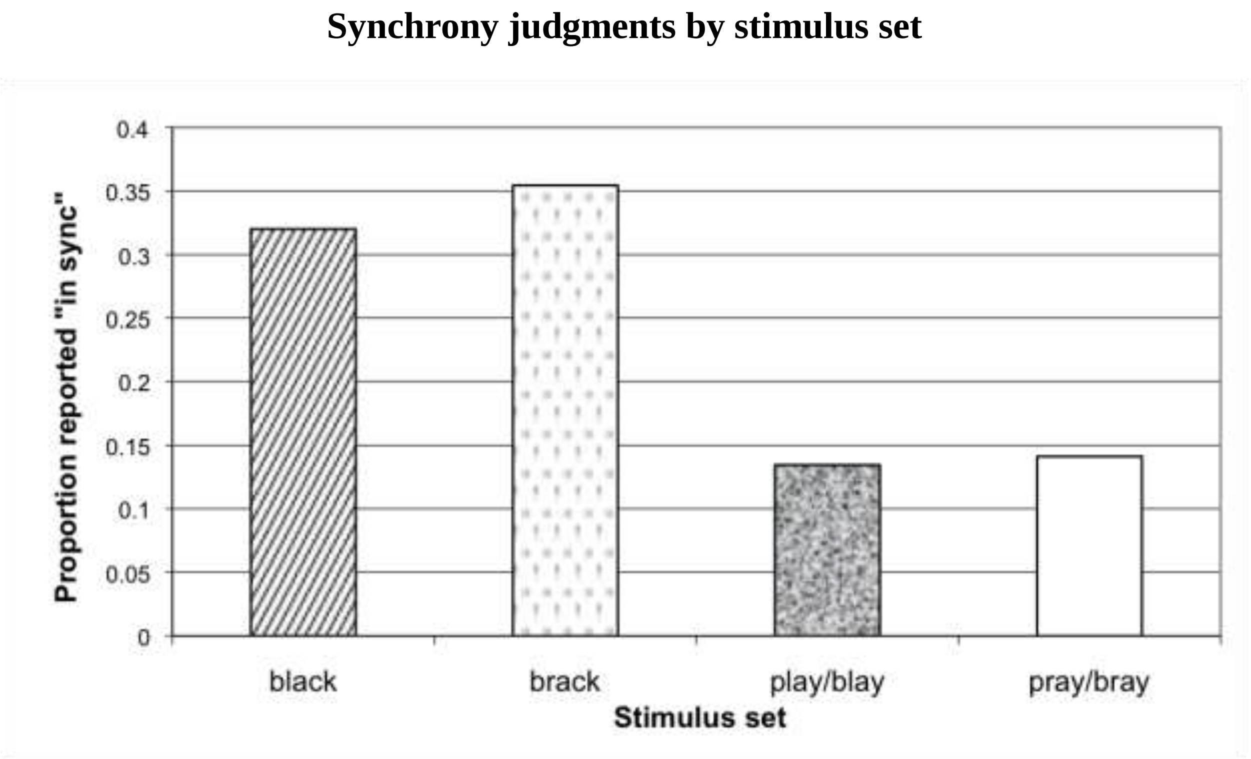

Figure 3. Proportion of synchrony judgments (i.e., "in sync" responses) are given above for each stimulus set. The two sets containing words ending in *-ack* were reported "in sync" significantly more often than the sets containing words ending in *-ay*.

Furthermore, subjects reported that the experimental stimulus items (like visual *back* and auditory *lack*) were "in sync" much less frequently than the control items (where the visual and auditory tokens were the same word). In addition, for both control and experimental items, the proportion of items reported "in sync" was always highest at smaller asynchrony levels and decreased at longer asynchrony levels.

The proportion of items reported "in sync" at each asynchrony level for both control and experimental stimuli is shown in Figure 5 and in Table 2. The mean proportion reported "in sync" was .601 (*SD* = .490) for control items and .252 (*SD* = .434) for experimental items. A univariate ANOVA with stimulus type (control vs. experimental) and asynchrony level as factors was carried out. The ANOVA revealed a significant main effect of stimulus type on synchrony judgments (*F*(24, 8950) = 1649.59, *p* < .001). Subjects reported the control stimuli (where the auditory token was the same word as the visual token) as being synchronous more often than the experimental stimuli across asynchrony levels. There was also a significant main effect of asynchrony level on synchrony judgments (*F*(1, 8950) = 1649.592, *p* < .001). Across stimulus types, subjects were more likely to respond "in sync" at lower asynchrony levels and less likely to respond "in sync" at greater asynchrony levels. Finally, a significant interaction was observed between asynchrony level and stimulus type (*F*(24, 8950) = 14.832, *p* < .001). The relative windows of synchrony judgments differed for control and experimental stimuli, and the threshold for stimuli being reported as "in sync" was smaller for experimental than for control stimuli.

Asynchrony detection by experimental item and asynchrony level

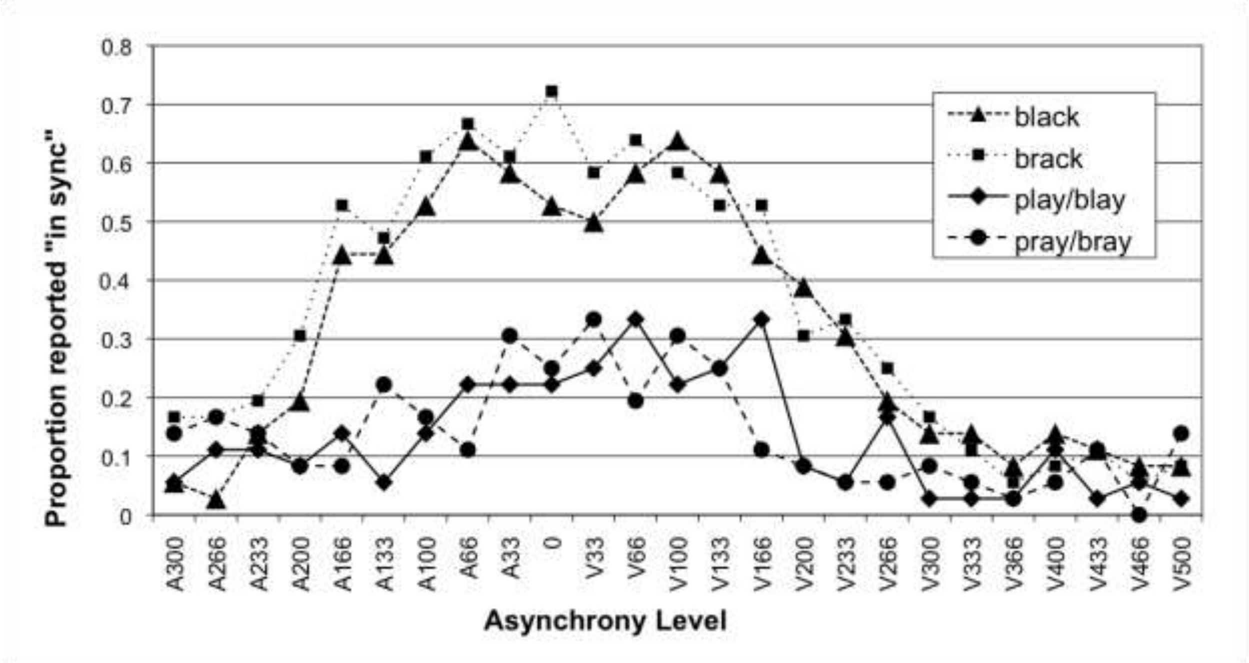

Figure 4. Asynchrony detection performance by experimental item is given for all asynchronies. The fused targets *black* and *brack* were fused more often than the targets *play/blay* and *pray/bray*.

Proportion of synchrony judgments by stimulus type

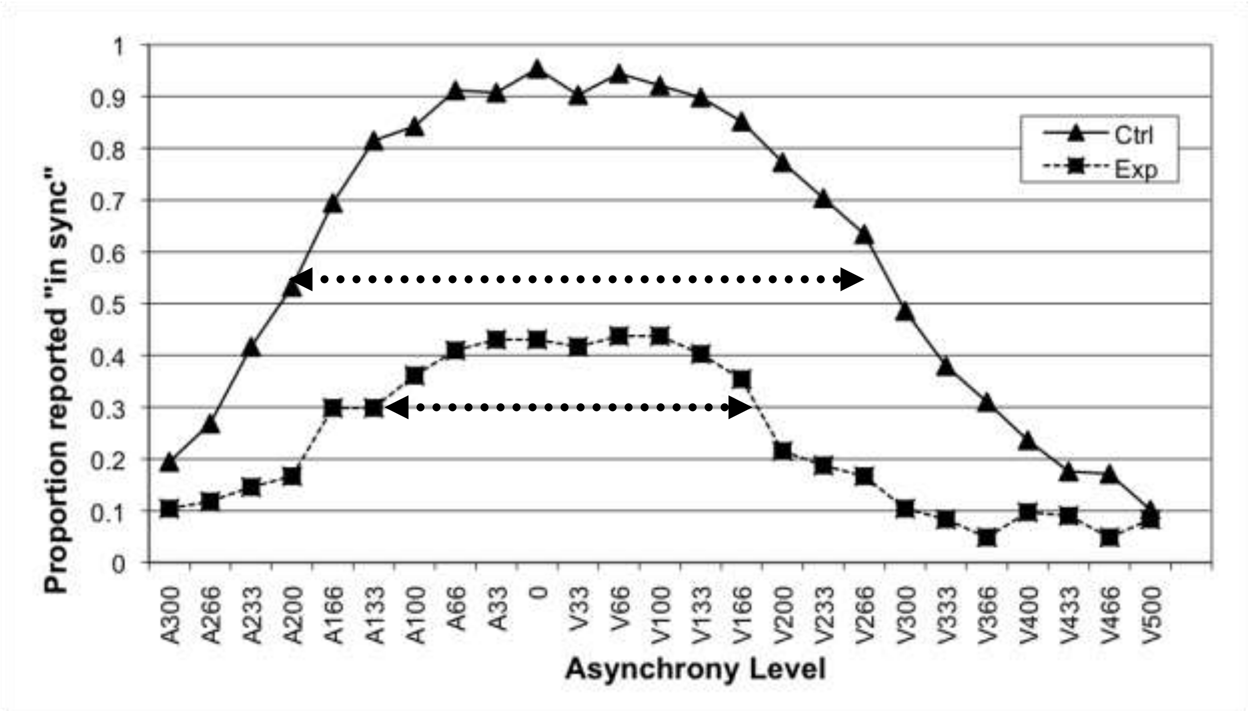

Figure 5. The proportion of synchrony judgments by stimuli type is given for all asynchrony levels. There is an effect of stimulus type, and experimental items are reported as being "in sync" less often than control items. In

addition, there is an effect of asynchrony level, and there are fewer synchrony judgments at larger levels of asynchrony.

Asynchrony Level	Control Mean (SD)	Experimental Mean (SD)
A300	0.19 (0.393)	0.104 (0.307)
A266	0.269 (0.444)	0.139 (0.347)
A233	0.412 (0.493)	0.153 (0.361)
A200	0.537 (0.5)	0.188 (0.392)
A166	0.704 (0.458)	0.319 (0.468)
A133	0.81 (0.393)	0.313 (0.465)
A100	0.838 (0.369)	0.389 (0.489)
A66	0.912 (0.284)	0.431 (0.497)
A33	0.907 (0.291)	0.451 (0.499)
0	0.954 (0.211)	0.458 (0.5)
V33	0.907 (0.291)	0.431 (0.497)
V66	0.944 (0.23)	0.451 (0.499)
V100	0.921 (0.27)	0.458 (0.5)
V133	0.898 (0.303)	0.431 (0.497)
V166	0.852 (0.356)	0.375 (0.486)
V200	0.778 (0.417)	0.236 (0.426)
V233	0.704 (0.458)	0.201 (0.402)
V266	0.634 (0.483)	0.188 (0.392)
V300	0.486 (0.501)	0.104 (0.307)
V333	0.37 (0.484)	0.097 (0.297)
V366	0.319 (0.467)	0.056 (0.23)
V400	0.241 (0.429)	0.097 (0.297)
V433	0.167 (0.374)	0.097 (0.297)
V466	0.167 (0.374)	0.049 (0.216)
V500	0.102 (0.303)	0.083 (0.277)

Table 2. Mean proportion of synchrony judgments for control and experimental stimuli.

For control stimuli, the maximum proportion of words reported “in sync” at a particular asynchrony level was .95 at 0 ms of asynchrony. For experimental stimuli, the maximum proportion reported “in sync” at a particular asynchrony level was .44 at both 66 ms and 100 ms visual lead. Using a threshold of 50% of the maximum for each of the stimulus types, the boundaries for reporting items as synchronous were from 200 ms auditory lead to 300 ms visual lead for control stimuli and from 166 ms auditory lead to 166 ms visual lead for experimental stimuli.

A univariate ANOVA with stimulus type and lead type (visual versus auditory) as factors was performed on the data. The ANOVA indicated that subjects were more likely to report items as synchronous if there was visual lead as opposed to an auditory lead (F(1, 8996) = 59.438, p < .001). A significant main effect of stimulus type was also observed (F(3, 8996) = 1170.419, p <.001), indicating that the proportion of items reported “in sync” was higher for control items than for experimental items. There was no significant interaction between stimulus type and lead type (F(1, 8996) = .637, p < .001). In addition, a univariate ANOVA with stimulus type and threshold matching (i.e., whether or not a particular asynchrony level met the threshold of proportion “in sync” responses as defined by 50% of the maximum reported “in sync”) as factors was conducted on the data. The ANOVA revealed that subjects reported all stimuli as synchronous more often when the stimuli fell between their established boundaries

(200 ms auditory lead/300 ms visual lead for control stimuli and 166 ms visual lead/166 ms visual lead for experimental stimuli) (*F*(1, 8996) = 2116.024, *p* < .001). There was also a significant main effect of stimulus type on proportion responded "in sync" (*F*(1, 8996) = 794.010, *p* < .001). Finally, a significant interaction between stimulus type and threshold matching was observed (*F*(1, 8996) = 220.921, *p* < .001). These findings reveal an advantage for visual lead for both experimental and control items and that the thresholds for synchrony judgments (as defined by 50% of the maximum proportion "in sync") are a reasonable estimate of the boundaries for robust synchrony (vs. asynchrony) judgments. Finally, the interaction between stimulus type and threshold matching further demonstrate that the boundaries of asynchrony detection for the two types of stimuli are not equivalent since there is a smaller boundary for asynchrony judgments.

Comparing the Two Tasks

Using the previously established thresholds (defined to be 50% of the maximum for both proportion fused and proportion reported "in sync"), a difference was observed in the size of the boundary windows for fusion and asynchrony detection for experimental stimuli. The window for asynchrony detection was larger than the window for fusion responses. The window for asynchrony detection included stimuli from 166 ms auditory lead to 166 ms visual lead whereas the window for fusion included stimuli from 66 ms auditory lead to 200 ms visual lead. The distributions of the responses for each task are plotted on multiple axes in Figure 6. Moreover, there was a weak but significant correlation between proportion of experimental stimuli fused in the perceptual fusion task and the proportion of experimental stimuli reported "in sync" in the asynchrony detection task (*r*(3600) = .125, *p* < .001). This result indicates that a subject's likelihood of fusing an item predicted his or her likelihood of responding that the item was synchronous.

Comparing fusion and asynchrony detection

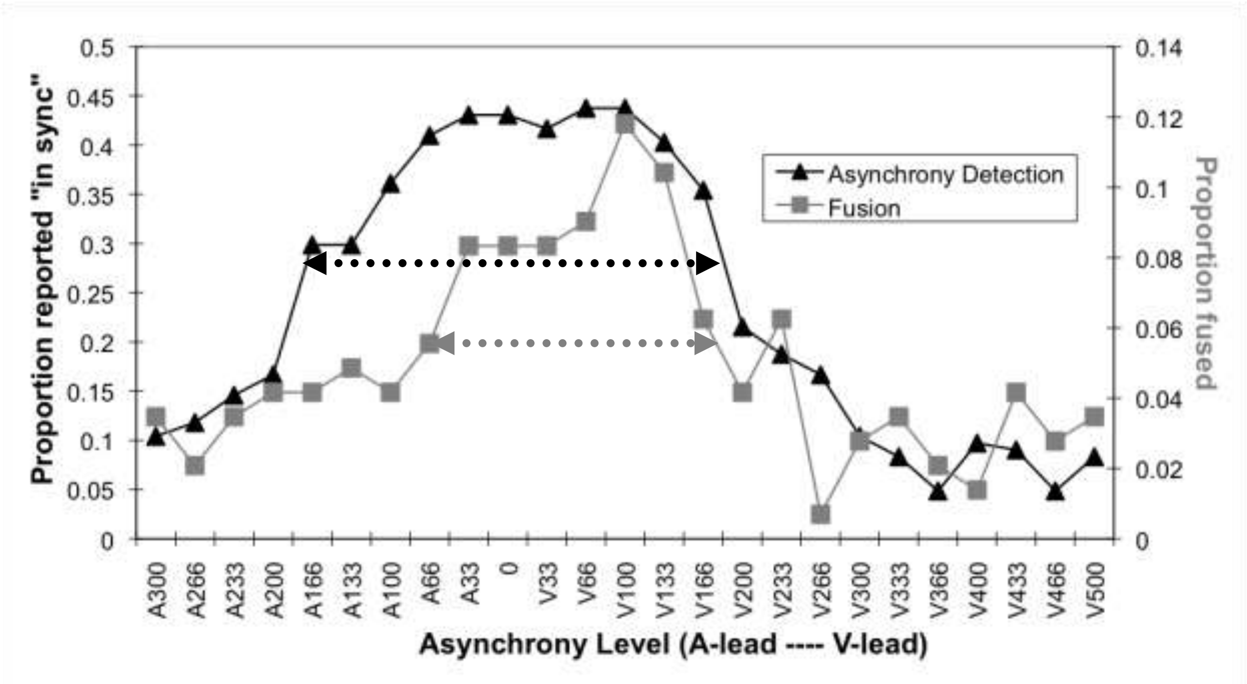

Figure 6. The proportion fused and proportion reported "in sync" for experimental stimuli are displayed. The

window where threshold is met for fusion (from 66 ms auditory lead to 200 ms visual lead) is smaller than the window for asynchrony detection (from 166 ms auditory lead to 200 ms visual lead).

Effects of Task Ordering on Perceptual Fusion

A non-significant trend was observed for the group who performed the fusion task first (FA) to perform differently on the perceptual fusion task from individuals in the group who performed the asynchrony detection task first (AF). The fusion rates for the FA group were slightly higher than those of the AF group at shorter asynchrony levels but somewhat higher at longer asynchrony levels, at least for the auditory lead stimuli.

Performance on the perceptual fusion task for each group is plotted in Figure 7, and the means for each group at each asynchrony level are given in Table 3. The overall fusion rate for each group was quite low for both the FA group (*M* = .050, *SD* = .218) and the AF group (*M* = .051, *SD* = .219) (*F*(1,34) = .034, *p* = .855). This result indicates that subjects in both groups showed similar performance, averaged across all asynchrony levels.

There was a significant main effect of asynchrony level on proportion fused, *F*(24, 816) = 3.121, *p* < .001. Subjects were more likely to give fused responses at shorter asynchrony levels and less likely to give fused responses at longer asynchrony levels. There was also a significant interaction between subject group and asynchrony level, (*F*(24, 816) = 1.591, *p* = .036), suggesting that the asynchrony level influences the effect of the subject group on fusion rates to some extent.

Perceptual fusion by group

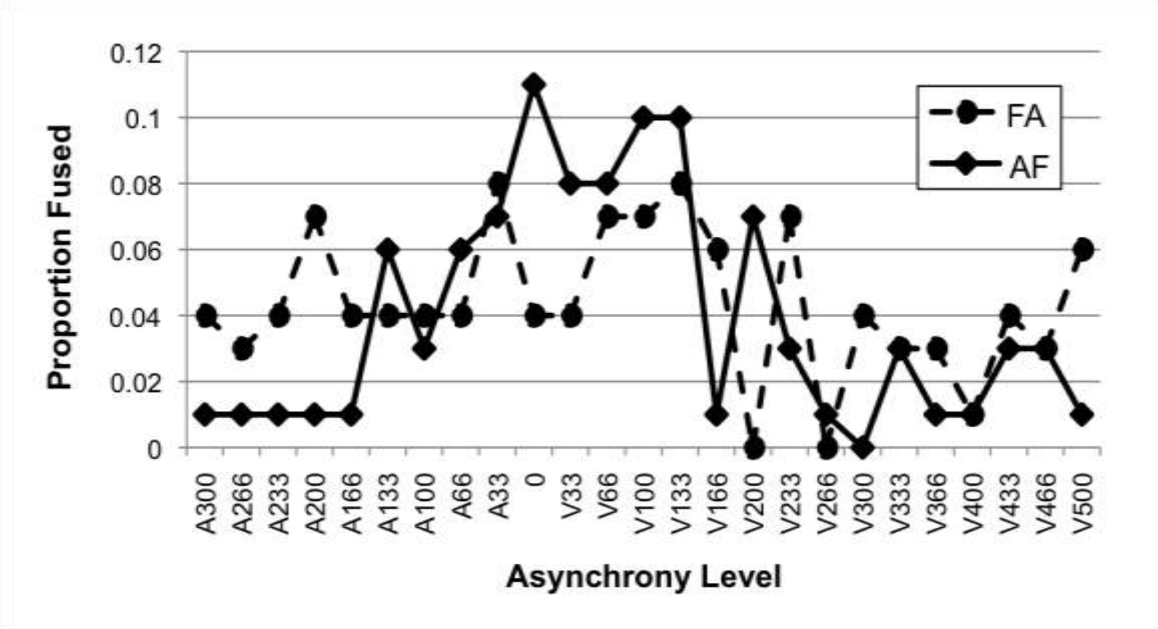

Figure 7. The proportion fused at each asynchrony level for each group is shown. Though there was no significant difference between groups, it can be seen that overall, the AF group was less likely to perceive fusion when the asynchrony level was higher and more likely to perceive fusion when the asynchrony level was lower.

Asynchrony Level	FA Mean (SD)	AF Mean (SD)
A300	0.04 (0.13)	0.01 (0.06)
A266	0.03 (0.08)	0.01 (0.06)
A233	0.04 (0.1)	0.01 (0.06)
A200	0.07 (0.17)	0.01 (0.06)
A166	0.04 (0.1)	0.01 (0.06)
A133	0.04 (0.13)	0.06 (0.11)
A100	0.04 (0.1)	0.03 (0.08)
A66	0.04 (0.1)	0.06 (0.11)
A33	0.08 (0.15)	0.07 (0.14)
0	0.04 (0.1)	0.11 (0.13)
V33	0.04 (0.13)	0.08 (0.12)
V66	0.07 (0.14)	0.08 (0.12)
V100	0.07 (0.12)	0.1 (0.13)
V133	0.08 (0.15)	0.1 (0.13)
V166	0.06 (0.11)	0.01 (0.06)
V200	0 (0)	0.07 (0.12)
V233	0.07 (0.12)	0.03 (0.08)
V266	0 (0)	0.01 (0.06)
V300	0.04 (0.1)	0 (0)
V333	0.03 (0.08)	0.03 (0.08)
V366	0.03 (0.08)	0.01 (0.06)
V400	0.01 (0.06)	0.01 (0.06)
V433	0.04 (0.1)	0.03 (0.08)
V466	0.03 (0.08)	0.03 (0.08)
V500	0.06 (0.14)	0.01 (0.06)

Table 3. Mean proportion of fusion by subject group (task ordering).

Effects of Task Ordering on Asynchrony Detection

The mean of items judged as “in sync” by subjects in the FA group was .281 (*SD* = .450) and the mean of items judged as “in sync” by subjects in the AF group was .194 (*SD* = .395). Performance on the asynchrony detection task by group is plotted in Figure 8 and displayed in Table 4. A univariate ANOVA with asynchrony level and subject group as factors revealed a significant main effect of asynchrony level on proportion fused, $F(24, 3550) = 19.154$, $p < .001$. In addition, there was a significant main effect of group type (FA vs. AF) on synchrony judgments, $F(24, 3550) = 42.73$, $p < .001$. Post-hoc independent samples t-tests revealed that the FA group reported a higher proportion of items “in sync” at the following asynchrony levels: 200-233 ms auditory lead, 200-366 ms visual lead, and 433 ms visual lead ($p < .05$ for all of these asynchrony levels). The trend suggests that individuals in the FA group were more likely to report asynchronous items as “in sync” at longer visual-lead asynchronies and some longer auditory-lead asynchronies.

Synchrony judgments by subject group

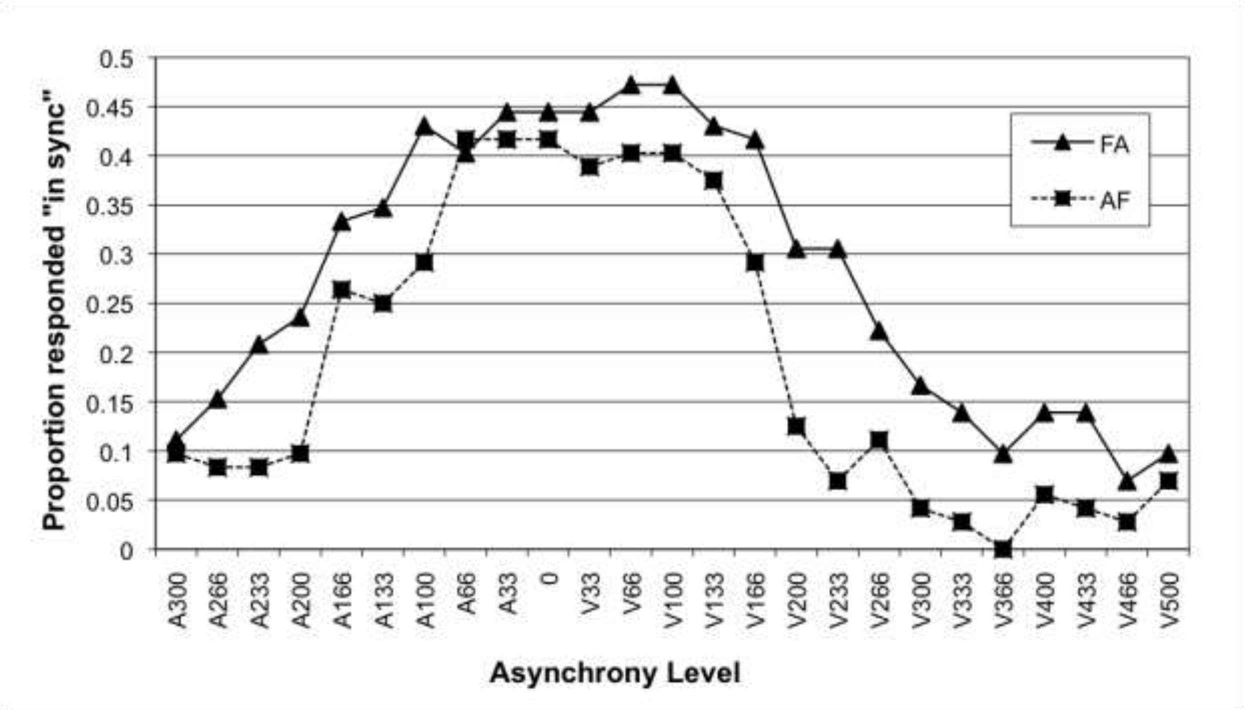

Figure 8. Synchrony judgments for each subject group (i.e., for each task order) is shown. Overall, subjects who performed the fusion task first reported experimental items "in sync" more often than subjects in the group who performed the asynchrony detection task first.

Asynchrony Level	FA Mean (SD)	AF Mean (SD)
A300	0.11 (0.32)	0.1 (0.3)
A266	0.15 (0.36)	0.08 (0.28)
A233	0.21 (0.41)	0.08 (0.28)
A200	0.24 (0.43)	0.1 (0.3)
A166	0.33 (0.47)	0.26 (0.44)
A133	0.35 (0.48)	0.25 (0.44)
A100	0.43 (0.5)	0.29 (0.46)
A66	0.4 (0.49)	0.42 (0.5)
A33	0.44 (0.5)	0.42 (0.5)
0	0.44 (0.5)	0.42 (0.5)
V33	0.44 (0.5)	0.39 (0.49)
V66	0.47 (0.5)	0.4 (0.49)
V100	0.47 (0.5)	0.4 (0.49)
V133	0.43 (0.5)	0.38 (0.49)
V166	0.42 (0.5)	0.29 (0.46)
V200	0.31 (0.46)	0.13 (0.33)
V233	0.31 (0.46)	0.07 (0.26)
V266	0.22 (0.42)	0.11 (0.32)
V300	0.17 (0.38)	0.04 (0.2)
V333	0.14 (0.35)	0.03 (0.17)
V366	0.1 (0.3)	0 (0)
V400	0.14 (0.35)	0.06 (0.23)
V433	0.14 (0.35)	0.04 (0.2)
V466	0.07 (0.26)	0.03 (0.17)
V500	0.1 (0.3)	0.07 (0.26)

Table 4. Mean proportion of synchrony judgments by subject group (task ordering).

Individual Differences

A wide range of variability was observed across subjects in both tasks. The overall proportion of fusion for each subject is shown in Figure 9. As previously mentioned, no significant effect of group (FA or AF) was observed on the fusion task. The minimum proportion fused was .00 (13 subjects never reported a fused response) and the maximum proportion fused was .34 (*M* = .050, *SD* = .219). From inspection of the data, it therefore seems that there are at least three groups of subjects: those who never fused (*N* = 13), those who fused, but very infrequently (with proportions of fusion from .01 to .05; *N* = 14), and those who fused more often (with a fusion rate of .1 or higher; *N* =7).

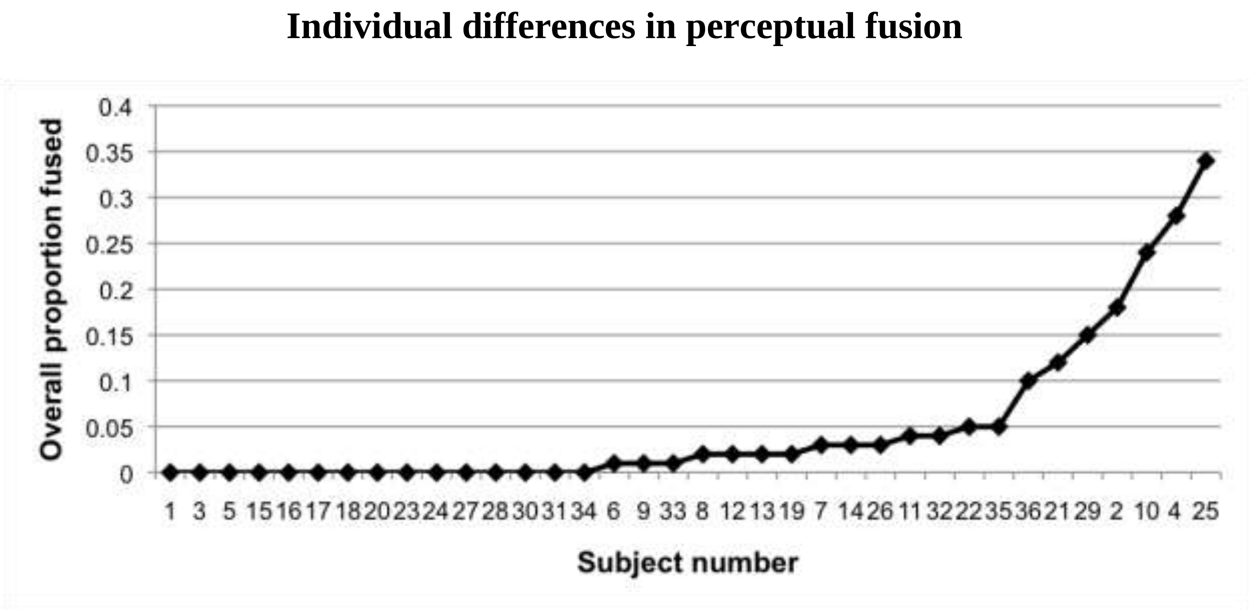

Figure 9. Individual differences in the perceptual fusion task are shown (in ascending order). There is wide variability across subjects. Thirteen of the subjects never fused incongruous AVPF items whereas seven subjects fused items at a rate of 10% or more.

The distribution of variability across subjects is very different for the asynchrony detection task from the fusion task described above. The data yield a linear trend, and it is difficult to divide subjects into discrete groups based on proportion of synchrony judgments. The overall proportion of experimental stimuli reported “in sync” by subject is shown in Figure 10. The maximum proportion reported “in sync” by a subject was .6 and the minimum proportion reported “in sync” by a subject was .02 (*M* = .238, *SD* =.163).

The overall proportion fused and proportion synchrony judgments for each subject are shown in Table 5. As previously mentioned, only a weak correlation was observed between proportion of experimental stimuli fused in the perceptual fusion task and the proportion of experimental stimuli reported “in sync” in the asynchrony detection task (*r*(3600) = .125, *p* < .001). However, we observed that the seven individuals who had the highest overall fusion scores (ranging from .1 to .34) demonstrated a correlation between proportion fused and proportion reported “in sync” that was over twice as high as the correlation for all subjects (*r*(700) = .283, *p* < .001). This indicates that performance on one task is better predicted by the other for “good” fusers than for poor or non-fusers.

Individual differences in asynchrony detection

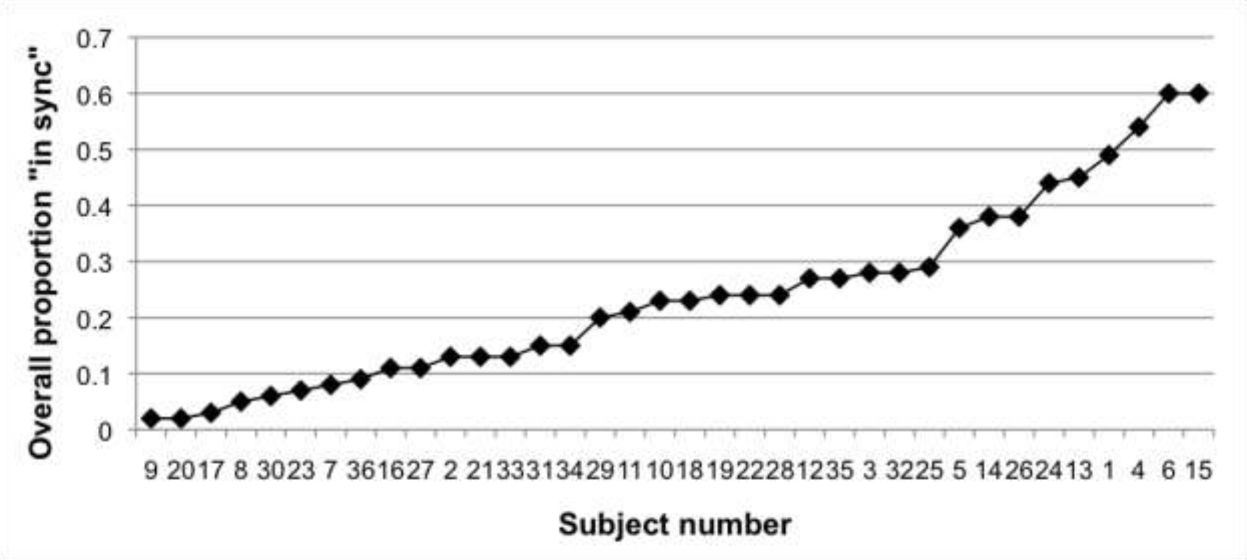

Figure 10. Individual differences in the perceptual task are shown (in ascending order). While there is a great amount of variability among subjects, the distribution approximates a linear continuum, unlike the variability demonstrated in the perceptual fusion task (shown in Figure 9 above).

Subject	Mean fused (SD)	Mean reported "in sync" (SD)
1	0 (0)	0.49 (0.35)
2	0.18 (0.21)	0.13 (0.18)
3	0 (0)	0.28 (0.29)
4	0.28 (0.2)	0.54 (0.26)
5	0 (0)	0.36 (0.31)
6	0.01 (0.05)	0.6 (0.38)
7	0.03 (0.08)	0.08 (0.14)
8	0.02 (0.07)	0.05 (0.1)
9	0.01 (0.05)	0.02 (0.07)
10	0.24 (0.18)	0.23 (0.24)
11	0.04 (0.09)	0.21 (0.2)
12	0.02 (0.07)	0.27 (0.25)
13	0.02 (0.07)	0.45 (0.24)
14	0.03 (0.08)	0.38 (0.29)
15	0 (0.07)	0.6 (0.3)
16	0 (0)	0.11 (0.16)
17	0 (0)	0.03 (0.08)
18	0 (0)	0.23 (0.2)
19	0.02 (0.07)	0.24 (0.23)
20	0 (0)	0.02 (0.07)
21	0.12 (0.15)	0.13 (0.22)
22	0.05 (0.1)	0.24 (0.26)
23	0 (0)	0.07 (0.11)
24	0 (0)	0.44 (0.38)
25	0.34 (0.19)	0.29 (0.35)
26	0.03 (0.08)	0.38 (0.31)
27	0 (0)	0.11 (0.18)
28	0 (0)	0.24 (0.21)
29	0.15 (0.19)	0.2 (0.23)
30	0 (0)	0.06 (0.11)
31	0 (0)	0.15 (0.19)
32	0.04 (0.09)	0.28 (0.29)
33	0.01 (0.05)	0.13 (0.19)
34	0 (0)	0.15 (0.18)
35	0.05 (0.1)	0.27 (0.31)
36	0.1 (0.13)	0.09 (0.14)

Table 5. Individual differences in the perceptual fusion and asynchrony detection tasks by subject.

Discussion

In this study, we observed a wide range of variability both among subjects and across stimulus sets for both the perceptual fusion and asynchrony detections tasks. Fusion rates and synchrony judgments were higher at shorter asynchrony levels than at longer asynchrony levels. Moreover, synchrony judgments were more robust than AVPF (i.e., fusion responses) to temporal asynchrony. In addition, task ordering affected performance, at least on the asynchrony detection task.

Perceptual Fusion Task

Considerable variability was observed in fusion rates across stimulus sets, which may be explained by several factors. First, the stimulus sets containing the liquid /l/ (*back* + *lack* and *pay* + *lay*) in the auditory signal were fused more frequently than the stimulus sets containing the liquid /r/. This replicates earlier findings by Radicke (2007) that fusion occurs more frequently with stop-liquid clusters containing /l/ than with stop-liquid clusters containing /r/.

Furthermore, the *back* + *lack* stimulus set yielded more fusions than the *pay* + *lay* set. One explanation for this pattern is that the fused responses for *back* + *lack* were always a real word, *black*, whereas the fused responses for *pay* + *lay* were often the pseudoword *blay* (rather than the real word *play*). Because voicing is not well specified by visual information, it is not surprising that subjects often reported hearing a voiced bilabial stop (likely assimilating the voicing of the bilabial stop to the voicing information they received in the auditory signal). Some evidence has been reported for the contribution of lexical effects in audiovisual integration. Brancazio (2004) found that subjects are more likely to fuse conflicting audiovisual stimuli and report a McGurk response when the fused response constitutes a word but the auditory signal does not. However, Barutchu and colleagues (2008) found conflicting results regarding lexical effects on McGurk fusions. When CVC words and pseudowords mismatched audiovisually in their final consonant, subjects were more likely to report McGurk fusions for words than to pseudowords. However, when an audiovisual mismatch occurred in the first consonant of the word, there was no lexical effect on McGurk fusions. Based on these results, the authors suggested that audiovisual integration is an early process that may be modulated at later stages by a lexical effect.

Another explanation for the difference in fusion rates for *back* + *lack* and *pay* + *lay* may involve integration difficulty. Because most of the stimuli (96%) in the experiment were temporally asynchronous, subjects were doing some "extra" integration for nearly every stimulus that they were presented with. Given that the *back* + *lack* pair has a consonant-final rhyme whereas the *pay* + *lay* pair ends in a vowel, it may be that subjects had to do more integration for pairs ending in -*ack* than for the pairs ending in -*ay*, which has no consonant in its coda. It is possible that the "extra" qualitative conflict of a visual stop and auditory liquid was better fused when other, more difficult integration of a word-final consonant was also occurring.

Wide variability of fusion rates was also observed across asynchrony levels, with higher fusion rates at shorter asynchrony levels and lower fusion rates at longer asynchrony levels. As predicted, there was a slight skew to the visual-lead side and the maximum fusion level occurred at 100 ms of visual lead. Furthermore, using 50% of the maximum fusion as a threshold for robust fusion, the window of fusion was found between 66 ms of auditory and 200 ms of visual lead. This is similar to the window of fusion for McGurk stimuli presented with audiovisual timing asynchronies. Robust fusion for McGurk stimuli has been found to occur at longer levels of visual than auditory lead (Munhall et al., 1996).

Asynchrony Detection Task

As in the perceptual fusion task, a great deal of variability was also observed in the synchronous judgments across stimuli. The stimuli sets which were reported "in sync" most often were the two sets ending in the rhyme -*ack*, *back* + *lack* and *back* + *rack*. These sets did not differ from each other but did differ from the two sets ending in -*ay* (*pay* + *lay* and *pay* + *ray*). In a previous study by Conrey and Pisoni (2006), a great deal of variability was also observed across words in an asynchrony detection task (where the same word was presented in both auditory and visual modalities).

The reason for these differences is not well understood; however, the explanation for the difference between the two sets of stimuli in the present study may also involve integration difficulty. In the sets with no final consonant, there is less information given at the end of the word which might inform the subject about temporal asynchrony; however, rather than being more likely to report such words as "in sync," subjects actually gave fewer synchronous responses for these items. It may be that the perceptual system first tries to integrate the entire stimulus (from both modalities) but when there is not as much information (as in the *pay* + *lay* and *pay* + *ray* cases) to integrate, subjects are better at detecting asynchrony. For instance, when subjects hear and see asynchronous *back* + *lack*, they get three cues of mismatch: the qualitative /b/ + /l/ mismatch, the temporal mismatch in the onset of the auditory and visual words, and the temporal mismatch in the offset of the auditory and visual words. However, when subjects hear and see asynchronous *pay* + *lay*, they get only two cues of mismatch: the qualitative /p/ + /l/ mismatch and the temporal mismatch in the onset of the auditory and visual words. It appears that the threshold for temporal asynchrony may actually be greater when there are more segmental cues that must be integrated.

As predicted, subjects reported that both control and experimental items were synchronous more often at shorter asynchronies and less often at longer asynchronies; furthermore, they were more likely to report that visual-lead stimuli were "in sync" than auditory-lead stimuli. This pattern is not surprising; previous studies have found more tolerance for visual-lead stimuli, and the nature of English phonotactics also predicts more tolerance for visual-lead stimuli. Visual /b/ precedes /l/ in the permissible consonant cluster /bl/; however, the cluster /*lb/ is not legal in English. These findings replicate other studies of audiovisual temporal asynchrony perception in which subjects are less likely to detect asynchrony with visual than with auditory lead (Conrey & Pisoni, 2006; Dixon & Spitz, 1980).

Subjects were nearly twice as likely to report that control stimuli were "in sync" as they were to report that experimental stimuli were "in sync." In addition, the thresholds (as defined by 50% of the maximum) for robust synchrony judgments differed for control and experimental items. For control items, the boundaries were located at 200 ms auditory lead and 300 ms visual lead. For experimental items, the boundaries were located at 166 ms auditory lead and 166 ms visual lead.[6] Critically, the window of synchrony judgments for experimental items was found to be smaller than for control items. It is possible that subjects are less likely to temporally integrate items when the auditory and visual streams are qualitatively different. However, given that fusion rates were generally low (approximately 5% overall), subjects may have reported items as being "out of sync" because they did not fuse them and noticed that the visual stream did not qualitatively match the auditory stream.

Comparing the Two Tasks

A difference was observed in the size of the window for robust fusion and robust synchrony judgments. Whereas the window for phonological fusion was largely skewed to the visual side, from 66 ms auditory lead to 200 ms visual lead, the window for synchrony judgments was symmetric about zero (from 166 ms auditory lead to 166 ms visual lead). Although the boundaries were different, a small correlation was found between fusion rate and synchrony judgments. Taken together, these findings suggest a relation between the mechanisms used for the explicit task of asynchrony detection and the more implicit integration that occurs during the perceptual fusion task. However, it is also clear that AVPF may disappear when there is too much temporal asynchrony. These findings indicate that the window of AVPF is smaller than the window of synchrony judgments, and hence synchrony judgments are more robust than AVPF.

In a recent study comparing closed-set responses (including visual, auditory, and fusion responses) to McGurk stimuli and synchrony judgments of the same stimuli, Van Wassenhove, Grant, and

[6] However, we also found that for both control and experimental items, visual-lead stimuli were reported as "in sync" more often than auditory-lead stimuli.

Poeppel (2007a) found similar results. The temporal windows over which items were reported as fused occurred from around 133 ms auditory lead to 266 ms visual lead (compared to 66 ms auditory lead and 200 ms visual lead in the present study), and the temporal windows in which items were reported as synchronous occurred between around 36 ms auditory lead to around 121 ms visual lead for incongruous pairs (compared to 166 ms auditory lead to 166 ms visual lead in the present study). Although the limits of the windows were defined somewhat differently in the study by Van Wassenhove and colleagues, the present study replicates the general pattern of results they reported.

Effects of Task Ordering on Perceptual Fusion

While we did not observe a significant effect of task ordering (i.e., subject group) on performance on the perceptual fusion task, there was a small but significant interaction between subject group and asynchrony level. Upon close inspection of Figure 7, it can be seen that subjects in the AF group tended to fuse items more often at shorter asynchrony levels and less often at longer asynchrony levels. This may indicate that if subjects have performed the asynchrony detection task first, they are more likely to fuse items that they were previously more likely to rate as being “in sync” and less likely to fuse items that they were previously less likely to rate as being “in sync.” Thus, the discrimination carried out during synchrony judgments in the asynchrony detection task may influence perception during the perceptual fusion task.

Effects of Task Ordering on Asynchrony Detection

Subjects were more likely to report that items were “in sync” if they had previously performed the perceptual fusion task (FA group). This finding suggests that, even though the overall fusion rate in the perceptual fusion was very low (about 5%) for both groups, completing the fusion task with conflicting stimuli results in less sensitivity to temporal asynchrony in those stimuli later. It also suggests that even if subjects do not fuse conflicting auditory and visual information in the perceptual task, they may still be integrating the information at some level, making them less likely to detect asynchrony later.

Individual Differences

It is not surprising that a great deal of variability in performance was observed in the perceptual fusion task. Enormous individual differences in lip-reading are frequently reported (Summerfield, 1992; Bernstein, Demorest, and Tucker, 2000). However, there is some evidence that speed of low-level visual processing is correlated with lip-reading (Shepherd, Delavergne, Fruch, and Clobridge, 1977; Shepherd, 1982); that is, subjects who had shorter peak-latencies in electroencephalic responses to a visual light flash were better at lip-reading than subjects with higher peak-latencies. These findings indicate that there may be a low-level visual component to audiovisual integration that is physiologically “hard-wired” (Summerfield, 1992). Although the correlation between performance on the perceptual fusion task and the asynchrony detection task was weak, a stronger correlation was found between the two tasks for the seven individuals who had the highest proportion of fusions. It is possible that this correlation may be attributable to faster low-level visual processing by individuals who are better fusers. However, it should be noted that the differences in performance on the perceptual fusion task were larger than the differences in performance on the asynchrony detection task, which varied along a more linear continuum. All subjects in the asynchrony detection task reported that at least some of the experimental items were “in sync,” even if they never reported fusion in the perceptual task. More research is needed in order to elucidate the underlying causes of individual differences in audiovisual integration tasks such as those performed in the present study.

Summary of findings

Findings from the present study suggest that (1) AVPF is only partially robust to temporal asynchrony; (2) synchrony judgments for AVPF stimuli are more robust than perceptual fusion is; and (3) even when AVPF stimuli are not often fused in a perceptual task, exposure to the stimuli may involve some implicit integration which effects an increased likelihood that the asynchronous stimuli will subsequently be judged as synchronous. These findings suggest that audiovisual fusions (in particular, AVPF) do not reflect the operation of impenetrable, encapsulated modules but rather can be modulated both by bottom-up (in the case of temporal asynchrony) and top-down (in the case of modulations due to previous exposure to AVPF stimuli) processes. In addition, the present findings support an integrative network model of audiovisual speech perception in which the probability of integration of the auditory and visual streams is modulated by both stimulus-driven and cognitive constraints based on prior experience and knowledge.

References

Barutchu, A., Crewther, S.G., Kiely, P., Murphy, M.J., & Crewther, D.P. (2008). When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Pyschology, 20*(1), 1-11.

Berger, K.W. (1972). Visemes and homophenous words. *Teacher of the Deaf, 70,* 396-399.

Bernstein, L.E., Demorest, M.E., & Tucker, P.E. (2000). Speech perception without hearing. *Perception & Psychophysics, 62*(2), 233-252.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 30(3), 445-463.

Calvert, G.A., Brammer, M., Bullmore, E., Campbell, R., Iversen, S.D., David, A. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport, 10*, 2619-2623.

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

Calvert, G.A., Spence, C., & Stein, B.E. (eds.) (2004). *The Handbook of Multisensory Processes*. Cambridge, Mass.: The MIT Press.

Conrey, B. and D.B. Pisoni (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America, 119*(6), 4065-4073.

Cutting, J. E. (1975). Aspects of Phonological Fusion. *Journal of Experimental Psychology, 104*(2), 105-120.

Cutting, J.E. and R.S. Day. (1975). The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics, 3,* 99-113.

Cutting, J. E. (1976). Auditory and Linguistic Processes in Speech Perception: Inferences from Six Fusions in Dichotic Listening. *Psychological Review 83*(2), 114-140.

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

Green, K.P., Kuhl, P.K., Meltzoff, A.N. & Stevens, E.B. (1991). Integrating speech information across talkers, genders, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50*(6), 524-536.

Levi, S.V., Radicke, J.L., Loebach, J.L. & Pisoni, D.B. (January, 2008). *Beyond the McGurk effect: Audiovisual consonant cluster formation.* Paper presented at the 82nd annual meeting of the Linguistics Society of America, Chicago, IL.

Lidestam, B. & Beskow, J. Visual phonemic ambiguity and speechreading. *Journal of Speech, Language, and Hearing Research, 49*(4), 835-47.

MacDonald, J. & H. McGurk. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253-257.

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

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

Miller, L.M. & D'Esposito, M. (2005). Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech. *The Journal of Neuroscience*, *25*(25), 5884-5893.

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

Radicke, J. L. (2007). *Audiovisual Phonological Fusion*. Unpublished master's thesis, Indiana University, Bloomington, IN.

Sams, M., Aulanko, R., Hamalainene, M. Hari, R. Rounasmaa, O.V., Lu, S.T., Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex.

Shepherd, D.C. (1982). Visual-neural correlate of speechreading ability in normal-hearing adults: Reliability. *Journal of Speech and Hearing Research*, *25*, 521-527.

Shepherd, D.C., Delavergne, R.W., Fruch, F.X., and Clobridge, C. (1977). Visual-neural correlate of speech reading ability in normal-hearing adults. *Journal of Speech and Hearing Research, 20*, 752-765.

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

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences*, *335*, 71-78.

Van Wassenhove, V., Grant, K.W., & Poeppel, D. (2007a). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*, 598-607.

Van Wassenhove, V., Grant, K.W., & Poeppel, D. (2007b). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102*(4), 1181-1186.

Vatakis, A. & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744-756.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: familiar faces and voices in the McGurk effect. *Perception & Psychophysics*, *57*(8), 1124-1133.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Error Analysis of Spoken Word Recognition[1]

Robert A. Felty, Adam Buchwald[2] and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH grant DC-00111 and DC-00012. We would also like to thank Wes Alford (stimulus selection); Melissa Troyer, Ryder Timberlake, Sophie Parkison, Katie Scherschel (running subjects and transcription); and audiences at MCWOP 13, and the Mental Lexicon 6 for their input.

[2] currently in the Department of Speech-Language Pathology and Audiology, Steinhardt School of Culture, Education and Human Development New York University, New York, NY 10012

Error Analysis of Spoken Word Recognition

Abstract. This study reports the analysis of incorrect responses from a spoken word recognition experiment of 1,428 words, designed to be a representative sample of the entire English lexicon. The stimuli were presented in six-talker babble to 193 normal-hearing listeners at three signal-to-noise ratios (0, 5, and 10 dB). The results reveal several patterns: (1) Errors tend to be of higher frequency than the target word, and response frequency is significantly correlated with target frequency; (2) Incorrect responses were very close to the target words in terms of number of phonemes and syllables, but had a mean edit distance of 3; and (3) NonWord responses were phonetically more similar to the target words than word responses. Implications for theories of spoken word recognition are discussed.

Background

The majority of studies on spoken word recognition report only percent correct analyses. Traditionally, the goal of such studies was to measure intelligibility, often for the means of testing radio and telephony equipment (e.g. Fletcher & Steinberg, 1929; Hirsh et al., 1954; House et al., 1965). While percent correct (and other measures such as Articulation Index) are appropriate for such purposes, more recent research on spoken word recognition has also aimed at providing more information about how words are stored and accessed in the mental lexicon (e.g. Luce & Pisoni, 1998). For such purposes, it is advantageous to investigate the nature of incorrect responses.

While several studies have performed error analyses of spoken word recognition, they have either focused solely on monosyllabic stimuli (e.g. Pollack et al., 1960), or have been concerned with automated speech recognition (e.g. Kawai & Higuchi, 1998; Siohan et al., 2004). Many models of speech recognition have assumed that the findings on monosyllabic words also apply to multisyllabic words, but few studies have actually investigated this. The present study seeks to fill this gap by analyzing errors in spoken word recognition from a stimulus list designed to be a representative sample of the entire English lexicon.

Neighbors as competitors

One key finding in research on spoken word recognition is that accuracy can be affected not only by the phonetic composition and frequency of the stimulus word, but also by its phonetic similarity to other words (Luce, 1986; Luce & Pisoni, 1998; Benkí, 2003). This characteristic has been termed neighborhood density, and refers to the number of neighbors for a given word, where the definition of neighbor has traditionally been that two words are neighbors if they differ by only one phoneme (addition, substitution, or deletion). Accuracy in spoken word recognition is less for words with many neighbors than for words with few neighbors. This finding has also been incorporated into most theories of spoken word recognition, in that they assume that multiple words are activated during the recognition process, and these words compete with each other. Under this assumption, one would expect that when a listener misperceives a word, that the misperception would be a neighbor of the target word. These competitors are usually related to the target word, and therefore can also be thought of as neighbors. However, it is an only an assumption that all competitors are neighbors. By looking at recognition errors, we can empirically determine the relationship between targets and competitors.

Method

Materials

1,428 English words chosen from the Hoosier Mental Lexicon, designed to be a representative sample of the entire English lexicon, based on: (1) Number of phonemes (2-11); (2) Number of syllables (1-5); (3) Syllable structure; (4) Initial phoneme; and (5) Lexical frequency.

The materials were recorded by 2 native speakers of American English (1 male, 1 female) in an IAC booth directly into digital format sampled at 22.5 KHz. Six-talker babble taken from the Connected Speech Test (Cox et al., 1987) was added to the stimuli at 3 different signal-to-noise ratios (S/N): 0, 5, and 10 dB.

Procedure

193 native English-speaking undergraduates from Indiana University heard the recorded materials over headphones and entered responses via keyboard. The stimuli were presented in isolation at 77 dB SPL. Each listener heard only 1/4 of the stimulus list, spoken by either the male or the female talker; 1/3 of the stimuli were presented at each S/N ratio.

Analysis

A total of 68,901 trials were presented to listeners. Of these, 597 trials were discarded, because the listener did not enter a response, or entered a random response such as 'asdf'. This left 68,304 trials for analysis. Responses were converted into phonetic transcriptions semi-automatically. We did so by checking the responses with a custom version of the CELEX (Baayen et al., 1993). The English portion of the CELEX database is based on British English sources, and has British English phonetic transcriptions. Since we are using American English talkers and listeners, American English transcriptions are preferred. To do so, we took phonetic transcriptions from the HML (Nusbaum et al., 1984) and the CMU pronouncing dictionary. For responses that were not in our database, the responses were checked manually by at least one laboratory assistant and the lead author. We classified the responses into the following categories.

- MISSPELL (e.g. *plian* for *plain*)
- MISSING (e.g. *google*)
- FOREIGN (e.g. bjorn, puedes)
- MULTIPLE (e.g. *and then, both men*
- NEOLOGISM (e.g. *untypical, righten*)
- NONWORD (e.g. *nisc, vicundity*)

Responses were categorized as misspellings if the response was not a real word, and a simple change would make it a real word. If there was any doubt, the response was treated as a nonword.

To determine whether a response was a word, we performed a search in our customized version of the CELEX database. In some cases, we determined that a response which was missing from the database should be considered a real word, e.g. *google, laptop*. Since the stimulus list contained some proper nouns, we also added other missing proper nouns, e.g. *Michigan, Stephen*. For these words, phonetic transcriptions were created using native speaker knowledge from the lead author as well as one or two assistants. Other

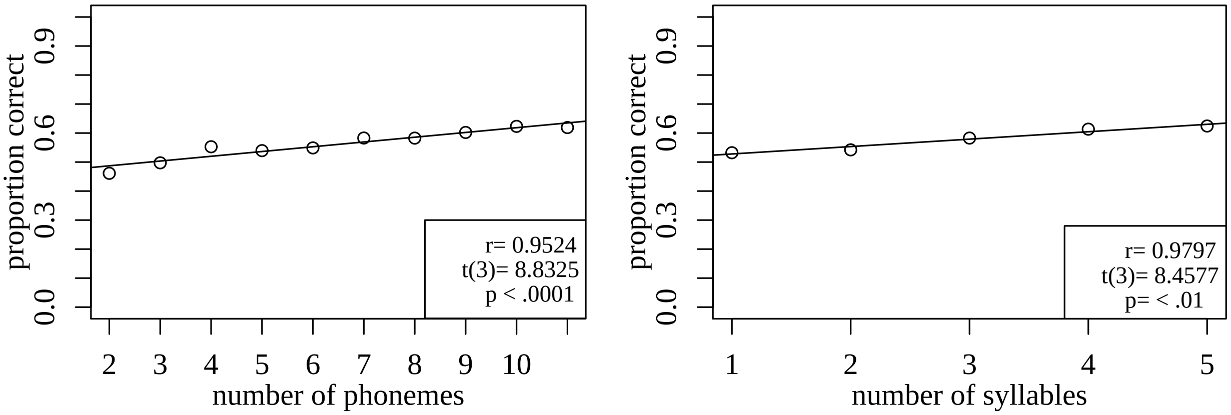

Figure 1. Percent correct as a function of the number of phonemes (left panel) and syllables (right panel) in the target word

sources such as the American Heritage Dictionary were also consulted. In addition, a word frequency estimate was calculated based on Google page count. To do this, the google page count for the 100 most frequent words in the CELEX database was compared with the CELEX frequency, and the mean ratio was calculated to be 94,778.56. For each missing word, the google page count was divided by 94,778.56, in order to achieve a reasonable comparison.

Results

Out of the 68,304 trials for analysis, 38,068 (55.7%) trials were classified as correct, and 30,236 (44.3%) as incorrect. Of the incorrect responses, 8,354 (27.6%) were nonwords.

Percent Correct

Figure 1 shows percent correct increases as a function of word length, whether measured in terms of the number of phonemes or syllables in the word. However, the effect here seems to be linear as opposed to previous Wiener & Miller (1946), who found a logarithmic pattern. This could be due to the fact that Wiener & Miller had a greater range of percent correct values.

Figure 2 shows the percent correct distribution for subjects and for items. The distribution for items is fairly flat, with approximately equal proportions of items that were highly intelligible, as well as items that extremely unintelligible. It should be noted that there is slight bias towards unintelligible items for the male speaker, and a bias for intelligible items for the female speaker. Overall the female speaker was more intelligible than the male speaker, which explains this effect. The even distribution of item intelligibility is consistent with our intentions of sampling the entire English lexicon, since the lexicon includes both “easy” as well as “difficult” words, and everything in between.

The percent correct distribution for subjects is quite different. All of the subjects scored a total percent correct between 26.9% and 78.2%. However, the range is actually smaller, when taking into account that the female talker was substantially more intelligible than the male talker. Listeners who heard the male talker scored between 26.8% and 56.6%, while listeners who heard the female talker scored between 53.9%

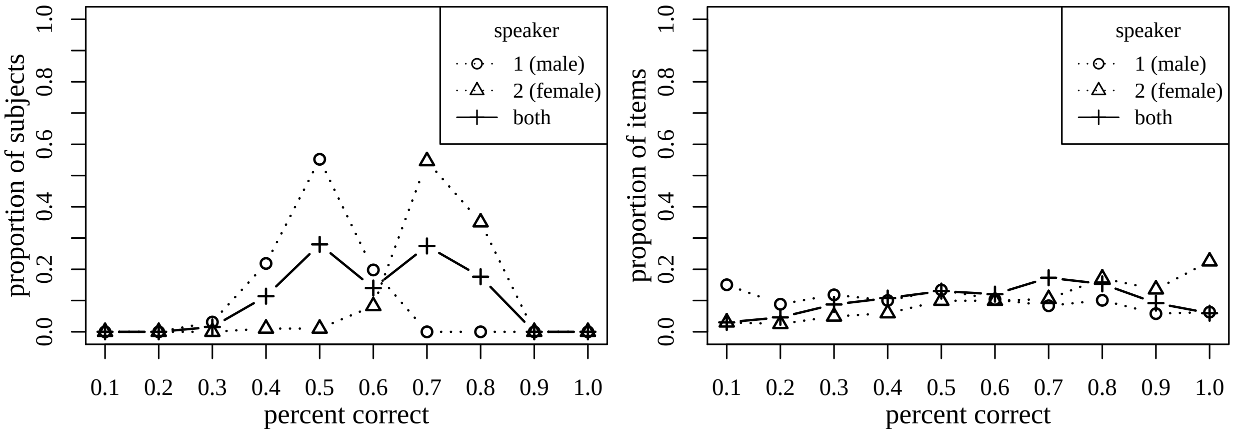

Figure 2. Percent correct distribution by subjects (left panel) and items (right panel)

and 78.2% (excluding two listeners, who scored more than 1.5 times the interquartile range below the first quartile).

Edit Distance

One way of comparing the incorrect responses to the target words is to measure the edit distance (also called Levenshtein distance) between the target and the response. Edit distance is defined as the number of edits — additions, deletions, and substitutions — to change one string into another. For example, the edit distance between *cat* and *cats* is 1; between *cat* and *cuts* is 2; between *cat* and *its* is 3.

Figure 3 shows the edit distance distribution for incorrect responses. The most striking result from the edit distance analysis is that less than 23% (< 26% of non-hermit stimuli) of errors are "neighbors" under the traditional definition (edit distance 1). Nearly 40% of the incorrect responses for one syllable words are neighbors, suggesting that the traditional definition of neighbor captures a fairly large portion of the variance for shorter words, but is not adequate for describing longer words.

Our results also indicate that edit distance increases as the S/N ratio becomes less favorable. This means that as the ambient noise increases, not only do listeners misperceive the target word more frequently, but the misperceptions are also phonetically less similar to the target word.

Results were also analyzed using an additional variant on edit distance which incorporates a notion of phonetic similarity. For this we used a slightly modified version of a computer program written by Adam Albright to compute phonetic dissimilarity. The program uses a similarity metric based on phonetic features (see Frisch et al., 1997). The program returns the minimum phonetic distance based on the number of edits to change one word into another; however, each edit is weighted by the number of features which differ between the two phonemes. For our analyses, we have used the default weight of 0.6 for insertions and deletions. Table 1 shows several examples of the calculations.

As shown in Figure 4, the phonetically-weighted edit distance analysis is largely consistent with the edit distance analysis. Edit distance increases as the S/N ratio decreases; edit distance also increases with target word length.

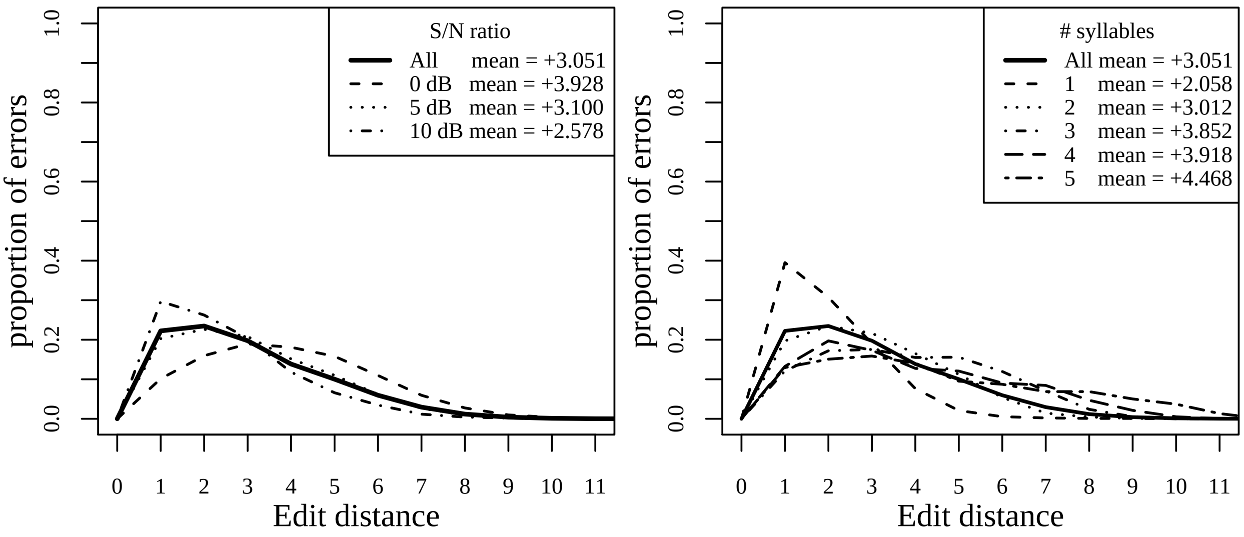

Figure 3. Edit distance distribution of incorrect responses compared to target words. The left panel shows the distributions broken down by S/N ratio. The right panel shows the distributions broken down by the number of syllables in the target word.

acclaim—*crane* = 1.755				
ə	k	l	e	m
	|		|	
-	k	r	e	n
0.60	0.00	0.52	0.00	0.64

ability—*important* = 5.617									
ə	-	b	ɪ	l	-	ɪ	-	t	i
	-			-			-	|	
ɪ	m	p	ɔ	ɹ	t	ə	n	t	-
0.71	0.60	0.52	0.75	0.52	0.60	0.71	0.60	0.00	0.60

Table 1. Examples of phonetically weighted edit distance calculations

Phoneme Difference

We also analyzed incorrect responses in terms of the difference in the number of phonemes between the response and the target word. Figure 5 shows the distribution of the phoneme differences by S/N ratio and by word length. Overall, incorrect responses tend to have roughly the same number of phonemes as the target word, with a slight bias towards deleting phonemes. The bias to delete phonemes increases with word length; however, even for five syllable words, most incorrect responses are only one phoneme shorter than the target word.

The difference in phonemes is highly consistent across S/N ratios, with a slight tendency for shorter responses as the noise level increases. This result, combined with the fact that edit distance tends to increases substantially as S/N ratio decreases, underscores the fact that listeners tend to make mostly substitution errors, with fewer deletion errors, and addition errors are the most uncommon.

In order to investigate whether this effect was due to chance, we performed a Monte-Carlo style simulation. The simulation used the following algorithm:

1. Pick a random word from the CELEX database which has the same edit distance from the target word as the real error. This is referred to as a pseudo-error.
2. Calculate the phoneme, frequency, and syllable difference between the pseudo-error and the target.

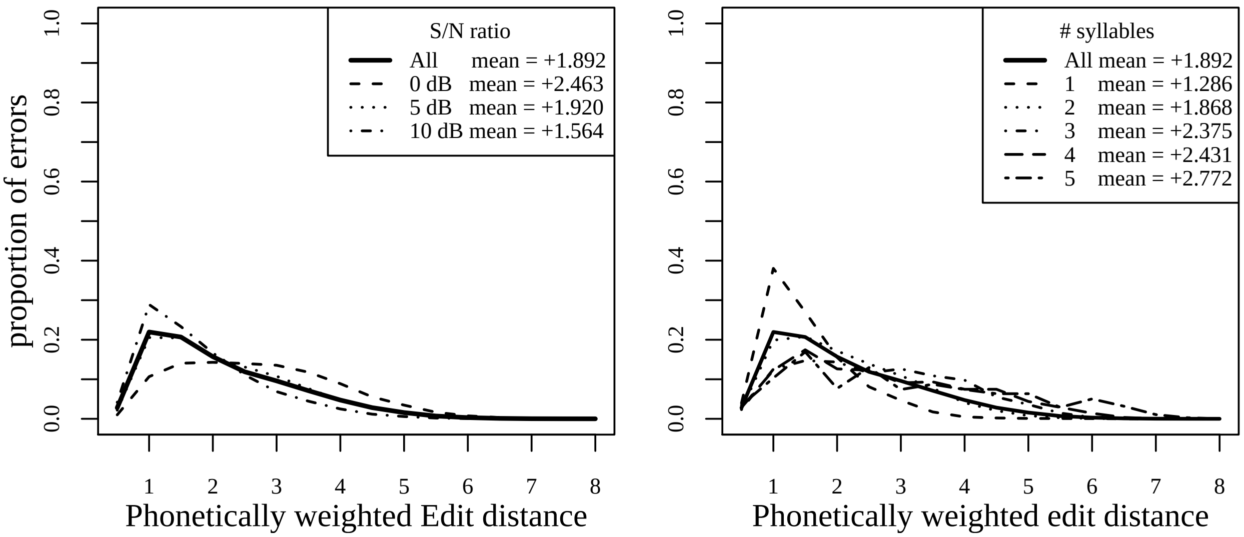

Figure 4. Phonetically-weighted edit distance distribution of errors for word and nonword responses

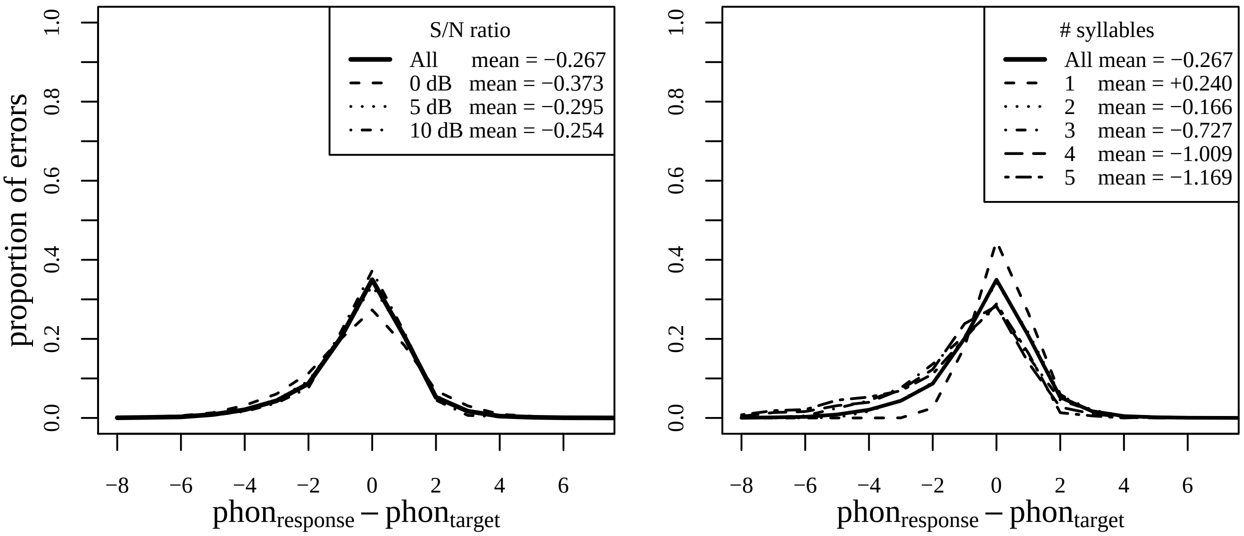

Figure 5. Difference in the number of phonemes between incorrect responses and the target words, broken down by S/N ratio (left panel) and target word length (right panel).

3. Calculate the mean phoneme difference for the pseudo-errors.

Our simulation executed this algorithm 10,000 times, yielding a range of possible values for the mean phoneme difference. If our actual number falls outside this range, we can say that our results are significant with $p < .0001$. In the case that our actual result falls within the range, we calculate the probability that the actual result would be in the range. Table 2 shows that our result does lie within the range of values in the simulation. Therefore we calculated that there were 15 simulation trials in which the mean phoneme difference was less than or equal to the actual result, and adjusted our p-value to $\frac{15}{10{,}000}$, or $p < .01$. The simulation indicates that listeners are more likely than chance to drop phonemes.

	Monte-Carlo			
	min	max	actual	p
mean phoneme difference	−0.2702	−0.2321	−0.267	< 0.01

Table 2. Monte-carlo simulation results for phoneme difference.

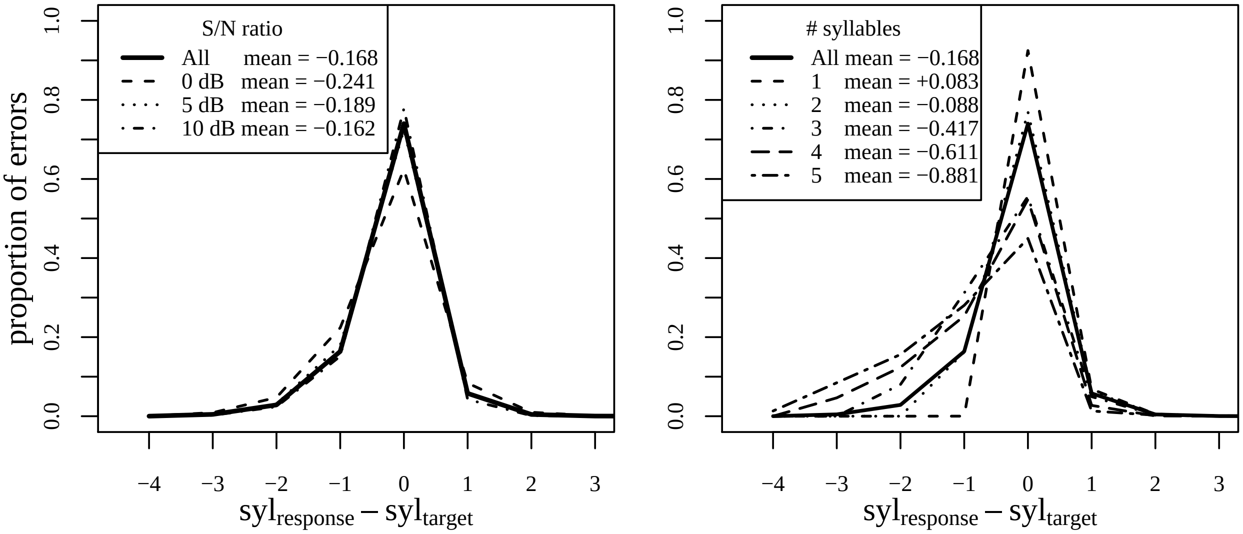

Figure 6. Distribution of syllable differences between response and target words, broken down by S/N ratio (left panel) and word length (right panel).

Syllable Difference

The incorrect responses were also analyzed by comparing the number of syllables in the response to the target, as shown in Figure 6. In agreement with the phoneme difference analysis, responses largely had the same number of syllables as the target word, with a slight bias towards dropping syllables. The tendency to drop syllables increases with word length, although most responses differ by less than one syllable from the target word, even for five syllable words. The difference in syllables is also highly consistent across S/N ratio, with a slight tendency for listeners to respond with shorter words as the S/N ratio decreases. Once again, this highlights the fact that the majority of errors are substitutions, rather than additions or deletions.

We also ran a simulation to determine whether the difference in syllables is due to chance, using the same algorithm as described above for the phoneme difference simulation. The results in Table 3 indicates that listeners are less likely than chance to drop syllables.

	Monte-Carlo			
	min	max	actual	p
difference	−0.2029	−0.1834	−0.168	< 0.0001

Table 3. Monte-carlo simulation results for difference in the number of syllables between incorrect responses and target words.

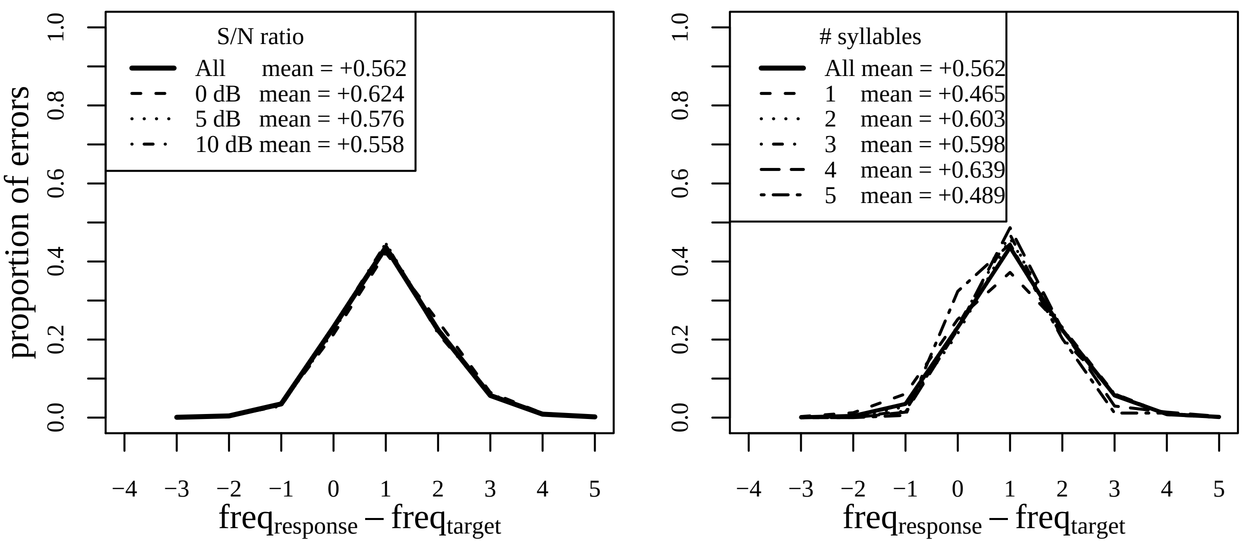

Figure 7. Distribution of differences between response and target word log-frequencies, broken down by S/N ratio (left panel) and word length (right panel).

Lexical Frequency

An additional factor to consider in error analysis is the frequency of the responses. For this analysis, we consider only the real word responses. Figure 7 shows the distribution of the differences in log frequency between response and target. Responses are higher in frequency than targets, which is consistent both across S/N ratio and word length. In addition, a Pearson test of correlation shows a modest correlation between the log-based frequency of the target words and the incorrect responses ($r = .296, p < .0001$), which is in contrast to the findings of Pollack et al. (1960). This discrepancy is partially due to the fact that the present study includes words ranging in length from one to five syllables, while the Pollack et al. study only used monosyllabic words. However, even when only considering monosyllabic target words, there is still a small correlation between the log frequency of target and response ($r = .173, p < .0001$).

As for the phoneme and syllable difference analyses, we also performed a monte-carlo simulation for frequency differences using the same algorithm. The results in Table 4 show that listeners are much more likely to respond with higher frequency words than would be predicted by chance. This is the expected result, given that there are many more low-frequency words in the language than high frequency.

	Monte-Carlo			
	min	max	actual	p
difference	−0.1197	−0.0842	0.562	< 0.0001

Table 4. Monte-carlo simulation results for difference in the log-frequency between incorrect responses and target words.

Word vs. nonword responses

Up until now, we have been treating all incorrect responses equally. However, it is also useful to distinguish between real word responses and nonword responses. In checking our data, we have divided the incorrect responses into the five categories; the proportion of responses per category is shown in table 5.

Category	#Responses	%Errors	Examples
WORD	21882	72.5	purse, skew
NONWORD	8100	26.7	nisc, vicundity
FOREIGN	23	0.1	bjorn, puedes
MULTIPLE	128	0.4	and then, both men
NEOLOGISM	103	0.3	untypical, righten

Table 5. Proportion of responses in each category

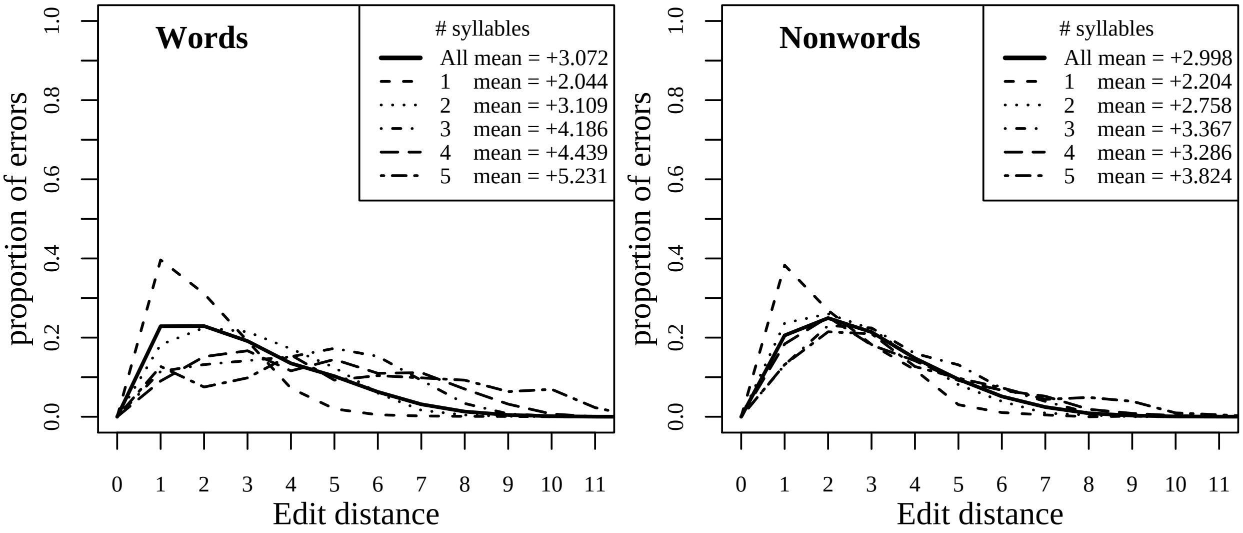

Figure 8. Edit distance distribution of errors for word and nonword responses

Since the foreign, multiple, and neologism categories are so small, these responses have been grouped into the nonword category for the remaining analyses.

Figure 8 shows the edit distance distribution separately for word and nonword responses. The mean edit distance for word and nonword responses does not differ very much (3.072*vs*2.998). However, differences do arise when viewing the edit distance distribution according to word length. For monosyllabic words, nonword responses have a higher edit distance than word responses, but for multi-syllabic target words the opposite is true.

We also analyzed the difference in phonemes separately for word and nonword responses, shown in Figure 9. Word responses have a greater phoneme difference than nonword responses at all word lengths. This suggests that when listeners provide nonword responses, they are responding in a more bottom-up fashion, and that they are using more top-down information when giving word responses.

As shown in Figure 10, the syllable difference for words and nonwords is consistent with the phoneme difference. Listeners' nonword responses tend to be phonetically more similar to the target word than word responses.

We have also replicated some of the analyses in Pollack et al. (1960), in order to test whether their results hold when analyzing a more representative sample of the English language. As in their study, we

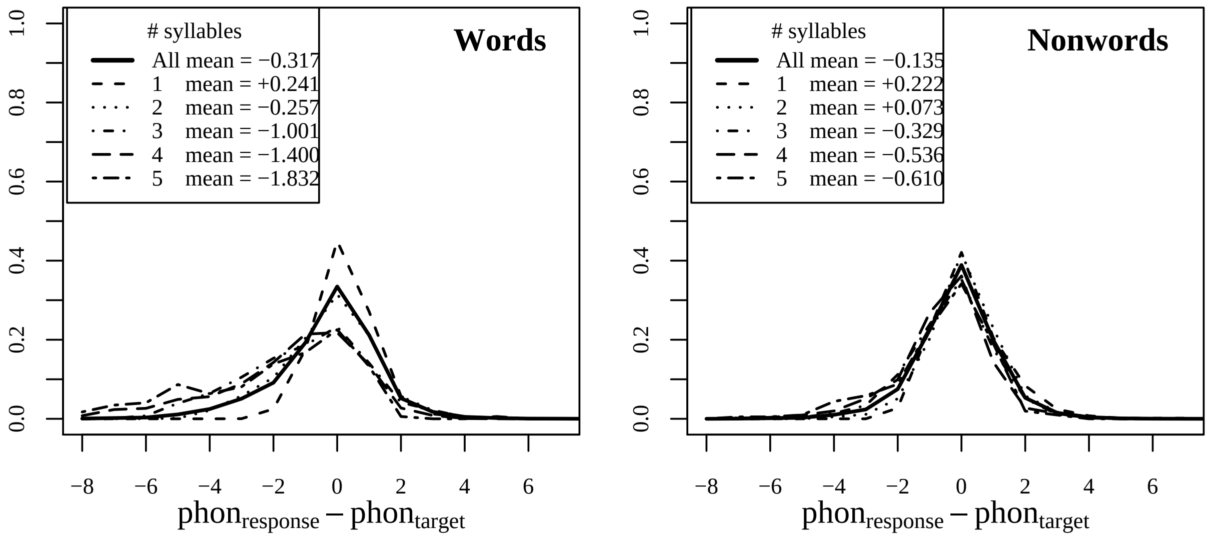

Figure 9. Phoneme difference distribution of errors for word and nonword responses

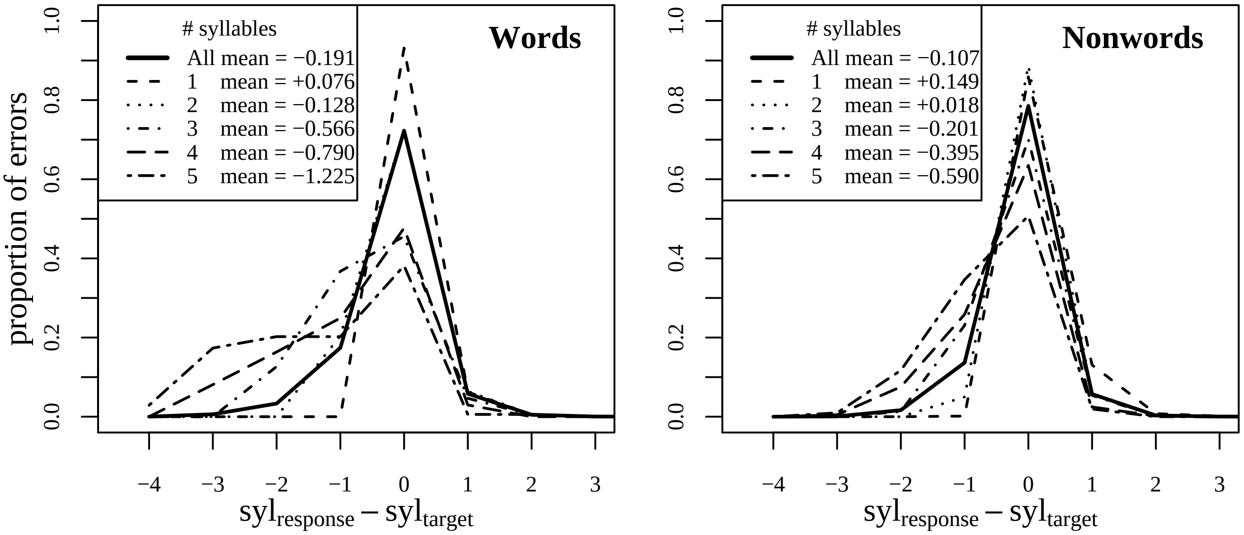

Figure 10. Syllable difference distribution of errors for word and nonword responses

have grouped the target words into eight different frequency classes. We have tried to keep the median frequency of each class approximately the same as those Pollack et al. used, while also keeping the number of words in each class fairly balanced. Table 6 shows the total number of responses in each frequency class at each S/N ratio, as well as the number of incorrect responses for each frequency class.[3]

Figure 11 shows the proportion of nonword responses at each S/N ratio by frequency class. In the figure, the eight frequency classes have been combined into four (1&2, 3&4, 5&6, 7&8). Our results show several differences from Pollack et al.. Overall, our results show a much larger proportion of nonword responses. Pollack et al. report a maximum of approximately 14% nonword responses, while we report a

[3] Note that the frequency classes are in reverse order (1 is most frequent; 8 is least frequent). This scheme has been preserved from Pollack et al. (1960) for more direct comparison

Class	mean Freq	median Freq	Number of Responses					
			S/N=0		S/N=5		S/N=10	
			incorrect	total	incorrect	total	incorrect	total
1	8871.8	4219	191	642	142	642	85	640
2	1055.7	1041	440	1322	237	1344	148	1347
3	316.3	289	1171	3460	725	3526	400	3538
4	71.5	62	1833	5111	1148	5160	757	5170
5	16.4	15	2734	6885	2070	6991	1408	7022
6	3.2	3	1133	2667	883	2713	638	2730
7	1.0	1	428	984	370	1005	286	1011
8	0.0	0	658	1445	562	1472	451	1477

Table 6. Number of incorrect responses in each frequency class.

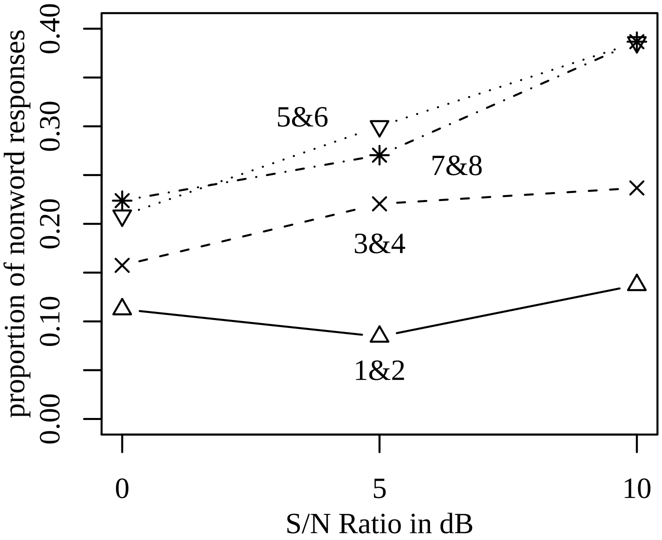

Figure 11. Proportion of nonword responses as a function of S/N ratio and word frequency class. Compare to Fig. 1 in Pollack et al. (1960)

maximum of nearly 40% nonword responses for words in frequency classes 5–8 at a S/N ratio of 10 dB. This is likely due to the fact that we have included multisyllabic words, while Pollack et al. used only monosyllabic words.

In addition, Pollack et al. report that they found significant effects for S/N ratio, but not for frequency class. Results from a logistic regression show significant effects for frequency class and S/N ratio, though no significant effect of interaction. As S/N ratio increases, the proportion of nonword responses increases. This may seem counterintuitive at first, but it is logical given that some of the target words are quite rare and long. When the level of the background noise is high, listeners are more likely to guess a real word which has the same general acoustic shape (especially the number of syllables) as the target word. In contrast, when the signal is fairly clear, listeners are more likely to respond based on the acoustic input, though they may make minor misperceptions. The proportion of nonword responses decreases with increasing frequency of the target word. This is also logical, since high frequency words tend to have high frequency neighbors, and since word frequency is highly correlated with word length. For example, given a CVC target word, there is a 31% chance that any randomly selected, phonotactically legal CVC sequence would be an actual English word ($\frac{2180 CVC English words in CELEX}{22 initial C * 15 V * 21 final C}$). This is in stark contrast to even CVCVC words, in which there is only a 3% chance that a random sequence would be a real word ($\frac{9857 CVC English words in CELEX}{22 initial C * 15 V * 21 final C * 2 V * 21 final C}$).[4]

	Estimate	Std. Error	z value	*p*
(Intercept)	1.3325	0.0498	26.78	<0.0001
S/N ratio	−0.0705	0.0084	−8.43	<0.0001
frequency Class	0.0002	0.0001	3.12	<0.01
S/N ratio:frequency Class	−0.0000	0.0000	−0.71	>0.1

Table 7. Logistic regression results comparing effects of S/N ratio and frequency class on the proportion of nonword responses.

Discussion

In this study we have tested various assumptions about the nature of incorrect responses which have been drawn largely on the basis of studies of monosyllabic words (e.g. Pollack et al., 1960). While some of these assumptions have been borne out, we have also found several differences.

Consistent with previous studies such as Wiener & Miller (1946), we found that longer words are perceived more accurately than shorter words, though we found that accuracy increased linearly with word length, as opposed to logarithmically. We also found that errors tend to be slightly higher in frequency than target words, and are significantly correlated, contrary to Pollack et al. (1960). In addition, our results indicate that as word length increases, errors become increasingly different from the targets (as estimated by edit distance), but that the difference in phonemes and syllables between the response and the target only increases slightly. Also, listeners are more likely to make responses which drop phonemes than predicted by chance, but are less likely to make responses which drop syllables. Finally, we find that listeners are more likely to provide nonword responses as S/N ratio increases, and as target word frequency decreases.

The effects we have found paint a somewhat modified view of the process of word recognition compared to previous models. Pollack et al. (1960) picture word recognition as akin to choosing balls from an urn, in which each ball represents word tokens. That is, there will be 779 balls representing *cat* out of a total of 17.9 million balls (the total word count in the COBUILD corpus used by the CELEX database). In this model, acoustic input effectively decreases the size of the urn, including only balls that are reasonable acoustic matches for the input. Pollack et al. claim that this model predicts that word frequency of the incorrect responses will be independent of the stimulus word frequency. We believe that this erroneous conclusion stems from the fact that they only used monosyllabic stimuli. Given that acoustic input narrows the set of candidates to phonetically similar words, and that word frequency is correlated with word length, the set of candidates should have word frequencies similar to the stimulus word, which is what we have found. Moreover, the result that 74% of incorrect responses had the same number of syllables as the target word suggests that the narrowing of candidates may very well start on a syllabic basis, and then be narrowed further on a phoneme and perhaps a featural level. This is also consistent with a view that the candidate pool is not constructed in a completely linear fashion as in some models (e.g. Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Zwitserlood, 1989).

In conclusion, our results underscore the fact that future models of word recognition should be

[4] This assumes that one of the two vowels in the word is a reduced vowel (either /ɪ/ or /ə/

designed to explain variance for a representative sample of the lexicon, and that the assumption that models based primarily on monosyllabic data do not hold for multisyllabic words.

References

Baayen, H. R., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database (cd-rom). Philadelphia: Linguistics Data Consortium, University of Pennsylvania.

Benkí, J. (2003). Quantitative evaluation of lexical status, word frequency and neighborhood density as context effects in spoken word recognition. *Journal of the Acoustical Society of America*, 113(3), 1689–1705.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the connected speech test (cst). *Ear And Hearing*, 8, 119S–126S.

Fletcher, H. & Steinberg, J. C. (1929). Articulation testing methods. *Bell Systems Technical Journal*, 7, 806–854.

Frisch, S., Broe, M., & Pierrehumbert, J. (1997). Similarity and phonotactics in arabic. URL `http://roa.rutgers.edu/view.php3?roa=223`.

Hirsh, I., Reynolds, E. G., & Joseph, M. (1954). Intelligibility of different speech materials. *Journal of the Acoustical Society of America*, 26(4), 530–538.

House, A. S., Williams, C. E., Hecker, M. H. L., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *The Journal of the Acoustical Society of America*, 37(1), 158–166.

Kawai, H. & Higuchi, N. (1998). Recognition of connected digit speech in japanese collected over the telephone network. In *Proceedings of the International Conference on Speech and Language Processing*.

Luce, P. (1986). *Neighborhoods of words in the mental lexicon*. Ph.D. thesis, Indiana University.

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

Marslen-Wilson, W. & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology*, 15(3), 576–585.

Marslen-Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1–71.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report 10*, Speech Research Laboratory, Psychology Department, Indiana University, Bloomington.

Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *Journal of the Acoustical Society of America*, 32(4), 454–457.

Siohan, O., Ramabhadran, B., & Zweig, G. (2004). Speech recognition error analysis on the english malach corpus. In *Proceedings of the International Conference on Speech and Language Processing*.

Wiener, F. & Miller, G. (1946). Some characteristics of human speech. *Transmission and Reception of Sounds under Combat Conditions*, 3, 58–68.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Working Memory Training and Implicit Learning[1]

Althea Bauernschmidt[2], Christopher M. Conway[3] and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This study was supported by NIH-NIDCD Research Grant DC00111 and NIH-NIDCD Training Grant DC00012.. We wish to thank Luis Hernandez for his help with technical aspects of the project as well as providing numerous constructive discussions.

[2] Department of Psychological Sciences, Purdue University, West Lafayette, IN

[3] Department of Psychology, Saint Louis University, St. Louis, MO

Working Memory Training and Implicit Learning

Abstract. Recent studies have shown that working memory can be improved through training and this improvement generalizes to other cognitive measures. Working memory studies typically focus on the retention of random sequences; however, much of what working memory is used for is not random. This study investigated what effect probabilistic structure has on adaptive training of working memory and how differing levels of structure affect generalization to other cognitive tasks. Participants received four days of working memory training that was either adaptive or non adaptive with sequences that were constrained or probabilistic. A battery of cognitive measures was taken at pre- and post- test. Probabilistic structure provided beneficial effects on adaptive training of working memory in as little as four hours of training. Moreover, these benefits carried over to improvements on measures of non-verbal reasoning and executive function.

Introduction

The concept of "mental exercise" has become more prominent in popular culture and recent media attention has been placed on memory training and improvement. Companies like Nintendo have begun to market games like Brain Age 2 to "help stimulate your brain and give it the workout it needs". Memory training has received attention within the scientific community as well with the recent findings that working memory canbe improved through adaptive training (Klingberg et al., 2002; Klingberg et al., 2005; Verhaeghen et al., 2004).

However, these tasks typically involve the retention of random sequences while much of what working memory is used for in real-world settings is not random. A more accurate view might be that working memory is most commonly used for the retention and manipulation of complex, probabilistic information. For example, working memory capacity is critical to the ability to carry on a conversation or read a text. Both rely on the ability to hold information recently received and to update it with new information. This information, language that is spoken or written, is not random. It is highly complex and probabilistic and requires language users to make constant use of implicit knowledge concerning syntax or semantics. If the final word in the sentence "They tracked the lion to his ___" was covered by noise or smudged on a paper it would still be possible, even easy, to figure out what the last word was. Based on knowledge about the syntax of the sentence, the final word in this case can be predicted to be a noun. In this way people are able to carry on conversations despite less than perfect auditory conditions, like talking over the phone or in a crowded room.

The present experiment seeks to further understand the relationship between working memory and implicit learning. Specifically, to determine what effect probabilistic structure has on the adaptive training of working memory and how differing levels of probabilistic structure affect generalization and transfer to other cognitive tasks

A Brief History of the Study of Immediate Memory

Different temporal forms of memory were dissociated at early as 1890 in William James's monumental work *Principles of Psychology*. James posited a dissociation between primary and secondary memory. Primary memory is so called because of the reliance on it to hold items in consciousness at the present time; whereas secondary memory is the unlimited long-term store of information.

Broadbent (1958) later combined the idea of primary memory with the computer metaphor of memory, resulting in the concept of short-term and long-term memory. Broadbent's conceptualization of primary memory had several tenets that were included in many subsequent models; that primary and secondary memory are two separate memory systems, the primary memory has a limited capacity, that information fades quickly from primary memory and must therefore be rehearsed to retain it.

The most popular model of short-term memory is that of Atkinson and Shiffrin (1968). The Atkinson and Shiffrin model treats memory as a somewhat linear process within information flowing from sensory memory through short-term memory and finally into long-term memory. In this model, the information that passes from sensory memory to short-term memory is modulated by attention. In turn, information is transferred into long-term memory from short-term memory by a process of rehearsal and maintenance. Information can be recalled from long-term memory into short-term memory by a process of retrieval.

A distinction that is present both within the primary v. secondary model and the short-term v. long-term model is capacity. Short-term, or primary, memory is thought to have a limited capacity, while long-term, or secondary, memory is thought to have a limitless capacity. Miller (1956) famously theorized that the capacity of short-term memory is 7 plus or minus 2 items. Miller compared data from a variety of different paradigms and concluded that people's ability to process information is limited by the amount of information they are able to hold in consciousness.

Working Memory

Working memory is the ability to retain and manipulate information in memory during a short period of time (Miller, Galanter, & Pribram, 1960). Though it can be used interchangeably with short term memory, the two are distinguished by the degree to which information is manipulated. Short term memory refers to the more passive store of information in immediate memory. Working memory refers to the more active store of information in immediate memory. A task that is frequently used to measure both short term memory and working memory is digit span; forward digit span is a measure of short term memory, while backward digit span is a measure of working memory. In a forward digit span task a participant hears a list of digits and asked to simply repeat the sequence aloud. In a backward digit span task a participant hears a list of digits and is asked to repeat the list in the reverse order. The simple manipulation in the backward digit span task highlights the difference between the ability to remember information in the short term and the ability to remember and manipulate that information.

The most prominent model of working memory is that of Baddeley and colleagues (Baddeley, 1986, 2000: Baddeley & Hitch, 1974; Baddeley & Logie, 1999). Within Baddeley's working memory model are the central executive, the visuo-spatial sketch pad, phonological loop, and episodic buffer. The central executive coordinates activities in the phonological loop, the episodic buffer, and the visuo-spatial sketchpad. The phonological loop and the visuo-spatial sketchpad are domain specific components that are responsible for auditory and visual information respectively. The episodic buffer is a recent addition (Baddeley, 2000) that links information across domains and is linked to long-term and semantic memory.

Several other cognitive abilities, such as attention and inhibition, are thought to interact with working memory. A classic example of the role of attention and inhibition is the cocktail party phenomenon (Moray, 1959). The cocktail party phenomenon refers to the ability to attend to only part of a noisy environment, such as hearing one's name in a noisy cocktail party. Conway, Cowan, and Bunting (2001) found that people with lower working memory capacity were more subject to the cocktail party phenomenon, suggesting that they have difficulty inhibiting distracting information.

Working memory is thought to play a role in a language processing (Baddeley, 2003; Baddeley & Gathercole, 1993; Daneman & Merikle, 1996). The connection between working memory and language processing is strongly supported by evidence for the phonological loop within working memory. Within the phonological loop there are two main components: the phonological store and the articulatory control process. The phonological store is a temporary memory store for speech-based information. The articulatory control process is responsible for translating visual information into a speech-based code (e.g. reading) and deposits it into the phonological store; and it refreshes items that are being held in the phonological store so that they can continue to be held in memory.

Such phenomena as phonological similarity effect, articulatory suppression, and word length effect all provide support for this sub-vocal rehearsal maintenance system within working memory. The phonological similarity effect is the finding that words that sound alike, that are more phonologically similar, are more difficult to correctly recall than words that do not sound alike. The phonological similarity effect happens for words that are presented auditorially or visually, providing evidence that the words are subvocally rehearsed (Baddeley, 1992).

Articulatory suppression refers to the inability of the articulatory control process to translate visual information into a speech-based code when it is being actively used. For example, if the word *the* is being repeatedly said aloud the phonological similarity effect does not happen for words that are presented visually, but the phonological similarity effect still occurs for words that are presented auditorially. Articulatory suppression has been explained as blocking covert rehearsal for words being transferred from visual to speech-like information by the articulatory control process (Peterson & Johnson, 1971).

The word length effect is the observation that the number of words that can be held in working memory decreases as the length of the word increases. A possible explanation of the word length effect is that longer words take longer to sub-vocally rehearse and therefore take up more capacity within the phonological loop (Baddeley, Thomson, & Buchanan, 1975).

Deficits in working memory ability have been found in deaf children. Campbell and Wright (1990) found that deaf children are susceptible to the word length effect, suggesting that they process input within the short-term memory store similarly to hearing children. Though deaf children use memory strategies that are similar to their hearing peers, they perform poorly on phonological memory tasks, particularly tasks that involve encoding and retrieval of sequential information (Banks, Gray, & Fyfe, 1990).

Deficits in working memory ability in deaf children exceed simple deficits in the phonological loop, which is to be expected given the somewhat modality specific nature of the phonological loop. Deaf children with cochlear implants have been found to have shorter memory spans than hearing children in a visual memory task in which the stimuli are presented sequentially (Cleary, Pisoni, & Geers, 2001).

There is also evidence that the visuo-spatial sketchpad may play a role in language comprehension (Phillips, Jarrold, Baddeley, Grant, & Karmiloff-Smith, 2004). Phillips et al. (2004) found that individuals with Williams syndrome showed impaired comprehension of spoken spatial terms.

Williams syndrome is typically marked by relatively strong language abilities an impoverished visual and spatial abilities. Phillips et al. concluded that the spatial difficulties experienced by individuals with Williams syndrome may constrain their language abilities in certain circumstances.

Working memory has been strongly linked to measures of intelligence and reasoning. Some have even argued that general reasoning ability is little more than working memory capacity (Kyllonen & Christal, 1990; Colom, Rebollo, Palacios, Espinosa & Kyllonen, 2004). Kyllonen & Christal (1990) performed a factor analysis on four different large studies (N=723, 412, 415, and 594) in which reasoning ability and working-memory capacity were assessed. They found consistently high estimates of the correlation between working-memory capacity and reasoning ability. Furthermore, it has been argued that individual differences in reasoning ability and intelligence reflect differences in working memory capacity. (Engle, Kane, & Tuholski, 1999; Conway, Kane, & Engle, 2002).

Working memory capacity is typically thought of as a fixed trait of an individual that remains constant throughout adult life. While the developmental course may change the capacity, healthy young adults are thought to have relatively consistent working memory ability (Cowan, 2001). However, recent studies have shown that working memory can be improved through training (Klingberg et al., 2002; Klingberg et al., 2005; Verhaeghen, et al., 2004). Klingerberg et al (2002) used a computerized training program as an intervention for children with ADHD. Subjects were trained using an adaptive staircase method that adjusted difficulty on a trial-by-trial basis. Subjects preformed the computerized training program for at least 20 minutes 4-6 days a week for 5 weeks. The training program consisted of 4 different working memory tasks. The training program led to significant improvements on both trained and non-trained WM tasks as well as parent and teacher reports of ADHD symptoms.

Verhaeghen et al. (2004) had participants complete a practice study that spanned over 10 1-hour sessions on a self-paced identity-judgment *n*-back test. An *n*-back test is a measure of working memory in which a sequence of stimuli are presented and subjects are asked to respond when the current stimulus matches a stimulus that was n-back from it. For example, in a 2-back task subjects are asked to respond when the current card matches one that was seen 2 cards previously. This task is said to employ working memory because it requires continual updating and manipulation of working memory in order to perform well. Over the 10 sessions reaction time improved dramatically and serial attention increased from 1 to 4 items, suggesting that working memory ability can be improved in a relatively short amount of time.

These and other studies of working memory typically focus on the fixed capacity of working memory ability. As such, experiments and stimuli have been designed with the goal of obtaining a “pure” estimate of capacity in mind. This has led to a focus on the use of random sequences, because they are thought to provide fewer confounds with the measure of working memory capacity. However, we believe this may be an oversight. A lot of what working memory is used for in real-world settings, such as language processing, is not random. A more accurate view of the world would be to say that it is governed by complex or probabilistic patterns.

Implicit Learning

Memory for complex or probabilistic patterns is usually discussed under the heading of implicit cognition. Implicit memory is distinguished from explicit memory by the level of explicit reference to the original study phase. A test of explicit memory makes explicit reference back to the original study phase. Recognition memory tests, for example, require subjects to first study a list of items and then to later identify whether a given item was on the original study list. In this type of task the test phase makes explicit reference back to the study phase. In a test of implicit memory, however, the test phase does not make explicit reference back to the study phase. In a word fragment completion task, for example, subjects study a list of words and later are asked to identify words that are presented in a degraded or altered form. Subjects tend to perform better on the words that were presented both in the study and in the testing phase, a phenomenon referred to as priming (Tulving & Schacter, 1990).

Similarly, explicit and implicit learning are distinguished by the degree of conscious or deliberate processes by which underlying complex structure is discovered. An example of an explicit learning task is the application of hypothesis-testing model to concept leaning (Bruner, Goodnow, & Austin, 1956). Participants were given pictures and asked to decide if it belonged in a category. They received feedback as to whether or not they were correct, but not as to what the rule determining the category was. An example of a rule used to determine category membership would be "If a flower is red and has three petals, it is a member; otherwise it is not." The rules used in this experiment were simple enough that participants could learn them after several trials and use them to assign category membership.

By contrast, implicit learning is the ability to learn complex or probabilistic patterns without conscious awareness. An example of implicit learning is the "weather prediction" task (Knowlton, Squire, & Gluck, 1994). In this task four cards were probabilistically associated with one of two outcomes, sunshine or rain. Participants were presented with one or more cards from the set and asked to determine what the weather would be. Similar to the Bruner et al. (1956) experiment, they were given feedback as to whether the response was correct or incorrect, but not why it was correct or incorrect. Unlike the Bruner, et al. experiment, however, the rules were probabilistic. In this task participants are able to improve their performance after many trials without being able to explicitly state the probabilistic rules governing their decision making.

Another type of implicit learning involves artificial grammars. An artificial grammar is a set of rules governing sequence production. Artificial grammars were first developed by Chomsky & Miller (1958) in an attempt to represent a small-scale model of linguistic rules. Miller (1958) tested whether grammar learning could involve the application of hypothesis formation testing, as in Bruner et al. (1956). To test this he presented participants with 9 letter strings, ranging from 4 to 7 letters in length. One list contained strings that were randomly generated, while another list contained strings that were generated by an artificial grammar. He found that participants learned the list of grammatical strings much more quickly than the list of random strings. He concluded that improved performance on the grammatical strings was due to redundancy within the sets, making it easier for participants to recode the grammatical sequences in to chunks and thus aid their performance.

Reber (1967) tested Miller's explicit encoding explanation and found that participants had no conscious or explicit awareness knowledge of the underlying grammar. He presented subjects with grammatical and non-grammatical strings and replicated Miller's finding that grammatical strings were easier to learn and remember than random strings. However, when he informed participants that the strings were rule governed and asked them to describe what they knew about the rules, he found that they were unable to verbally express explicit knowledge about the rules of the grammar.

To further test whether the rules of the grammar could be explicitly encoded Reber conducted a second experiment. There were two phases within the experiment, a learning phase and a test phase. In the learning phase, participants were asked to memorize 20 grammatical letter strings. The strings were presented in four sets of five strings and learning was said to have occurred when participants could correctly reproduce the set two consecutive times. Participants were not told that the strings were produced according to any rules or that there would be a testing phase later.

During the testing phase, the participants were informed that the 20 strings they had just learned were made according to a complex system of grammatical rules. They were then presented with a series of novel strings and asked to decide if these strings were made following the same rules as the ones they had just learned. Half of the strings that they received were randomly generated and half were novel grammatical strings. Participants were not given any feedback as to whether or not their responses were correct. Reber found that participants were able to classify the novel strings as grammatical or non-grammatical well above chance. Despite their ability to discern the grammatical from the non-grammatical, however, they were unable to verbally or explicitly express knowledge of the rules of the grammar when questioned after the experiment.

Few studies in the intervening decades have used immediate recall of grammatical sequences as a dependent measure. One exception is Karpicke and Pisoni (2004). They found that participants with implicit knowledge of an artificial grammar improved immediate memory span for novel sequences generated by that grammar. There were two phases, undifferentiated to the participants, an “acquisition phase” and a “test phase”. The task was the same throughout: to simply reproduce the sequence. Participants were presented with sequences of colors that lit up on a custom response box (modeled after the commercial game “Simon” manufactured by Milton Bradley) and asked to reproduce the sequence using the custom response box. Sequences presented in the acquisition phase were generated by a grammar. Half of the sequences presented in the test phase were novel sequences generated by the same grammar and the other half were generated by a different grammar.

Memory span was then obtained by adding the total number of items correctly reproduced on each perfectly recalled trial. Learning scores were calculated by subtracting the span for the strings generated by the untrained grammar from the span for the novel strings generated by the trained grammar in the test phase (trained-not trained). Overall, participants showed learning. However, a wide range of individual differences was reported. These individual differences covaried with measures of auditory digit span, suggesting that participants with greater immediate memory capacity are better able to learn and subsequently exploit the information available within the grammatical sequences. Furthermore, using a similar implicit learning task, Conway et al. (2007) found that participants’ learning scores were correlated with their ability to use context to aid their understanding of degraded sentences in a speech perception task.

Present Experiment

Findings that working memory can be improved through training (Klingberg et al., 2002; Klingberg et al., 2005; Verhaeghen et al., 2004) taken in conjunction with previous studies of implicit learning that have demonstrated beneficial effects of structure on recall (Miller, 1958; Karpicke & Pisoni, 2004) suggest that probabilistic structure should have beneficial effects on adaptive training of working memory. We therefore predict that training with probabilistic structure will not only improve working memory better than using random sequences, but will also lead to better generalization and transfer to other cognitive tasks.

Participants took part in a 6 day study in which they received a battery of cognitive tests on the first day, then received four days of working memory training, and received the same batter of cognitive tests on the sixth and final day. The type of working memory training that participants received was either adaptive training in which the sequences that were used were produced according to a constrained structure, adaptive training in which the sequences that were used were produced pseudo-randomly, or non-adaptive training in which the sequences that were used were produced pseudo-randomly.

Methods

Participants

31 participants, 7 males and 24 females, recruited through Indiana University (ages 18-33) participated in this study for monetary compensation.

Apparatus

The following were used in one or more tasks: A *Magic Touch* touch-sensitive monitor, Macintosh Power PC G4, a desk mounted Electric Voice dynamic cardiod model 664 microphone, a Marantz solid state recorder PMD 660, and Beyer dynamic DT100 headphones.

Materials and Procedures

Participants took part in the experiment for 6 days, with no more than 2 intervening days between sessions. On the first day they were assessed on a number of cognitive measures; an Implicit Learning task, a measure of spoken sentence perception, the Stroop Color and Word Test, Raven's Progressive Matrices, and measures of forwards and backwards digit span. During the next 4 sessions participants received a Working Memory Training program. In the final session they were re-assessed on the same measures as the first day.

Day 1 (Pre-Test)	Days 2-5 (Working Memory Training)	Day 6 (Post-Test)
• Spoken Sentence Perception • Stroop Color and Word Test • Forwards Digit Span • Backwards Digit Span • Raven's Standard Progressive Matrices • Implicit Learning	Group 1: Adaptive, Constrained	• Spoken Sentence Perception • Stroop Color and Word Test • Forwards Digit Span • Backwards Digit Span • Raven's Standard Progressive Matrices • Implicit Learning
	Group 2: Adaptive, Pseudo-random	
	Group 3: Non-Adaptive, Pseudo-random	

Table 1. Overview of Procedure

Pre- and Post- Test Measures

Spoken Sentence Perception Task. Participants were presented with sentences under degraded listening conditions and asked to write down the last word that they heard in the sentence. Speech Perception in Noise sentences by Kalikow et al (1977) that were modified by Clopper and Pisoni (2006) were used. Sentences were spoken by a female and male speaker, life-time residents of the "midland" region of the United States, whose spoken recordings were chosen from amongst a set of recordings taken from multiple speakers developed as part of the "Nationwide Speech Project" (see Clopper & Pisoni, 2006). The sentences were then degraded by processing them with a sinewave vocoder (www.tigerspeech.com) that simulates listening conditions for a user of a cochlear implant with 6 spectral channels. The sentences vary in terms of the final word's predictability so that they can be of 3 types: high-predictability sentences, low-predictability sentences, and anomalous sentences. The sentences were played over Beyer Dynamic headphones.

Participants received a different set of 54 sentences, with mixed predictability and speakers, in pre and post test. Responses were scored on the number of correct target words for each sentence type; anomalous, low predictability, and high predictability.

Stroop Color and Word Test. In this task participants were asked to read three pages (Word, Color, and then Color-Word page) of 100 items aloud. Each page was presented in 5 columns of 20 items. The Word page consisted of the words "red", "green, and "blue" arranged randomly and printed in black ink on a white 8.5" x 11" page. No word followed itself within a column. The Color page consisted of 100 items written as XXXX, printed in either red, green, or blue ink. Again, no color was allowed to follow itself in a column or to match the corresponding item in the Word page. The Color-Word page consisted of the words from the Word page printed in the colors from the Colors page. The two pages were blended item for item: Item 1 from the Word page was printed in the color of the Item 1 from the Color page. (Golden and Freshwater, 2002)

The responses were scored on how many items in the list were said aloud in 45 seconds. Participants were instructed to start again at the beginning if they reached the end of the 100 item list before the 45 seconds were done. Raw scores were then converted to T-scores based on Age and Education (Golden and Freshwater, 2002).

Measure of Digit Span. Measures of digit span were based on the digit span task within the WISC-III (Wechsler, 1991). In the Forward digit span task subjects were presented with lists of digits at progressive lengths and asked to repeat the sequence aloud. In the Backwards digit span task subjects were presented with lists of digits at progressive lengths and asked to repeat the sequence *in reverse order* aloud. For example, if they heard "1, 8, 3" then they were to say aloud "3, 8, 1". Lists of digits ranging in length from 2 to 10 from the WISC (Wechsler, 1991) were used. Digits were played over Beyer Dynamic headphones and recorded by a desk-mounted microphone. Subjects were scored on the longest sequence that they correctly recalled in each digit span task.

Raven Standard Progressive Matrices. In this task participants were asked to identify which of the given pictures will best complete the larger pattern in the matrix. The difficulty of the test item increases as the test goes on, so that each of the 5 subsets is progressively more difficult than the last. Subjects received either the odd half or the even half of a 60 item set taken from Raven's Standard Progressive Matrices (Raven, Raven, & Court, 2000). Responses were scored on total number correct.

Implicit Learning Task. In this task subjects saw a series of colored squares light up on the screen and were asked to reproduce the sequence that they had just seen. The red, blue, green, and yellow squares were presented one at a time and appeared in the upper left, upper right, lower left, and lower right quadrants of the screen. The mapping of color to screen location was randomly determined for each participant, as was the mapping between the four sequence elements (1-4) to each of the four quadrants/colors. For each subject, the mapping remained consistent across all trials. The sequences were presented in one continuous run so that the learning phase and the test phase were undistinguished to the participant. In the learning phase, the 48 learning sequences were presented once each, in random order. In the test phase, the 20 novel constrained and 20 unconstrained test sequences were presented in random order, once each.

The stimuli were presented using the *Magic Touch* touch-sensitive monitor and the Macintosh Power PC G4 and responses were made on the *Magic Touch* touch-sensitive monitor. The grammars and procedures used in this task were similar to Conway and Pisoni (2007). Two artificial grammars were used to generate the stimuli (c.f., Jamieson & Mewhort, 2005). These grammars, as shown in Table 2, specify the probability of a particular element occurring given the preceding element. The grammar on the left was used to create constrained sequences whereas the control grammar on the right was used to create pseudorandom (unconstrained) sequences. For each sequence, the starting element (1-4) was randomly determined and then the probabilities were used to determine each subsequent element, until the desired length is reached.

The constrained grammar was used to generate 48 unique sequences for the learning phase and 20 sequences for the test phase. The control grammar was used to generate twenty sequences for the test phase as well.

A learning score was obtained by subtracting the span score for the ungrammatical sequences in the test phase from the grammatical sequences in the test phase.

	Constrained grammar (n+1)				Control grammar (n+1)			
Colors/locations (n)	1	2	3	4	1	2	3	4
1	0.0	0.5	0.5	0.0	0.0	0.33	0.33	0.33
2	0.0	0.0	1.0	0.0	0.33	0.0	0.33	0.33
3	0.5	0.0	0.0	0.5	0.33	0.33	0.0	0.33
4	1.0	0.0	0.0	0.0	0.33	0.33	0.33	0.0

Table 2. Constrained and Control Grammars used in Implicit Learning Task
Grammars show transition probabilities from position n of a sequence to position n+1 of a sequence for four colors labeled 1-4.

Working Memory Training

Overview. In this task participants saw a series of green circles light up and were asked to reproduce the sequence they had just seen (Figure 1).

Materials. The stimuli were presented using the *Magic Touch* touch-sensitive monitor and the Macintosh Power PC G4 and responses were made on the *Magic Touch* touch-sensitive monitor. The circles were arranged in a 4x4 grid, inspired by a computerized training program used by Klingberg et al (2002).

The order that the circles lit up in was either constrained or pseudo-random. Sequences that were constrained were generated so that each element in the sequence could be followed by 3 others in the set with equal likelihood. The pseudo-random sequences were generated so that each element in the sequence could be followed by any other in the set with equal likelihood. Sequence lengths ranged from 4 to 16. 200 unique lists were produced at each length.

Each element in the sequence can be represented by a number, so that in the constrained sequences the 1st element could be followed with equal probability by elements 2, 3, and 4. In turn, 2 can be followed with equal probability by 3, 4, and 5, and so on. The elements in these sequences were randomly mapped onto 1 of the 16 circle locations for each subject. The mapping was consistent for all blocks and sessions for each subject. So, while the underlying sequences were the same for each subject, the spatial representation was different between subjects.

Procedure. Subjects were randomly assigned to 1 of 3 groups; constrained sequences at adaptive lengths, pseudo-random sequences at adaptive lengths, or pseudo-random sequences at random lengths.

> **Constrained sequences at Adaptive Lengths**: In this condition participants received constrained sequences (as described above) at adaptive lengths. Sequence lengths in the adaptive conditions were based on a 2 up 2 down metric. For example, if a subject starts at sequence length 4 and correctly reproduces all items in that sequence then their next trial will be a sequence of length 4. If the subject correctly reproduces all elements in the second sequence of length 4 then they will move up to a sequence of length 5 in the next trial. If they *incorrectly* reproduce this sequence of length 5 then their next trial will be of a sequence of length 5. If they respond incorrectly to this sequence as well, then their next sequence will be moved down to length 4.
>
> **Pseudo-Random Sequences at Adaptive Lengths:** In this condition participants received pseudo-random sequences (as described above) at adaptive lengths. This condition used the same 2 up 2 down metric as previously described, and differed from the constrained sequences at adaptive lengths condition only in the type of sequences it used.
>
> **Pseudo-Random Sequences at Non-Adaptive Lengths:** In this condition participants received pseudo-random sequences (as described above) at non-adaptive lengths. Participants received sequences varying in length from 4 to 16 elements. The sequence length was randomly determined at each trail.

The sequences were presented as a persistent 4x4 grid of dark green circles in which individual circles lit up for 250ms and were off for 250ms between elements in the sequence. At the end of the presentation of a sequence if the participant did not make a response within 2 seconds a new sequence was presented. As participants made their responses, the circle that they pressed lit up for 250ms.

Each Working Memory Training session consisted of 3 blocks of 50 trials. Subjects received feedback at the end of each block as to what their highest correctly reproduced sequence length was for that block as well as what their “High Score” was. The High Score was calculated by the total number of correct single button responses made in that block. For example, if a subject received a sequence at length 7 but only correctly reproduced 5 of 7 items they would gain 5 points onto their High Score.

Responses were scored by average length received per block in the adaptive conditions, as the length that they received was reflective of their ability. Responses in the non-adaptive condition were scored by the longest length that was correctly reproduced at least 50% of the time within each block. For example, if a participant received seven sequences at length 5 and correctly reproduced four of the seven, then their score for that block would be 5.

Results and Discussion

Working Memory Training

Overall, participants improved on the Working Memory Training task. As shown in Table 3, the mean performance increased from day to day within each condition. Within each day of training the Adaptive, Constrained condition preformed better than the Adaptive, Pseudo-Random condition which in turn was better than the Non-Adaptive, Pseudo-Random condition. A one-way ANOVA with the factor of condition showed these differences to be significant by the third day of training, F(2,25) = 4.48, *p* < .05, as well as the fourth day of training, F(2,25) = 5.83, *p* < .01.

Condition	Averages				Difference Day 1 - Day 4
	Day 1	Day 2	Day 3	Day 4	
Adaptive, Constrained	5.69	5.94	6.15	6.37	0.75**
Adaptive, Pseudo-Random	5.28	5.54	5.56	5.79	0.59**
Non-Adaptive, Pseudo-Random	4.88	5.03	5	5.17	0.44*

* = sig. to .05, ** = sig to .01

Table 3. Average Length Across Days.

When analyzed by factor of adaptive or non-adaptive, a one-way ANOVA shows that the adaptive conditions, Adaptive, Constrained and Adaptive, Pseudo-Random, significantly out perform the non-adaptive condition within each day of training. When analyzed by factor of structure, a one-way ANOVA shows that the constrained condition out performs the pseudo-random conditions. These results are summarized in (Table 4). It can bee seen by the significance values reported on each day, that the differences between these groups becomes more distinct by the fourth day of training, with a significance of p < .01.

However, a one-way ANOVA revealed that there were no significant differences in overall improvement from day 1 of training to day 4 of training between groups, *F*(2,24) = .773, *p* = .47.

	Adaptive vs. Non-Adaptive			Constrained vs. Pseudo-Random		
	Means		F	Means		F
	Adaptive	Non-Adaptive		Constrained	Pseudo-Random	
Day 1	5.48	4.88	2.31*	5.69	5.08	4.47*
Day 2	5.74	5.03	5.33*	5.94	5.28	4.59*
Day 3	5.85	5	6.59*	6.15	5.28	6.83*
Day 4	6.08	5.17	8.98*	6.37	5.48	8.11**

* = sig. to .05, ** = sig. to .01

Table 4. Differences within Training Days

Within each condition there were considerable individual differences. However, a Binomial test (under the assumption that improvement could be obtained by chance) revealed that participants improved significantly in the Adaptive, Constrained condition, *p* < .05, and in the Adaptive, Pseudo-

Random condition, *p* < .01. However, participants in the Non-Adaptive, Pseudo-Random condition did not perform above chance, *p*= 1.00. This suggests that, despite individual differences, adaptive training reliably leads to improvement in memory span, while non-adaptive training does not.

Taken as a whole, these results suggest that probabilistic structure has a beneficial effect on adaptive training of working memory. Performance is best for adaptive conditions both in terms of improvement across days as well as performance on individual days. Performance better constrained probabilistic structure is compared to pseudo-random structure within the adaptive conditions. Again, this trend holds both for improvement across days as well as performance on individual days. These results are somewhat striking given the amount of individual differences within each condition.

Generalizations to other cognitive tasks

Spoken Sentence Perception Task. A one-sample T test found that there were significant improvements on nearly all measures in the spoken sentence perception task, as summarized in Table 5. Performance within the anomalous sentences improved significantly from pre- to post- test in all conditions. Performance within the low-predictability sentences improved significantly from pre- to post-test as well. Within the high-predictability sentences, however, performance is the relatively the same in pre- and post-test in the Adaptive, Pseudo-Random and the Non-Adaptive, Pseudo-Random conditions. Though it seems that they in fact worsen in these two conditions, the difference is not significant.

The drop in performance from pre- to post-test is significant in the Adaptive, Constrained condition that all groups performed the similarly at post-test. The abnormally high performance on the pre-test for those within the Adaptive, Constrained conditions makes the drop in performance significant. As the table shows, performance was high in the Adaptive, Constrained condition at pre-test on all sentence types, making it difficult to draw conclusions about the effects of the Working Memory training on the spoken sentence perception task.

The improvements across the board in the anomalous and low predictability sentence types suggests task specific training effects as the participants become more familiar with the CI simulation.

Sentence Type	Working Memory Training Condition	Pre-Test	Post-Test	Difference
Anomalous	Adaptive, Constrained	10.66	13.74	3.07**
	Adaptive, Pseudo-Random	9	12.20	3.20**
	Non-Adaptive, Pseudo-Random	9.9	12.9	3**
Low Predictability	Adaptive, Constrained	10.38	13	2.62**
	Adaptive, Pseudo-Random	9.16	12.83	3.66*
	Non-Adaptive, Pseudo-Random	9.9	12.7	2.8*
High Predictability	Adaptive, Constrained	15.17	13.4	-1.77*
	Adaptive, Pseudo-Random	14.16	13.83	-0.33
	Non-Adaptive, Pseudo-Random	14	13.5	-0.5

* = sig. to .05, ** = sig. to .01

Table 5. Differences on Speech Perception in Noise Task.

Stroop Color and Word Test. A one-sample T test found that there were significant improvements on nearly all measures in the Stroop Color and Word task, as summarized in Table 6. As the table shows, there was significant improvement within the Word list for all the training groups. Within the Color list there was significant improvement for only the Adaptive, Constrained and the Non-Adaptive, Pseudo-Random condition. However, the improvement in the Adaptive, Pseudo-Random condition, 4.14, was similar to the improvement in the Adaptive, Constrained condition, 4.00. The lack of statistical significance for this improvement suggests that there were a range of individual differences.

As Table 6 shows, the Color-Word scores improve significantly in all conditions with the most improvement is seen in the adaptive conditions, with the constrained condition being better than the pseudo-random condition, following the same trend as seen within the Working Memory Training. A similar trend is seen within the Color-Word interference scores, with the most improvement is seen in the adaptive conditions, with the constrained condition being better than the pseudo-random condition. The improvement is only significant within the Adaptive, Pseudo-Random condition, however.

Stroop Measure	Working Memory Training Condition	Pre-Test	Post-Test	Difference
Word	Adaptive, Constrained	55.67	63.22	7.44*
	Adaptive, Pseudo-Random	52.00	57.43	7.71*
	Non-Adaptive, Pseudo-Random	52.10	59.80	7.70*
Color	Adaptive, Constrained	51.89	55.89	4.00*
	Adaptive, Pseudo-Random	48.67	54.71	4.14
	Non-Adaptive, Pseudo-Random	52.00	57.40	6.00*
Color-Word	Adaptive, Constrained	54.33	63.33	7.89*
	Adaptive, Pseudo-Random	54.22	62.14	6.57*
	Non-Adaptive, Pseudo-Random	55.20	60.50	4.30*
Color-Word Interference	Adaptive, Constrained	53.78	58.56	4.78
	Adaptive, Pseudo-Random	55.44	60.57	4.29*
	Non-Adaptive, Pseudo-Random	55.60	58.10	2.30

* = sig. to .05

Table 6. Differences on Stroop Sub-Scales

Digit Span. A One-Sample Test revealed that there were no significant differences on digit span performance. As Table 7 shows, there were only modest improvements in digit span performance. In the forward digit span task participants improved by only .30, *t*(9) = 1.4, *p* = .193, .28,*t*(6) = 1.00, *p* = .35, and .3, *t*(9) = 1.96, *p* = .081 in the Adaptive, Constrained, Adaptive, Pseudo-Random and Non-Adaptive, Pseudo-Random conditions respectively. In the backward digit span task participants improved by .30, *t*(9) = .75, *p* = .46, .14, *t*(6) = .35, *p* = .73, and .10, *t*(9) = .28, *p* = .78, in the Adaptive, Constrained, Adaptive, Pseudo-Random and Non-Adaptive, Pseudo-Random conditions respectively. However, an inspection of the pre- test values in Table 7 shows that the Adaptive, Constrained condition performed much better at pre-test than the other two conditions. This makes it difficult to attribute any improvement in performance or lack thereof solely to the Working Memory Training. Overall, no interesting results were found on digit span performance.

Digit Span Task	Working Memory Training Condition	Pre	Post	Difference
Forward	Adaptive, Constrained	7.5	7.8	0.3
	Adaptive, Pseudo-Random	6.22	6.57	0.28
	Non-Adaptive, Pseudo-Random	6.5	6.8	0.3
Backward	Adaptive, Constrained	6.1	6.4	0.3
	Adaptive, Pseudo-Random	5.22	5.71	0.14
	Non-Adaptive, Pseudo-Random	5.1	5.2	0.1

Table 7. Performance on Digit Span Task by Training Condition.

Raven Standard Progressive Matrices. Within Raven Standard Progressive Matrices participants preformed at ceiling in subsets A, B, and C at pre-test, as shown in Table 8. Ceiling performance on these subscales indicates that all participants were on task, as Raven's Standard Progressive Matrices are sets of progressively more difficult reasoning tasks. As D and E are the final two, and therefore most difficult, subsets it is expected that these are the two subsets that would be most sensitive to differences.

Excluding participants who were at ceiling in subsets D and E, improvements were found. In subset D participants improved in the Adaptive, Constrained condition, *t*(4) = 3.162, *p* < .05, the Adaptive, Pseudo-Random condition, *t*(2) = 4.00, *p* = .057, and the Non-Adaptive, Pseudo-Random condition, *t*(6) = 1.549, *p* = .172. In subset E participants improved in the Adaptive, Constrained condition, *t*(7) = 1.080, *p* = .316. No improvements were found in subset E for the Adaptive, Pseudo-Random and the Non-Adaptive, Pseudo-Random conditions.

Raven's Matrices Subset	Working Memory Training Condition	Pre	Post	Diff
Set A	Adaptive, Constrained	5.9	5.9	0
	Adaptive, Pseudo-Random	5.9	6	0
	Non-Adaptive, Pseudo-Random	5.9	6	0
Set B	Adaptive, Constrained	5.7	5.5	-0
	Adaptive, Pseudo-Random	5.7	6	0
	Non-Adaptive, Pseudo-Random	5.8	5.7	-0
Set C	Adaptive, Constrained	5.4	5.5	0
	Adaptive, Pseudo-Random	5.7	5.57	-0
	Non-Adaptive, Pseudo-Random	5.5	5.6	0
Set D	Adaptive, Constrained	5.1	5.2	0
	Adaptive, Pseudo-Random	5.4	6	1
	Non-Adaptive, Pseudo-Random	5.2	5.4	0
Set E	Adaptive, Constrained	3.8	4.2	0
	Adaptive, Pseudo-Random	4.6	4.43	-0
	Non-Adaptive, Pseudo-Random	4.3	4.2	-0
Set D: Corrected	Adaptive, Constrained	4.2	5.2	1
	Adaptive, Pseudo-Random	4.7	6	1
	Non-Adaptive, Pseudo-Random	4.9	5.14	0
Set E: Corrected	Adaptive, Constrained	3.3	3.75	1
	Adaptive, Pseudo-Random	4.3	4.33	0
	Non-Adaptive, Pseudo-Random	3.9	3.88	0

Table 8. Performance on Raven's Standard Progressive Matrices by Training Condition.

When results were collapsed across the adaptive conditions, the improvement in subset D was significant for adaptive conditions, $t(7) = 4.965$, $p < .01$, as shown in Figure 4.

Implicit Learning Task. Differences within the Implicit Learning task are shown in Table 8. As shown in the table, there was a range of performance at pre-test between the different conditions. The Adaptive, Pseudo-Random condition specifically, has a much higher Grammatical span than the other two conditions at pre-test. As this differences was found before training it makes it difficult to compare the affects of the Adaptive, Pseudo-Random training to the affects of the other two conditions.

An inspection of the table shows that within the Grammatical span the Adaptive, Constrained and the Non-Adaptive, Pseudo-Random conditions have much more comparable pre-test means (86.5 and 86.7, respectively), making it easier to attribute the difference in performance at post-test to the differences in Working Memory Training Condition. A One-Sample Test showed that only the improvement within the Grammatical span by the Adaptive, Constrained group was significant. This improvement in performance is notable because it is only occurring in the group that received the Constrained, or grammatical, sequences. This shows that there was a generalization from one grammar-learning task to another.

Within the Ungrammatical span there is a considerable range of pre-test performance between the groups, as well. Again, this makes it difficult to attribute differences to the differing performance at post-test to the intervening Working Memory Training.

Implicit Learning Measure	Working Memory Training Condition	Pre-Test	Post-Test	Difference
Grammatical	Adaptive, Constrained	86.5	98.8	13.8*
	Adaptive, Pseudo-Random	94.5	93.8	-2.33
	Non-Adaptive, Pseudo-Random	86.7	88.2	2.5
Ungrammatical	Adaptive, Constrained	61.4	70.5	9.1
	Adaptive, Pseudo-Random	78	61	-16
	Non-Adaptive, Pseudo-Random	72.8	67	-8

General Discussion

Within the Working Memory Training task performance improved in every condition. There was greater improvement in the adaptive conditions as compared to the non-adaptive condition and within the adaptive conditions there was greater improvement still in the constrained condition as compared to the pseudo-random condition. These results suggest that probabilistic structure has beneficial effects on the adaptive training of working memory.

Given that the differences on daily performance between each condition, particularly the Adaptive, Constrained and the Non-Adaptive, Pseudo-Random conditions, became more statistically pronounced as training progressed, it can be assumed that further training would yield a significant difference in overall improvement between conditions. These results are particularly striking given the amount of individual differences within the task and suggest that given a larger sample size, these differences might become statistically significant between all groups.

These improvements, and this pattern of improvement, generalized and transferred to several cognitive measures. Significant improvements were found on the spoken sentence perception task, the Stroop Color and Word Test, Raven’s Standard Progressive Matrices, and the implicit learning task. Improvement on Raven’s Progressive Matrices and the Stroop Color and Word Test replicates previous findings of transfer of improvement on working memory training to these tasks (Klingberg et al., 2002). Improvement on the implicit learning task, particularly the improvement in Grammatical span within the Adaptive, Constrained condition, are particularly interesting because it shows generalization from one implicit learning task to another.

The beneficial effects of probabilistic structure on the transfer and generalization to these tasks has implications for using working memory training as an intervention, as Klingberg et al. (2002, 2005) did for children with ADHD. Using probabilistic sequences capitalizes on the observation that everyday life is not random. The results of this study indicate that using probabilistic sequences may be more ecologically valid when studying working memory, especially if working memory training is to be used as an intervention with clinical populations such as children with ADHD or deaf children with cochlear implants.

It should be noted that Klingberg et al. (2002, 2005) improvements in working memory over roughly 25 hours of training across several weeks, while the improvements reported here were obtained with under 4 hours of training across several days. This is shorter even than the findings of Verhaegan et al., (2004) who found improvements with 10 hours of training. Moreover, differences within training became apparent as early as the third day of training.

References

Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon.

Baddeley, A. (1992). Is working memory working? *Quarterly Journal of Experimental Psychology, 44A*, 1-31.

Baddeley, A. (2000). The episodic buffer: A new component to working memory?. *Trends in Cognitive Science, 4(11),* 417-423.

Baddeley, A., & Gathercole, S. (1993). Working memory and language. *Essays in cognitive psychology*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Baddeley, A., & Hitch G., (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol.8). New York: Academic Press.

Baddeley, A., & Logie, R. (1999). *Working memory: The multiple-component model*. New York: Cambridge University Press

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

Bruner, J., Goodnow, J., & Austin, G., (1956). *A study of thinking*. New York: Wiley.

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

Chomsky, N., & Miller, G. A., (1958), Finite state languages. *Information and Control, 1*, 91-112.

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.

Colom, R., Rebollo, I., Palacios, A. Juan-Espinosa, M., & Kyllonen P. (2004). Working memory is (almost) perfectly predicted by g. *Intelligenc, 32(3),* 277-296.

Conway, A., Cowan, N., & Bunting, M. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8(2)*, 331-335.

Conway, A., Kane, M., & Engle, R. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7(12),* 547-552.

Conway, C., Karpicke, J., & Pisoni, D. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education*, 12(3), 317-334.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Golden, C., & Freshwater, S. (2002). The Stroop Color and Word Test. Wood Dale, IL: Stoelting Co.

Daneman, M., Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3(4),* 422-433.

Engle, R., Kane, M., & Tuholski, S. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. *Models of working memory: Mechanisms of active maintenance and executive control.* 102-134. Miyake, Akira; Shah, Priti. New York: Cambridge University Press

Kalikow, D., Stevens, K., & Elliot, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acourstical Society of America*, 61, 1337-1251.

Karpicke, J. D. & Pisoni, D. (2004). Using immediate memory span to measure implicit learning. *Memory & Cognition*, 32(6), 956-964.

Klingberg, T., Forssberg, H., Westerberg, H., (2002). Training of working memory in children with ADHD. *Journal of Clinical Experimental Neuropsychology*, *24*, 781-791.

Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C., Forssberg, H., Westerberg, H. (2005). Computerized training of working memory in children with ADHD – A randomized, controlled trial. *Journal of the American Academy of Child Adolescent Psychiatry*, 44(2), 177-186.

Knowlton, B., L. Squire, and M. Gluck. 1994. Probabilistic classification learning in amnesia. *Learn. Mem.* **1:** 106-120.

Kyllonen, P. & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14(4),* 389-433.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology*, 56, 485-491.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.* New York: Henry Holt and Co.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56-60.

Peterson, L., & Johnson, S. (1971). Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, *10*, 346-354.

Phillips, C., Jarrold, C., Baddeley, A., Grant, J., & Karmiloff-Smith, A. (2004). Comprehension of spatial language terms in Williams syndrome: Evidence for an interaction between domains of strength and weakness. *Cortex, 40(1),* 85-101.

Raven, J., Rave, J. C., & Court, J. (2000). Standard Progressive Matrices. San Antonio, TX: Harcourt Assessment.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855-863.

Tulving, E., & Schacter, D. (1990). Priming an dhuman memory systems. *Science, 247,* 301-306.

Verhaegan, P., Cerella, J., & Chandramallika, B. (2004). A working memory workout: How to expand the focus of attention from on to four items in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(6), 1322-1337.

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

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Spoken Word Recognition Deficits in Mild Cognitive Impairment: Some Preliminary Findings Using a Sentence Repetition Task[1]

Lauren M. Grove, Vanessa Taler[2], and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, IN 47405

[1] This research was supported by funds from NIDCD Research Grant DC-00111 and NIDCD T32 Training Grant DC-00012. We would also like to thank Luis Hernandez for his technical advice.

[2] Center for Neuroimaging, Department of Radiology, Indiana University School of Medicine, 950 W. Walnut St., Indianapolis, IN 46202

Spoken Word Recognition Deficits in Mild Cognitive Impairment: Some Preliminary Findings Using a Sentence Repetition Task

Abstract. Word frequency and neighborhood density (ND) are known to be reliable predictors of spoken word recognition. Mild cognitive impairment (MCI) has recently been defined as a stage between normal aging and dementia; even in individuals with MCI, language has been found to typically be altered. The current study examined frequency and ND effects on spoken word recognition in MCI and compared the results to healthy controls. Four MCI patients (mean age: 76.75 ± 6.24), nine young adults (mean age: 24.11 ± 6.01), and four healthy elderly adults (mean age: 68.75 ± 10.18) completed a lexical discrimination task measuring the accuracy of word recognition in sentences masked by multi-talker babble (-3 SNR). All participants also completed a battery of neuropsychological tests. Young adults displayed the highest accuracy on the lexical discrimination task while the healthy elderly adults outperformed MCI patients. In all participants, a word frequency effect was observed, and accuracy on high-ND words was slightly lower than accuracy on low-ND words. The results of this preliminary study suggest that aging and cognitive impairment leads to a decline in spoken word recognition, i.e., the cognitive abilities used to recognize words in sentences.

Introduction

Spoken language is unique to humans, and the underlying components that allow us to understand spoken language must be investigated if we are to understand why some individuals have trouble understanding spoken language. Lexical discrimination refers to the process of distinguishing a stimulus word from other phonologically similar words (Luce & Pisoni, 1998). The current study attempts to understand declines in lexical discrimination abilities in healthy aging adults and mild cognitive impairment (MCI) patients. Three groups of participants, healthy young controls, healthy elderly controls, and MCI patients, were compared on an experimental task assessing lexical discrimination abilities with sentence materials.

Spoken Word Recognition

Spoken word recognition is important for humans to communicate successfully. Investigations of the mental lexicon, humans' psychological knowledge of words, have found that the frequency of word use in the English language and the number of similar neighbors, or neighborhood density (ND), play an important role in spoken word recognition when background noise is present (Luce & Pisoni, 1998). Luce and Pisoni defined lexical discrimination as the process of correctly identifying words in the mental lexicon to match the phonological input of a stimulus.

Many earlier studies have demonstrated processing advantages for high-frequency words over low-frequency words (Broadbent, 1967; Catlin, 1969; Goldiamond & Hawkins, 1958; Newbigging, 1961; Pollack, Rubenstein & Decker, 1960; Savin, 1963; Soloman & Postman, 1952; Triesman, 1971, 1978a, b). The criterion-bias theory states that the cause of the word-frequency effect is a response bias, which is a tendency to respond with common words rather than rare ones (Broadbent, 1967). Savin (1963) found that when participants gave incorrect responses, they were likely to give the same incorrect responses as one another. The number of times that a word is encountered, that is its experienced frequency, does not solely account for the word frequency effect. The ND of a word and its similarity to other words combines with the word frequency effect in spoken word recognition tasks (Luce & Pisoni, 1998).

Luce and Pisoni (1998) developed the neighborhood activation model (NAM), which states that a word is recognized relationally in the context of other words in long-term memory and that, when a stimulus is recognized, the lexical neighborhood containing similar-sounding words is activated. Luce and Pisoni defined the neighborhood of a target word (i.e. the word being presented to the participant) as the set of words that differ from it by the addition, deletion, or substitution of a single phoneme. The number of words occurring in a target word's neighborhood is referred to as the neighborhood density of the target word. NAM predicts that words with a high ND are more difficult to process (perceive and recognize) than words with a low ND. The phonetic similarity in neighboring competing words makes them easily confusable between one another.

Neighborhood competition has been shown to play a major role in spoken word recognition (Vitevich & Luce, 1998; Goldinger et al, 1989; Luce and Pisoni, 1998; Dirks et al, 2001). Bell and Wilson (2001) found a processing advantage for low ND words when they were presented in sentences under both degraded and quiet conditions. Metsala (1997) found that both adults and children showed higher error rates for words with many lexical neighbors. Sommers (1996) found elderly adults to be at a disadvantage when recognizing lexically hard words relative to young adults. These findings suggest that age differences among adults could affect lexical discrimination in sentences. However, a recent study conducted in our lab found no significant differences in lexical competition perception or production among healthy young adults and healthy elderly adults (Taler, Aaron, & Pisoni, submitted). Taler et al.'s findings are consistent with previous studies showing that language production is similar across the lifespan (Newman & German, 2005; Vitevitch & Sommers, 2003). Other previous studies found that elderly adults do not do as well as young adults with the perception of high ND isolated words, which is likely due to inhibitory decline (Sommers, 1996; Sommers & Danielson, 1999). However, there has been little research examining lexical competition deficits in patients with age-related disorders, such as MCI. The present study was carried out to examine lexical discrimination in individuals with MCI. Below, we discuss MCI and the deficits associated with this cognitive decline.

Mild Cognitive Impairment (MCI)

MCI is a term used to describe the clinical cognitive state between that of healthy aging adults and adults with dementia. Petersen and colleagues (1999) were the first researchers to introduce formal criteria for the diagnosis of MCI. The formal criteria are as follows:

1. An individual's subjective complaint of memory decline
2. An objective impairment of memory ability
3. No depression or other major psychological disorders
4. No significant decline in daily living activities
5. Individual does not meet the criteria for dementia

Petersen (2003) suggests the etiology of MCI to be heterogeneous. This may be in part because of the wide range of annual conversion rates to Alzheimer's disease (AD), or that MCI may be an early stage of a different type of dementia other than AD. Different subtypes of MCI have been identified, either with or without memory impairment (i.e. amnestic MCI and non-amnestic MCI respectively) (Petersen, et al., 2001a).

The Mini Mental State Examination (MMSE) is a physicians' screening tool for separating individuals with normal cognitive function from those with dementia, and clinicians commonly use this screening tool. A score of 24 or below (out of 30) is considered a sign that the patient is demented (Folstein, Folstein, & McHugh, 1975). Many individuals who meet the clinical criteria for MCI scored above a 26 on the MMSE (Ihl, et al., 1992; Nasreddine et al., 2005); therefore Nasreddine and colleagues (2005) created a brief ten-minute screening tool, the Montreal Cognitive Assessment (MoCA), which was specifically designed to screen patients with MCI that would score within the normal range on the MMSE. They determined that a cutoff score of 26 provided the best specificity (100%) and sensitivity (90%) balance for the AD and MCI groups.

Neuropsychological tests of memory and verbal fluency that are sensitive and specific in the differential diagnosis of various types of dementia may be useful for the detection of MCI (DeJager, Hogervorst, Combrinck, & Budge, 2003). The MoCA tests the cognitive domains of attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation.

Previous research suggests that multiple areas of cognition are affected in MCI patients. Bozoki et al. (2001) found that memory loss alone was a risk factor for dementia, but when memory loss was combined with other clear cognitive impairments, the risk of dementia significantly increased. Scores on all cognitive domains are significantly lower at baseline in persons with MCI (Bennett, et al., 2002). Memory impairment is necessary to predict dementia, but the presence of impairments in other cognitive domains such as language, visuospatial cognition, and executive function are the best positive predictors for dementia (Sacuiu et al., 2005). Multiple-domain MCI may represent a more advanced stage of prodromal dementia because these individuals are the most likely clinical subtypes to convert to dementia (Alexopoulos et al., 2006). A steep decline on memory and executive function tasks has been found in individuals who will soon convert to AD (Chen et al., 2001).

Language impairment begins early in the course of AD and is common among individuals with AD (Taler & Phillips, 2008). Assessing the performance of MCI individuals on language tasks may help assess early signs of cognitive impairment in AD. Verbal fluency tasks are commonly used to assess language abilities in individuals with dementia (Henry, Crawford, & Phillips, 2004). Letter and category fluency are two common types of verbal fluency tasks. Letter fluency tasks require participants to name as many words as they can that begin with a particular letter of the alphabet; category fluency tasks require participants to name as many words as they can from a particular semantic category (e.g. animals, foods, or vehicles). Letter fluency requires phonemic search strategies whereas category fluency requires semantic search strategies (Rohrer, Salmon, Wixted, & Paulsen, 1999). Verbal fluency deficits have been found in individuals with AD, especially deficits in category fluency (Taler & Phillips, 2008). In one study, letter fluency was also found to reliably predict conversion to dementia (Small et al., 2001), but other studies have shown that this is not a reliable predictor of dementia onset (Chen et al., 2001; & Goldman et al., 2001).

MCI and AD patients also show declines in other language abilities, particularly naming abilities. MCI patients show deficits on the Boston Naming Test (BNT; Kaplan, Goodglass, &Weintrab, 1983), and decline more rapidly on other semantic abilities than healthy elderly adults (Bennett et al., 2002). As cognitive impairment progresses, there is a more rapid rate of decline on the BNT (Storandt, Grant, Miller, & Morris, 2002). In contrast, Rubin et al. (1998) found that individuals show no differences in naming until after diagnosis of dementia. In a more recent study, Beinhoff et al. (2005) found no differences on the 15-item BNT between MCI, normal controls, and individuals diagnosed with major depressive disorder.

Lexical discrimination tasks that have been used to assess spoken word recognition in healthy young adults have not often been used to test AD patients and have never been used with MCI patients. Sommers (1998) found that individuals with mild AD and healthy aging adults did not differ on lexical discrimination tasks, but those with a more severe form of AD performed significantly worse than healthy aging adults. Sommers' results may be affected by declines in inhibitory function, which are known to occur in individuals with MCI (Traykov et al., 2007).

Most researchers agree that MCI presents risk factors for AD, but there is still controversy over whether all cases of MCI represent prodromal AD (Taler & Phillips, 2008). Chertkow (2002) suggested that the variability in findings of the MCI conversion rate to AD results from the different clinical criteria used to diagnose MCI. Intervention with preventive therapies may be ideal during the early MCI stages; even though there has been controversy, the risk factors associated with the progression of MCI to AD are becoming more evident. Subtypes of MCI have recently been identified with respect to conversion rate to AD and severity of the disorder. Amnestic MCI, the subtype used in the current study, is thought to be the most likely subtype to convert to AD (Peteresen, 2003), although Fischer et al. (2007) found subtypes to be poor predictors of conversion to specific dementia types.

Petersen and colleagues have studied conversion rates to AD in several studies. In one study, they found an average annual conversion rate of 14% (Petersen et al., 2001b). In another study, they found an average conversion rate of 12% per year, and after 6 years the total conversion was 80% (Petersen & Morris, 2003). Geslani et al. (2005) reported a much higher conversion rate of 41% after one year and 64% after two years. Chertkow et al. (2001) found inconsistent results with regards to conversion rate; 25% of MCI individuals did not convert to AD even ten years after the onset of their memory problems. These findings may be contrasted to the more recent findings of Morris et al. (2007), who found a 100% long-term conversion rate, where long-term was defined as 9.5 years.

Current Study

In the current study, we tested the hypothesis that lexical discrimination is compromised during spoken word recognition in MCI patients relative to healthy elderly adults. We also predicted that younger adults would perform better than both healthy elderly adults and MCI patients on lexical discrimination tasks due both to age and to declines in several core cognitive functions. We assessed this hypothesis by measuring the effects of lexical competition on spoken word recognition in healthy young adult controls, healthy elderly adult controls, and MCI patients.

To assess lexical discrimination participants were given a sentence repetition task. We used a modified version of an auditory sentence repetition task based on the concepts from the NAM, the Veterans Affair Sentence Test (VAST, Bell & Wilson, 2001). The VAST has been found to provide accurate measures of speech intelligibility in laboratory settings (Bell & Wilson, 2001; Bell, 1996; Lin, 2000).

Spoken word recognition was examined by looking at deficits in lexical discrimination, which may be caused by alterations in language processing. Declines in lexical discrimination would demonstrate alterations in the effects of lexical competition, which is of clinical and theoretical significance for our understanding of language function in preclinical AD. We hope that this work will lead to early prediction of the development of AD from assessing MCI patients' spoken word recognition. Understanding communication declines in AD populations will help families and loved ones better understand AD patients' struggles to communicate and may suggest novel behaviorally based interventions to prevent further decline.

Methods

Participants

Study participants were native speakers of American English. All participants provided written informed consent and were compensated $10 per hour for their time. In order to eliminate the influence of hearing impairment on spoken word recognition performance, the experimenter gave participants a hearing screening before testing began. Young adults whose thresholds exceeded 20 dB and elderly adults whose thresholds exceeded 25 dB for 500, 1000, and 2000 Hz or exceeded 45 dB for 4000 Hz were excluded. All participants had no history of neurological or psychiatric disorders. Age and years of education did not differ between any of the groups.

Young Control Participants: Ten healthy young adults were recruited through flyers and email at Indiana University in the Psychological and Brain Sciences Department. Using the SPSS outlier function, we found one young adult to be an outlier and therefore excluded him/her from further analyses, yielding a final young adult control group of nine participants (mean age: 24.11 ± 6.01; 1 male, 8 females).

Healthy Elderly Control Participants: Six healthy elderly adults were recruited using flyers through local retirement communities, exercise classes, community service organizations, and the Neurology Clinic at Indiana University Hospital. One healthy elderly adult did not meet the requirements for hearing thresholds and therefore was excluded from further analyses. Using the SPSS outlier function, we found one healthy elderly adult to be an outlier and he/she was excluded from further analyses, yielding a final elderly adult control group of four participants (mean age: 68.75 ± 10.18; 4 females).

MCI Participants: Seven MCI patients were recruited through the Neurology Clinic at Indiana University Hospital. Three MCI participants did not meet the hearing screening requirements and therefore were excluded from further analyses, yielding a MCI group of four participants (mean age: 76.75 ± 6.24; 1 male, 3 females). MCI patients also underwent a comprehensive clinical assessment including physical and neurological examination by a physician, informant interview, neuropsychological testing, and laboratory studies. All MCI participants met criteria for amnestic MCI similar to that of Petersen and colleagues (1999) listed in Table 1.

Apparatus

An AMBCO AB audiometer (Model 650A) was used for the audiometric screening. A Shure MX 185 Microflex Lavalier microphone was used to record the neuropsychological assessments and the lexical discrimination task. A TASCAM digital tape recorder was used for the recording. The following standardized neuropsychological tests were administered to all participants by the experimenter: the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintrab, 1983), the Stroop Color and Word Test (Golden & Freshwater, 2002), forward and backward digit span taken from the Wechsler Memory Scale-Third Edition (WMS-III; Wechsler, 1997), and the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). For the lexical discrimination sentence task, sentences were played through Beyer Dynamic DT-100 headphones using PsyScript 5.1d3 (Bates & D'Oliveira, 2003) on a laptop computer (Apple PowerBook G4, Cupertino, CA). Table 2 displays a summary of the neuropsychological performance of all three groups.

	YC (Mean ± SD)	HEC (Mean ± SD)	MCI (Mean ± SD)
N	9	4	4
Age	24.11 ± 6.01	68.75 ± 10.18	76.75 ± 6.24
Education	15.44 ± 1.74	16.00 ± 2.45	16.75 ± 3.78
Sex	1 M, 8 F	4 F	1 M, 3 F
BNT (/60)	54.56 ± 2.83	57.50 ± 3.70	51.25 ± 5.85
Stroop – Word	104.89 ± 10.71	53.67 ± 39.17	84.50 ± 27.39
Stroop – Color	82.22 ± 7.97	65.00 ± 10.54	51.25 ± 15.97
Stroop – Color-word	55.33 ± 10.83	32.00 ± 1.73	24.00 ± 8.83
Stroop – Interference	59.67 ± 8.26	46.00 ± 5.20	42.50 ± 7.14
Forward DS	11.56 ± 2.79	11.25 ± 1.71	8.75 ± 2.22
Backward DS	10.00 ± 2.69	8.25 ± 3.20	6.00 ± 0.82
MoCA	28.78 ± 1.56	27.00 ± 1.41	21.25 ± 1.50
PTA (dB), 500 Hz	8.57 ± 5.56	16.25 ± 7.50	22.50 ± 5.00
PTA (dB), 1000 Hz	10.00 ± 2.89	15.00 ± 12.91	7.50 ± 5.00
PTA (dB), 2000 Hz	4.29 ± 6.07	7.50 ± 9.57	11.25 ± 11.09
PTA (dB), 4000 Hz	10.71 ± 5.35	10.00 ± 8.16	31.25 ± 14.93

Table 2. Study participant information. Note: Hearing thresholds are for the better ear at each frequency. Hearing thresholds were not available for two young adults, although all participants met the minimum threshold requirements stated in the text. One MCI participant reported having tinnitus and was unable to hear the pure-tones at 4000 Hz, but still performed well on all tasks and all other pure tone thresholds met requirements. BNT: Boston Naming Test; MoCA: Montreal Cognitive Assessment; PTA: pure tone average; M: males; F: females.

Stimulus Materials

Lexical discrimination was assessed by a sentence repetition task that was based on the materials in the Veterans Affairs Sentence Test (VAST; Bell & Wilson, 2001). The VAST is a spoken word recognition test that was designed to examine lexical competition effects. The test is based on the principles of the NAM. The VAST consists of 320 sentences, each containing 3 target words distributed in 4 conditions: high frequency/ high competition, high frequency/ low competition, low frequency/ high competition, and low frequency/ low competition. All target words were rated as highly familiar in a norming study using students from Indiana University (Nusbaum, Pisoni, & Davis 1984).

Because the original VAST sentences were not normed for cloze probability, we had 400 undergraduate college students complete a cloze norming task and we calculated the average cloze probability for each sentence (i.e., the average across the 3 keywords). All sentences had low cloze probability. The current study included the 80 lowest and the 80 highest cloze probability sentences, for a total of 160 sentences distributed across the 8 conditions, as illustrated in Table 3. One female native English speaker produced all of the sentences. In order to avoid ceiling effects in the lexical discrimination task, we masked the sentences using multi-talker babble at a signal-to-noise ratio (SNR) of -3dB. Multi-talker babble is believed to have greater validity than other types of noise, such as Gaussian white noise or speech-shaped noise (e.g. Kalikow, Stevens, & Elliot, 1997).

	High Cloze Probability	Low Cloze Probability
High frequency, High ND	The *map* is on the *deck* of the *ship.*	*Pile* the *load* into the *hut.*
High frequency, Low ND	The *couple* gathered *nuts* in the *woods.*	He *stole* each *check* issued to the *gang.*
Low frequency, High ND	Hang the *cage* on a *hook* in the in the *yacht.*	The *tag* on the *vest* made it *sag.*
Low frequency, Low ND	The *wind* will *rustle* the flap on the *bib.*	*Prop* up the *trash* to keep in the *sludge*

Table 3. The conditions included in the experiment and corresponding sample sentences. Target words are italicized.

Procedure

All participants completed the neuropsychological battery immediately following the audiological screening. The testing occurred in a quiet room inside the Speech Research Laboratory for most control participants; those healthy elderly adult participants (N=2) and MCI patients (N=7) who were not able to travel to the laboratory were tested in their homes by the experimenter. In order to record the entire session, the participants wore a lapel microphone, which was connected to a portable DAT recorder.

The experimenter administered the tests in the following order: (1) The BNT (Kaplan et al., 1983), in which participants produced names for the items presented. The 60-item standard form was used, presenting first all the odd numbers and then all the even numbers. Participants were instructed to say the name of the picture presented on the page. (2) The Stroop test (Golden & Freshwater, 2002), in which participants read color words (red, green, blue), named the color of X's printed on the page, and named the font color of color words where the ink color and the word do not match (e.g., the word "red" printed in green ink). The participants were given 45 seconds to read as many stimuli as they could in each category. (3) An immediate serial recall test (digit span; Wechsler, 1997) was given to assess short term and working memory. The experimenter read a list of numbers aloud at a rate of one number per second. For the forward digit span, the participants were instructed to repeat back what they heard in the exact same order. Backward digit span, which assessed working memory capacity, was given by instructing participants to repeat back the numbers they heard in reverse order. For both forward and backward digit spans, testing stopped once the participant responded incorrectly on two consecutive lists of the same length. (4) Category fluency, in which participants were given categories (animals, occupations, cities, vehicles, and foods) and asked to name as many words as they could within that category in two minutes. (5) Category fluency (switching) where participants were asked to switch between the categories vegetables/ musical instruments and weapons/tools (e.g., produce one item from the first category, one from the second, and continue switching for the full two minutes). (6) The MoCA (Nasreddine et al., 2005), a brief neuropsychological battery that has high sensitivity and specificity for MCI. The component tasks included the following: trail-making (joining letters and numbers in ascending order), copying a cube, clock drawing, naming (pictures of three animals), delayed recall (five words), digit span (repeating a list of 5 digits forward and 3 digits backward), attention (hearing a list of letters and tapping the table each time the letter A was heard), serial subtraction (counting backward from 100 in increments of 7), language (repeating sentences; naming as many words starting with F as possible in one minute), similarities (stating how two words are semantically similar), and orientation (stating the date and where the participant was).

The lexical discrimination task was administered immediately following the MoCA. The experimenter and participants used headphones to listen to the sentences. The sentences were played through the headphones from a laptop computer and the microphone was used to record the participants' responses. Participants were instructed to listen to each sentence and simply repeat back what they heard as quickly and accurately as possible. They were also told to repeat back any words they heard, even if they did not understand the entire sentence. The sentence stimuli were divided into two lists, one containing all high-cloze stimuli (Block A) and the other containing all low-cloze stimuli (Block B). Participants received both blocks in counterbalanced order, and stimuli within each block were presented in randomized order. A break was given between each block (i.e., after the first 80 sentences).

Results

Repeated-measures ANOVAs were conducted on accuracy scores, with participant group as a between-subjects variable, and frequency and ND as within-subjects variables. Significant effects and interactions were decomposed using least square difference (LSD) post hoc analyses. A one-way ANOVA was used to assess differences in difficulty cost between groups. All analyses were conducted using SPSS Version 12.0.1 for Windows (SPSS, 2003).

Demographics

Education did not differ significantly between any of the groups (p = 0.23; mean: 16 years ± 2.45). Healthy elderly adult and MCI participants did not significantly differ in age (p = 0.75; mean: 73 years ± 8.91). There were more females (N = 15) than males (N = 2) overall, although no group differences were observed (χ^2(2, N = 17) = 1.21, p = 0.55).

Total Word Accuracy Between Groups

Figure 1 presents word recognition accuracy for each group as a function of condition for the four sentence conditions. A repeated-measures ANOVA revealed main effects of group (F(1,14) = 31.59, p < .001), word frequency (F(1,14) = 172.41, p < .001), and ND (F(1,14) = 24.40, p < .001). Higher accuracy was observed in young adults than healthy elderly adults and MCI participants, and in healthy elderly adults than MCI participants; in high than low frequency items; and in low than high ND items. The interaction between ND and group was significant (F(2,14) = 5.40, p < .05) and is shown separately in Figure 2. The ND effect was stronger in the healthy elderly adults and MCI patients than in the young adults.

Difficulty Cost

A difficulty cost index was calculated for each participant by subtracting accuracy scores for the most difficult condition (low frequency/ high ND) from the score for the easiest condition (high frequency/ low ND). Difficulty cost is presented in Figure 3. The groups differed significantly from one another on difficulty cost (F(2,14) = 7.36, p < .01). The post hoc LSD revealed significant differences when young adults were compared to the healthy elderly adults and MCI patients (p < .01), but not between healthy elderly adults and MCI patients (p = .92).

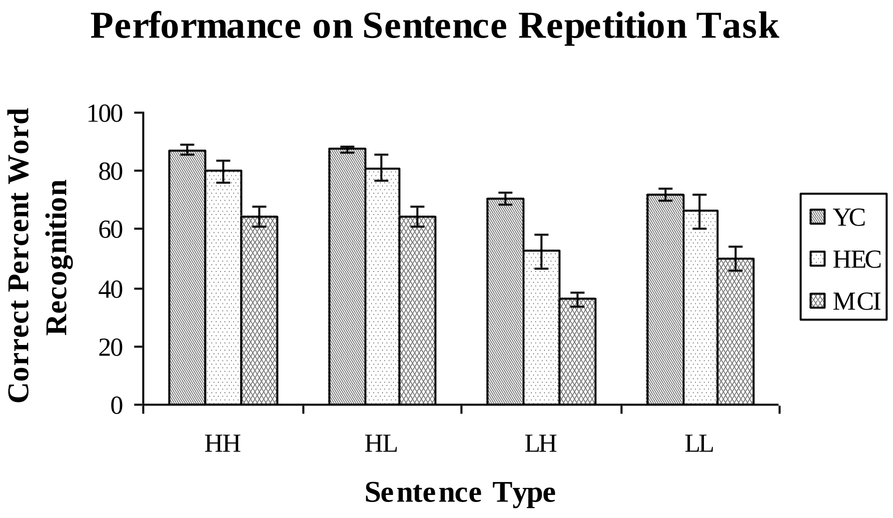

Figure 1. YC and HEC groups differ on the LH sentence type. HEC and MCI groups differ on all sentence types. Note: Error bars represent standard error; YC: young controls; HEC: healthy elderly controls; HH: high frequency/ high ND; HL: high frequency/ low ND; LH: low frequency/ high ND; LL: low frequency/ low ND.

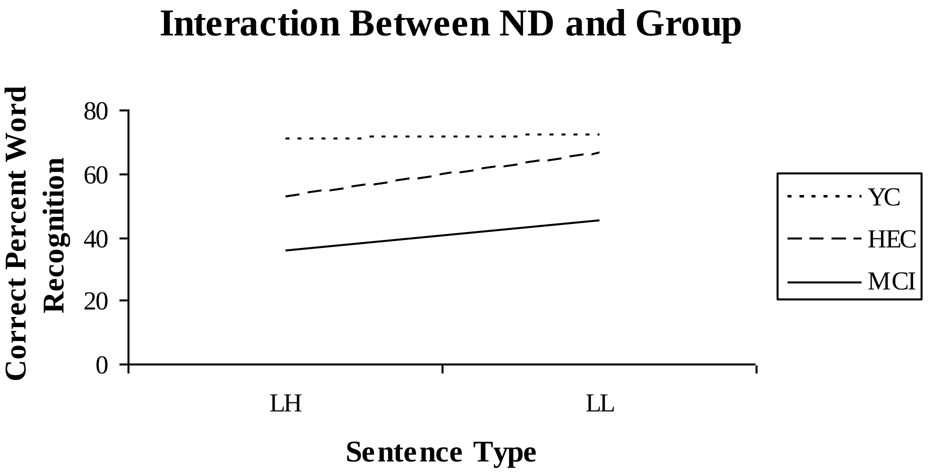

Figure 2. The significant group by ND interaction. High ND items were harder to understand than low ND items. Note: YC: young controls; HEC: healthy elderly controls; LH: low frequency/ high ND; LL: low frequency/ low density.

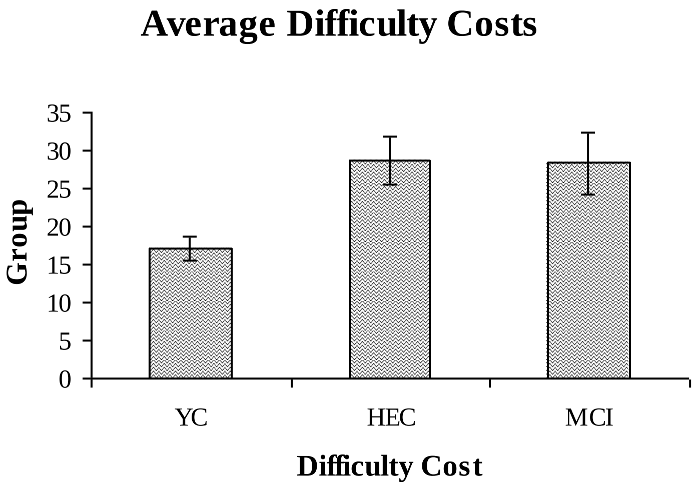

Figure 3. Difficulty cost is a measure of lexical discrimination. Young adults differ significantly from healthy elderly adults (*p* = .007) and MCI patients (*p* = .009). Healthy elderly controls and MCI patients did not differ significantly (*p* = .920). Note: YC: young controls; HEC: healthy elderly controls; Error bars represent standard error.

Correlations with Neuropsychological Measures

The difficulty cost was correlated with the demographics and the measures of neuropsychological function collected for all of the participants combined together. The correlations are shown in Table 4. Difficulty cost correlated significantly with age and scores on the MoCA and Stroop color/word scores. Difficulty cost did not correlate significantly with any of the other neuropsychological measures.

	1	2	3	4	5	6	7	8
1. Age	1	.284	-.852**	-.104	-.757**	-.433	-.509*	.773**
2. Education		1	-.240	-.306	-.159	.299	.151	.469
3. Stroop			1	.284	.709**	.454	.653**	-.551*
4. BNT			1	.370	.145	.134	-.142	
5. MoCA					1	.533*	.510*	-.576*
6. ForwDS						1	.611**	-.266
7. BackDS							1	-.239
8. Difficulty Cost								1

Table 4. Correlations between performance on the experimental task, demographic information, and neuropsychological measures. Note: **p* < .05; ***p* < .01; Stroop measures are color/word raw scores; BNT: Boston Naming Task; MoCA: Montreal Cognitive Assessment; ForwDS: Forward Digit Span; BackDS: Backward Digit Span.

Discussion

The current study found that overall, spoken word recognition in sentences was lowest in MCI patients, intermediate in healthy elderly adults, and highest in healthy young adults. This pattern of results supports our hypothesis that both age and cognitive function play a role in spoken word recognition. It appears that age most strongly affects lexical discrimination because the young adults have the lowest difficulty cost, but the healthy elderly adults and the MCI patients have nearly the same difficulty cost. This does not support our hypothesis that lexical discrimination is compromised during spoken word recognition in MCI patients.

The total accuracy on the lexical discrimination task differed between all groups. This suggests that age plays a role in spoken word recognition, and the ability to distinguish words declines with age. Age differences were not found between the healthy elderly adults and MCI patients, and this suggests that age plays a role in spoken word recognition. These findings are consistent with previous research (Newman & German, 2005; Vitevitch & Sommers, 2003). Since we found differences between healthy elderly adults' and MCI patients' overall accuracy, cognitive decline likely affects spoken word recognition.

The word frequency effect found in the current study is consistent with the predictions of the NAM (Luce & Pisoni, 1998) and other previous studies (Broadbent, 1967; Catlin, 1969; Goldiamond & Hawkins, 1958; Newbigging, 1961; Pollack, Rubenstein & Decker, 1960; Savin, 1963; Soloman & Postman, 1952; Triesman, 1971, 1978a, b). All participants identified high frequency words in sentences more accurately when presented in background noise than low frequency words. The participants performed better on words with low ND than high ND, which is also consistent with the NAM (Luce & Pisoni, 1998). The finding that low ND words are easier to identify than high ND words replicates previous findings (Vitevich & Luce, 1998; Goldinger et al, 1989; Luce and Pisoni, 1998; Dirks et al, 2001).

In this study, we also measured lexical discrimination abilities using a difficulty cost measure – the difference between the hardest and easiest sentences. This was done to normalize for overall group differences in spoken word recognition. We found lower difficulty costs in young adults than in the other two groups. However, we failed to find a difference between healthy elderly adults and MCI patients using this index. Thus, lexical discrimination does not appear to be strongly impacted by declining cognitive ability, although it decreases with age.

Although our previous research did not find age to be a factor in lexical discrimination (Taler, Aaron, & Pisoni, submitted), the current study found correlations between the difficulty cost index and age. This strengthens the idea that age is a factor in the declining abilities of lexical discrimination. Scores on the MoCA neuropsychological assessment were also correlated with the difficulty cost even though we did not find lexical discrimination differences between healthy elderly adults and MCI patients. The Stroop color/word scores correlated with difficulty cost, which would be expected since inhibitory function has been found to decline in individuals with MCI (Traykov, 2007).

Although the MCI and healthy elderly groups did not differ on difficulty cost measures, we did find correlations between cognitive measures and lexical difficulty costs when all subjects were combined. The small sample size likely has an impact on the results, especially in the healthy elderly adult and MCI groups. Recruitment is ongoing for this study.

The significant difference in lexical discrimination between young and older adults (with or without cognitive impairment) raises questions about whether the sentence repetition task was appropriate for measuring the lexical discrimination abilities between the three groups. The SNR was set to -3dB for this study. This may not have been the optimal SNR for MCI patients. Their mean total accuracy was below 50%, meaning that floor effects may be present for certain participants. Another limitation to the current study is the disproportionate female-to-male ratio. There were no males in the healthy elderly adult group and only one male in the MCI group. Cognitive decline in the aging population is as common in males as females (Barnes, 2003); therefore, this sample is not representative of the total population.

Future research should be undertaken to improve our understanding of spoken word recognition performance in MCI patients. Declines in spoken word recognition are known to occur in AD (Sommers, 1998). Adding a fourth group in the present study, AD patients, to compare to the other three groups would allow us to assess how quickly spoken word recognition declines in individuals suffering from cognitive impairment. The conversion rate in the MCI patients to AD should be studied to help determine what factors of spoken word recognition might be early predictors of AD.

To better assess lexical discrimination in these populations, a study should be conducted to find the optimal SNR for MCI and AD patients. Another possible avenue for future research is to explore the links between lexical discrimination and everyday communication skills. Future work will include an analysis of reaction times and the collection of more data from healthy elderly adults, MCI patients, and AD patients.

References

Alexopoulos, P., Grimmer, T., Perneczky, R., Domes, G., & Kurz, A. (2006). Progression to dementia in clinical subtypes of mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders, 22,* 27-34.

Beinhoff, U., Hilbert, V., Bittner, D., Grön, G., & Riepe, M. W. (2005). Screening for cognitive impairment: A triage for outpatient care. *Dementia and Geriatric Cognitive Disorders, 20,* 278-285.

Barnes L. L., Wilson R. S., Schneider J. A., Bienias J. L., Evans D. A., & Bennett D. A. (2003). Gender, cognitive decline, and risk of AD in older persons. *Neurology*, *60*(11), 1777-81.

Bates, T. C., & D'Oliveria, L., (2003). PsyScript: a Macintosh application for scripting experiments. *Behavior Research Methods, Instruments, & Computers, 35,* 565-576.

Bell, T. S. (1996). A new measure of word recognition. *Sound & Video Contractor, October 20*, 28-33.

Bell, T. S., & Wilson, R. H. (2001). Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology, 12,* 514-522.

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Beckett, L. A., Aggarwal, N. T., et al. (2002). Natural history of mild cognitive impairment in older persons. *Neurology, 59*, 198-205.

Bozoki, A., Giordani, B., Heidebrink, J. L., Berent, S., & Foster, N. L. (2001). Mild cognitive impairments predict dementia in nondemented elderly patients with memory loss. *Archives of Neurology, 58,* 411-416.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74,* 1-15.

Catlin, J. (1969). On the word-frequency effect. *Psychological Review, 76,* 504-506.

Chen, P., Ratcliff, R., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2001). Patterns of cognitive decline in pre-symptomatic Alzheimer's disease: A prospective community study. *Archives of General Psychiatry, 58,* 853-858.

Chertkow, H. (2002). Mild cognitive impairment. *Current Opinion in Neurology, 15*(4), 401.

Chertkow, H., Verret, L., Bergman, H., Wolfson, C., & McKelvey, R. (2001). *Predicting progression to dementia in elderly subjects with mild cognitive impairment: A multidisciplinary approach.* Paper presented at 53rd Annual Meeting of the American Academy of Neurology, Contemporary Clinical Issues Plenary Session. Philadelphia, PA, USA.

De Jager, C. A., Hogervorst, E., Combrinck, M., & Budge, M. M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine, 33,* 1039-1050.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test-second edition: adult version manual*. San Antonio, TX: The Psychological Corporation.

Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsiner, D., Fausti, S. A. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing, 22,* 1-13.

Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., et al. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology, 23,* 288-291.

Folstein, M., Folstein, S., & McHugh, P. (1975). " Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Geslani, D., Tierney, M., Herrmann, N., & Szalai, J. (2005). Mild cognitive impairment: An operational definition and its conversion rate to Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 19,* 383-389.

Golden, C., & Freshwater, S. (2002). The Stroop Color and Word Test. Wood Dale, IL: Stoelting Co.

Goldiamond, I., & Hawkins, W. F. (1958). Vexierversuch: The logarithmic relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology, 56,* 457-463.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501-518.

Goldman, W. P., Price, J. L., Storandt, M., Grant, E. A., McKeel, D. W., Rubin, E. H., et al. (2001). Absence of cognitive impairment or decline in preclinical Alzheimer's disease. *Neurology, 56,* 361-367.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia, 42,* 1212-1222.

Ihl R., Frölich L., Dierks T., Martin E. M., & Maurer K. (1992). Differential validity of psychometric tests in dementia of the Alzheimer type. *Psychiatry Research, 44* (2), 93-106.

Kalikow, D. N., Stevens, K. N., & Elliot, L. L. (1977). Development of a test of speech intelligibility in noise using sentences with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). Boston Naming Test (Revised 60-item version). *Philadelphia: Lea & Febiger*.

Lin, A. (2000). *Speech discrimination: The reliability and validity of the Veterans Affairs Sentence Test.* California State University, Los Angelos.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Doctoral dissertation, Indiana University, Bloomington, IN.

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

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & cognition, 25,* 47-56.

Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., et al. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology, 58,* 397-405.

Nasreddine, Z., Phillips, N., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699.

Newbigging, P. L. (1961). The perceptual redintegration of frequent and infrequent words. *Canadian Journal of Psychology, 15,* 123-132.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech, 48,* 123-156.

Nusbaum, H., Pisoni, D., & Davis, C. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words (Research on Speech Perception Progress Report No. 10). *Bloomington: Indiana University, Psychology Department, Speech Research Laboratory*.

Petersen, R. C. (2003). *Mild cognitive impairment.* New York: Oxford University Press.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., et al. (2001a). Current concepts in mild cognitive impairment. *Archives of Neurology, 58,* 1985-1992.

Petersen R. C., & Morris J. C. (2003). Clinical features. In R. C. Petersen (Ed.), *Mild cognitive impairment: Aging to Alzheimer's disease* (pp. 15-40), New York: Oxford University Press, Inc.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild Cognitive Impairment Clinical Characterization and Outcome. *Archives of Neurology, 56*(3), 303-308.

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001b). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology, 56,* 1133-1142.

Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *Journal of the Acoustical Society of America, 32,* 340-353.

Rohrer, D., Salmon, D. P., Wixted, J. T., & Paulsen, J. S. (1999). The disparate effects of Alzheimer’s disease and Huntington’s disease on semantic memory. *Neuropsychology, 13,* 381-388.

Rubin, E. H., Storandt, M., Miller, J. P., Kinscherf, D. A., Grant, E. A., Morris, J. C., et al. (1998). A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of Neurology, 55,* 395-401.

Sacuiu, S., Sjögren, M., Johansson, B., Gustafson, D., & Skoog, I. (2005). Prodromal cognitive signs of dementia in 85-year-olds using four sources of information. *Neurology, 65,* 1894-1900.

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America, 35,* 200-206.

Small, B. J., Herlitz, A., Fratiglioni, L., Almkvist, O., & Bäckman, L. (1997). Cognitive predictors of incident Alzheimer’s disease: A prospective longitudinal study. *Neuropsychology, 11,* 413-420.

Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology, 43,* 195-201.

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related changes in spoken word recognition. *Psychology and Aging, 11,* 333–341.

Sommers, M. S. (1998). Spoken word recognition in individuals with dementia of the Alzheimer’s type: Changes in talker normalization and lexical discrimination. *Psychology and Aging, 13,* 631-646.

Sommers, M. S., & Danelson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14,* 458-472.

SPSS (2003). *SPSS version 12.0.1 for Windows,* SPSS Inc.: Chicago, IL.

Storandt, M., Grant, E. A., Miller, P., & Morris, J. C. (2002). Rates of progression in mild cognitive impairment and early Alzheimer’s disease. *Neurology, 59,* 1034-1041.

Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer’s disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology, 30*(5), 501-556.

Taler, V., Aaron, G. P., & Pisoni, D. B. (Submitted). Neighborhood density effects in spoken word recognition and production in healthy aging.

Traykov, L., Raoux, N., Latour, F., Gallo, L., Hanon, O. (2007). Executive functions deficit in mild cognitive impairment. *Cognitive and Behavioral Neurology, 20,* 219-224.

Triesman, M. (1971). On the word frequency effect: Comments on the papers by J Catlin and L. H. Nakatani. *Psychological Review, 78,* 420-425.

Triesman, M. (1978a). A theory of the identification of complex stimuli with an application to word recognition. *Psychological Review, 85,* 525-570.

Triesman, M. (1978b). Space or lexicon? The word frequency effect and the error response frequency effect. *Journal of Verbal Learning and Verbal Behavior*, *17*, 37-59.

Vitevitch, M., & Luce, P. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science, 9*(4), 325-329.

Vitevich, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition, 31,* 491-504.

Wechsler, D. (1997). *Wechsler Memory Scale (WMS-III)*. San Antonio, Texas: The Psychological Corporation.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Speech Perception and Production[1]

Elizabeth D. Casserly[2] and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH DC00111 and DC00012 grants to Indiana University and an NSF Graduate Research Fellowship to the first author.

[2] Department of Linguistics, Indiana University, Bloomington, IN 47405

Speech Perception and Production

Abstract. Until recently, research in speech perception and speech production has largely focused on the search for psychological and phonetic evidence of discrete, abstract, context-free symbolic units corresponding to phonological segments or phonemes. Despite this common conceptual goal and intimately related objects of study, however, research in these two domains of speech communication has progressed more or less independently for more than sixty years. In this chapter, we present a brief overview of the foundational works and current trends in the two fields separately, specifically discussing the progress made in both lines of inquiry as well as the basic fundamental issues that neither has been able to resolve satisfactorily thus far. We then turn to discussion of theoretical models and recent experimental evidence that point to the deep, pervasive connections between speech perception and production. We conclude that although research focusing on each domain individually has been vital in increasing our basic understanding of spoken language processing, the human capacity for speech communication is so complex that gaining a full understanding will not be possible until speech perception and production are conceptually reunited in a joint approach to common problems.

Introduction

Historically, language research focusing on the spoken (as opposed to written) word has split into two distinct fields: speech perception and speech production. Psychologists and psycholinguists worked on problems of phoneme perception, while phoneticians examined and modeled articulation and speech acoustics. Despite their common goal of discovering the nature of the human capacity for spoken language communication, the two broad lines of inquiry have experienced limited mutual influence or cross-talk. The split has been partially practical, since methodologies and analysis are necessarily quite different when aimed at direct observation of overt behavior, as in speech production, or examination of hidden cognitive and neurological function, as in speech perception. Academic specialization has also played a part, since there is an overwhelming volume of knowledge available, but single researchers can only learn and use a small piece. In keeping with the goal of this series, however, we argue that the greatest prospects for progress in speech research over the next few years lie in the combination of insights from research on speech perception and production, and in investigation of the inherent links between these two processes.

In this chapter, therefore, we will discuss the major theoretical and conceptual issues in research dedicated first to speech perception and then to speech production, as well as the successes and lingering problems in these domains. Then we will turn to exciting new directions in experimental evidence and theoretical models which begin to close the gap between the two research focuses by suggesting ways in which they may work together in everyday speech communication and highlighting the inherent links between speaking and listening.

Speech Perception

Before the advent of modern signal processing technology, linguists and psychologists assumed that speech perception was a fairly uncomplicated process. Theoretical linguistics' description of spoken language relied on sequential strings of abstract, context-invariant segments or phonemes which provided the mechanism of contrast between lexical items (e.g. distinguishing *pat* from *bat*) (Chomsky & Miller, 1963). The immense analytic success and relative ease of approaches using such symbolic structures led language researchers to believe that the physical implementation of speech would adhere to the segmental "linearity condition," so that the acoustics corresponding to consecutive phonemes would concatenate like an acoustic alphabet or a string of beads stretched out in time. If that were the case, perception of the linguistic message in spoken utterances would be a trivial matching process – simple.

Understanding the true nature of the physical speech signal, however, has turned out to be far from easy. Early signal processing technologies, prior to the 1940's, could detect and display time-varying acoustic amplitudes in speech, resulting in the familiar waveform seen in Figure 1. Phoneticians have long known that the component frequencies encoded within speech acoustics, and how they vary over time, serve to distinguish one speech percept from another, and waveforms do not readily provide access to this key information. A major breakthrough, then, came in 1946, when Ralph Potter and his colleagues at Bell Laboratories developed the speech spectrogram, a representation which uses the mathematical Fourier transform to uncover the strength of the speech signal hidden in the waveform amplitudes of Figure 1 at a wide range of possible component frequencies (Potter, 1945). Each calculation finds the signal strength through the frequency spectrum of a small time window of the speech waveform; stringing the results of these time-window analyses together yields a spectrogram or "voiceprint," representing the dynamic frequency characteristics of the spoken signal (Figure 2).

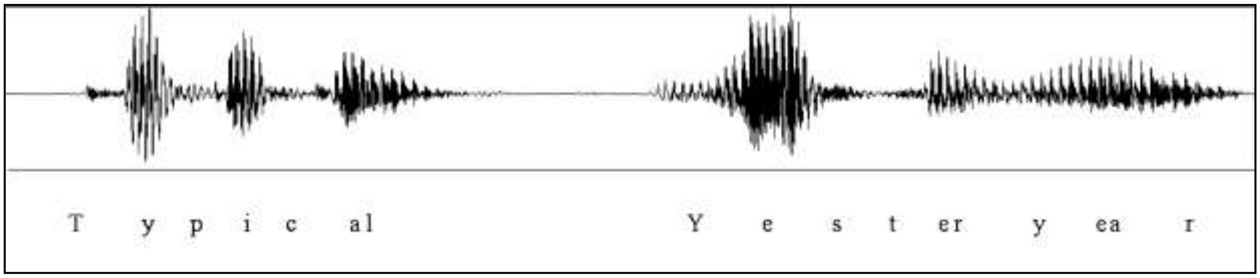

Figure 1. Speech waveform of the words *typical* and *yesteryear* as produced by an adult male speaker, representing variations in amplitude over time. Vowels are generally the most resonant speech component, corresponding to the highest amplitude levels seen here. The identifying formant frequency information in the acoustics is not readily accessible from visual inspection of waveforms such as these.

Phonemes – An Unexpected Lack of Evidence

As can be seen in Figure 2, the content of a spectrogram does not visually correspond to the discrete segmental units perceived in speech in a straightforward manner. Although vowels stand out due to their relatively high amplitudes (darkness) and clear internal frequency structure, reflecting harmonic resonances or "formant frequencies" in the vocal tract, their exact beginning and ending points are not immediately obvious to the eye. Even the seemingly clear-cut amplitude rises after stop consonant closures, such as for the [p] in *typical*, do not directly correlate with the beginning of a discrete vowel segment, since these acoustics simultaneously provide critical information about the identity of the consonant and the following vowel. Demarcating consonant/vowel separation is even more difficult in the case of highly sonorant (or resonant) consonants such as [w] or [r].

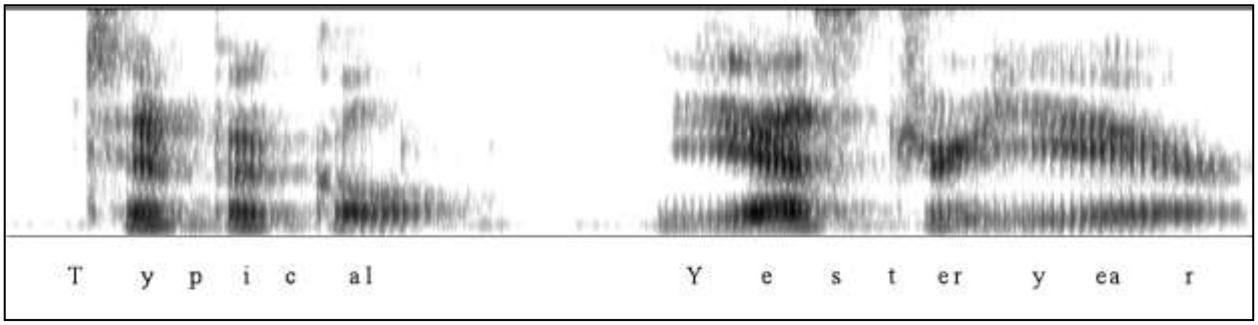

Figure 2. A wide-band speech spectrogram of the same utterance as in Fig.1, showing the change in component frequencies over time. Frequency is represented along the y-axis, time on the x-axis. Darkness corresponds to greater signal strength at the corresponding frequency and time.

The "acoustic alphabet" view of speech received another set-back in the 1960's, when Franklin Cooper and colleagues reported that acoustic signals composed of strictly serial, discrete units corresponding to segments are actually impossible for listeners to process at speeds near normal speech processing (1969). No degree of signal simplicity, contrast between units, or user training with the context-free concatenation system could produce natural rates of speech perception for listeners. Therefore, Cooper's research team concluded that speech must transmit information in parallel, through use of the contextual overlap observed in spectrograms of the physical signal. Speech does not look like a string of discrete, context-invariant acoustic segments, and in order for listeners to process its message as quickly as they do, it cannot be such an system. Instead, as Alvin Liberman concludes, speech is a "code," taking advantage of parallel transmission through co-articulation (see *Variation in Invariants*, below) on a massive scale (1967).

In light of these discoveries, many researchers began wondering: If it is true that phonemes are a genuine property of linguistic systems, as phonological evidence implies; and it is also true that the acoustic speech signal does not directly correspond to phonological segments; then how do listeners actually recover the linguistic content of the continuously varying acoustics? Hockett's famous Easter egg analogy succinctly demonstrates the bewilderment of speech scientists at this early stage:

> Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between two rollers of a wringer, which quite effectively smash them and rub more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter. At a subsequent point, we have an inspector whose task it is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread out albumen, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer.
> (Hockett, 1955, pg. 210)

For many years, researchers in the field of speech perception focused their efforts on trying to solve this enigma, believing that the heart of the speech perception problem lay in the seemingly impossible task of phoneme recognition – putting the Easter eggs back together.

Synthetic Speech & the Haskins Pattern Playback

Soon after the spectrogram enabled researchers to visualize the spectral content of speech acoustics and its changes over time, that knowledge was put to use in the development of technology able to generate speech synthetically. One of the early research synthesizers was the Pattern Playback, developed by scientists and engineers, including Cooper and Liberman, at Haskins Laboratories (Cooper et al., 1949). This device could take simplified sound spectrograms like those shown in Figure 3 and use the component frequency information to produce highly intelligible corresponding speech acoustics. Hand-painted spectrographic patterns such as those shown in Figure 3 allowed researchers tight experimental control over the content of this synthetic, very simplified Pattern Playback speech. By varying its frequency content and transitions over time, the investigators were able to determine many of the specific aspects in spoken language which are essential to particular speech percepts, and many which are not.

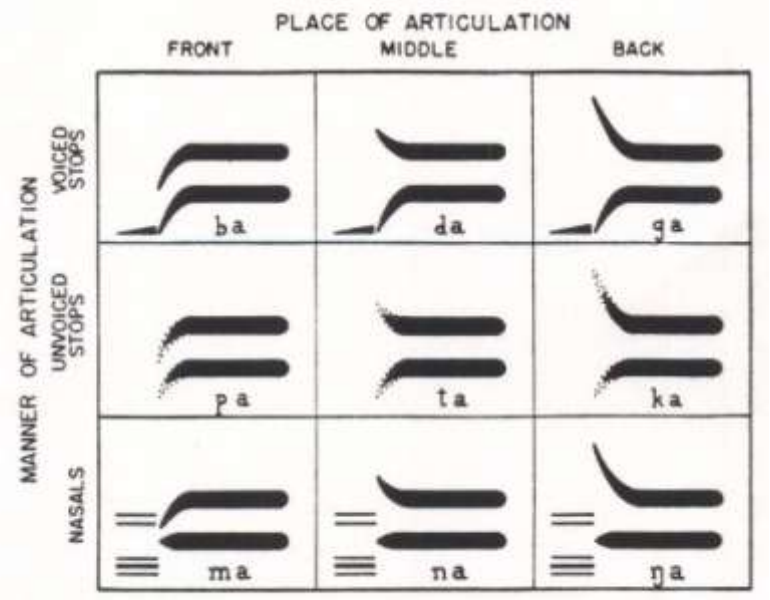

Figure 3. A series of hand-painted schematic spectrographic patterns used as input to the Haskins Pattern Playback speech synthesizer in early research on perceptual “speech cues.” (From AM Liberman, “Some results of research on speech perception.” Journal of the Acoustical Society of America. 29(1);117-123. 1957. Reprinted with permission.)

Perceptual experiments with the Haskins Pattern Playback and other speech synthesizers revealed, for example, the pattern of complex acoustics that signals the place of articulation of English stop consonants, such as [b], [t] and [k] (Liberman et al., 1967). For voiced stops ([b], [d], [g]) the transitions of the formant frequencies from silence to the vowel following the consonant largely determine the resulting percept. For voiceless stops ([p], [t], [k]) however, the acoustic frequency of the burst of air following the release of the consonant plays the largest role in identification. The experimental control gained from the Pattern Playback allowed researchers to alter and eliminate many aspects of naturally-produced speech signals, making many such discoveries regarding identification of the sufficient and necessary acoustic cues for a given speech percept. This early work attempted to pair speech down to its bare essentials, hoping to reveal the mechanisms of speech perception. While largely successful in identifying perceptually crucial aspects of speech acoustics and greatly increasing our fundamental understanding of speech perception, however, these efforts did not yield invariant, context-independent features corresponding to segments or phonemes. If anything, this research program suggested alternative bases for the communication of linguistic content (Liberman & Mattingly, 1985; Goldstein & Fowler, 2003).

Phoneme Perception – Positive Evidence

Some of the research conducted with aim of understanding phoneme perception, however, did lead to results suggesting the reality of psychological particulate units such as phonemes. For instance, in some cases listeners show evidence of "perceptual constancy," or abstraction from signal variation to more generalized representations –possibly phonemes. Various types of such abstraction appear to occur in speech perception, but we will address two of the most influential here.

Categorical Perception Effects. Phoneme representations split potential acoustic continuums into discrete categories. The duration of aspiration occurring after the release of a stop consonant, for example, constitutes a potential continuum ranging from zero milliseconds, where vocalic resonance beginning simultaneously with release of the stop, to an indefinitely long period between the stop release and the start of the following vowel. Yet stops in English falling along this continuum are split cleanly into two groups, voiced ([b], [d], [g]) or voiceless ([p], [t], [k]), based on the length of this "voice onset time." In general, this phenomenon is not so strange – categories often serve to break continuous variation into manageable chunks.

Speech categories appear to be unique in one aspect, however: listener deafness to differences between two members of the same category. That is, although we may assign two different colors both to the category "red," we can easily distinguish between the two shades in most cases. When speech scientists give listeners stimuli varying along an acoustic continuum, however, their discrimination between different tokens of the same category (analogous to two shades of red) is very close to chance (Liberman et al., 1957). They are highly accurate at discriminating tokens spanning category boundaries, on the other hand. The combination of sharp category boundaries in listeners' labeling of stimuli and their within-category "deafness" in discrimination, as shown in Figure 4, appears to be unique to human speech perception, and constitutes some of the strongest evidence in favor of robust segmental categories underlying speech perception.

According to this evidence, listeners sacrifice sensitivity to acoustic detail in order to make speech category distinctions more automatic and perhaps also less subject to the influence of variability. This type of category robustness is observed more strongly in perception of consonants than vowels. Not coincidentally, as discussed briefly above and in more detail in the *Acoustic Phonetics* section below, the stop consonants which are perceived with such "deafness" also prove the greatest challenge to define in terms of invariant acoustic cues (Stevens, 2005).

Perceptual Constancy. Categorical perception effects are not the only case of such abstraction or constancy in speech perception; listeners also appear to "translate" the speech they hear into more symbolic idealized forms based on expectations of gender and accent. Niedzielski, for example, found that listeners identified recorded vowel stimuli differently when they were told that the original speaker was from their own versus another dialect group (1999). For these listeners, therefore, the mapping from physical speech characteristics to linguistic categories was not absolute, but mediated by some abstract conceptual unit. Johnson summarizes the results of a variety of studies showing similar

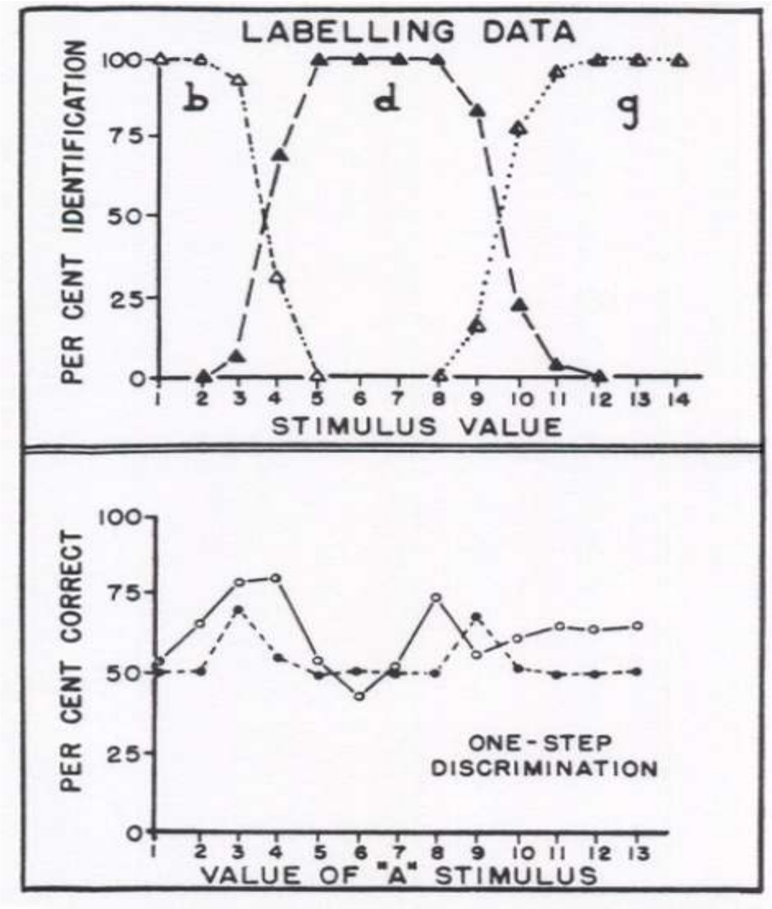

Figure 4. Data for a single subject from a categorical perception experiment. The upper panel gives labeling or identification data for each step on a [b]/[g] place-of-articulation continuum. The lower graph gives this subject's ABX discrimination data (filled circles) for the same stimuli with one step difference between pairs, as well as the predicted discrimination performance (open circles). Discrimination accuracy is high at category boundaries and low within categories, as predicted. (From AM Liberman, KS Harris, HS Hoffman, BC Griffith. "The discrimination of speech sounds within and across phoneme boundaries." Journal of Experimental Psychology, 54(5);358-68. 1957. American Psychological Association. Reprinted with permission.)

behavior (2005), which corroborates the observation that, although indexical or "extra-linguistic" information such as speaker gender and dialect are not inert in speech perception, more abstract linguistic units play a role in the process as well.

Far from being exotic, this type of "perceptual equivalence" corresponds very well with language users' intuitions about speech. Although listeners are aware that individuals often sound drastically different, the feeling remains that something holds constant across talkers and speech tokens. After all, *cat* is still *cat* no matter who says it. Given the signal variability and complexity observed in speech acoustics, such consistency certainly seems to implicate the influence of some abstract unit in speech perception, possibly contrastive phonemes or segments.

Phoneme Perception – Shortcomings & Roadblocks

From the discussion above, it should be evident that speech perception research with the traditional focus on phoneme identification and discrimination has been unable either to confirm or deny the psychological reality of context-free symbolic units such as phonemes. Listeners' "deafness" to stimuli differences within a linguistic category and their reference to an abstract ideal identification support the cognitive role of such units, while synthetic speech manipulation has simultaneously demonstrated that linguistic percepts simply do not depend on invariant, context-free acoustic cues corresponding to segments. This paradoxical relationship between signal variance and perceptual invariance constitutes one of the fundamental issues in speech perception research.

Crucially, however, the research discussed up until now focused exclusively on the phoneme as the locus of language users' perceptual invariance. This approach stemmed from the assumption that speech perception can essentially be reduced to phoneme identification, relating yet again back to theoretical linguistics' analysis of language. Especially given the roadblocks and contradictions emerging in the field, however, speech scientists began to question the validity of those foundational assumptions. By attempting to control variability and isolate perceptual effects on the level of the phoneme, experimenters were asking listeners to perform tasks that bore little resemblance to typical speech communication. Interest in the field began to shift towards the influence of larger linguistic units such as words and how speech perception processes are affected by them, if at all.

Beyond the Phoneme – Spoken Word Recognition Processes. Both new and revised experimental evidence readily confirmed that the characteristics of word-level units do exert massive influence in speech perception. The lexical status (word versus non-word) of experimental stimuli, for example, biases listeners' phoneme identification such that they hear more tokens as [d] in a dish/tish continuum, where that percept creates a real word, than a da/ta continuum where both perceptual options are non-words (Ganong, 1980). Since then, research into listeners' perception of words has gone well beyond their effects on standard phoneme perception tasks to discover many factors which play a major role in word recognition but which almost never factor into phoneme perception work.

Perhaps the most fundamental of these factors is word frequency: how often a lexical item tends to be used. The more frequently a listener encounters a word over the course of their daily lives, the more quickly and accurately they are able to recognize it, and the better they are at remembering it in a recall task (e.g. Howes, 1957; Oldfield, 1966). High-frequency words are more robust in noisy listening conditions, and whenever listeners are unsure what they have heard through such interference, they are more likely to report hearing a high-frequency lexical item than a low-frequency one (Goldiamond & Hawkins, 1958). In fact, the effects of lexical status mentioned above are actually only extreme cases of frequency effects; phonotactically legal non-words (i.e. non-words which seem as though they could be real words) are treated psychologically like real words with a frequency of zero. Like cockroaches, these so-called "frequency effects" pop up everywhere in speech research.

The nature of a word's "lexical neighborhood" also plays a pervasive role its recognition. If a word is highly similar to many other words, such as *cat* is in English, then listeners will be slower and less accurate to identify it, whereas a comparably high-frequency word with fewer "neighbors" to compete with it will be recognized more easily. "Lexical hermits" such as *Episcopalian* and *chrysanthemum*, therefore, are particularly easy to recognize despite their low frequencies. As further evidence of frequency effects' ubiquitous presence, however, the frequencies of a word's neighbors also influence perception: a word with a dense neighborhood of high-frequency items is more difficult to recognize than a word with a dense neighborhood of relatively low-frequency items, which has weaker competition (Luce & Pisoni, 1998; Luce et al., 2000).

Particularly troublesome for abstract, phoneme-based views of speech perception, however, is the discovery that the indexical properties of speech (see *Perceptual Constancy* above) also influence word recognition. Goldinger, for example, has shown that listeners are more accurate at word recall when stimuli are repeated by the same versus different talkers in an experiment (1998). If speech perception were mediated only by linguistic abstractions, such “extra-linguistic” detail should not be able to exert this influence. In fact, this and an increasing number of similar results have caused some speech scientists to abandon theories of phoneme-based linguistic representation altogether, instead positing that lexical items are composed of maximally detailed “episodic” memory traces (Goldinger, 1998; Port, 2007).

Conclusion – Speech Perception

Regardless of the success or failure of these episodic representational theories, a host of new and re-shaped research questions remain open in speech perception. The variable signal/common percept paradox remains a fundamental issue: what accounts for the perceptual constancy across highly diverse contexts, speech styles and speakers? From a job interview in a quiet room to a reunion with an old friend at a cocktail party, from a southern belle to a Detroit body builder, what makes communication possible? Answers to these questions may lie in discovering the extent to which the speech perception processes tapped by experiments in word recognition and phoneme perception are related, and uncovering the nature of the neural substrates of language that allow it to adapt to such diverse situations. Deeply connected to these issues, Goldinger, Johnson and others’ results have prompted us to wonder: what is the representational specificity of speech knowledge and how does it relate to perceptual constancy?

Although speech perception research over the last sixty years has made substantial progress in increasing our understanding of perceptual challenges and particularly the ways in which they are *not* solved by human listeners, it is clear that a great deal of work remains to be done before even this one aspect of speech communication is truly understood.

Speech Production

Speech production research serves as the compliment to the work on speech perception described above. Where investigations of speech perception are necessarily indirect, using listener response time latencies or recall accuracies to draw conclusions about underlying linguistic processing, research on speech production can be extremely direct. In typical production studies, speech scientists observe articulation or acoustics as they occur, then analyze this concrete evidence of the speech production process. Conversely, where speech perception studies give researchers exact experimental control over the nature of their stimuli and the inputs to a subject’s perceptual system, research on speech production severely limits experimental control, making the investigators observe more or less passively while speakers do as they will in response to their prompts.

Such fundamentally different experimental conditions, along with focus on the opposite side of the perceptual coin, allows speech production research to ask different questions and draw different conclusions about spoken language use and speech communication. As we will discuss below, in some ways this “divide and conquer” approach has been very successful in expanding our understanding of speech as a whole. In other ways, however, it has met with many of the same roadblocks as its perceptual compliment and leaves many critical questions unanswered in the end.

A Different Approach

When the advent of the speech spectrogram made it obvious that the speech signal does not straightforwardly mirror phonemic units, researchers responded in different ways. Some, as discussed above, chose to question where the perceptual source of phoneme intuitions, trying to define the acoustics necessary and sufficient for an identifiable speech percept. Others, however, began separate lines of work, aiming to observe the behavior of speakers more directly. They wanted to know what made the speech signal as fluid and seamless as it appeared, whether the observed overlap and contextual dependence followed regular patterns or rules, and what evidence speakers might show in support of the reality of the phonemic units. In short, they wanted to demystify the puzzling acoustics seen on spectrograms by taking them in the context of their source.

The Continuing Search. It may seem odd, perhaps, that psychologists, speech scientists, engineers, phoneticians and linguists were not ready to abandon the idea of phonemes as soon as it became apparent that the physical speech signal did not straightforwardly support their psychological reality. However, one must take into account the incalculable gains the concept of segments and phonemes gave linguistic analysis, stretching back to Panini grammarian study of Sanskrit. Phonemes appear to capture the domain of many phonological processes, for example, and enable linguists to make sense of the multitude of distribution patterns of speech sounds across the world's languages. It has even been argued (Abler, 1989) that their discrete, particulate nature underlies humanity's immense potential for linguistic innovation, making "infinite use of finite means" (Humboldt, 1836).

Beyond these theoretical gains, phonemes found empirical support in research on speech errors or "slips of the tongue," which always appeared to operate over phonemic units. That is, the kinds of errors observed during speech production, such as anticipations ("a leading list"), perseverations ("pulled a pantrum"), reversals ("heft lemisphere"), additions ("moptimal number"), and deletions ("chrysanthemum p_ants"), appear to involve errors in the ordering and selection of whole segmental units, and always result in legal phonological combinations, whose domain is typically described as the segment (Fromkin, 1973).

In light of the particulate nature of linguistic systems, the enhanced understanding gained with the assumption of segmental units, and the empirical evidence observed in speech planning errors, therefore, researchers were and are reluctant to give up the search for the basis of phonemic intuitions in physically observable speech production.

Acoustic Phonetics

One of the most immediately fruitful lines of research into speech production focused on the acoustics of speech. This body of work, part of "Acoustic Phonetics," examines the speech signals speakers produce in great detail, searching for regularities, invariant properties and simply a better understanding of the human speech capacity. Although the speech spectrograph did not immediately show the invariants researchers anticipated, they reasoned that such technology would also allow them to investigate the speech signal at an unprecedented level of scientific detail. Since speech acoustics are so complex, invariant cues corresponding to phonemes may be present, but difficult to pinpoint.

While psychologists and phoneticians in speech perception were manipulating synthesized speech in an effort to discover the acoustic "speech cues," therefore, researchers in speech production refined signal processing techniques enabling them to analyze the content of naturally produced speech acoustics. Many phoneticians and engineers took on this problem, but perhaps none has been as tenacious and successful as Kenneth Stevens of MIT.

An electrical engineer by training, Stevens took the problem of phoneme-level invariant classification and downsized it, capitalizing on the phonological theories of Jackobson, Fant and Halle (1952) and Chomsky and Halle's *Sound Patterns of English* (1968) which postulated linguistic units below the level of the phoneme called distinctive features. Binary values of universal features such as [sonorant] [continuant] and [high], these linguists argued, constituted the basis of phonemes. Stevens and his colleagues thought that perhaps invariant acoustic signals corresponded to distinctive features rather than phonemes (Stevens, 1986; 2005). Since phonemes often share features (e.g. /s/ and /z/ share values for all distinctive features except [voice]), it would make sense that their acoustics are not as unique as we might otherwise expect.

Stevens, therefore, began a thorough search for invariant feature correlates that continued until his retirement in 2007. He enjoyed several notable successes: many phonological features, it turns out, can be reliably specified by one or two signal characteristics. Features specifying English vowels, for example, correspond closely with the relative spacing or distance between the first and second harmonic resonances of the vocal tract (or “formants”) during their production (Stevens, 2005).

Some features, however, remained more difficult to define acoustically. Specifically, the acoustics corresponding to consonant place of articulation seemed to depend heavily on context – the exact same burst of noise transitioning from a voiceless stop to a steady vowel might result from the lip closure of a [p] or the tongue-dorsum occlusion of a [k], depending on the vowel following the consonant. Equally problematic, the acoustics signaling the alveolar ridge closure location of coronal stop [t] are completely different before different vowels (Fowler, 1996). The articulation/acoustic mismatch is represented in Figure 5.

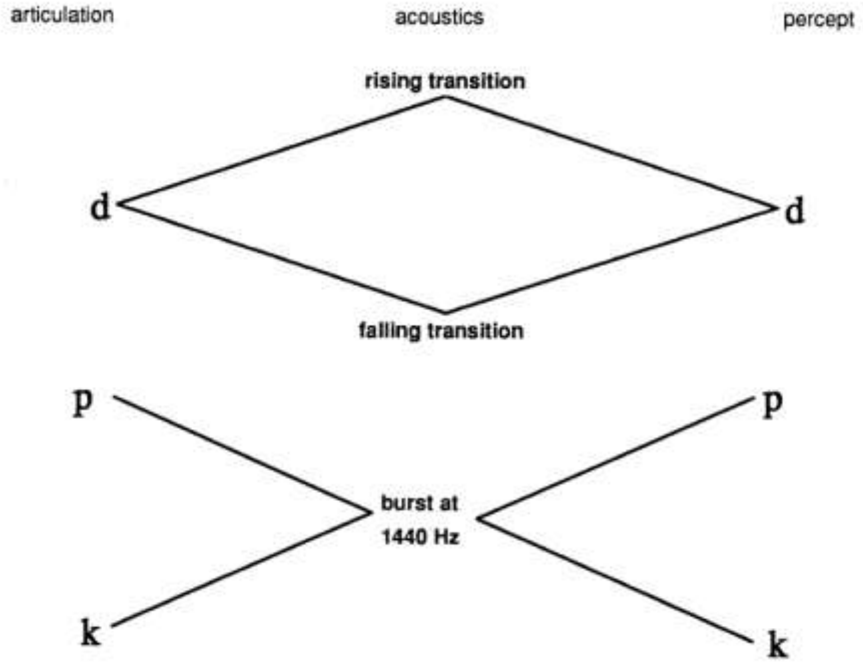

Figure 4. Observations from early perceptual speech cue studies. In the first case, two different acoustic signals (consonant/vowel formant frequency transitions) result in the same percept. In the latter case, identical acoustics (release burst at 1440 Hz) results in two different percepts, depending on the vocalic context. In both cases, however, perception reflects articulatory, rather than acoustic, contrast. (From CA Fowler. “Listeners do hear sounds, not tongues.” Journal of the Acoustical Society of America. 99(3);1730-41. 1996. Reprinted with permission.)

Variation in Invariants. Why do the acoustics and articulation follow these separate paths? Perhaps the answer becomes more intuitive when we consider that even the most reliable acoustic invariants described by Stevens and his colleagues tend to be somewhat broad, dealing in relative distances between formant frequencies in vowels and relative abruptness of shifts in amplitude and so on.

This dependence on relative measures comes from two major sources: individual differences among talkers and contextual variation due to co-articulation. Individual speakers' vocal tracts are shaped and sized differently, and therefore they resonant differently (just as different resonating sounds are produced by blowing over the necks of differently sized and shaped bottles), making the absolute formant frequencies corresponding to different vowels, for instance, impossible to generalize across individuals.

Perhaps more obviously problematic, though, is the second source: speech acoustics' sensitivity to phonetic context. Not only do the acoustics cues for [p], [t] or [k] depend on the vowel following the stop closure, for example, but because the consonant and vowel are produced somewhat simultaneously, the identity of the consonant reciprocally affects the acoustics of the vowel. Such co-articulatory effects are extremely robust, even operating across syllable and word boundaries. This extensive interdependence makes the possibility of identifying reliable invariance in the acoustic speech signal highly remote.

While some researchers, such as Stevens, attempted to factor out or neutralize these co-articulatory effects, others believed that they are central to the functionality of speech communication. Alvin Liberman, Franklin Cooper and their colleagues at Haskins Laboratories pointed out that co-articulation of consonants and vowels allows the speech signal to transfer information in parallel, transmitting messages more quickly than it could if spoken language consisted of concatenated strings of context-free discrete units (Liberman et al., 1967). Co-articulation therefore enhances the efficiency of the system, rather than being a destructive or communication-hampering force. Partially as a result of this view, some speech scientists focused on articulation as a potential key to understanding the reliability of phonemic intuitions, rather than on its acoustic consequences. They developed the research program called "articulatory phonetics," aimed at the study of the visible and hidden movements of the speech organs.

Articulatory Phonetics

Techniques. In many ways articulatory phonetics constitutes as much of an engineering challenge as a linguistic one. Since the majority of the vocal tract "machinery" lies hidden from view (see Figure 6), direct observation of the mechanics of speech production requires technology, creativity or both. And, any potential solution to the problem of observation cannot disrupt natural articulation too extensively if its results are to be useful in understanding natural production of speech. The challenge, therefore, is to investigate aspects of speech articulation accurately and to a high level of detail, while keeping interference with the speaker's normal production as minor as possible.

Various techniques have been developed that manage to satisfy these requirements, spanning from the broadly applicable to the highly specialized. Electromyography (EMG), for instance, allows researchers to measure directly the activity of muscles within the vocal tract during articulation via insertion of tiny electrode pins. These recordings have broad applications in articulatory phonetics, from determining the relative timing of tongue movements during syllable production to measures of pulmonary function from activity in speakers' diaphragms, to examining tone production strategies via raising and lowering of speakers' larynxes. EMG electrode placement can significantly impact articulation, however, which does impose limits on its use. More specialized techniques are typically still more disruptive of typical speech production, but interfere minimally with their particular investigational target. In transillumination, for example, a bundle of fiberoptic lights is fed through a speaker's nose until the light source is positioned just above their larynx (Lisker et al., 1969). A light-sensitive photocell is then placed on the neck just below the glottis to detect the amount of light passing through the vocal folds at any given moment, which correlates directly with the size of glottal opening over time. While transillumination is clearly not an ideal method to study the majority of speech articulation, however, it provides a highly accurate measure of various glottal states during speech production.

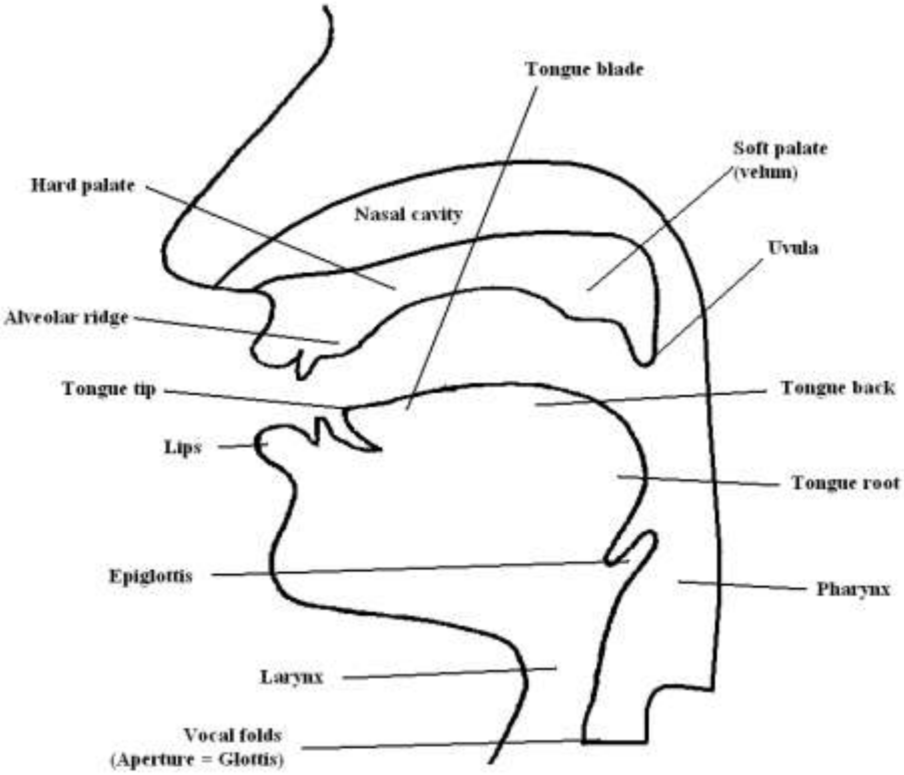

Figure 5. A sagittal view of the human vocal tract showing the main speech articulators as labeled. (From MC Cleary & DB Pisoni. "Speech Perception and Spoken Word Recognition." In Goldstein (ed.) Blackwell Handbook of Perception. 499-534. 2001. Reproduced by permission of: John Wiley & Sons, Inc.)

Perhaps the most currently celebrated articulatory phonetics methods are also the least disruptive to speakers' natural articulation. Simply filming speech production in real time via x-ray provided excellent, complete "snapshots" of moments of unobstructed articulation, but for health and safety reasons can no longer be used to collect new data. Methods such as x-ray microbeam and electromagnetic mid-sagittal articulometer (EMMA) attempt to surpass past "x-ray vision" by safely tracking the movements of speech articulators in real time. The former uses a tiny stream of x-ray energy to monitor the shadows created by radio-opaque pellets attached to a speaker's lips, teeth and tongue as they speak. The later, EMMA, generates similar positional data using the current generated by small electromagnetic coils attached to speaker's articulators as they pass through a magnetic field focused on the talker during speech. Both methods track the movements of articulators despite their inaccessibility to visible light, providing highly comparable position-over-time data for each point, and minimally disrupt natural speech (Nadler et al., 1987; Perkell et al., 1992). However, comparison across subjects can be difficult due to inconsistent placement of tracking nodes from one subject to another and simple anatomical differences between subjects.

Ultrasound provides another, even less invasive, articulatory phonetics technique and has been gaining popularity in recent years (e.g. Gick et al., 2005; Pouplier, 2008). Using portable machinery that does nothing more invasive than send sound waves through a speaker's tissue and monitor their reflections, speech scientists can track movements of the tongue body, tongue root and pharynx that even x-ray microbeam and EMMA cannot capture, as these articulators are all but completely manually inaccessible. By placing an ultrasound wand at the juncture of the head and neck below the jaw, however, images of the tongue from its root in the larynx to its tip can be viewed in real time during speech production, with virtually no interference to the speech act itself. The tracking cannot extend beyond cavities of open air, making this method inappropriate for studies of precise place of articulation against the hard palate or of velum movements, for example, but these are areas in which x-ray microbeam and EMMA excel. The data recently captured using these techniques is beginning to give speech scientists a more complete picture of speech articulation than ever before.

Impact on the Search for Phonemes. Unfortunately for phoneme-based theories of speech production and planning, the results of a recent articulatory study of speech errors do not seem to paint a compatible picture. As discussed above, the categorical nature of speech errors has served as important support for the use of phonemic units in speech production. Goldstein, Pouplier and their colleagues, however, used EMMA to track speakers' production of errors in repetition task similar to a tongue twister. They found that while speakers' articulation sometimes followed a categorically "correct" or "errorful" gestural pattern, it was more frequently somewhere between two opposing articulations. In these cases, small "incorrect" movements of the articulators would intrude upon the target speech gesture; both gestures would be executed simultaneously; or the errorful gesture would completely overshadow the target articulation. Only the latter reliably resulted in the acoustic percept of a speech error (Goldstein et al., 2007). As Goldstein and Pouplier point out, such non-categorical, gradient speech errors cannot constitute support for discrete phonemic units in speech planning.

Importantly, this finding was not isolated in the articulatory phonetics literature: speakers frequently appear to execute articulatory movements that do not result in any acoustic consequences. Specifically, x-ray microbeam tracking of speakers' tongue tip, tongue dorsum and lip closures during casual pronunciation of phrases such as *perfect memory* reveals that speakers raise their tongue tips for [t]-closure, despite the fact that the preceding [k] and following [m] typically obscure the acoustic realization of the [t] completely (Browman & Goldstein, 1990). Although they could minimize their articulatory effort by not articulating the [t] where it will not be heard, speakers faithfully proceed with their complete articulation, even in casual speech.

Beyond the Phoneme

So far we have seen that, while technological, methodological and theoretical advances have enabled speech scientists to understand the speech signal and its physical production better than ever before, the underlying source of spoken language's systematic nature remains largely mysterious. New research questions continue to be formulated, however, using results that were problematic under old hypotheses to motivate new theories and approaches to the study of speech production.

The theory of "Articulatory Phonology" stands as a prominent example; its proponents took the combination of gradient speech error data, speakers' faithfulness to articulation despite a lack of acoustic transmission, and the lack of invariant acoustic speech cues as converging evidence that speech is composed of articulatory, rather than acoustic, fundamental units (Browman & Goldstein, 1990; Goldstein & Fowler, 2003). Under this theory, linguistic invariants are underlyingly motor-based articulatory gestures which specify the degree and location of constrictions in the vocal tract. Constellations of these gestures constitute syllables and words, without reference to segmental or phonemic units. Speech perception, then, consists of determining the speech gestures responsible for the received acoustic signal, possibly through extension of experiential mapping between the perceiver's own gestures and their acoustic consequences, as in Liberman and Mattingly's Motor Theory of Speech Perception (1985) or Fowler's Direct Realist approach (1986). Recent evidence from neuroscience (Rizzolatti et al., 1996) may provide a biological mechanism for this process (see *Neurological Evidence – Mirror Neurons,* below).

And although researchers such as Stevens continued to focus on speech acoustics as opposed to articulation, the separate lines of inquiry actually appear to be converging on the same fundamental solution to the invariance problem. The most recent instantiation of Stevens' theory posits that some distinctive phonological features are represented by sets of redundant invariant acoustic cues, only a subset of which are necessary for recognition in any single token. As Stevens recently wrote, however, the distinction between this most recent feature-based account and theories of speech based on gestures may no longer be clear:

> The acoustic cues that are used to identify the underlying distinctive features are cues that provide evidence for the gestures that produced the acoustic pattern. This view that a listener focuses on acoustic cues that provide evidence for articulatory gestures suggests a close link between the perceptually relevant aspects of the acoustic pattern for a distinctive feature in speech and the articulatory gestures that give rise to this pattern.
> (Stevens, 2005, p. 142)

Just as in speech perception research, however, some speech scientists are beginning to wonder if the invariance question was even the right question to be asking in the first place. In the spirit of Lindblom's (1990) hyper- and hypo-articulation theory (see *Perception-Driven Adaptation in Speech Production*, below), these researchers have begun investigating control and variability in production as a means of pining down the nature of the underlying system. Speakers are asked to produce the same sentence in various contextual scenarios such that a target elicited word occurs as the main element of focus, as a carrier of stress, as a largely unstressed element, and as though a particular component of the word was misheard (e.g. in an exchange such as "Boy?" "No, *toy*"), while their articulation and speech acoustics are recorded. Then the data are examined for regularities. If the relative lengths of the onset consonant and following vowel remain constant across emphasized, focused, stressed and unstressed conditions, for example, that relationship may be specified in the representation of CV syllables, while the absolute closure and vocalic durations vary freely, and therefore must not be subject to linguistic constraint. Research of this type seeks to determine the articulatory variables under active, regular control and which (if any) are mere derivatives or side effects of deliberate actions (de Jong & Zawaydeh, 2002; de Jong, 2004; Silbert & de Jong, 2008).

Conclusion – Speech Production

Despite targeting a directly observable, overt linguistic behavior, speech production research has had no more success than its compliment in speech perception at discovering decisive answers to the foundational questions of linguistic representational structure or the processes governing spoken language use. Due to the joint endeavors of acoustic and articulatory phonetics, our understanding of the nature of the acoustic speech signal and how it is produced has increased tremendously, and each new discovery points to new questions. If the basic units of speech are gesture-based, what methods and strategies do listeners use in order to perceive them from acoustics? Are there testable differences between acoustic and articulatory theories of representation? What aspects of speech production are under demonstrable active control, and how do the many components of the linguistic and biological systems work together across speakers and social and linguistic contexts? Although new lines of inquiry are promising, speech production research seems to have only begun to scratch the surface of the complexities of speech communication.

Speech Perception and Production Links

As Roger Moore recently pointed out (2007), the nature of the standard scientific method is such that “it leads inevitably to greater and greater knowledge about smaller and smaller aspects of a problem” (p. 419). Speech scientists followed good scientific practice when they effectively split the speech communication problem, one of the most complex behaviors of a highly complex species, into more manageable chunks. And the perceptual and productive aspects of speech each provided enough of a challenge, as we have seen, that researchers had plenty to work on without adding anything. Yet we have also seen that neither discipline on its own has been able to answer fundamental questions regarding linguistic knowledge, representation and processing.

While the scientific method serves to separate aspects of single phenomena, however, the ultimate goal of any scientific enterprise is to unify individual discoveries, uncovering connections and regularities that were previously hidden. One of the great scientific breakthroughs of the nineteenth century, for example, brought together the physics of electricity and magnetism, previously separate fields, and revealed them to be variations of the same basic underlying principles. Similarly, where research isolated to either speech perception or production has failed to find success, progress may lie in the unification of the disciplines. And unlike electricity and magnetism the apriori connection between speech perception and speech production is clear: they are two sides of the same process, two links in Denes and Pinson’s famous “speech chain” (Figure 7) (1963). Moreover, information theory demands that whatever signals generated in speech production match those received in perception, a criteria known as “signal parity” which must be met for successful communication to take place; therefore, the two processes must at some point even deal in the same linguistic currency (Liberman & Whalen, 2000).

In this final section we will discuss theories and experimental evidence that highlight the deep, inherent links between speech perception and production. Perhaps by bringing together the insights won within each separate line of inquiry, the recent evidence pointing to the nature of the connection between them, and several theories of how they may work together in speech communication, we can point to where the most exciting new research questions lie in the future.

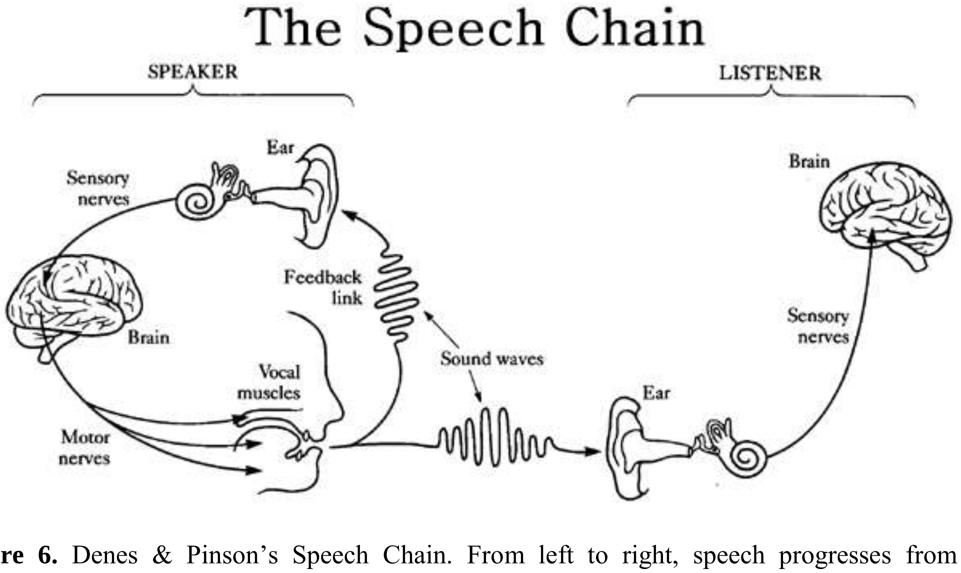

Figure 6. Denes & Pinson’s Speech Chain. From left to right, speech progresses from the “linguistic level” of the speaker, to his or her “physiological level,” to the “acoustic level,” into the

"physiological level" of the listener, and finally to the listener's "linguistic level." (From P Denes & E Pinson. The Speech Chain. 1963. Anchor Press/Doubleday. Reprinted by permission of W.H. Freeman & Company.)

Subtle Links – Phonetic Convergence

Recent work by Pardo builds on the literature of linguistic "alignment" to find evidence of an active link between speech perception and production in "real-time," typical communicative tasks (2006). She had pairs of speakers play a communication game called the "map task," where they must cooperate to copy a path marked on one speaker's map to the other's blank map without seeing one another. The speakers refer repeatedly to certain landmarks on the map, and Pardo examined their productions of these target words over time. She asked naïve listeners to compare a word from one speaker at both the beginning and end of the game with a single recording of the same word said by the other speaker. Consistently across pairs, the recordings from the end of the task were judged to be more similar than those from the beginning. Previous studies have shown that speakers align in their patterns of intonation (Gregory & Webster, 1996), for example, but Pardo's are the first results demonstrating such alignment at the phonetic level in an ecologically valid speech setting.

This "phonetic convergence" phenomenon defies explanation unless the processes of speech perception and subsequent production are linked within an individual. Otherwise, what a speaker hears his or her partner say could not affect subsequent productions. Further implications of the phenomenon become apparent in light of the categorical perception literature (see *Categorical Perception Effects*, above). In these robust speech perception experiments, listeners appear to be "deaf" to differences in acoustic realization of particular segments (Liberman et al., 1957). Yet the convergence observed in Pardo's work seems to operate at the sub-phonemic level, affecting changes within linguistic categories (i.e. convergence results do not depend on whole-segment substitutions, but much more subtle effects).

As Pardo's results show, the examination of links between speech perception and production has already pointed towards new answers to some old questions. Perhaps we do not understand categorical perception effects as well as we thought – if the speech listeners hear can have these gradient within-category effects on their own speech production, then why is it that they cannot access these details in the discrimination tasks of classic categorical perception experiments? And what are the impacts of the answer for various representational theories of speech?

Perception-Driven Adaptation in Speech Production

Despite the typical separation between speech perception and production, the idea that the two processes interact within individual speakers is not new. In 1990, Björn Lindblom introduced his "hyper- and hypo-articulation" (H&H) theory, which postulated that speakers' production of speech is subject to two conflicting forces: economy of effort and communicative contrast. The first pressures speech to be "hypoarticulated," with maximally reduced articulatory movements and maximal overlap between movements. In keeping with the theory's roots in speech production research, this force stems from a speaker's motor system. The contrasting pressure for communicative distinctiveness pushes speakers towards "hyperarticulated" speech, executed so as to be maximally clear and intelligible, with minimal co-articulatory overlap. Crucially, this force stems from listener-oriented motivation. Circumstances that make listeners less likely to correctly perceive a speaker's intended message – ranging from physical factors like presence of background noise, to psychological factors such as the lexical neighborhood density of a target word, to social factors such as a lack of shared background between the speaker and listener – cause speakers to hyperarticulate, expending greater articulatory effort to ensure transmission of their linguistic message.

For nearly a hundred years, speech scientists have known that physical conditions such as background noise do affect speakers' production. As Lane and Tranel neatly summarized, a series of experiments stemming from the work Etienne Lombard in 1911 unequivocally showed that the presence of background noise causes speakers not only to raise the level of their speech relative to the amplitude of the noise, but also to alter their articulation style in ways similar to those predicted by H&H theory (1971). No matter the eventual status of H&H theory in all its facets, however, this "Lombard speech" phenomenon empirically demonstrates a real and immediate link between what a speaker is hearing and the speech they produce. As even this very early work demonstrates, speech production does not operate in a vacuum, free from the influences of its perceptual counterpart; the two processes are closely linked.

Much more recent experimental work has demonstrated that speakers' perception of their own speech can be subject to direct manipulation, as opposed to the more passive introduction of noise used in inducing Lombard speech, and that the resulting changes in production are immediate and extremely powerful. In one experiment, for example, speakers repeatedly produced a target vowel [e] while hearing their speech only through headphones. The researchers ran the speech through a signal processing program which calculated the formant frequencies of the vowel and shifted them incrementally towards the frequencies characteristic of [æ], raising the first formant and lowering the second. Speakers were completely unaware of the real-time alteration of the acoustics corresponding to their speech production, but they incrementally shifted their articulation of [e] to compensate for the researchers' manipulation: they began producing lower first formants and higher second formants. This compensation was so dramatic that speakers who began by producing [e] ended the experiment by saying vowels much closer to [i] (when heard outside the formant-shifting influence of the manipulation) (Villacorta et al., 2007).

The researchers who conducted this experiment, Villacorta, Perkell and Guenther, point out that such "sensorimotor adaptation" phenomena demonstrate an extremely powerful and constantly active feedback system in operation during speech production. Apparently, a speaker's perception of his or her own speech plays a significant role in the planning and execution of future speech production.

The Role of Feedback – Modeling Spoken Language Use. In his influential theory of speech production planning and execution, Levelt make explicit use of such perceptual feedback systems in production (1999). In contrast to Lindblom's H&H theory, Levelt's model (WEAVER++) was designed primarily to provide an account of how lexical items are selected from memory and translated into articulation, along with how failures in the system might result in typical speech errors. In Levelt's model, a speaker's perception of their own speech allows them to monitor for errors and execute repairs. The model goes a step further, however, to posit another feedback loop entirely internal to the speaker, based on their experience with mappings between articulation and acoustics.

According to Levelt's model, then, for any given utterance a speaker has several levels of verification and feedback. If, for example, a speaker decides to say the word *day* the underlying representation of the lexical item is selected and prepared for articulation, presumably following the various steps of the model not directly relevant here. Once the articulation has been planned, the same "orders" are sent to both the real speech organs and a mental emulator or "synthesizer" of the speaker's vocal tract. This emulator generates the acoustics that would be expected from the articulatory instructions it received, based on the speaker's past experience with the mapping. The expected acoustics feed back to the underlying representation of *day* to check for a match with remembered instances of the word. Simultaneous to this process, the articulators are actually executing their movements and generating acoustics. That signal enters the speaker's auditory pathway, where the resulting speech percept feeds back to the same underlying representation, once again checking for a match.

Such a system may seem redundant, but each component has important properties. As Moore points out for his own model (see below), internal feedback loops of the type described in Levelt's work allow speakers to repair errors much more quickly than reliance on external feedback would permit, which translates to significant evolutionary advantages (2007). Without external loops backing up the internal systems, however, speakers might miss changes to their speech imposed by physical conditions (e.g. noise). Certainly the adaptation observed in Villacorta and colleagues' work demonstrates active external feedback control over speech production: only an external loop could catch the disparity between the acoustics a speaker actually perceives and his or her underlying representation.

As suggested above, Moore has recently proposed a model of speech communication that also incorporates multiple feedback loops, both internal and external (2007). His PRESENCE model ('PREdictive SENsorimotor Control and Emulation') goes far beyond the specifications of Levelt's production model, however, to incorporate additional feedback loops that allow the speaker to emulate the *listener's emulation of the speaker*, and active roles for traditionally "extra-linguistic" systems such as the speaker's affective or emotional state. In designing his model, Moore attempts to take the first step in what he argues is the necessary unification of not just research on speech perception and production, but the work related to speech in many other fields as well, such as neuroscience, automated speech recognition, text-to-speech synthesis and biology, to name just a few (2007).

Perhaps most fundamental to our discussion here, however, is the role of productive emulation or feedback during speech perception in the model. Where Levelt's model deals primarily with speech production, Moore's PRESENCE incorporates both speech perception and production, deliberately emphasizing their interdependence and mutually constraining relationship. According to his model, speech perception takes place with the aid of listener-internal emulation of the acoustic-to-articulatory mapping potentially responsible for the received signal. As Moore puts it, speech perception in his model is essentially a revisiting of the idea of "recognition-by-synthesis" (Halle & Stevens, 1962), while speech production is (as in Levelt) "synthesis by recognition."

Neurological Evidence – Mirror Neurons

The experimental evidence we considered above demonstrates pervasive links between what listeners hear and the speech they produce. Conversational partners converge in their production of within-category phonetic detail, speakers alter their speech styles in adverse listening conditions, and manipulation of speakers' acoustic feedback from their own speech can dramatically change the speech they produce in response. As we also considered, various theoretical models of spoken language use have been proposed to account for these phenomena and the observed perceptual and productive links. Until recently, however, very little neurobiological evidence supported these proposals. The idea of a speaker-internal vocal tract emulator, for instance, seemed implausible to many speech scientists; how would the brain possibly implement such a structure?

Cortical populations of newly-discovered mirror neurons, however, seem to provide a plausible neural substrate for proposals of direct, automatic and pervasive links between speech perception and production. These neurons "mirror" in the sense that they fire both when a person performs an action themselves and when they perceive someone else performing the same action, either visually or through some other (auditory, tactile) perceptual mode. Human mirror neuron populations appear to be clustered in several cortical areas, including the pre-frontal cortex, which is often implicated in behavioral inhibition and other executive function, and areas typically recognized as centers of speech processing, such as Broca's area (for in-depth review of the literature and implications, see Rizzolatti & Sinigaglia, 2008).

Neurons which physically equate (or at least directly link) an actor's production and perception of a specific action have definite implications for theories linking speech perception and production: they provide a potential biological mechanism The internal feedback emulators hypothesized most recently by Levelt and Moore could potentially be realized in mirror neuron populations, which would emulate articulatory-to-acoustic mappings (and vice versa) via their mutual sensitivity to both processes and their connectivity to both sensory and motor areas. Regardless of whether their specific applicability to Levelt and Moore's models, however, these neurons do appear to be active during speech perception, as one study using transcranial magnetic stimulation (TMS) demonstrates elegantly. TMS allows researchers to temporarily either attenuate or elevate the background activation of a specific brain area, respectively inducing a state similar to the brain damage caused by a stroke or lesion or making it so that any slight increase in the activity of the area causes overt behavior where its consequences would not normally be observable. The later excitation technique was used by Fadiga and colleagues, who raised the background activity of specific motor areas controlling the tongue tip. When the "excited" subjects then listened to speech containing consonants which curled the tongue upward, their tongues twitched correspondingly (Fadiga et al., 2002). Perceiving the speech caused activation of the motor plans that would be used in producing the same speech – direct evidence of the link between speech perception and production.

Perception/Production Links – Conclusion

Clearly, the links between speech perception and production are inherent in our use of spoken language. They are active during typical speech perception (TMS mirror neuron study), are extremely powerful, automatic and rapid (sensorimotor adaptation), and influence even highly ecologically valid communication tasks (phonetic convergence). Speech processing, therefore, seems to represent a linking of sensory and motor control systems; speech perception is not just sensory interpretation and speech production is not just motor execution. Rather, both processes draw on common resources, using them in tandem to accomplish remarkable tasks such as generalization from talker to talker, and acquiring new lexical items. As new information regarding these links comes to light, models such as Lindblom's H&H, Levelt's WEAVER++ and Moore's PRESENCE will both develop greater reflection of the actual capabilities of language users (simultaneous speakers and listeners) and be subject to greater constraint in their hypotheses and mechanisms. And hopefully, theory and experimental evidence will converge to discover how speech perception and production interact in the highly complex act of vocal communication.

Conclusion

Despite the strong intuitions of linguists and psychologists alike, spoken language does not appear to straightforwardly consist of linear sequences of discrete, abstract, context-free symbols such as phonemes or segments. This discovery begs the question, however: how does speech convey equivalent information across talkers, dialects and contexts? And how do language users mentally represent both the variability and constancy in the speech they hear?

New directions in research on speech perception include theories of exemplar-based representation of speech and experiments designed to discover the specificity, generalized application and flexibility of listeners' perceptual representations. Efforts to focus on more ecologically valid tasks such as spoken word recognition also promise fruitful progress in coming years, particularly those which provide tests of theoretical and computational models. In speech production, meanwhile, the apparent convergence of acoustic cue and articulatory theories of representation points to the emerging potential for exciting new lines of research combining their individual successes. At the same time, more and more speech scientists are turning their research focus towards variability in speech, and what patterns of variation can reveal about speakers' language-motivated control and linguistic knowledge.

Perhaps the greatest potential for progress and discovery, however, lies in continuing to explore the behavioral and neurobiological links between speech perception and production. Although made separate by practical and conventional scientific considerations, these two processes are inherently and intimately connected, and it seems that we will never truly be able to understand the human capacity for spoken communication until they have been conceptually reunited.

References

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social & Biological Structures*, *12*, 1-13.

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 341-376). Cambridge: Cambridge University Press.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper and Row.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 2, pp. 269-321). New York: Wiley.

Cooper, F. S., Borst, J. M., & Liberman, A. M. (1949). Analysis and synthesis of speech-like sounds. *Journal of the Acoustical Society of America, 21*(4), 461-461.

Cooper, F. S., Gaitenby, J. H., Mattingly, I. G., & Umeda, N. (1969, June 9-11). *Reading aids for the blind: a special case of machine-to-man communication.* Paper presented at the IEEE International Conference on Communications, Boulder, Co.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics, 32*(4), 493-516.

de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*(1), 53-75.

Denes, P., & Pinson, E. (1963). *The Speech Chain*. Garden City, NY: Anchor Press/Doubleday.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15*, 399-402.

Fowler, C. A. (1986). An event approach to the study of speech-perception from a direct realist perspective. *Journal of Phonetics, 14*(1), 3-28.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America, 99*(3), 1730-1741.

Fromkin, V. A. (1973). Slips of tongue. *Scientific American, 229*(6), 110-117.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110-125.

Gick, B., Bird, S., & Wilson, I. (2005). Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics & Phonetics, 19*(6-7), 503-514.

Goldiamond, I., & Hawkins, W. F. (1958). Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology, 56*, 457-463.

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

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: a phonology for public language use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production* (pp. 159-207). Berlin: Mouton de Gruyter.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units in speech production errors. *Cognition, 103*(3), 386-412.

Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accomodation and social status perceptions. *Journal of Personality and Social Psychology, 70*(6), 1231-1240.

Halle, M., & Stevens, K. N. (1962). Speech recognition: a model and a program for research. *IRE Transactions on Information Theory, 8*(2), 155-159.

Hockett, C. F. (1955). *A Manual of Phonology*. Baltimore: Waverly Press.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29*, 296-305.

Humboldt, W. v. (1836). *Linguistic Variability and Intellectual Development*. Baltimore, MD: University of Miami Press.

Jackobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to Speech Analysis: The Distinctive Features*. Cambridge, MA: MIT.

Johnson, K. A. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception*. Malden, MA: Blackwell Publishing, Ltd.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 4*(4), 677-709.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187-196.

Lindblom, B. E. F. (1990). Explaning phonetic variation: a sketch of H&H theory. In H. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling, NATO ASI Series D: Behavioural and Social Sciences* (Vol. 55, pp. 403-439). Dordrecht: Kluwer A.P.

Lisker, L., Abramson, A. S., Cooper, F. S., & Schvery, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America, 45*(6), 1544-1546.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62*(3), 615-625.

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

Moore, R. K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication, 49*, 418-435.

Nadler, R., Abbs, J. H., & Fujimora, O. (1987). *Speech movement research using the new x-ray microbeam system.* Paper presented at the 11th International Congress of Phonetic Sciences, Tallinn.

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

Oldfield, R. C. (1966). Things, words and the brain. *Quarterly Journal of Experimental Psychology, 18*, 340-353.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382-2393.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electromagnetic mid-sagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America, 92*, 3078-3096.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25*, 143-170.

Potter, R. K. (1945). Visible patterns of sound. *Science, 9*, 463-470.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics, 36*(1), 114-140.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain: How our Minds Share Actions and Emotions* (F. Anderson, Trans.). Oxford: Oxford University Press.

Silbert, N., & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America, 123*(5), 2769-2779.

Stevens, K. N. (1986). *Models of phonetic recognition II: a feature-based model of speech recognition.* Paper presented at the Montreal Satellite Symposium on Speech Recognition (Twelfth International Congress of Acoustics).

Stevens, K. N. (2005). Features in speech perception and lexical access. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception* (pp. 125-155). Malden, MA: Blackwell Publishing Ltd.

Villacorta, V. M., Perkell, J., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America, 122*(4), 2306-2319.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

The Relation Between Early Word Stress Discrimination and Later Lexical Development[1]

Danielle Elder[2,3], Carolyn Richie[2], and Derek M. Houston[3]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by a National Institute for Deafness and Other Communication Disorders (NIDCD) Research Grant (R01 DC006235) to Derek Houston and a Holcomb Undergraduate Grant for scientific research to Danielle Elder. Thanks are also due to the Infant Language Lab.

[2] Butler University, 4600 Sunset Ave. Indianapolis, IN 46208

[3] Indiana University School of Medicine, 699 West Drive, Indianapolis, IN 46202

The Relation Between Word Stress Discrimination and Later Lexical Development

Abstract: A key component of early intervention for children with delayed language acquisition is early assessment. Previous research has shown that English-learning infants' sensitivity to lexical stress plays a role in their segmentation of words from fluent speech – a critical step to developing a lexicon. This study investigated the possibility that performance on a word stress discrimination task predicts later lexical development for normal-hearing infants. A version of the visual habituation procedure (VHP) was used to test infants' ability to discriminate two-syllable words with trochaic versus iambic stress patterns (trochaic being most common in English). Infants were first habituated to recorded lists of words with either trochaic or iambic stress patterns and then presented with words of the habituated stress pattern and words of the novel stress pattern. Infants' attention towards the novel stress pattern was measured by looking duration in seconds and compared to their looking durations during the presentation of the habituated stress pattern. Longer looking times to the novel stress pattern suggests discrimination of the stress patterns. To assess lexical development, the MacArthur Bates Communicative Development Inventory (CDI) was administered at twelve months of age. Overall, infants showed a high degree of variability on both the discrimination task and the CDI measures. As a group, infants habituated to words following trochaic and iambic stress patterns but did not demonstrate the ability to discriminate between the novel and familiar stress patterns. There was no significant relationship between the ability to discriminate and the CDI measures. These results suggest that there is not a relationship between early word stress discrimination and later lexical development. Future work will aim to determine the degree of influence word stress discrimination has on later language acquisition.More specifically, future work will compare word stress discrimination ability in infant directed speech to later language development.

Introduction

There is growing consensus among language development experts that early assessment and intervention is key to reducing the long-term impact of language acquisition challenges. For example, Yoshinaga-Itano et al (1998) found that when identification of hearing loss occurred before six months of age, hearing-impaired children developed better language skills than when assessment occurred after six months. These findings suggest that, at least for hearing loss, "early intervention" means intervention during prior to six months of age. For infants who are diagnosed with severe-to-profound hearing loss, early intervention may involve cochlear implantation and intensive speech and language therapy. It also involves close evaluation of the effectiveness of the intervention strategy for developing age-appropriate speech perception and language skills. However, while there are reliable methodologies for assessing hearing loss in infants, there are no established methodologies for assessing infants' speech perception and language skills directly. Currently, clinicians must rely on parental reports for assessment, which are limited by the evaluation skills of the parents. Thus, developing methodologies for direct assessment of speech perception and language during infancy is crucial for equipping clinicians with tools for tailoring early intervention strategies to the specific needs of each child.

Recent research has identified some early speech perception skills that appear to be important for later language development. Tsao, Lie, & Kuhl (2004) examined the relation between phonetic discrimination at six months of age and later vocabulary. Using a conditioned head-turn paradigm, Tsao et al. measured English infants' ability to discriminate two computer-synthesized Finnish vowels, /u/ and

/y/ similar to the English vowels /u/ and /i/, and then assessed their vocabulary at 13, 16, and 24 months of age. They found that infants who required fewer trials to reach a criterion for phonetic discrimination had larger vocabularies at the later ages. In a follow-up study, Kuhl et al (2005) investigated the relation between 7-month-olds' discrimination of nonnative (Chinese) contrasts and later vocabulary growth. Previous research has shown that 6- to 10-month-olds begin to show more sensitivity to phonetic contrasts of their native language and less sensitivity to nonnative contrasts. They found that infants who showed a more mature pattern of discrimination – i.e., discriminated native but not nonnative contrasts – had greater vocabulary growth than infants who discriminated both types of contrasts. These findings suggest that milestones of speech perception development during infancy may predict later language outcomes and may be useful for identifying infants who are at risk for language disorders.

There are several potential speech perception skills acquired during infancy that may be useful for predicting later language difficulties. Newman and her colleagues recently compared several speech perception measures assessed during infancy to later language outcomes (Newman et al., 2006). They found that speech segmentation was a particularly important predictor; infants who demonstrated the ability to recognize familiarized words in fluent speech developed better language skills than infants who did not. These findings are consistent with mounting evidence that infants' ability to segment words from fluent speech is an important milestone in early language development. Thus, speech perception skills that are linked to speech segmentation may serve as important indicators of later language skills.

Speech Segmentation by Infants

Spoken language does not have consistent markers, such as pauses, that inform the listener of individual word boundaries. Therefore, the infant must somehow segment the continuous speech stream in order to comprehend individual words and develop a lexicon. Recently, researchers have attempted to identify the cues used by infants to segment the continuous speech stream including rhythmic, phonotactic, allophonic, and distributional cues (Houston & Jusczyk, 2000; Morgan & Saffran, 1995; Newman et al., 2006). Past research indicates that English-learning infants use rhythmic cues, especially word-level stress, to segment fluent speech. For example, Juscyzk, Cutler, & Redanz (1993) found that infants show a preference to trochaic (stressed-unstressed), bisyllabic words when contrasted with iambic (unstressed-stressed) bisyllabic words. In a follow up study by Jusczyk, Houston, and Newsome (1999) infants demonstrate a marked sensitivity to trochaic, bisyllabic words when segmenting fluent speech.

This sensitivity to lexical stress and preference for trochaic stress plays a role in infants' segmentation of words from fluent speech – a critical step to developing a lexicon. During participation in a heard-turn preference paradigm, 7.5-month-old infants were able to segment trochaic bisyllabic words from fluent speech following a short familiarization period. Errors were, however, made on iambic words because infants attended to the stressed second syllable. By 10-months-old, infants were correctly segmenting trochaic and iambic words from fluent speech (Jusczyk, Houston, & Newsome, 1999) suggesting that during development infants recognize lexical stress as an indicator of word onset. However, word stress is not the only useful cue to speech segmentation and research suggests that no one cue appears to carry more or less weight with regard to usefulness (Morgan & Saffran, 1995; Thiessen & Saffran, 2003; Turk, Jusczyk, & Gerken, 1995).

In a study by Turk, Jusczyk, & Gerken (1995), infants' use of syllable weight (tense vs. lax vowels) was compared to their use of word stress in speech segmentation. 9-month-old infants showed a preference for trochaic words over iambic words when syllable weight was not a factor. When syllable weight was included, infants continued to prefer trochaic words over iambic words. Infants also demonstrated a preference for stressed syllables with a tense vowel, but not with a lax vowel. The study suggests that syllable weight and word stress function independently, but infants do recognize trochaic words and heavy syllables as indicators of word onset. Another study explored speech segmentation while controlling for word stress and statistical cues (Thiessen & Saffran, 2003). At 6.5 months of age, infants attended equally to both trochaic and iambic words. However, by 9 months of age, infants segmented words at the stressed syllable whether it was the onset of a word (trochaic) or of the second syllable (iambic). Statistical familiarity and word stress proved useful to infants at different ages. These studies suggest that no one cue is more or less useful, but infants attend to a variety of cues at different points of language acquisition.

In summary, some researchers have shown that early abilities, such as phonetic discrimination, are predictive of later language development. Others have emphasized the usefulness of word stress in infant speech segmentation, and it has been determined that although various cues are useful at different points in language development no one cue stands out as more or less useful. However, in American English, word stress is a fairly reliable cue because of its frequent occurrence. Research has not explored the relation between word stress as a cue to speech segmentation and later language development.

This study was designed to explore two specific questions. First, do nine-month-old infants possess the perceptual ability to discriminate between two-syllable, American English (AE) words varying in stress? Second, is there a significant relationship between the ability to discriminate stress patterns of two-syllable words at nine months and later language production?

The purpose of this study was to provide new information about language development by focusing on the use of word stress as a cue used by infants to segment speech during the first year of life. Specifically, this study explored the relationship between infants' ability to discriminate between AE words with trochaic versus iambic stress and lexical growth. This study predicted that infants who were successful in discriminating different stress patterns, an ability related to word segmentation, would be more likely to develop a larger vocabulary more quickly when compared with their peers who did not demonstrate this discriminatory ability.

Methodology

Participants

Participants included 24 normal-hearing typically developing infants, aged 8 to 10 months (M=8.96) at the study onset; 13 were male (M=8.97) and 11 were female (M=8.91). Data collected from six of the participants are not reported, two participants completed a hearing test that revealed fluid in the ear and a mild hearing loss, and four infants did not complete the speech perception task for affective reasons. Thus, data from eighteen participants is reported for the speech perception task.

An adjusted birth date for any child born three or more weeks early or three weeks late was used to encourage developmental similarity among participants. For example, a child who was three weeks old is considered developmentally similar to a child who is three weeks late at birth. This adjusted birth date describes the child at his or her developmental age rather than chronological age.

In addition participants were determined to be normally developing with no physical, cognitive, or developmental delays expressed by the family pediatrician. All participants were found to have normal hearing as evidenced by a newborn hearing screening given at birth and a questionnaire given before testing. Infants with four or more ear infections since birth completed a hearing test performed by an audiologist at the time of study to confirm normal hearing ability and rule out a temporary conductive hearing loss.

Design

This study was completed in two parts. The first part was a speech perception task completed by the infant participants at nine months of age. The speech perception task tested infants' ability to discriminate between two different word stress patterns. The second part was a language development survey, the MacArthur-Bates Communicative Development Inventory (CDI) (Fenson et al., 1994) completed by the primary caregiver when the infant was twelve months of age. The CDI assessed infant participants' language acquisition according to the following categories: phrases understood, words understood or words understood and produced, early gesturing, and later gesturing.

Speech Perception Task

Stimuli

The stimuli for the speech perception task consisted of 552 words recorded in a professional studio by a linguistics student with previous experience recording similar stimuli. The talker was in her mid 20s at the time of recording. She was a native speaker of American English with no obvious regional accent. There were two types of stimuli: words with trochaic stress and words with iambic stress. The trochaic stimuli consisted of two-syllable words following a stressed-unstressed pattern. The iambic stimuli consisted of two-syllable words following an unstressed-stressed pattern. Each word within the trochaic condition had a counterbalanced word within the iambic condition. Words counterbalanced each other by having identical vowel sounds in the stressed syllable. For example, the word p*ea*nut in the trochaic condition was balanced with the word app*ea*se in the iambic condition. Frequency of word use was controlled by these methods; a sample of one-fourth of the stimuli was analyzed for frequency of use in infant-directed speech by using the CHILDES (child language data exchange system) database online at http://childes.psy.cmu.edu/. The CHILDES database provides the frequencies of words spoken to children by their parents. The frequency of the trochaic words was compared to the frequency of the iambic words using an unpaired t-test to determine that the relative frequency of occurrence was similar (Seshadri & Houston, 2004).

In the speech perception task, infants were habituated to words with trochaic or iambic stress, and then tested on discrimination between words with trochaic and iambic stress. Stimuli used in the habituation phase included 18 lists of 12 words with the trochaic stress pattern and 18 lists of 12 words with the iambic stress pattern. Stimuli used in the discrimination test phase included 14 lists of 12 words. Of these 14 word lists, 10 lists followed the habituated stress pattern (tokens with the same stress pattern from the pattern presented during the habituation phase). Each word list was 22 seconds in length allowing 1.83 seconds per word. The stimuli order within the lists was randomized using www.random.org, and each word was presented only once in the habituation phase. There were nine participants randomly assigned to each habituation condition, words with trochaic stress (participants SW1 through SW9) or words with iambic stress (participants WS1 through WS9). During the discrimination test, the words in the habituated stress pattern condition alternated between habituated words and novel words. The remaining four lists followed the novel stress pattern (words with a different stress pattern from the pattern presented during habituation) and alternated with words following the habituated stress pattern. Again, similar to the habituation phase, no word was presented twice within the discrimination phase.

Procedure

During the speech perception task, the infant participant and caregiver sat in a sound booth. The caregiver was asked to hold her child on her lap the entire time and asked to remain neutral during testing by not speaking, pointing, or moving themselves or their child around. The caregiver wore earplugs and headphones with music to reduce opportunities for biasing the infant's responses.

The walls of the sound booth were covered with a curtain façade, designed to ensure that the infant participant would not be distracted by the details of the larger room. Centered below a television screen was an audio speaker used to present acoustic stimuli. Slightly above the television was a small hole fitted with a camera, and the camera was hooded with a black cloth to minimize its appearance and reduce distraction. The infant was recorded with the video camera, which transferred the image to a computer in the control room where the primary experimenter was seated. The image was stored on the computer via Hack TV software, and the participants' looking time in seconds was recorded by software for Macintosh called Habit. The looking times were recorded as a result of the experimenter pressing the designated computer key when the participant looked in the direction of the acoustic stimuli, but the experimenter was blind to the stimuli of the study.

During the test, a smiling baby appeared on the screen between trials as the attention getter. The caregiver was instructed to use this time to situate herself and her child. This was also intended to help the child orient to the television if he or she became distracted during the study. During this time, the caregiver could point or say, "Look at the baby".

The habituation phase began with the spacey, a spinning figure played with noise, to attract the infant's attention. Once the child became bored with the spacey, the habituation trials began (a trial is one word list). During the acoustic stimuli, a blue and white checkerboard pattern appeared on the television screen. Attention towards the stimulus was measured in seconds of eye gaze towards the television. Habituation occurred when the child looked for one half the time as he or she did over the three longest trials. A trial ended during the experiment when the experimenter indicated that the child looked away from the television screen for more than one second.

Once habituated a participant moved on to the discrimination phase. Looking time was also measured in seconds during the discrimination phase. The length of time spent looking was recorded in seconds for each trial. The discrimination phase presented fourteen word lists to the participant, and a trial ended when a participant looked away for more than one second. The order of the word lists was randomized, and no two novel stress pattern lists were presented consecutively.

Language Development Survey

Following the speech perception task, caregivers were contacted again when infants reached 12 months of age and asked to complete the CDI. The CDI may be used to assess Early Words and Actions and Gestures. Early words include Phrases Understood, which reflect an infants' understanding of daily phrases and routines such as *get up*. Early words also included Words Understood and Words Produced, such as *mommy*, *doggie*, *milk*. The Actions and Gesture category assessed infants' early communicative and representational skills not dependent on verbal expression. These included early gestures, such as shaking head yes or no, and later gestures, such as pretending to drink from a toy cup or putting a baby doll to bed. Studies have shown the CDI to have reliability and validity (Fenson et al., 1994). To date, data has been collected for 13 participants who have completed the speech perception task and the survey.

Results

Habituation to words with trochaic or iambic stress

In order to determine whether infants became habituated to the stress patterns they were exposed to, habituation was measured over at least four trials. Each participant attended to at least 4 trials but up to as many as 18. As predicted all participants showed habituation (Figure 1) as demonstrated by reduced looking time across the habituation phase. The mean number of trials needed to become habituated to the stimuli was calculated for both groups. Participants in the trochaic condition needed on average 7.5 trials to become habituated, whereas participants in the iambic condition needed on average 8.8 trials to become habituated. The difference between groups, however, was not significant, which will be explained below.

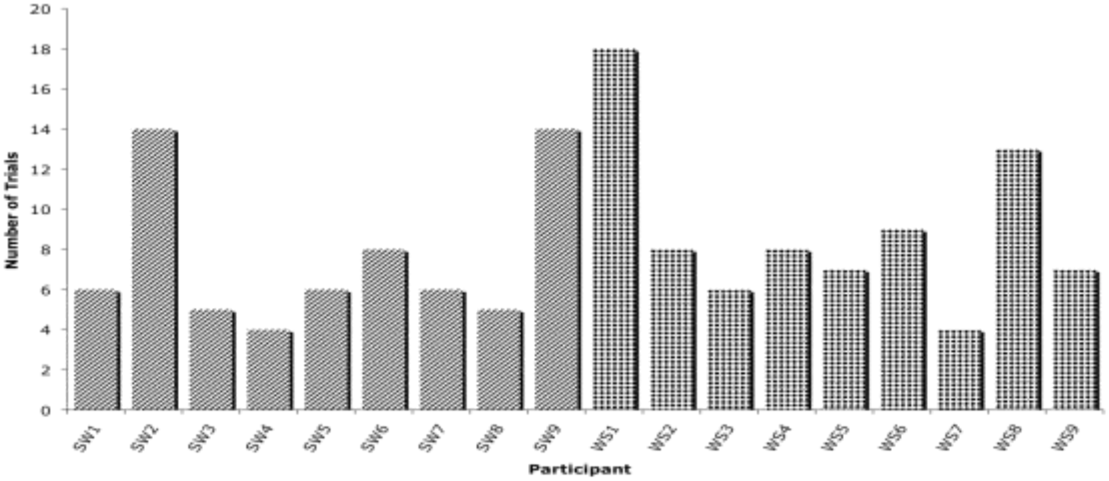

Figure 1. The number of trials needed to complete the habituation phase for each participant.

In order to compare across groups and determine whether or not the habituation condition affected the rate of habituation, the average looking time over the first four trials was measured in seconds (Figure 2). It was predicted that infants presented with words of trochaic stress would habituate more quickly than infants presented with words of iambic stress, because of the higher frequency of words with trochaic stress in AE. Participants in the trochaic condition looked on average over the first four habituation trials for 7.61 seconds, and participants in the iambic condition looked on average over the first four habituation trials for 10.67 seconds. However, the results of an independent t-test revealed the difference between groups to be non-significant.

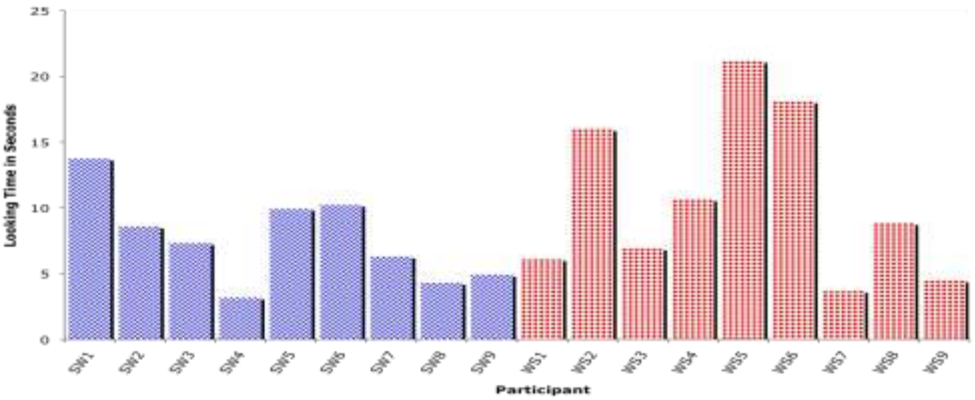

Figure 2. The rate of habituation as measured by average looking time in seconds over the first four habituation trials for each participant.

A one-way analysis of variance (ANOVA) was completed to determine if the habituation condition (habituation to words with trochaic or iambic stress) was related to the number of trials needed to become habituated to the stimuli, or the average looking time in seconds over the first four trials. Results of the ANOVA showed that there was no significant main effect of group on the number of trials needed to habituate, F(1,16)= .498, *p*= .491, or the average looking time in seconds over the first four habituation trials, F(1,16)=1.644, *p*= .218. Overall, the results showed that there were no significant differences between groups during habituation, and there were also no significant interactions between condition and habituation.

Discrimination between novel stress patterns and habituated stress patterns

In order to measure the infants' ability to discriminate between words with trochaic and iambic stress, the average looking time in seconds for words with the novel stress pattern and words with the habituated stress pattern was calculated. It was predicted that infants who discriminated between the different stress patterns would look longer to words with the novel stress pattern than the habituated stress pattern.

During the discrimination phase, infants habituated to words following the trochaic stress pattern were presented with novel words following the habituated trochaic stress pattern and words following the novel iambic stress pattern. They looked for 4.39 seconds on average to stimuli of the novel stress pattern compared to 3.57 seconds on average to stimuli of the habituated stress pattern (Figure 3). Six of nine participants who were habituated to the trochaic stress pattern looked longer to words with the novel, iambic stress pattern compared to the habituated, trochaic stress pattern. However, the difference in looking time between the habituated conditions was not significant, indicating that discrimination did not occur.

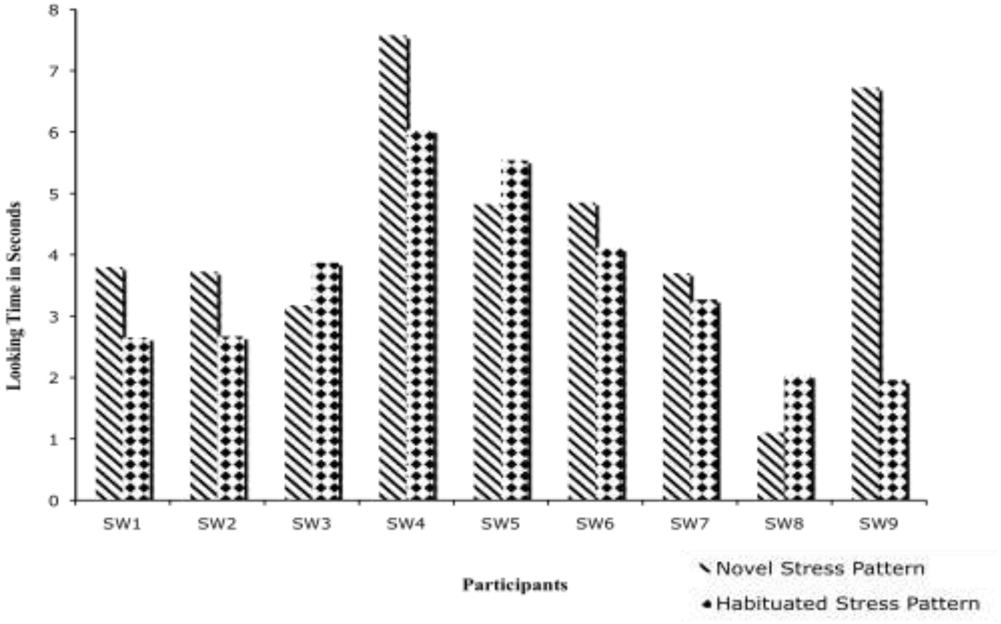

Figure 3. Average looking time in seconds for participants habituated to words with trochaic stress, for test stimuli with novel iambic stress (WS) and stimuli with habituated trochaic stress (SW).

Participants habituated to words with the iambic stress pattern were presented with words of the habituated iambic pattern, and words of the novel trochaic pattern. Unexpectedly, the group looked for 5.03 seconds on average to the novel, trochaic stimuli compared to 5.21 seconds on average to the trials containing the habituated, iambic stimuli (Figure 4). Only three of nine participants who were habituated to the iambic stress pattern looked longer to stimuli of the novel, trochaic stress pattern during the discrimination phase compared to the habituated iambic stress pattern. Again the difference in looking times between the conditions was not significant, which showed that infants did not discriminate between the different stress patterns.

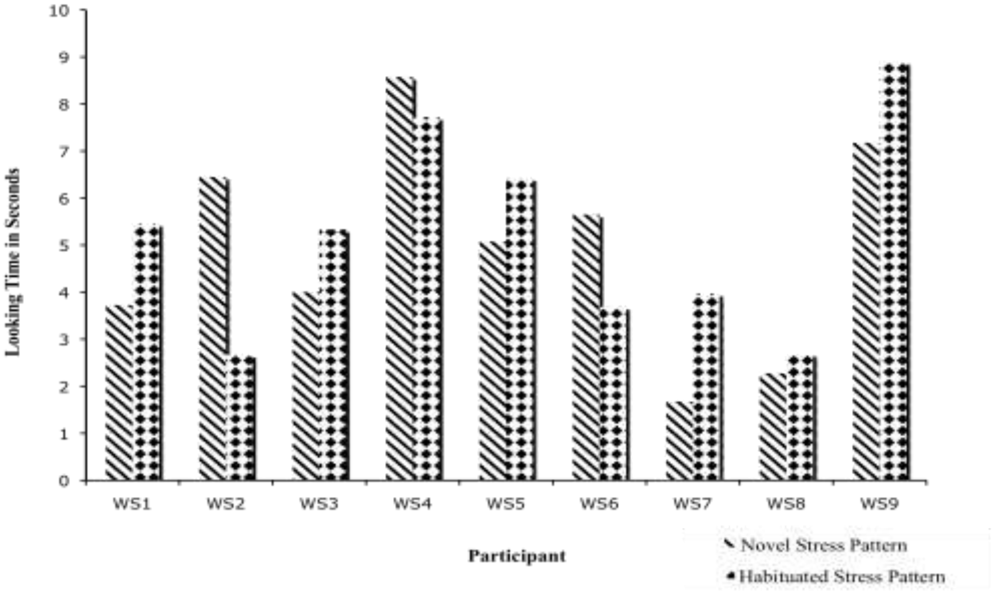

Figure 4. Average looking time in seconds for participants habituated to words with iambic stress for stimuli with novel trochaic stress (SW) and stimuli with habituated iambic stress (WS).

Differences between the average looking time for stimuli with novel stress and stimuli with habituated stress over all eighteen participants showed that half of the participants looked longer at the novel stimuli and half looked longer at the habituated stimuli. Two-thirds of the participants who looked longer towards words of the novel stress pattern compared to the habituated stress pattern were habituated to words with the trochaic stress (Figure 5).

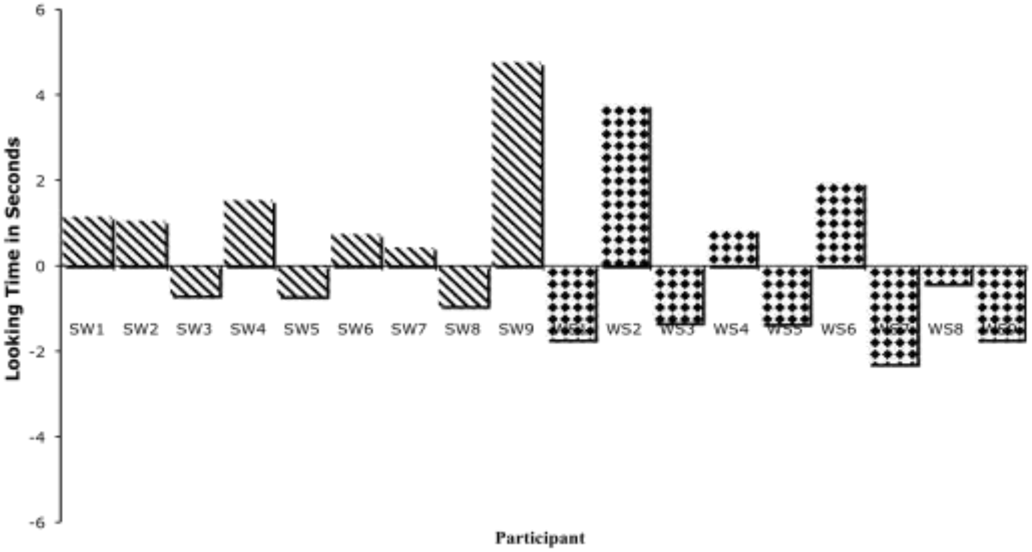

Figure 5. The difference between the average looking time in seconds for the stimuli with novel stress and the stimuli with habituated stress.

It was predicted that participants who discriminated would look longer towards stimuli following the novel stress pattern compared to stimuli following the habituated stress pattern. A paired t-test was performed using the average looking time in seconds to the stimuli with novel stress and the average looking time in seconds to the stimuli with habituated stress in order to determine discrimination. The paired t-test included all eighteen participants regardless of habituated condition. Results showed that as a group, infant participants did not look significantly longer to words with the novel stress pattern (M= 4.67, SD=2.06), compared to the habituated stress pattern (M=4.39, SD= 1.99), t(17)=.631, p>.05. A within groups paired t-test also showed no significant difference in looking times between the stimuli with novel stress and stimuli with habituated stress. It was concluded that infants did not discriminate between trochaic and iambic words.

MacArthur-Bates Communicative Development Inventory Data

These participants, ages 11.81 months to 13.06 months (M= 12.24) at the time the CDI data were collected, included seven males and seven females. Four infants who completed the speech perception task had not yet reached one year of age, and therefore their caregivers have not been given the CDI.

Parent reports on the CDI indicate that infant participants demonstrated a wide range of ability. Infants' raw scores for phrases understood ranged from 6 to 27 out of a possible total of 28 phrases. The range for words understood was from 16 to 388 out of a possible total of 396 words. Within the category of words produced there was a possible total of 396 words, and the infants' raw scores ranged from 1 to 52 words. Early gestures included 18 possibilities, and infants' scores were from 8 to 14. Later gestures included 45 possibilities, and infants' scores were from 2 to 25. The range for total gestures was 14 to 39 out of a possible total of 63.

Data from the present study were compared to normative data for twelve-month-olds from a study by Fenson, Thal, & Bates (1990). Comparisons are shown in Table 1 below. While the means were similar between studies the variance was very different between the two samples. For example, the category "phrases understood" has means of 16.63 in the present study and 15.8 in Fenson et al., but the standard deviations were 7.54 and 5.5 respectively. In the categories words understood, early gestures, and later gestures, both the means and the standard deviation were very different between the two samples. It seems that the data/participants tested in this study varied more than the data/participants tested in Fenson et al., perhaps for reasons such as a much smaller sample in the present study (n=13).

Participant	Phrases Understood	Words Understood	Words Produced	Early Gestures	Later Gestures	Total Gestures
SW1	15	7	40	30	23	20
SW2	40	96	72	55	55	50
SW3	78	83	96	65	55	55
SW4	60	59	93	75	83	80
SW5	93	63	55	65	73	73
SW6	25	20	53	20	23	18
SW7	80	55	62	85	50	60
WS1	30	29	43	75	75	75
WS2	40	75	50	55	23	30
WS3	98	86	86	95	60	73
WS4	35	27	43	55	20	25
WS5	99	99	62	75	80	78
WS6	15	14	25	65	6	13

Table 1. Individual percentile ranks for infant participants based on the CDI parent report.

Table 2 below shows the sample's mean raw score for each category assessed by the CDI. It demonstrates the wide range of inter-participant variability, means, and standard deviation for each measure assessed by the CDI.

	Current Study				Fenson et al.			
Category (Total possible)	Min.	Max.	M	SD	Min.	Max.	M	SD
Phrases Understood (28)	6	28	16.63	7.54	4	28	15.8	5.5
Words Understood (396)	16	388	118.2	106.21	7	263	84.9	52.5
Words Produced (396)	0	52	12.71	17.61	0	66	10.0	12.0
Early Gestures (18)	8	16	11.93	2.16	NA	NA	54.8	13.3
Later Gestures (45)	2	25	13.43	6.47	NA	NA	36	15.9

Table 2. Raw scores for infants in the present study and the normative data by Fenson et al. (1990).

Correlations between discrimination phase measures and CDI percentile ranks

It was originally predicted that there would be a significant correlation between infant participants' looking time difference in terms of effect size and the percentile ranks on the CDI. In order to compare the individuals' looking time differences between novel and habituated stimuli with other infants', the looking time differences were transformed to a measure of effect size, which accounted for variance and made all of the measures positive. Effect size is a measure of the strength of the relationship between attention to the novel stimuli and habituated stimuli. More specifically, it is an indicator of the amount of influence the novel stress pattern had on the looking time difference during the discrimination phase.

In order to calculate effect size, a resampling program designed by David Howell was downloaded from http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html. The resampling program randomized the obtained data from the sample 10,000 times and drew potential samples. The means from these 10,000 randomized sets of data were used to produce a histogram of the potential sampling distribution that could be drawn from the population. The t-test statistics were calculated for each participant's sampling distribution and provided the t score, p-value, and effect size for each participant. These statistics were similar to those of the sample. The effect size is shown in Figure 6.

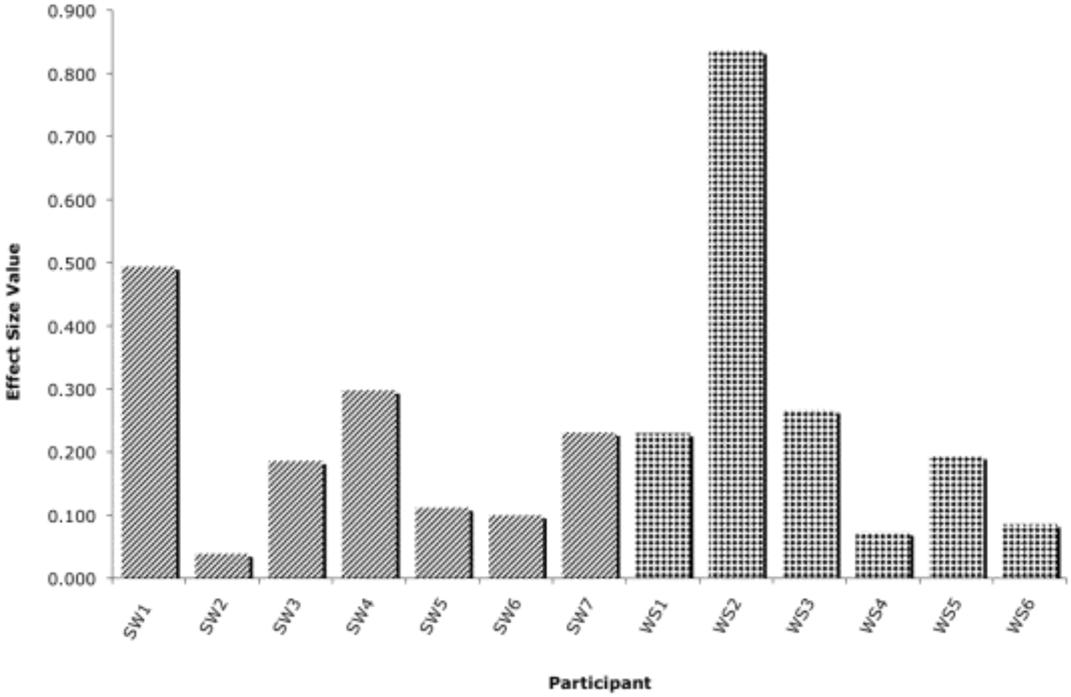

Figure 6. The looking time difference in seconds in terms of effect size for participants who completed the speech perception task and language survey.

Results showed no significant correlations between effect size and the following CDI measures: phrases understood (r= -.352), words understood (r= -.185), words produced (r= -.275), early gestures (r= -.287), later gestures (r= -.353), and total gestures (r= -.335). There was not a significant correlation between this early ability to discriminate and later lexical development, which is likely a result of infants' inability to discriminate between words of trochaic and iambic stress.

Discussion

This study examined whether nine-month-old infants' ability to discriminate between two-syllable AE words following trochaic and iambic stress patterns is predictive of later lexical acquisition. The study predicted that infants who demonstrated the ability to discriminate would develop a larger vocabulary more quickly than their peers who did not demonstrate this discrimination ability. The motivation for the prediction came from previous research that supported the relationship between early linguistic abilities and later language development. Specifically, a study by Juszcyk, Cutler, & Redanz (1993) showed that nine-month-old infants showed a preference for words following a trochaic stress pattern. The results of the present study, however, did not support the hypothesis that infants who discriminated between words of trochaic stress and words of iambic stress would develop a larger vocabulary more quickly than their peers who did not discriminate.

While participants in both groups habituated to trochaic or iambic lexical stress, there were no group differences in terms of the rate of habituation. Both groups habituated after a similar number of trials and with similar looking times in seconds. These measures demonstrated a range of variability within both conditions, but the differences between conditions were not significant.

Unexpectedly, the infants in both groups did not discriminate between the novel stimuli and the habituated stimuli, and the correlations between the discrimination ability and the CDI measures were also non-significant. Bivariate correlations were, however, completed to determine if there was a relationship between infant participants' looking times to iambic and trochaic word stress patterns and their language acquisition as measured by the CDI. Measures included the difference between the average looking time to words with novel stress and the average looking time to words with the habituated stress pattern, and the CDI percentile ranks.

Post-hoc analyses for discrimination ability and CDI measures

The inability to discriminate between word stress patterns seen in the present study was inconsistent with Jusczyk et al.'s study (1993) that suggested nine-month-old infants prefer the trochaic stress pattern. In order for infants to prefer one stress pattern they would first need to demonstrate the ability to discriminate between two stress patterns. Boredom may have presented as a possible confound. It was eliminated as a confounding variable by analyzing the first two relevant trials only, but still no significant difference was obtained. Infants did not discriminate between the novel stress pattern (M= 5.49, SD= 5.41) and the habituated stress pattern (M= 5.61, SD= 5.03), t(17)= -.074, p= .942.

Conclusions

Three main conclusions were drawn from this research study. First, nine-month-old infants became habituated to AE words following trochaic and iambic stress patterns relatively quickly and without between group differences. However, when presented with words of a novel stress pattern following habituation, infants could discriminate between words of the trochaic and iambic stress patterns. Finally, early word stress discrimination was not predictive of later lexical development for the entire group. These unexpected results may have been due to boredom, a small sample, or stimuli presentation methods. Future work will aim to determine whether a larger sample will yield different results and whether replicating the study by Jusczyk et al., (1993), which demonstrated preference for the trochaic stress in nine-month-olds, will show a relationship between early word stress discrimination and later lexical development.

Future Directives

The sample for this study was small and increasing the sample has the potential to change the results, and make them more reflective of the population and more similar to the normative data with regard to the CDI. Providing a larger sample in a study reduces variability effects on results. Also, as infant participants get older, more CDI measures will be collected providing longitudinal data and allowing the opportunity for developmental patterns to emerge.

This study did not cover a large time span. The age of infant participants at the study onset was nine-months-old, and the first CDI measures were taken at 12-months-old. CDI measures are currently being collected from caregivers when infant participants reach 15 months of age. More distinct developmental patterns may emerge in the CDI measures following this additional data collection.

The results derived from the infants who looked longer in seconds to words of novel stress were inconsistent with the prior results that included all infant participants, which suggested that there was no relationship. Modifying the methodology and re-running the experiment may yield different results. For instance, replicating Jusczyk, Cutler, & Redanz (1993) experiment that showed nine-month-old infants' preference for trochaic patterned words, completing CDI measures, and running correlations could lead to a different outcome. Replicating the study by Jusczyk et al. (1993) would involve presenting infants with words of both trochaic and iambic stress patterns during habituation and then novel words of both stress patterns during the test phase to determine whether or not infants demonstrate preference. The correlations would include infants who showed a preference for words of trochaic stress, infants who showed a preference for iambic stress, and infants who did not show a preference for either stress, related to their CDI percentile ranks.

Finally, work is currently underway to partially replicate the present study using infant directed speech as well as adult directed speech. Infant directed speech is when the talker raises the pitch and increases the vowel duration of her speech making it more animated and interesting to the infant listener. It is predicted that infants will respond differently to the infant directed speech compared to the adult directed speech. More specifically, it is predicted that infants will discriminate between trochaic and iambic word stress when spoken in infant directed speech. If infants show the ability to discriminate between trochaic and iambic stress in infant directed speech, then a CDI language survey will be completed to test for any correlation between word stress discrimination and later lexical development.

References

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59(5),* i-185.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64,* 675-687.

Jusczyk, P.W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159-207.

Kuhl, P.K., Conboy, B.T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the "critical period". *Language Learning and Development, 1(3&4)*, 237-264

Morgan, J.L., & Saffran, J.R. (2005). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Development*, *66*, 911-936.

Newsom, M. & Jusczyk, P.W. 1995. Do infants use stress as a cue in segmenting fluent speech? D. Maclaughlin & S. McEwen (Eds.) *Proceedings of the 19th Annual Boston University Conference on Language Development, 2*, 415-426 Somerville, MA: Cascadilla Press.

Newman, Ratner, Jusczyk, A., Jusczyk, P., & Dow. 2006. Infants' early ability to segment the conversational speech signal predicts later language development: retrospective analysis. *Developmental Psychology*, *42(4)*, 643-655.

Seshadri, P. & Houston, D. (year). Sensitivity to rhythmic properties of words in normal hearing infants and deaf infants who use cochlear implants. Indiana University School of Medicine, Indianapolis, IN.

Thiessen and Saffran. 2003. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39(4),* 706-716.

Tsao, F., Liu, H., & Kuhl, P.K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development, 75(4),* 1067-1084.

Turk, Jusczyk, & Gerken. 1995. Do English-learning infants use syllable weight to determine stress? *Language and Speech, 38(2),* 143-158.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Dynamic Modeling Approaches for Audiovisual Speech Perception and Multisensory Integration[1]

Nicholas Altieri

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] The research described in this chapter was supported by NIH-NIDCD Research Grant DC00111 and NIH-NIDCD Training Grant DC00012.

Dynamic Modeling Approaches for Audiovisual Speech Perception and Multisensory Integration

Abstract. Multimodal information including auditory, visual and even haptic information is integrated during speech perception. Articulatory information provided by a talker's face enhances speech intelligibility in congruent and temporally coincident signals, and produces a perceptual fusion (e.g. the "McGurk effect") when the auditory and visual signals are incongruent. This paper focuses on promising dynamical approaches for modeling representational as well as time-based accounts of audiovisual speech integration. In the first part of the paper, the articulatory dynamic approach of Browman and Goldstein is reviewed as a theoretical account of how perceptual representations are operated on by the cognitive system during integration. In the second part, a linear dynamic model (a two dimensional linear differential equation) is proposed as a model for 'early' versus 'late' integration in audiovisual speech perception. Simulations of the linear dynamic model provide a valuable tool for constraining and falsifying broad classes of models of integration while additionally allowing researchers to test *processing capacity*—a statistical measure specifying integration efficiency when visual information is presented in conjunction with auditory information.

Introduction

The dominant modality in speech perception for normal hearing listeners is the auditory modality. In spite of the dominance of the auditory modality, communication between talkers is often face-to-face where listeners are able to obtain relevant information from the visual modality (e.g., Erber, 1969; Sumby & Pollack, 1954). In a pioneering study, Sumby and Pollack (1954) showed that listeners obtain a gain of up to 15 dB in speech intelligibility in low signal-to-noise (S/N) ratios when they are able to perceive the talker's face.

In addition to providing redundant signal information and filling in missing information when the auditory signal is degraded or otherwise incomplete, incongruent auditory and visual signals contribute to a perceptual fusion. McGurk and MacDonald (1976) demonstrated that when a visually articulated /ga/ is combined with an auditory /ba/, subjects experience the perceptual fusion of /da/. Fowler and Dekle (1991) also observed the McGurk illusion when incongruent auditory and haptic stimuli were presented to participants. In another application of incongruent speech using temporally incongruent auditory and visual stimuli, Green and Miller (1985) demonstrated that the rate of visually articulated speech influences the perception of stop consonants.

Although phenomena in audiovisual perception are ubiquitous and interest in multimodal perception has increased in recent years, there is little agreement, as to how the auditory and visual components of the speech signal interact during the integration process. While speech is a continuous physical stream unfolding in real time, little is known about how the dynamics involved in audiovisual speech integration be accounted for within a rigorous mathematical framework. One pertinent question concerns how the information in the auditory and visual streams is represented by the cognitive system during the integration process. Another important yet unanswered question related to the time course of integration concerns whether integration occurs 'early'–prior to syllable or word recognition, or at a later stage in the recognition process (see Bernstein, 2005; Rosenblum, 2005; and Summerfield, 1987, for a theoretical treatment of that issue). This paper will assess dynamic modeling strategies pertaining to both issues.

Summerfield (1987) discussed five possible theories of audiovisual speech integration: (1) integration of phonetic features (VPAM), (2) the filter function of the vocal tract (2), vectors describing the values of independent acoustical and optical parameters, (4) static articulatory configurations, and (5) articulatory dynamics of the vocal tract. The first part of this paper will focus on evaluating the merits of *articulatory dynamics of the vocal tract* as a parsimonious metric of audiovisual integration in speech perception. The dynamical approach to speech perception will be reviewed and compared with phonology based approaches. Next, the theoretical implications of articulatory dynamics in the context of audiovisual speech perception will be discussed. In the course of the discussion, the mass spring system will be presented as a generic and illustrative model of modality independent perception.

In the second part of the paper, a two dimensional linear dynamic parallel model with separate auditory and visual channels, as well as a coactive model (where the channels are pooled together), are proposed as a models of early and late integration.

Articulatory Dynamics: Speech Perception as a Dynamical System

Speech perception has traditionally been conceptualized as a process where the multidimensional waveform from an utterance is translated by the listener's cognitive system into a series of discrete, context free symbols. The idea is that the cognitive machinery involved in perception operates like a digital computer or symbol processor on discrete phonological representations. An analogy used to describe the translation of the waveform into phonological units during spoken word recognition is Hockett's famous Easter egg analogy (Hockett, 1955). Imagine Easter eggs on a conveyor belt approaching a wringer, which smashes the eggs and mixes them together. In the analogy, the eggs represent the speaker's discrete and symbolic phonological representations and the wringer represents the physical processes involved in taking the representations and translating them into a continuous time varying acoustic signal. The listener must reconstruct the phonological representations from the intermixed mess of colored eggshells. Reconstructing the eggs—or the speech signal in its original form is strikingly difficult because the representations cannot be seen or directly inferred from the continuous, time variable acoustics.

The Easter egg analogy provides a naïve account of the cognitive processes involved in continuous spoken word recognition. In Hockett's analogy, a listener must infer the configuration of the discrete phonological structure involved in producing the utterance from a continuous and multi-dimensional speech waveform. Hence, there is a fundamental distinction between the physical mechanism that produces speech, and the cognitive system that decodes and comprehends the utterance. Browman and Goldstein (1995) have pointed out that research investigating the cognitive aspects of speech perception has proceeded largely independent from research on the physical aspects of speech perception. A unified theory of speech perception, they argue, requires finding a way to *translate* between the physical and cognitive realms.

In spite of the popularity of formalist approaches in the field of linguistics (e.g., Chomsky & Halle, 1968), Browman and Goldstein (1995), among others, have argued that it suffers from several flaws. For one, the phonetic and phonological descriptions of language ignore important dynamical aspects, including timing. The general character of arguments against the formalist approach have to do with the fact that the time variable speech signal, and internal representations of lexical items, share little in common with symbolic time invariant phonemes (Port & Leary, 2005; see also Elman, 1995, for a dynamical account of lexical access).

Speech perception and production occur in real time, and a theoretical account should, *a priori*, account for timing mechanisms in cognitive processes and linguistic representations (at least in the production process). Browman and Goldstein's *Articulatory Phonology* diverges from the formalist approach by assuming that the fundamental units of speech are dynamic rather than static (Browman & Goldstein, 1986; 1989; 1995). According to this framework, the fundamental unit of speech perception is the articulatory gesture (see Gentilucci & Corballis, 2006). Unlike static time-invariant symbols however, articulatory gestures are specified via the parameters of a dynamical system. The dynamics of the vocal tract include opening and closing of the lips, movements of the tongue, as well as the velum and glottis. The system of equations required to adequately describe a vocal tract would require numerous parameters and dimensions, where approximating or solving the system would be computationally intensive if not entirely impossible. That is why, as we shall see, a lower dimensional description of a speech signal is required.

The theory of articulatory phonology was developed to unify the physical and cognitive aspects of the system by arguing that the fundamental difference between the cognitive and physical components of the system is a question of dimensionality. Speech production and comprehension, on one hand, involve the input signal produced by a complex dynamical system. The internal cognitive representations of speech appear to comprise fewer degrees of freedom than the physical mechanisms required for production. The crux of Browman and Goldtein's theory of articulatory phonology is construing a relation between the physical and the cognitive components of speech. They argued that the physical and cognitive aspects of speech are two components of the same system, or two different levels of description—a *macroscopic* and *microscopic* (Browman & Goldstein, 1995). The macroscopic aspects of the system are low dimensional constructs with few degrees of freedom from which phonological units arise. The microscopic aspects of the utterance are higher dimensional, and contain far more degrees of freedom since it contains detailed characteristics regarding the physical mechanisms that produced the utterance. Therefore, the phonetic description of language does not contain enough physical or dynamical information.

Extending Articulatory Dynamics to the Visual Modality

As mentioned earlier, the dominant modality used to perceive speech in normal hearing adults is the auditory modality. However, there is a considerable body of evidence showing that speech perception is a multimodal function relying heavily, at times, on the visual modality (Erber, 1969; Sumby & Pollack, 1954; Summerfield, 1987).

The theory that the primitives or representations of speech perception are articulatory in nature can be naturally extended to the visual domain. In the theoretical treatment of speech developed thus far, speech perception involves the perception of articulatory gestures unfolding in real time, and accessing stored (lower dimensional) representations that describe properties of the dynamical system responsible for producing the utterance. Extending the dynamical theory of speech perception to the visual domain requires a description of how the system encodes time varying information provided by movements from a talker's face.

Summerfield (1987) described an account of how listeners integrate auditory and visual information obtained from the *articulatory dynamics* of a talker. In Summerfield's account of *articulatory dynamics*, listeners extract information from both the auditory and visual modality by accessing *modality free* information afforded by the kinematic patterns of the talker's vocal tract. *Modality free* information refers to the fact that whether speech is heard auditorially or perceived visually, the dynamics of the talker's vocal tract is perceived rather than abstract modality dependent information such as auditory phonological representations or visemes in the visual domain (see Summerfield, 1987 for a more in depth analysis of the *visual place, auditory manner* and other alternative accounts of integration).

To conceptualize the notion of modality free information, including how articulatory dynamics provides a theoretical foundation for audiovisual integration, consider a hypothetical talker with a simple vocal tract. Imagine the hypothetical vocal tract moving its lips sinusoidally at a frequency of 4 Hz. This would sound like a sequence of /ma/s produced at a rate of 4 per second. The frequency of oscillation is accessible in both the auditory and visual channels, and therefore, 'modality free' according to Summerfield (1987). The rate of amplitude modulation and the rate at which the first formant F1 peaks are accessed through the auditory channel. The oscillations made by the simplified vocal tract are, of course, visible to the listener.

While the vocal tract described above is vastly over-simplified and contains few parameters (i.e., no parameters for tongue position, or glottal opening), it is nonetheless an appropriate model for describing the dynamics of real talkers, at least in a very limited sense. The model takes into account the close correspondence between the lip-movements of real talkers and the amplitude of the speech waveform. Lip movements, for example, can be sinusoidal (Kelso, Bateson, Salzman, & Kay, 1985). Syllabic rhythm produced by this simple device is provided to the listeners both auditorially and visually as a sinusoidal pattern.

The dynamics of the hypothetical talker can be readily elucidated by the second order linear differential equation used to describe the mass spring system (see e.g., Summerfield, 1987). Parameters including *mass, force, viscosity,* and *stiffness* are controlled and manipulated by a talker to carry out certain movements of the vocal tract. The second order equation can be written as:

$$m*y'' + b*y' + k(0 - y) = F(t) \qquad (1)$$

where *m* represents the mass, *F(t)* is the impulse response function, *b* is the viscosity, and *k* is the stiffness. The mass can be divided out of the system to yield a simplified, mass normalized second order equation given below, where the parameters A and B represent the mass normalized viscosity and stiffness respectively:

$$y'' + A*y' + B(0 - y) = G(t) \qquad (2)$$

The critical point of this second order equation can behave either as a periodic attracter, or a point attracter depending on the values of the parameters. Oscillatory movements for instance, can be produced by the movement of limbs between two target states, or by lips as described in our example (e.g. the hypothetical talker producing a series of /ma/s). This situation occurs when the forcing function (G(t) or F(t)) injects energy into the system which adequately compensates for the energy lost as a result of friction. In this scenario, the forcing function can be eliminated and written as 0 yielding the simplified equation given below:

$$y'' + B(0 - y) = 0 \qquad (3)$$

If however, $0 < A/2 \leq B$, the system will tend toward its resting length with *oscillation* or *critical dampening*. In the former case, the talker's lips will oscillate at increasingly lower amplitudes before settling down to the resting state. In the latter case, the lips will not overshoot the target state and oscillate, but instead will tend directly toward the target point. Fowler, Rubin, Remez, and Turvey (1980) and Summerfield (1987) argued that a critically damped mass spring system is an appropriate model for describing complex tasks like speech production, which require the speaker to achieve different target states (determined by previous states in the system) in rapid succession.

The articulatory dynamic account of audiovisual speech perception serves as an extension of Browman and Goldstein's articulatory phonology to the visual domain by allowing for a hypothetical lip-reader to visually perceive the kinematic patterns of the vocal tract. Of course, the simplicity of the vocal tract in the hypothetical talker poses a potential problem for this account of audiovisual speech perception. Natural vocal tracts are far more complex containing numerous degrees of freedom where recovering the relevant dynamic parameters is potentially intractable.

Articulatory dynamics is nonetheless a useful theoretical construct inasmuch as it provides a parsimonious account of audiovisual speech perception and integration. The information specified in the auditory and visual channels contains identical dynamic information—which allows the speech perception function to have access to the same parameters during the conflux of auditory and visual speech information. There are numerous plausible alternatives to the articulatory dynamical account of audiovisual speech perception, although this theory is useful inasmuch as it provides a starting point for investigating dynamical aspects of audiovisual information processing.

Audiovisual Speech Integration as a Dynamical Process

Another important question related to audiovisual speech integration concerns how the cognitive system (or the *black box*) operates on the inputs and integrates information from the auditory and visual channels in real time. Time varying information in both the auditory and visual modalities enters the black box, which then formulates an output. However, the intermediate states where the auditory and visual channels operate on the continuous inputs, needs more careful consideration in terms of rigorous mathematical language.

One way to address the question of how the cognitive system operates on the information contained in the auditory and visual streams is to consider whether integration occurs *early* in processing (prior to phonetic categorization or word recognition), or *later* in processing (after phonetic categorization or word recognition). This debate has recently surfaced in the audiovisual perception literature in separate reviews and analyses carried out by Rosenblum (2005), and Bernstein (2005). Rosenblum reviewed a considerable body of work ranging from behavioral studies (e.g. Green & Miller, 1985) to neuroimaging studies, and argued that audiovisual integration occurs in the early stages of information processing, prior to phonetic categorization or word recognition. In this framework, the primitives of speech perception derived from the auditory and visual channels are generally conceptualized as amodal–meaning that modality independent gestural information is integrated. Cognitive processes, as a result of obtaining this amodal information, are able to readily combine the auditory and visual modalities in early stages of processing.

Bernstein (2005) however, interprets evidence obtained in behavioral and neuroimaging studies differently. For instance, Bernstein claims that neuroimaging studies (e.g. those conducted by Calvert, Campbell, and Brammer (2000)) do not have the temporal and spatial resolution that would be required to show that audiovisual speech is integrated early (prior to word recognition), or that large populations of neurons in language processing areas respond preferentially to audiovisual stimuli. Hence, rather than argue for early integration of the auditory and visual streams, Bernstein (2005) claims that current evidence suggests that the auditory and visual streams are processed simultaneously but separately, and integrated in later processing stages (i.e., if they are combined then perhaps it is post perceptual). That is, extensive unimodal processing occurs on the respective auditory and visual features in audiovisual speech perception.

Regardless of how the ensuing debate is resolved, it is important to begin couching the discussion in terms of mathematically rigorous terminology. The following section will consider a two dimensional linear dynamic model with hypothetical auditory and visual inputs. The model is capable of making unique architectural predictions for early integration of the auditory and visual channels (i.e., *coactive processing*), and late integration (*parallel processing* where channels are processed separately and simultaneously) by utilizing reaction time (RT) distributions collected from data in hypothetical audiovisual speech perception experiments. Specifically, different architectures predict different functional forms for the interaction contrast of RT distributions (described later). Interestingly, coactive and parallel models make unique predictions regarding the efficiency of having both channels operating (i.e., auditory and visual) relative to conditions with only one active channel (auditory only or visual only information). The issue of efficiency is yet another informative question that has not been subject to rigorous empirical testing.

A Linear Dynamic Approach for Investigating ‘Early’ Versus ‘Late’ Integration

Two classes of linear dynamic models will be discussed in this section. In parallel models, information is accrued in each channel and processing occurs simultaneously and separately on each channel, although parallel models with cross channel excitation or inhibition will be considered in the discussion. Decisions are also made separately in each channel. Parallel models, with separate decisions on each channel, represent mathematical framework for describing late (or post perceptual) audiovisual integration. Coactive models represent a special case of parallel models where activation from each channel is summed together into a common processor prior to a decision stage (Townsend & Nozawa, 1995). The summation of information in coactive models is a way to mathematically implement the concept of early integration. After describing the mechanics of the models, the predictions that parallel and coactive models make in terms of reaction time distributions will be discussed.

Linear dynamic models specify *a state space* describing the accumulation of perceptual or cognitive activation in the auditory or visual channel at each point in time. The process of accumulation begins when input enters the system from the environment or from another internal system. The state space can be though of as an internal representation of activity in real-time systems, and is used to describe phenomena in a wide range of fields from physics, to biology and muscle activation and language production as discussed in earlier sections (e.g., Browman & Goldstein, 1985; Fowler, Rubin, Remez, & Turvey, 1980; Saltzman & Kelso, 1983, for examples of dynamic systems used to describe articulatory dynamics).

The state space approach differs from an alternative approach where a probability distribution on processing time is placed on each channel of interest. In the non-state space approach, a distribution is placed on finishing times alone rather than on intermediate psychological states involved in processing the stimuli. Townsend and Wenger (2004) argued that the non-state space approach is often unconstrained. This is because the investigator can simply select the amount and type of dependency that exists between processing times of components in the system without worrying about other distributional aspects like the marginal distributions.

Model Specifications

The linear dynamic model and parameters contained in the model are similar to the one used by Townsend and Wenger (2004), where the accumulator channels act in parallel, or in the coactive case, can be summed together.

In the system of equations, let $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ be a vector representing the activation level in each of the two channels. The input functions to the two channels can be represented as the following vector:

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

At this point we can write a simplified version of the model, which allows for accumulation of activation based on the input u(t):

$$\frac{d}{dt}\mathbf{x}(t) = \quad \mathbf{x}(t) + \mathbf{u}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

Next, we add a matrix of coefficients determining how activations in each of the two channels interact. That is, we need a description of how the rate of evidence accumulation in one channel affects the rate of evidence accumulation in the other channel, if information sharing between channels exists. Let A be a matrix of weights describing the distribution of activation across channels:

$$\mathbf{A}(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix}$$

The parameter a_{12}(t) represents the amount of cross talk or information sharing from channel 1 (e.g., auditory channel) to channel 2 (e.g., visual channel), and a_{21}(t) similarly represents the amount of sharing from channel 2 to channel 1. Setting parameters a_{12}(t) and a_{21}(t) equal to zero means that cross-channel information sharing is turned off, which is a parallel model with independent channels (see Townsend & Nozawa, 1995). The diagonal elements a_{11}(t) and a_{22}(t) are terms that stabilize the activation accumulation rate for a particular channel. In particular, they represent a within-channel negative feedback, preventing the activation level from exponential explosion.

For our purposes, the terms a_{11}(t), a_{12}(t), a_{21}(t), and a_{22}(t) will be set to constants a_{11}, a_{12}, a_{21} ,and a_{22} rather than functions of time. This restricts our model to the class of linear dynamic models with constant coefficients rather than variable coefficients. Then, we can write the two-channel interactive parallel model:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

In addition to simplifying the equation by using constant coefficients, other simplifying assumptions utilized by Townsend and Wenger (2004) were introduced in the model described here. We also assume that the activation rates in each channel and across channels are symmetrical, $a_{11} = a_{22}$ and $a_{12} = a_{21}$. Furthermore, the values of the channel accumulation rates are constrained to maintain asymptotic stability thereby preventing the activation rates to run off to infinity. A matrix describing the distribution of inputs across channels is not being included since we are not considering the effects of channel interactions in the earliest stages of processing, but only at higher levels of cognitive processing. The general equation is a two dimensional linear dynamic system of the form:

$$x_1(t) = \frac{1}{a_{11}+a_{12}}\left(e^{(a_{11}+a_{12})t} - 1\right) \tag{4}$$

The solution to the system for each channel is an exponential expression. Equation (1) is *deterministic* because the solution is an explicit function and noise is not added prior to integration. In order for the system not to reach an activation level of infinity as t → ∞, the sum of the parameters in the exponential term must be less than 0. Thus, a_{11} and $a_{22} < 0$; and $|a_{12}|$ and $|a_{21}|$ must be less than a_{11} and a_{22} to prevent the sum from being positive.

Finally, in order to make the model stochastic rather than deterministic, Gaussian white noise of the form $\eta(t)$ is added to the inputs. Consequently, the final differential equation describing cross channel activation in a stochastic model with two parallel and (possibly) dependent channels is given below:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} u_1 + \eta_1(t) \\ u_2 + \eta_2(t) \end{bmatrix} \tag{5}$$

where symmetry constraints $a_{11} = a_{22}$ and $a_{12} = a_{21}$ still hold. In parallel independent models without cross channel excitation or inhibition we set $a_{12} = a_{21} = 0$.

This class of stochastic systems is produced by introducing Gaussian white noise to the inputs such that they are capable of producing a probability distribution on the processing times of each channel. A schematic diagram of the parallel model is presented in Figure 1.

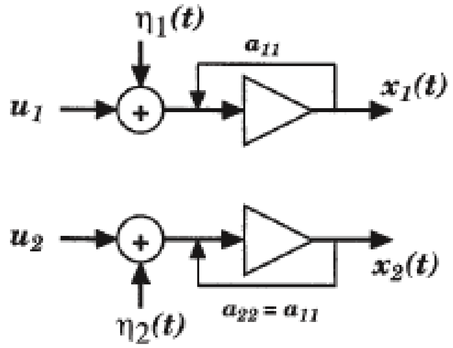

Figure 1 (from Townsend & Wenger, 2004). A parallel model with hypothetical auditory and visual inputs (u1 and u2). Noise is added to the inputs (triangles) and the channels are processed separately. The model could be readily modified to include cross channel excitation or inhibition.

The model presented thus far accumulates activation over time, but does not account for the production of responses, which would be required in models of word or phoneme recognition. To accomplish this, the system must be augmented in two ways. First, a threshold is needed to determine when enough evidence has accumulated in a specific channel to make a response. This is accomplished by adding accumulation thresholds to the auditory and visual channels, which we represent by $\gamma 1$ and $\gamma 2$. Since we have both auditory and visual channels accumulating evidence, it is necessary to decide whether a decision can be made when the first channel reaches threshold, or whether both channels need to reach their respective accumulation threshold. Logical AND or OR gates are imposed on the system as the auditory and visual channels reach their respective threshold. Thus, the logical gates represent the decisional aspects of the system in the sense that they decide whether processing finishes when one channel reaches a threshold (OR), or, whether the system needs to wait for both channels to finish processing (AND). The former decision rule characterizes a first-terminating parallel system, and the latter characterizes an exhaustive parallel system. The termination rule dictates the formulas for the cumulative distribution function, or *CDF*, of processing times. The CDF represents the probability that a process has finished at or before time *t*. The CDF for the AND rule is given by $P(\mathrm{RT} \leq \mathrm{t}) = P(\mathrm{T1} \leq \mathrm{t} \text{ AND } \mathrm{T2} \leq \mathrm{t}) = P(\mathrm{x1}^*(\mathrm{t}) > \gamma 1 \text{ AND } \mathrm{x2}^*(\mathrm{t}) > \gamma 2)$. The CDF for the OR rule is given by $P(\text{OR RT} \leq \mathrm{t}) = P(\mathrm{T1} \leq \mathrm{t} \text{ OR } \mathrm{T2} \leq \mathrm{t}) = P(\mathrm{x1}^*(\mathrm{t}) > \gamma 1 \text{ OR } \mathrm{x2}^*(\mathrm{t}) > \gamma 2)$, where T_1 and T_2 are the random variables for processing times on the auditory and visual channels.

As previously intimated when coactive models were introduced, the activation in the auditory and visual channels can be summed together, thereby producing a model that assumes “early” audiovisual integration. Coactive models have often been considered as a type of parallel model where activation is pooled into a common processor (see Townsend & Nozawa, 1995; Colonius & Townsend, 1997). For a coactive model, the probability that the system has completed processing by time *t* is equal to the probability that the summed activation of the auditory and visual channels exceeds the decision criterion γ. Townsend and Wenger represent the probability as: $P(T(\mathrm{a+v}) \leq t) = P(\max(\mathrm{xa}(\mathrm{t'}) + \mathrm{xv}(\mathrm{t'}), \mathrm{t'} < \mathrm{t}) > \gamma)$, where the quantity a + v denotes the summation of auditory and visual information during recognition. Figure 2 below shows a schematic diagram of a coactive model. The following section will review the predictions made by coactive and parallel models in terms of predicted reaction time distributions and *processing* c*apacity* (i.e., the overall efficiency of the system when the visual channel is added). Reaction time distributions derived from the parallel and coactive dynamic models can be used as diagnostic tools for determining the ‘type’ of processing in audiovisual perception tasks (e.g., early versus late integration).

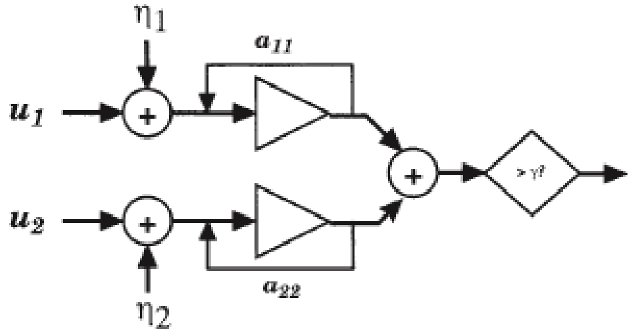

Figure 2 (From Townsend & Wenger, 2004). Coactive processing model of early integration in audiovisual speech perception. Auditory and visual inputs are summed together in a common processor prior to word or segment recognition.

Simulations of Parallel and Coactive Models

The double factorial design is an experimental paradigm that combines two factorial designs. The presence or absence of a certain stimulus or channel (e.g. auditory or visual) is one factorial manipulation. The other manipulation used in the paradigm concerns the saliency of the information in the auditory and visual channels. Consider an audiovisual detection task with four possible trial configurations: auditory only information, visual only information, and congruent audiovisual information. When both modalities are present (i.e., auditory and visual), a factorial manipulation on saliency can be performed, which gives four different trial types: high-high (HH), high-low (HL), low-high (LH), and low-low (LL). In the HH condition, the auditory and visual information appears salient (e.g., a high signal to noise ratio and clear view of a talker's face). In HL and LH trials, only one of the channels carries salient manipulation (auditory or visual) information (e.g., a high signal to noise ratio and an unclear view of the talker's face). In LL trials, both channels have low salience.

Each of the four factorial conditions (HH, HL, LH, LL) yields a reaction time distribution. The survivor function(s), which is the complement of the CDF (1 – F(t)) can be estimated empirically in each of the four conditions. The CDF indicates the probability that processing of a certain stimulus (in the auditory or visual channels) has finished, while the survivor function indicates the probability that processing has not finished. Townsend and Nozawa (1995) developed the *survivor interaction contrast*, which is an extension of the additive factors method developed by Sternberg (1969). The survivor interaction contrast is: $S_{IC}(t) = S_{HH}(t) - S_{HL}(t) - S_{LH}(t) + S_{LL}(t)$. Townsend and Nozawa showed that different processing architectures predicts different shapes, or functional forms of the $S_{IC}(t)$, making it a useful tool for differentiating between different models (i.e., coactive and parallel models).

The $S_{IC}(t)$ predictions, demonstrated by showing simulation of the parallel dynamic and coactive model, are shown in Figure 3. The parallel model includes both AND and OR designs (referred to as *Exhaustive* and *first-Terminating* respectively). As one can observe from the simulation results (which correspond to predictions derived mathematically by Townsend and Nozawa (1995)), parallel self-terminating models (late integration) predict an entirely overadditive $S_{IC}(t)$, parallel exhaustive models predict an entirely underadditive $S_{IC}(t)$ (late integration where both channels need to finish processing before a final decision is made), and coactive models predict an SIC(t) that is mostly overadditive, with a region of negativity for early processing times. Table 1 shows the parameters in the linear dynamic model used in these simulations. Model predictions are stable across a wide range of parameter values. Reaction time distributions obtained from real human participants can be compared to the $S_{IC}(t)$ distributions from the simulations and used as a diagnostic tool for determining parallel versus coactive processing architecture. Simulations are currently being carried out to determine the effects that positive and negative channel interactions have on $S_{IC}(t)$ and architecture.

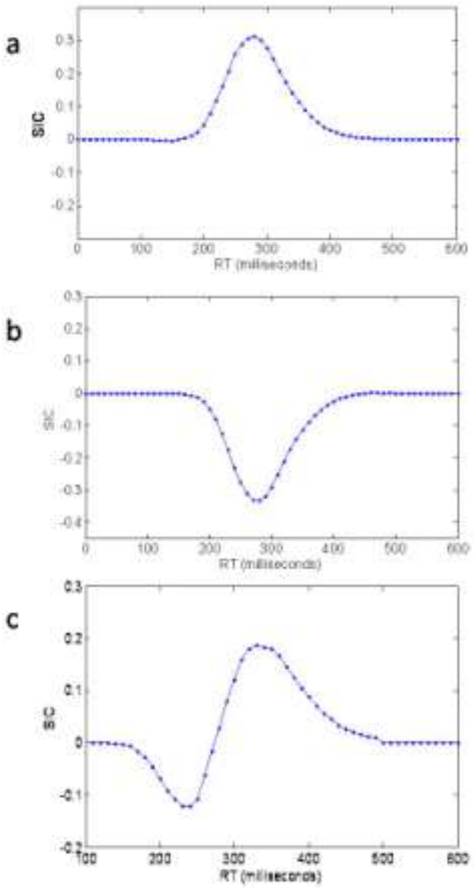

Figure 3. The top left panel (a) is a parallel model with a first-terminating stopping rule. The system finishes when the first channel completes processing. Panel (b) below shows the predictions of a parallel exhaustive model. The bottom panel (c) shows the predictions of a coactive model. The model predicts an overadditive $S_{IC}(t)$, with a region of negativity for early processing times (smaller than the positive region.

Notation and/or variable	Value	Explanation
$x_i(t), i = 1,2$	*	Level of state of each of the two channels
$u_i(t), i = 1,2$	L=1; H = 10	Level (deterministic) of the input to each of the two channels
n	1,000	Number of trials
$\eta_i(t), i = 1,2$	Var = 100	Gaussian white noise introduced to each of the two channels
$\gamma_i, i = 1,2$	1.10	Activation thresholds for each of the channels

Table 1. Parameter values used in simulations

Processing Capacity Predictions

A redoubtable body of evidence has demonstrated that seeing a talker's face improves the quality of speech across a considerable range of S/N ratios in both normal hearing and hearing impaired individuals (Sumby & Pollack, 1954; see also Braida, 1991, for a review). However, while visual obtained from lip-reading may boost accuracy scores in audiovisual phoneme and word identification tasks, there may be a cost in terms of efficiency. Thus, the addition of visual information during processing might boost accuracy, but the increased workload caused by having a second channel operating may not be efficient in terms of processing rate. The double factorial paradigm as described above can measure efficiency, or rather processing capacity indicated by C(t). A tutorial is available which explains in straightforward terms how to compute capacity from RT data (Wenger & Townsend, 2000).

Measuring processing capacity requires analyzing the ratio of the integrated Hazard functions. The hazard function, which is used to calculate the capacity coefficient C(t), is given below (Townsend & Nozawa, 1995):

$$h(t) = \frac{f(t)}{1 - F(t)} \qquad (6)$$

The quantity $f(t)$ is the probability density function on the finishing time, and 1 – F(t) is the survivor function ($S(t)$) indicating the probability that a process has not yet finished. The hazard function $h(t)$ indicates the probability that a process will terminate at the next moment (*t* + 1) in time given that it has not yet terminated at time *t*.

To calculate the capacity coefficient C(t) at each time point, we calculate the integrated hazard function for the conditions where the participant is presented with the redundant audiovisual information, and divide it by the sum of the integrated hazard functions of each single modal condition (auditory only plus visual only trials). The subscripts A and V indicate the audio and visual channels.

$$C(t) = \frac{H_{AV}}{H_A(t) + H_V(t)} \qquad (7)$$

Each integrated hazard function in the capacity coefficient $H(t)$ is equivalent to -log[1 – F(t)] or -log[S(t)], and in the field of physics it is used as a measure of the total energy consumed or work performed. The system operates at *super capacity* at a certain point in time t if C(t) is greater than 1, unlimited capacity if it equals 1, and limited capacity if it is less than 1 (Wenger & Townsend, 2000; Townsend & Nozawa, 1995). Capacity in parallel independent models (assuming stochastically independent processing times and a self-terminating stopping rule) is unlimited and therefore equal to 1. In coactive models capacity is considerably greater than 1 and hence referred to as 'super capacity'. Capacity predictions produced by model simulations are shown in Figure 4 for both parallel and coactive models. Interestingly, current and previous simulation work (e.g., Townsend & Wenger, 2004) demonstrate that facilitatory channel interactions in a parallel model lead to a capacity coefficient C(t) > 1, although not as high as coactive models predict, while inhibitory channel interaction yield a capacity coefficient C(t) < 1. Parallel models with facilitatory or inhibitory connections can be classified as models of late integration with information sharing between modalities. Such information sharing in the brain might occur due to projections from visual to auditory regions of the cortex.

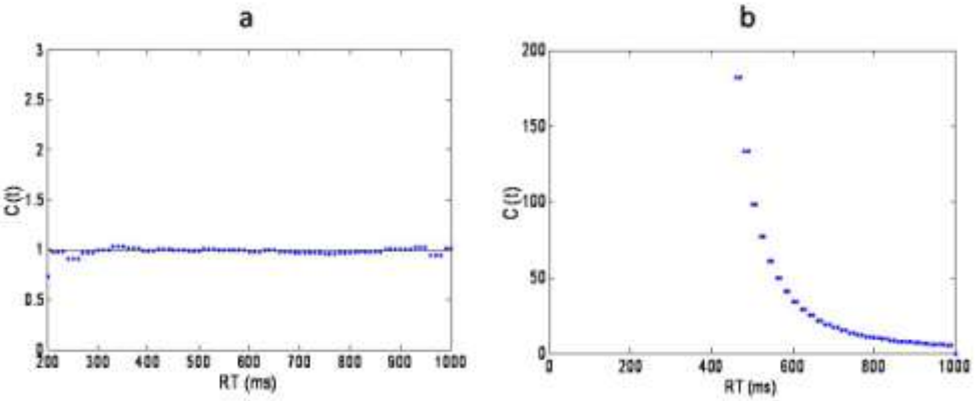

Figure 4. Capacity predictions derived from simulations of the linear dynamic model. The capacity predictions for the parallel model with independent channels are shown in the left panel (a), and the capacity predictions for coactive model, demonstrating super-capacity, are shown in the right panel (b).

Thus, the capacity coefficient predicted by parallel (dependent and independent) and coactive models is another useful diagnostic tool that can be applied to real human data as a mathematical measure of processing efficiency in audiovisual detection and discrimination tasks. First, the capacity coefficient C(t) can determine whether the addition of visual information in an audiovisual speech perception task yields a gain or loss in processing efficiency. Secondly, C(t) can provide some insightful supplementary information about whether processing is parallel or coactive, since coactive models predict super-capacity, and parallel independent models with unlimited capacity predict C(t) = 1 (Townsend & Wenger, 2004; Townsend & Nozawa, 1995).

Summary and Conclusions

This paper focused on the importance of audiovisual speech perception by highlighting a redoubtable body of evidence showing that speech perception is a multimodal phenomenon (e.g. Braida, 1991; MacGurk & MacDonald, 1976; Summerfield, 1987; Sumby & Pollack, 1954). Different theoretical accounts of the representations of speech perception in both the auditory and visual signal were introduced including: phonetic features, the filter function of the vocal tract, analysis of auditory and visual spectra, and finally articulatory dynamics (see Summerfield, 1987, for a more thorough theoretical treatment of these metrics). The report primarily focused on the account of articulatory dynamics as a metric or account of audiovisual speech perception for several reasons. First, as discussed in the introduction, Browman and Goldstein (1995) summarized some the shortcomings of the formalist approach to speech perception, while arguing that a unified theory of speech perception and production is necessary if the science is to progress. The assumption that the cognitive system operates on modality independent information is one way to incorporate dynamics and timing information into a representation based theory of audiovisual speech integration. In this framework, both auditory and visual channels consist of identical modality independent information that exists in the form of a common currency. This aspect of the theory might explain how the auditory and visual information could be combined into a common processor during integration if such a combination of information does in fact occur.

In addition to summarizing dynamic based theories of the cognitive representations relevant to audiovisual integration, a two dimensional linear dynamic model was introduced as a methodology for testing competing theories of the time course of integration. The model was introduced as a qualitative statistical tool for investigating whether the auditory and visual channels pool information into a common processor when speech is perceived audio-visually (early integration), or whether decisions can be made separately on each channel (late integration). The issue of processing or integration efficiency (i.e., capacity coefficient C(t)) was explored as a supplementary statistical tool. The dynamic models of integration proposed here provide a framework for collecting reaction time data, comparing the reaction time distributions to model predictions, and potentially ruling out broad classes of models. For instance, the methods proposed here provide a benchmark for inquiring if the addition of the visual channel leads to a reduction in integration efficiency (limited capacity), facilitation of integration efficiency (super capacity), or has a null effect on integration efficiency (unlimited capacity). Parallel and coactive models, as was shown in simulation results, moreover predict differences in the capacity coefficient C(t). In summary, the linear dynamic approach combined with the methodology of the double factorial paradigm provides powerful mathematical machinery for investigating how cognitive processes operate on auditory and visual inputs while adding considerable clarification to questions regarding the precise nature of ‘integration’ in speech perception.

References

Bernstein, L. E. (2005). Phonetic perception by the speech perceiving brain. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 79-98). Malden, MA: Blackwell Publishing.

Bernstein, L. E., Auer, E. T., & Moore, J. K. (2004). Audiovisual speech binding: Convergence or association? In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *Handbook of Multisensory Processing* (pp. 203-223). Cambridge, MA: MIT Press.

Bergeson, T.R., & Pisoni, D.B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 153-176). Cambridge, MA: The MIT Press.

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, 163,* monograph, 347-370.

Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. *The Quarterly Journal of Experimental Psychology, 43A*(3), 647-677

Browman, C.P., and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.

Browman, C.P., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201-251.

Browman, C.P., and Goldstein, L. (1995). Dynamics and Articulatory Phonology. In R.F. Port and Timothy Van Gelder (Eds.), *Explorations in the Dynamics of Cognition. Mind as Motion* (pp. 175-193). MIT Press, Cambridge, MA.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*, 649-657.

Chomsky, N., and Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Colonius, H., and Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 245–254). Hillsdale, NJ: Erlbaum.

Dodd, B., McIntosh, B., Erdener, D., & Burnham, D. (2008). Perception of the auditory-visual illusion in speech perception by children with phonological disorders. *Clinical Linguistic & Phonetics, 22*(1), 69-82.

Elman, J.L. (1995). Language as a Dynamical System. In R.F. Port and Timothy Van Gelder (Eds.), *Explorations in the Dynamics of Cognition. Mind as Motion* (pp. 195-225). MIT Press, Cambridge, MA.

Erber, N.P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423-425.

Fowler, C.A., Rubin, P., Remez, R.E., and Turvey, M.T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language Production*. New York: Academic Press.

Gentilucci, M., & Corballis, M.C., (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews, 30*, 949-960.

Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception and Psychophysics, 38*(3), 269-276.

Hockett, C. (1955). A manual of phonology. Chicago: University of Chicago.

Kelso, Batesman, Saltzman, and Kay (1985). As cited by Summerfield (1987).

Kugler, P.N., and Turvey, M.T. (1987). Information, natural Law, and the self-assembly of rhythmic movement. Hillsdale, NJ: Erlbaum.

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

Port, R.F., and Leary, A. (2005). Against formal phonology. *Language, 81*, 927-963.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.). *The Handbook of Speech Perception* (pp. 51-78). Malden, MA: Blackwell Publishing.

Saltzman E., and Kelso, J.A.S. (1987). Skilled actions: a task dynamic approach. *Psychological Review, 94*, 84-106.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica, 30*, 276-315.

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 12-15.

Summerfield, Q (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), The Psychology of Lip-Reading (pp. 3-50). Hillsdale, NJ: LEA.

Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*(4), 321-359.

Townsend, J. T., and Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*(4), 1003-1035.

Wenger, M. J., and Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *The Journal of General Psychology, 127*(1), 67-99.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Qualitative and Quantitative Aspects of Conversations of Deaf Peers: Some Preliminary Findings[1]

Jessica Beer[2]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by NIH-NIDCD T32 Postdoctoral Training Grant DC00012 to Indiana University. I would like to thank St. Joseph Institute for the Deaf in Indianapolis and the families and children who participated.

[2] Currently at Indiana University School of Medicine, Indianapolis.

Qualitative and Quantitative Aspects of Conversations of Deaf Peers: Some Preliminary Findings

Abstract. Hearing impaired and deaf infants and children experience a period of degraded auditory access to social linguistic interactions, and a lack of opportunity to actively participate in social linguistic interactions which impacts language acquisition and communicative competence. The present study investigates both quantitative (duration of conversational turns, duration of talk and silence) and qualitative (semantic connectivity) linguistic aspects of conversations of oral deaf peers during undirected play. Preliminary data will be presented regarding the frequency and duration of 1) conversational turns, 2) talk and silence, and 3) connected and failed turns in individual children and between children in a dyad. Comparisons to existing data with normal-hearing preschoolers and children with cochlear implants are made and directions for the next step in the analyses and methodology are proposed.

Introduction

Human infants are born into a richly structured socio-cultural environment of caregivers, objects, and routines that create affordances for the emergence of uniquely human capacities such as spoken language (Tomasello, 1992). American infants are immersed in Western practices of child-rearing that include close contact and frequent interactions with a caregiver, infant-inspired cultural artifacts such as bouncy seats, “Boppy’s”, pacifiers, and changing tables, and highly structured routines such as diapering, feeding, “tummy-time”, and sleeping. In addition, typically developing humans possess social-cognitive abilities such as social imitation, shared gaze, and joint attention that allow them to “tune-in” to others from in the first hours of life. Socio-cultural theories of development posit that the phylogenetic and ontogenetic development of higher psychological functions such as spoken language acquisition, reasoning, problem-solving, and concept development emerge from the intimate relationship between our socio-cultural environment and our social-cognitive and social learning abilities (Nelson, 1996; Tomasello, 1992). From this perspective, human cognition is socially—and linguistically—mediated through interactions with others using cultural artifacts within socially shared events that make up everyday life (Vygotsky, 1978).

A usage-based or functional theory of language acquisition suggests that language structure—lexical, grammatical, syntactical, and conversational—emerges from language function thus taking seriously a socio-cultural perspective of cognitive development (Tomasello, 2003). From this perspective, the acquisition of a new piece of language emerges naturally through a child’s active use of joint attention and imitative learning in prestructured events so that “…communicative uses of words and utterances…serve as resources for the individual language learner’s construction of structure” (Nelson, 2006). The functional perspective obviates the need for a priori linguistic knowledge on the part of the child as posited by formalist theories, and suggests rather that social learning and socio-cognitive abilities are sufficient for explaining language acquisition. Consideration of the social and linguistic experiences of an individual child is therefore essential to understanding language acquisition; differences in experience will result in individual differences in development (Nelson et al., 2003).

Deaf children of hearing parents experience degraded or limited opportunities to actively listen and participate in communicative exchanges; therefore their functional use of language is very different from normal-hearing children. The structure of language that emerges from these experiences will be very different as well. Differences in structure and function of language will affect higher cognitive processes that are mediated by language such as executive function, theory of mind (ToM), reasoning, and problem solving.

A large part of our knowledge of speech and language development of hearing impaired children and deaf children with cochlear implants is obtained through context free assessments of speech perception, production and standardized language measures. Although highly controlled testing procedures and normed measures are necessary for evaluating speech and language outcomes more proximal to hearing loss, they provide little information about more distal outcomes such as a child's functional use of language in everyday social interactions from which the lexical, syntactical, and grammatical structure of language emerge (Tomasello, 2003). Sociocultural theory and research suggests that from birth the social and linguistic environments of normal-hearing children and a child's active participation in social exchanges are integral to language and vocabulary acquisition (Akhtar, Jipson, & Callanan, 2001; Tomasello, 2003), socio-cognitive development (Ontai & Thompson, 2008; Racine & Carpendale, 2007), autobiographical memory (Fivush & Nelson, 2004) and communicative competence (Tomasello, 1992).

Atypical development of linguistic structure and impoverished communicative exchanges may contribute to some of the delays observed in social and emotional understanding of deaf of hearing children (Moeller & Schick, 2006; Peterson, 2004; Schick, de Villiers, de Villiers, & Hoffmeister, 2007; Woolfe, Want, & Siegal, 2002). In normal-hearing preschoolers, maternal talk about mental states such as cognitions, desires, and emotions is related to children's performance on tasks of emotion understanding (Denham, Zoller, & Couchoud, 1994; Taumoepeau & Ruffman, 2006). Mother's and children's mental state talk within semantically connected conversation at 2 years of age both independently predicted social understanding at 4 years of age (Ensor & Hughes, 2008). In a comparison study of maternal linguistic input, Moeller and Schick (2006) reported that hearing mothers of normal-hearing children referenced mental states more frequently than hearing mothers of deaf children, suggesting a functional explanation of the developmental lag in social understanding experienced by deaf children of hearing parents (Peterson, 2004).

There is evidence that both qualitative and quantitative linguistic maternal input is related to language ability in preschoolers who are CI (cochlear implant) users. DesJardin and Eisenberg (2007) found that mother's mean length of utterance (MLU) and facilitative language techniques such as open-ended questions and recasts were significantly correlated with their child's expressive and receptive language ability and also accounted for a significant portion of the variance in language beyond that explained by the child's age. Schick and her colleagues (2007) reported that vocabulary and comprehension of false complement clauses, which allow for a false proposition in a true sentence, were independent predictors of successful reasoning about false beliefs in deaf children who use oral language or American Sign Language.. Combined, this research suggests that rich conversational environments that include conversational scaffolding by parents and syntax for representing mental state concepts may explain some of the contribution of language to the development of social understanding in young deaf children.

During the preschool years children begin to forge friendships and relationships beyond their immediate family. These relationships expand the range of collaborative activities and conversational environments available to children. In normal-hearing preschoolers, the linguistic features of conversation between a child and her mother, sibling, or peer are different suggesting that children converse and interact with different interlocutors in different ways (Brown, Donelan-McCall, & Dunn, 1996; Cutting & Dunn, 2006). We know very little about the qualitative linguistic features of conversation between oral deaf *peers* such as ignoring, repeating, and elaborating. Nor do we know much about how these functional uses of social dialogue relate to speech, language and academic outcomes. Investigation of conversational interactions of deaf peers who use spoken language is important because a percentage of these children will likely be mainstreamed into classroom environments of typically developing children and mainstreamed classroom instruction. This classroom environment will require children to be competent *users* of spoken language so they may actively participate in social and linguistic interactions with peers and teachers.

One qualitative feature of conversational interactions is *connectedness* of alternating exchanges of two people engaged in a conversation. A speaker's utterance is defined as connected if it is semantically related to the interlocutor's prior utterance (Gottman, 1983). A measure of connectivity between two speakers engaged in a joint activity can give us information about how well the speakers are "tuned in" to one another. Tuning in requires general knowledge about the desires, beliefs, and intentions of one's interlocutor and linguistic ability to negotiate an activity using this knowledge. Connectedness of conversations is related to the development of social understanding in typically developing preschoolers. In their investigation of mother-child conversations with 2 year olds, Ensor and Hughes (2008) found that mothers referred to mental states most often during connected turns and that both children's and mother's references to mental states within connected turns predicted social understanding at age 4. In a study of the play conversations of 4 year old peers, Slomkowski and Dunn (1996) found that children spent the majority of the play session in connected communication and that the average percentage of connected turns was highest during pretend play. Connected communication was positively correlated with false belief performance and affective perspective-taking. These findings suggest that the ability to engage in connected conversations with an interlocutor may facilitate the development of social understanding.

The present study examines the conversations of oral deaf peer dyads during undirected play. Measures of both qualitative and quantitative aspects of conversational exchanges are assessed. As this is a work in progress, the measures analyzed thus far include the semantic connectivity of the exchanges between peers, the amount of time they spend in silence during a play session, and the duration of their conversational turns. Future analyses will examine children's use of mental state references within connected and failed exchanges and intelligibility of exchanges. These measures will be compared to those of normal-hearing age-matched peers engaged in the same activity. Normal-hearing data has been collected and is presently being transcribed. These qualitative aspects of conversation are hypothesized to be responsible for the contribution of everyday conversations to social understanding in normal-hearing children (Beer, In preparation; de Rosnay & Hughes, 2006; Ensor & Hughes, 2008; Harris, de Rosnay, & Pons, 2005)

Methods

Participants

Children were recruited from an oral deaf school in Indianapolis. A speech/language pathologist familiar with the spoken language abilities of each child paired children into dyads based on her assessment of their ability to use spoken language to communicate and the availability of a familiar same gender peer. Criteria for inclusion to the study were that the child used oral language to communicate, had no disabilities that would impede him or her from engaging in pretend play with a peer, could be paired with a same gender peer, and had a monolingual English home environment. Eight children (4 dyads) participated in this study (*M* age = 5.58 yr *SD* = .92, *Range* = 4.5 to 7.25 yrs). Table 1 provides descriptive information regarding hearing loss. Hearing age refers to the number of years the child has experienced aided hearing, either through the use of a hearing aid or a cochlear implant. It is calculated by subtracting the age at which the child received aided hearing from the child's chronological age at testing.

Dyad	Child	Age	Hearing Age	Type of Aid	Degree of Loss
A	1	7.25	4.0	CI[a]	Sev-Prof[b]
	2	5.75	2.5	CI (bi)[c]	Sev-Prof
B	3	4.92	3.25	CI	Sev-Prof
	4	4.5	2.25	CI	Sev-Prof
C	5	6.17	4.25	CI	Sev-Prof
	6	5.92	1.33	CI	Sev-Prof
D	7	5.58	3.08	CI	Sev-Prof
	8	4.58	4.17	HA (bi)[d]	Mod-Sev[e]

Table 1. Participant Descriptives
[a]Cochlear implant. [b]Severe to profound hearing loss. [c]Bilateral cochlear implants. [d]Bilateral hearing aids. [e] Moderate to severe hearing loss.

Procedures

Each peer dyad played in a familiar room at their school for 15 minutes, after which they returned to their classroom. The interactions were recorded by a video camera. Each child wore a wireless microphone and mini audio recorder in a fanny pack. Total participation time took approximately 30 minutes from start to finish. Parents were informed of their childs' peer partner prior to the play session so they could prepare their child prior to participation. Toys for both boy-boy and girl-girl pairs included kitchen items and pretend food. Toys for boy-boy pairs also included a castle with knights and Rescue Hero's with their vehicles. Toys for girl-girl pairs included and a dollhouse with Polly Pockets and Barbie's with their vehicles. Children were introduced to the toys and encouraged to play. Once the children appeared to be comfortable with the room and the equipment, the experimenter told them to play independently while she did some work in another area of the room and assured them that she could still see and hear them should they need her. Children received a small toy for their participation.

Measures

Qualitative linguistic measures. Each videotape was transcribed into alternating conversational turns. A conversational turn included the utterances of one child bound by the utterances of the other child including non-words, talk to self, and silence within the turn (adapted from Shatz & Gelman, 1973). Nonverbal responses and elicitations consisting of gestures and shifts in gaze were coded as conversational turns. Each turn was coded for *semantic connectivity*. A connected turn is one that is semantically related to the partner's previous utterance. A failed turn is not semantically related to the partner's prior utterance.

Quantitative linguistic measures. After transcription of the play session was complete, two duration measures were calculated using the soundwave of the individual audio recordings. The *duration of each conversational turn* was measured and excluded? within-talker time in silence that measured over 5 seconds in duration. The second measure of duration was *total time in silence* for the 15 minute play session. This measure included silence between talkers and silence within a talker of more than 5 seconds in duration.

Language measures. Two measures of language were obtained for each child. Vocabulary ability was assessed using the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007). Overall language ability was assessed using the Clinical Evaluation of Language Fundamentals-Preschool (CELF-P; Wiig, Secord & Semel, 1992). Both the PPVT and the CELF-P have means of 100 and standard deviations of 15. Clinically significant scores are those that fall one standard deviation below the mean (85 or less). All but one child fell within the normal range in vocabulary whereas only two children fell within the normal range in overall language (See Figure 1).

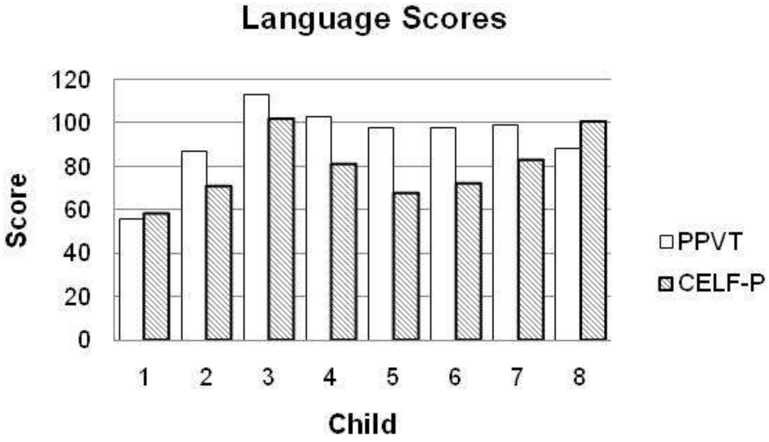

Figure 1. Language scores for each child.

Results

Connected Conversation

Each of the 8 children had more connected than failed turns during the 15 minute play session with a mean proportion of connected to total turns of .62. There was surprisingly little variability in the number of connected turns across children with the frequency of connected turns within the 15 minute play session ranging between 37 and 44 for all but one child (See Figure 2).

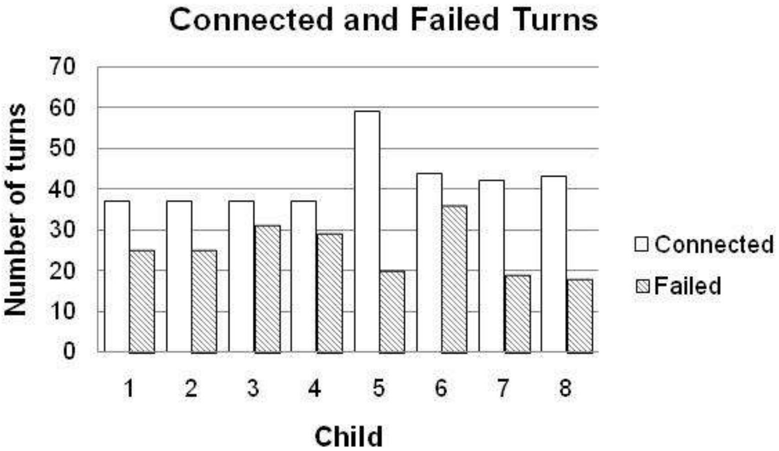

Figure 2. Number of connected and failed turns for each child.

Time in Silence

All but one of the four peer dyads spent more of the play session in silence than talking with a mean percent time in silence across all dyads of 34% (Range = 26% to 47%) (See Figure 3). Tye-Murray (2003) reported percent time in silence of 15% for oral deaf children and 5% for normal-hearing children age 8 and 9 in conversation with an unfamiliar adult.

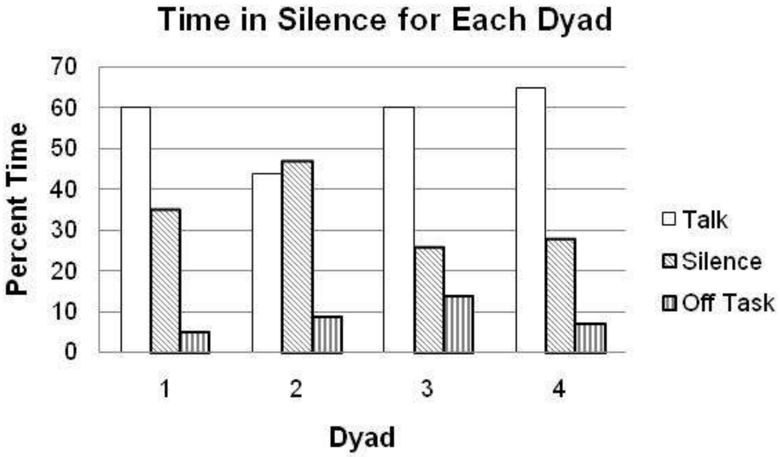

Figure 3. Percent time in silence and talk for each dyad. Off task refers to the percent of time children disengaged from one another to address the experimenter.

Duration of Conversational Turns

The durations of conversational turns across all children were highly variable with a mean of 3.83 seconds (*SD* = 6.09, Range = 0 to 59.99 seconds). Seventy-five percent of all conversational turns were between 0 and 4 seconds in duration. As shown in Figure 4, within individual children there was a large amount of variability in the duration of conversational turns. A large amount of individual variability in duration of speaker turns may lead to imbalance in the conversation when examined at the level of the dyad with one child doing most of the talking during the play session.

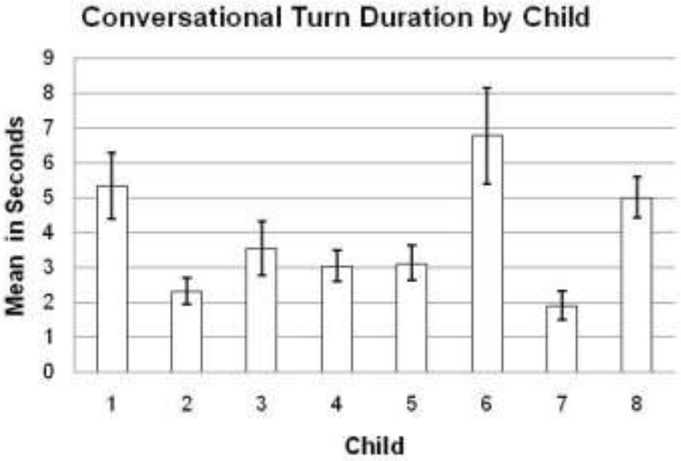

Figure 4. Mean durations of conversational turns for each child. Error bars represent standard error.

Duration of Connected and Failed Turns

In this analysis the duration of connected and failed turns were calculated, thus combining both a qualitative and quantitative measure of conversation. Results indicate that *failed* turns were longer than connected turns for six of the children, and that the mean duration of *failed* turns was greater than the mean duration of connected turns (*M* failed = 4.5, *SD* = 7.1 vs. *M* connected = 3.4 sec, *SD* = 5.4). Figure 5 shows durations of connected and failed turns for each child.

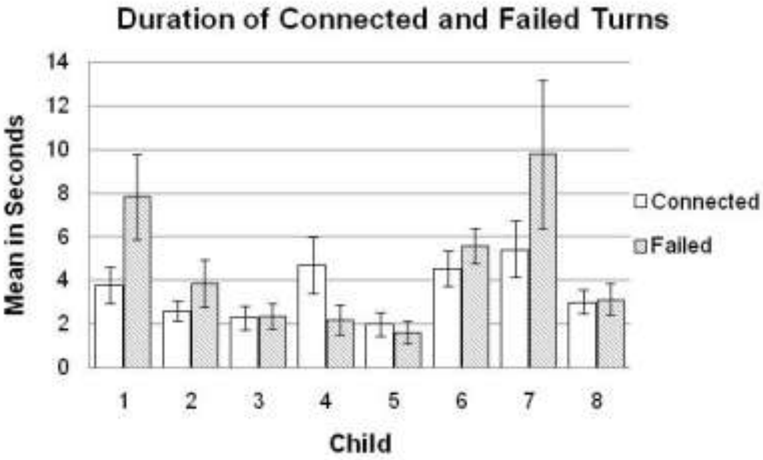

Figure 5. Durations of connected and failed turns for each child. Error bars represent standard error.

Discussion

The results of this study indicate that deaf children who use oral language are able to engage in connected conversation with a peer and do so in about 60% of their exchanges. This percentage is close to that reported by Slomkowski and Dunn (1996) in normal-hearing 4 year old peer play. The children in the present study are on average one and one-half years older than the children in the Slomkowski and Dunn study, suggesting a possible lag in the ability of deaf children to engage in connected conversation compared to normal-hearing same age peers. Whether or not there is a lag will be answered by comparing this data to normal-hearing age-matched controls. It is encouraging, however, to find that oral deaf children who have had on average three years of aided hearing are able to engage in connected exchanges with another oral deaf peer. This type of exchange can be thought of as the necessary foundation upon which more socially and cognitively complex exchanges may be co-constructed.

With regard to time in silence, findings show that 3 out of the 4 peer dyads spent more time talking than in silence. Time in silence is a measure used by Tye-Murray (2003) to assess conversational fluency in cochlear implant recipients. She reported that time in silence is negatively correlated with intelligibility, AV speech perception, and auditory only speech perception. The children in the present study spent more time in silence than the oral deaf children in the Tye-Murray study. However, a direct comparison should be made with caution because the children in the present study were 3 years younger than the children in the Tye-Murray study and played with a peer as opposed to conversing with an unfamiliar adult. A normal-hearing age-matched control group engaged in the same free play activity is necessary in order to identify group differences in conversational measures.

The durations of conversational turns were highly variable between children with some turns lasting a few seconds and others lasting almost one minute. Comparisons of the durations of connected versus failed turns suggest that failed turns are, in general, longer than connected turns. It is possible that failed turns are longer due to expressive language delays of oral deaf children or to socio-cognitive deficits that make it difficult to for these children to "tune in" to a peer using spoken language. By tuning in I am referring to a child's ability to use perspective taking and coordination skills to engage in collaborative activities with a peer.

These preliminary findings suggest that oral deaf children may have difficulty establishing and maintaining connected conversation with a peer—a necessary precursor to more sophisticated linguistic interactions that require collaborative co-construction among conversational partners. Connected talk and time in silence are key tools for assessing a child's ability to engage a person verbally and sustain successful and collaborative conversations. These measures speak directly to a child's *functional use of language*, or communicative competency. These types of measures are not routinely used to evaluate outcome in cochlear implant recipients, but are highly relevant to understanding the efficacy of cochlear implantation (Pisoni et al., 2008). This line of research will allow the identification of differences in linguistic features of conversation between normal-hearing peers and oral deaf peers. These differences may explain some of the variability in more distal outcomes related to hearing loss such as the delay in social understanding observed in deaf of hearing children (Peterson, 2004; Peterson & Siegal, 1999), as well as contribute to our understanding of new areas of research such as communicative competence of oral deaf children. In addition, assessment of both qualitative and quantitative aspects of conversations may explain other language-related delays in deaf of hearing children such as ToM (Theory of Mind), emotion understanding, and particular aspects of executive function (Pisoni et. al., In press; Figueras, 2008).

Future Directions

The data presented in this report are descriptive and represent preliminary findings. I am presently transcribing and coding data of age-matched normal-hearing peer dyads in order to identify differences in the conversations between the two groups of children engaged in the same 15 minute play activity. Preliminary comparisons between the four hearing impaired (HI) dyads and the four normal-hearing (NH) dyads indicate group differences in the mean number of exchanges per dyad (*M* (HI) = 136.5, *M* (NH) = 162.5), the percent of connected turns (*M* (HI) = 61%, *M* (NH) = 70%), and the mean number of references to mental states (*M* (HI) = 14.5, *M* (NH) = 26.5) during the 15 minute play episode. Both groups referred to mental states most often during connected turns and the differences between groups in the percent of mental state references that occur within connected turns was minimal (*M* (HI) = 67%, *M* (NH) = 71%). This preliminary data suggests differences in both the qualitative and quantitative aspects of conversations of hearing impaired and normal-hearing peers. Hearing impaired children may have difficulty engaging in and maintaining perspectively rich conversation with a peer. Correlations between language ability and these qualitative and quantitative aspects of conversations will be calculated.

References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning Words through Overhearing. *Child Development*, *72*(2), 416-430.

Beer, J. (In preparation). Negotiations during pretend play: Conversation, collaboration, and co-construction.

Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why talk about mental states? The significance of children's conversations with friends, siblings, and mothers. *Child Development*, *67*, 836-849.

Cutting, A., & Dunn, J. (2006). Conversations with siblings and with friends: Links between relationship quality and social understanding. *British Journal of Developmental Psychology, 24*, 73-87.

de Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding? *British Journal of Developmental Psychology, 24*(1), 7-37.

Denham, S. A., Zoller, D., & Couchoud, E. A. (1994). Socialization of preschooler's emotion understanding. *Developmental Psychology, 30,* 928-936.

DesJardin, J. L., & Eisenberg, L. S. (2007). Maternal contributions: Supporting language development in young children with cochlear implants. *Ear & Hearing, 28*(4), 456-469.

Dunn, L.M. & Dunn, D.M. (2007). The Peabody Picture Vocabulary Test, Fourth Edition. Bloomington, MN: NCS Pearson, Inc.

Ensor, R., & Hughes, C. (2008). Content or connectedness? Mother–child talk and early social understanding. *Child Development, 79*(1), 201-216.

Figueras, B., Edwards, L., & Langdon, D. (2008). Executive function and language in deaf children. *Journal of Deaf Studies and Deaf Education, 13*(3), 362-377.

Fivush, R., & Nelson, K. (2004). Culture and langauge in the emergence of autobiographical memory. *Psychological Science*, *15*(9), 573-577.

Gottman, J. (1983). How children become friends. *Monographs of the Society for Research in Child Development, 48*(Serial No. 201).

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and Children's Understanding of Mental States. *Current Directions in Psychological Science, 14*(2), 69-73.

Moeller, M. P., & Schick, B. (2006). Relations between maternal input and theory of mind understanding in deaf children. *Child Development, 77*(3), 751-766.

Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind.* New York: Cambridge University Press.

Nelson, K. (2006). Advances in pragmatic developmental theory: The case of language acquisition. *Human Development, 49*, 184-188.

Nelson, K., Plesa-Skwerer, D., Goldman, S., Henseler, S., Presler, N., & Fried-Walkenfeld, F. (2003). Entering a community of minds: An experiential approach to 'Theory of mind'. *Human Development, 46*, 24-46.

Ontai, L. L., & Thompson, R. A. (2008). Attachment, Parent-Child Discourse and Theory-of-Mind Development. *Social Development, 17*(1), 47-60.

Peterson, C. (2004). Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry, 45*(6), 1096-1106.

Peterson, C., & Siegal, M. (1999). Representing Inner Worlds: Theory of Mind in Autistic, Deaf, and Normal Hearing Children. *Psychological Science, 10*(2), 126-129.

Pisoni, D.B., Conway, C. M., Kronenberger, W. G., Horn, D. L., Karpicke, J., & Henning, S. C. (2008). Efficacy and Effectiveness of Cochlear Implants in Deaf Children. In Marschark, M. & Hauser, P.C. (Eds) Deaf Cognition: Foundations and Outcomes (pp. 52-102). NY, NY: Oxford University Press.

Pisoni, D. B., Conway, C. M., Kronenberger, W., Henning, S., & Anaya, E. (In press). *Executive Function and Cognitive Control in Deaf Children with Cochlear Implants*.

Racine, T. P., & Carpendale, J. I. M. (2007). The role of shared practice in joint attention. *British Journal of Developmental Psychology, 25*(1), 3-25.

Schick, B., de Villiers, P., de Villliers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development, 78*(2), 376-396.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of the listener. *Monographs of the Society for Research in Child Development*, *38*, 1-37.

Slomkowski, C., & Dunn, J. (1996). Young children's understanding of other people's beliefs and feelings and their connected communication with friends. *Developmental Psychology*, *32*(3), 442-447.

Taumoepeau, M., & Ruffman, T. (2006). Mother and Infant Talk About Mental States Relates to Desire Language and Emotion Understanding. *Child Development*, *77*(2), 465-481.

Tomasello, M. (1992). The social basis of language acquisition. *Social Development*, *1*(1), 67-87.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tye-Murray, N. (2003). Conversational fluency of children who use cochlear implants. *Ear & Hearing*, *24*(1S), 82S-89S.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: MIT Press.

Wiig, E., Secord, W., & Semel, E. (1992). Clinical evaluation of language fundamentals-preschool. San Antonio, TX: Psychological Corporation

Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to Development:Theory of Mind in Deaf Children. *Child Development*, *73*(3), 768-778.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Modularity in the Channel: A Response to Moreton (2008)[1]

Vsevolod Kapatsinski

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] Many thanks to the NIH for financial support through grants DC00111 and DC00012 and to Tom Gruenenfelder and David Pisoni for helpful comments.

Modularity in the Channel: A Response to Moreton (2008)

Abstract: Moreton (2008) argues for a distinction between analytic bias and channel bias in language learning. Analytic bias is defined as a set of cognitive predispositions for certain types of generalizations, which constrains the learner but does not influence perception and production. Channel bias is defined as 'phonetically systematic errors in transmission between speaker and hearer'. Moreton proposes a new type of analytic bias, the modularity bias, whereby dependencies between consonant dimensions, such as voicing features of different consonants within the same word, and dependencies between vowel dimensions, such as height features, are easier to learn than dependencies involving a vowel dimension and a consonant dimension, e.g., vowel height and consonant voicing. Moreton argues that the modularity bias does not come from either perception or articulation because the voicing of a consonant influences the height of the preceding vowel as much as the identity of the following vowel does, and the acoustic consequences of the voicing of a consonant on the voicing of a non-adjacent consonant in the same word are relatively minor. In this paper, I employ the Garner interference paradigm (Garner, 1974) to show that the modularity bias does in fact arise in perception and thus can be seen as a type of channel bias, arising from how acoustic cues are parsed into cognitive representations (Blevins, 2004: 151-153).

Introduction

Moreton (2008) and Wilson (2003) introduce a distinction between channel bias, "the effect of systematic errors in transmission between speaker and listener", and analytic bias, "cognitive predispositions which make learners more sensitive to some patterns than others" (Moreton, 2008: 83). Moreton (2008:87) distinguishes several subtypes of channel bias, including differences in magnitudes between phonetic precursors of various phonological patterns (e.g., the degree to which the F2 of a vowel is affected by the F2 of the following vowel as opposed to the voicing of the following consonant), "differences in perceptual similarity between sounds", "differences in auditory robustness of acoustic cues", and, crucially, "cognitive biases, specific to language, in how acoustic cues are parsed into phonological representations (Blevins 2004:151-153)". Thus, not all cognitive biases are analytic biases. In order to count as an analytic bias, the bias must be localized in the learner module, which is hypothesized to generalize over phonological representations that are the output of perception (Moreton, 2008: Figure 1). Perception and production are assumed to be part of the "channel". Thus in order to show that a certain cognitive predisposition is an analytic bias, the bias should only manifest itself when the learner generalizes across multiple perceived stimuli, and not during the perception of a single stimulus.

It is uncontroversial to say that every learner is biased in favor of certain hypotheses, since every set of observations is consistent with a vast, possibly infinite, set of hypotheses (Mitchell, 1997). For instance, if we observe that AB, AC, and AD are classified into Category X, while BA, BC, and BD are classified into Category Y, we can hypothesize that 1) if there is an A in first position, the stimulus is an X, while if there is a B, it is a Y, or being more specific, 2) if there is an A in first position, the stimulus is an X, while if there is a B, it is a Y, unless the items in the two positions are identical, or 3) the stimuli

AB, AC, and AD are X's while all other stimuli are Y's, etc. However, as Moreton (2008) points out, analytic biases influencing the learning of phonology have not been experimentally demonstrated.[2]

Moreton reports a new type of learning bias, which he calls 'modularity bias', and argues that it is an analytic bias, rather than a channel bias. Moreton finds that adult English-speaking learners presented with CVCV stimuli learn dependencies between voicing features of the two consonants or height features of the two vowels more easily than they learn a dependency between the height of the first vowel and the voicing of the second consonant. Through a meta-analysis of a number of phonetic studies, Moreton argues that the acoustic precursor for the hard-to-learn height-voice dependency is as large as the acoustic precursor for the easy-to-learn height-height dependency and larger than the precursor for the voice-voice dependency. Thus, he argues, modularity bias is an analytic bias, not a channel bias.

In this paper, I argue that this is not a valid conclusion if analytic biases must be localized in the learner module, which is assumed to operate on the output of perception, and not on raw acoustics. Even if the acoustic precursors for two patterns (the raising of a vowel's F1 before voiced consonants and height harmony) are equally robust, one still needs to account for the perceptual (encoding) processes mediating between the acoustics and the learner. In the remainder of the paper, I report two experiments showing that the features linked by easy-to-learn dependencies are less perceptually separable (Garner, 1974; Garner & Felfoldy, 1970) than the features linked by hard-to-learn dependencies.

Garner & Felfoldy (1970; Garner 1974) have shown that certain stimulus dimensions (e.g., hue and saturation) are more perceptually "integral" (less perceptually separable) than others (e.g., color and shape). The degree of perceptual separability between a pair of stimulus dimensions is measured by testing whether random variation on one dimension adversely influences the speed and/or accuracy of categorization of the stimuli along the other dimension. If perceivers are slower in classifying stimuli along dimension X when the stimuli vary randomly along dimension Y than when dimension Y is held constant, X and Y are said to be perceptually integral.

The working hypotheses in the present study are that, for a CVCV stimulus, 1) random variation in the height of the second vowel will slow down categorization of the height of the first vowel more than would random variation in the voicing of the second consonant, and 2) random variation in the voicing of the first consonant will slow down categorization of the voicing of the second consonant more than would random variation in the height of the first vowel. Thus, voicing features of different segments should be combined in perception, as should height features, while voicing would remain relatively separable from height. If this is the case, then modularity bias could be a type of channel bias, coming from "cognitive biases, specific to language, in how acoustic cues are parsed into phonological representations" (Blevins 2004: 151-153; Moreton, 2008: 87) and analytic biases in phonological learning would still remain to be demonstrated experimentally.

Methods

For these experiments, I used the stimuli constructed by Moreton (2008, available at http://www.unc.edu/~moreton/Stimuli/PhonologyInPress/). He describes the stimulus set as follows:

[2] One somewhat clear example of an analytic bias that has been defended on theoretical grounds is the Subset Principle, a hypothetical tendency to come up with the most specific possible generalization that fits the observed data (e.g., Berwick, 1986; Dell, 1981; Hale & Reiss, 2003; Langacker, 1987; see Finley, 2008; and Xu & Tenenbaum, 2007, for counterevidence).

> Stimuli were synthesised using the MBROLA diphone concatenative synthesiser (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken 1996), using the "US 3" voice (a male speaker of American English). Each "word" was synthesised individually. The nominal duration parameters for both consonants were set to 100 ms, while those for both vowels were set to 225 ms, with 150 ms of silence initially and finally. Intonation was left at the default monotone of 123 Hz. In order not to perturb the natural intensity difference between high and low vowels, no amplitude normalisation was applied. (Moreton 2008: 97)

Since the present study is concerned with perceptual integrality of consonant and vowel features, rather than response competition, the stimuli were constrained so that the consonants of each CVCV word differed in place of articulation, and vowels differed along the front/back dimension. The consonants were chosen from the set {[t], [d], [k], [g]}. The vowels were chosen from the sent {[i], [æ], [u], [o]}. These are the same vowel and consonant sets used in Moreton (2008).

In Experiment I, 36 listeners were asked to categorize the first vowel in each word as either the vowel in ‘beet’ ([i]) or the vowel in ‘bat’ ([æ]). There were several blocks of trials. In all blocks, the first consonant in a word could be either [t], [d], [k], or [g]. In some blocks, which I shall call the *vowel variation blocks,* the second consonant was held constant while the second vowel varied between [u] and [o]. In *consonant variation blocks* the second vowel was held constant while the second consonant varied between a voiced consonant and a voiceless consonant with the same place of articulation. The sequence of trials within a block was randomized separately for each subject. The sequence of blocks was counterbalanced across subjects, as shown in Table 1. The counterbalancing ensures that each position in the block sequence hosts consonant variation and vowel variation blocks equally often, thus if a difference between consonant variation and vowel variation blocks is found, it cannot be attributed to the block sequence.

Sequence 1		Sequence 2		Sequence 3		Sequence 4	
C_2	V_2	C_2	V_2	C_2	V_2	C_2	V_2
t	{u;o}	{t;d}	u	{t;d}	o	t	{u;o}
{t;d}	u	t	{u;o}	t	{u;o}	{t;d}	o
d	{u;o}	{t;d}	o	d	{u;o}	{t;d}	u
{t;d}	o	d	{u;o}	{t;d}	u	d	{u;o}
{k;g}	u	k	{u;o}	k	{u;o}	{k;g}	o
k	{u;o}	{k;g}	u	{k;g}	o	k	{u;o}
{k;g}	o	g	{u;o}	{k;g}	u	g	{u;o}
g	{u;o}	{k;g}	o	g	{u;o}	{k;g}	u

Table 1. The sequences of blocks presented to individual subjects in Experiment I. The C_1 in $C_1V_1C_2V_2$ is always {t;d;k;g}, V_1 is always {i;æ}. Each subject was exposed to one sequence.

In Experiment II, one group of 20 listeners was asked to categorize the second consonant in each word as either [k] or [g]. Another group of 24 listeners classified the second consonant as either [t] or [d]. The first vowel was either fixed (in consonant variation blocks) or varied between [æ] and [i]. For subjects who categorized velars, the first consonant could either be fixed or varied between [t] and [d]. For subjects who categorized alveolars, the first consonant could either be fixed or varied between [k] and [g]. The sequences of blocks presented to subjects are shown in Table 2.

Each block in each experiment consisted of 64 stimuli. Subjects were allowed to take a break between blocks. The entire experiment lasted 10-15 minutes depending on the individual subjects’ reaction times. The subjects were allowed up to 10 seconds to respond. To respond, the subject pressed a button on a button box. The buttons were labeled with the segments they corresponded to and an example word for each segment (the same as in the instructions shown in the Appendix). The subjects were instructed to respond as fast as possible without sacrificing accuracy (see the Appendix for instructions). If the subject did not respond within 10 seconds, the next stimulus was presented, and the trial was excluded from analysis. If the subject responded, the response time and the button pressed were recorded. Incorrect responses and responses with reaction times that were further than three standard deviations away from the grand mean were excluded. Overall, 333 responses, or 1.8% of total, were excluded from the analysis. Reaction times were logarithmically scaled for the purposes of analysis in order to reduce skewing.

Block	Sequence 1		Sequence 2		Sequence 3		Sequence 4			
	C_1	V_1	C_1	V_1	C_1	V_1	C_1	V_1	C_2	V_2
1	g	{i;æ}	{k;g}	æ	{k;g}	i	k	{i;æ}	{t;d}	o
	d		{t;d}		{t;d}		t		{k;g}	
2	{k;g}	æ	g	{i;æ}	k	{i;æ}	{k;g}	i	{t;d}	
	{t;d}		d		t		{t;d}		{k;g}	
3	{k;g}	i	k	{i;æ}	g	{i;æ}	{k;g}	æ	{t;d}	
	{t;d}		t		d		{t;d}		{k;g}	
4	k	{i;æ}	{k;g}	i	{k;g}	æ	g	{i;æ}	{t;d}	
	t		{t;d}		{t;d}		d		{k;g}	
5	{k;g}	æ	g	{i;æ}	k	{i;æ}	{k;g}	i	{t;d}	u
	{t;d}		d		t		{t;d}		{k;g}	
6	g	{i;æ}	{k;g}	æ	{k;g}	i	k	{i;æ}	{t;d}	
	d		{t;d}		{t;d}		t		{k;g}	
7	k	{i;æ}	{k;g}	i	{k;g}	æ	g	{i;æ}	{t;d}	
	t		{t;d}		{t;d}		d		{k;g}	
8	{k;g}	i	k	{i;æ}	g	{i;æ}	{k;g}	æ	{t;d}	
	{t;d}		t		d		{t;d}		{k;g}	

Table 2. The sequences of blocks presented to individual subjects in Experiment II.

The participants in both experiments were introductory psychology students who reported being native English speakers with no history of speech, language, or hearing impairments. The participants received course credit for participation. Nine participants were assigned to each block sequence. Participants were tested either individually or in pairs. Each participant was seated in a separate testing booth, with stimuli presented over headphones.

Results

Participants in Experiment I were slower at categorizing the first vowel in a CVCV as [i] vs. [æ] when the second vowel varied between [u] and [o] than when the second consonant varied between [k] and [g] or [t] and [d] (single-sample Wilcoxon signed rank test on the differences between mean reaction times for consonant variation vs. vowel variation blocks, p=.0007). There was no effect of block sequence (Kruskal-Wallis chi-squared = 2.7805, df = 3, p = 0.4267).[3] The histogram of reaction time differences is shown in Figure 1.

[3] The Wilcoxon and Kruskal-Wallis tests are nonparametric analogs to the t-test and the one-way ANOVA respectively. They are used here because they do not make the assumption of normality and are more conservative than the parametric analogs.

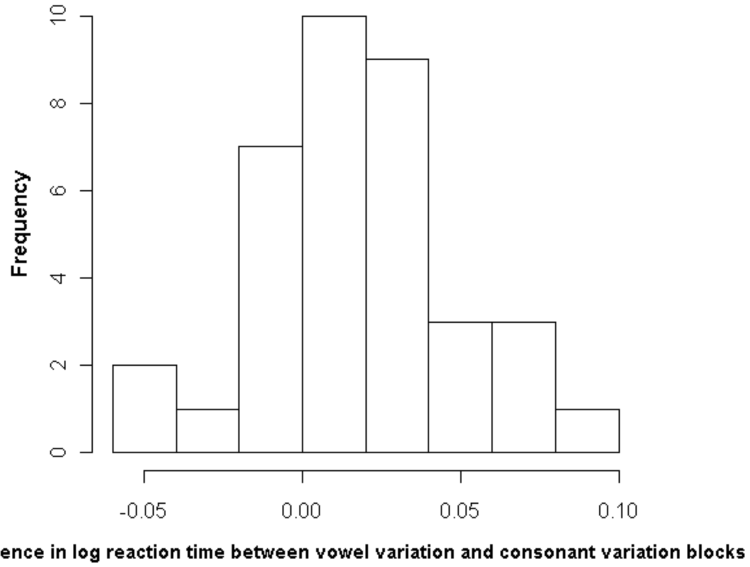

Figure 1. The histogram of differences in log reaction time for categorizing V_1 in a $C_1V_1C_2V_2$ between blocks in which C_2 was held constant and V_2 varied and blocks in which C_2 varied, and V_2 was held constant. A positive difference indicates that a subject had a slower mean reaction time on vowel variation blocks than on consonant variation blocks, i.e., variation in V_2 slowed down categorization of V_1 more than did variation in C_2.

The results of Experiment II are less clear. Subjects asked to classify /t/ vs. /d/ in intervocalic position showed a high error rate (10% compared to 4% for /k/ vs. /g/ classification and 1% for /æ/ vs. /i/ classification). This is not surprising given that both /t/ and /d/ are neutralized to a flap in intervocalic position following a stressed vowel in American English. Listeners varied widely in their error rate on intervocalic alveolar discrimination with one listener showing an accuracy of 55%, three listeners showing an accuracy of 70-80%, 6 listeners falling in the 80%-90% range and 14 listeners falling into 90%-100% range), suggesting that some English speakers lose the ability to discriminate /t/ and /d/ in the environment in which they normally reduce to a flap.

No significant difference between blocks with variation in the preceding consonant and blocks with variation in the preceding vowel was obtained on either reaction time or accuracy measures (Wilcoxon signed rank test, for reaction time, p=.55; for accuracy, p=.87) for the subjects who were asked to discriminate between /t/ and /d/ (no significant effect on reaction time is observed even if only subjects with accuracy above 80% are considered: Wilcoxon signed rank test, p=.43). Neither accuracy nor reaction time interacted significantly with block sequence (p>.05).

In contrast, as shown in Figure 2, subjects asked to discriminate between /k/ and /g/ showed a (barely) significant reaction time effect (Wilcoxon signed rank test, p=.0497) in the expected direction: the second consonant was classified faster and more accurately when the preceding vowel varied than when the preceding consonant varied. There was no significant difference in accuracy between the two block types (Wilcoxon signed rank test, p=.36), although the numerical trend is towards responses being less accurate when the preceding consonant varies. There was a significant effect of block sequence on reaction time differences (Kruskal-Wallis chi-squared = 7.8624, df = 3, p = 0.049): consonant variability slowed down reaction time more than vowel variability for block sequences 1-3 (Wilcoxon signed rank test, p = 0.003) but not for block sequence 4, which was one of the sequences in which the first block featured vowel variability.

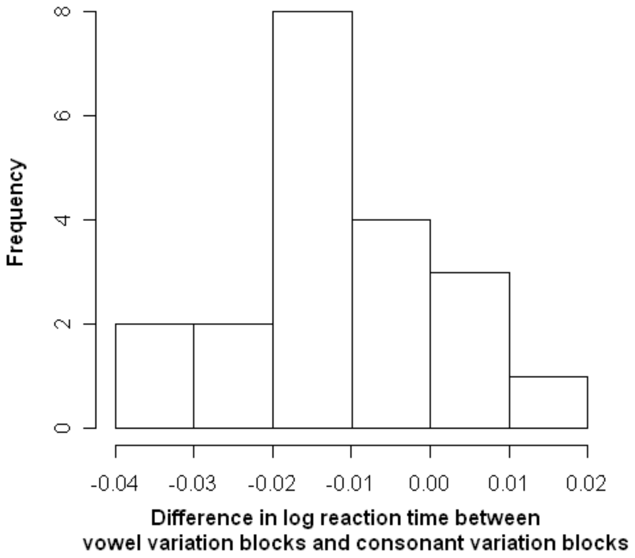

Figure 2. The histogram of differences in log reaction time for categorizing C_2 in a $C_1V_1C_2V_2$ between blocks in which C_1 was held constant and V_1 varied and blocks in which C_1 varied, and V_1 was held constant. A negative difference indicates that a subject had a slower mean reaction time on consonant variation blocks than on vowel variation blocks, i.e., variation in C_1 slowed down categorization of C_2 more than did variation in V_1.

Conclusion

Moreton (2008) reported that dependencies between voicing features of onset consonants of a CVCV word and dependencies between height features of the word 's vowels were easier to learn than height-voice dependencies. He argued that this finding is not due to channel bias, including 'cognitive biases, specific to language, in how acoustic cues are parsed into phonological representations' (Moreton 2008: 87). Rather, Moreton proposes that the bias is localized in the learner, which he assumes to be a separate module mediating between production and perception.

In Experiment I, categorization of a vowel along the height dimension was slowed down more by random variation in the following vowel's height than in the following consonant's voicing, at least for the stimuli used by Moreton (2008) and native English listeners. In Experiment II, categorization of a consonant along the voicing dimension was slowed down more by random variation in the preceding consonant's voicing than in the preceding vowel's height. Although the second result is more tentative, being obtained for [k] vs. [g] but not [t] vs. [d], it is worth noting that the corresponding result in Moreton (2008: 111-113) is also more tentative than the result corresponding to our Experiment I: in Moreton's Experiment II there was "better performance in the V[oice]-V[oice] condition than in the H[eight]-V[oice] condition; however, the coefficient was somewhat smaller, and its significance level much lower, than had been found for the H[eight]-H[eight] Condition in Experiment 1". The weakness of this effect is perhaps unsurprising because voice-voice relationships between non-adjacent consonants are typologically rare and there is little voicing coarticulation between non-adjacent consonants (Moreton 2008: 113).

The present set of results suggests that modularity bias is a channel bias localized in the perceptual processing of speech. Ever since the foundational studies of Posner (1964), Shepard (1964) and Garner and Felfoldy (1970), it has been suggested that certain physical dimensions are mapped onto smaller sets of psychological dimensions, e.g., combining brightness and saturation in Garner and Felfoldy (1970), or loudness and pitch in Grau and Kemler Nelson (1988). That is, perception features many-to-one and many-to-many mappings of physical dimensions onto perceptual dimensions (see Cheng & Pachella, 1984, Garner, 1974, Grau & Kemler Nelson, 1988, Posner, 1964, and Shepard, 1964, for discussion) in which a pair of physical dimensions may be mapped onto overlapping sets of perceptual dimensions. The present results suggest that acoustic correlates of voicing features of non-adjacent consonants may be mapped onto sets of perceptual representations that show a high degree of overlap relative to the degree of overlap between sets of representations representing vowel height and consonant voicing. The modularity bias thus influences how acoustic cues are parsed into phonological or perceptual representations, a type of bias Moreton (2008: 87) identifies as a channel bias. Thus, while there must be some analytic biases influencing phonological learning, these biases remain to be discovered.

References

Berwick, R.C. 1986. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.

Blevins, J. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

Cheng, P. W., & R. G. Pachella. 1984. A psychophysical approach to dimensional separability. *Cognitive Psychology, 16,* 279-284.

Dell, F. 1981. On the learnability of optional phonological rules. *Linguistic Inquiry*, *12,* 31-37.

Finley, S. 2008. Formal and cognitive restrictions on vowel harmony. Unpublished Ph.D. Dissertation: Johns Hopkins University.

Garner, W. R. 1974. *The processing of information and structure*. New York: Wiley.

Garner, W. R., & G. L. Felfoldy. 1970. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1,* 225-241.

Grau, J. W., & D. G. Kemler Nelson. 1988. The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, *117*, 347-370.

Hale, M., & C. Reiss. 2003. The Subset Principle in phonology: Why the *tabula* can't be *rasa*. *Journal of Linguistics, 39,* 219-244.

Langacker, R. W. 1987. *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford: CSLI.

Mitchell, T. M. 1997. *Machine learning*. McGraw-Hill.

Moreton, E. 2008. Analytic bias and phonological typology. *Phonology, 25,* 83-127.

Posner, M. I. 1964. Information reduction in the analysis of sequential tasks. *Psychological Review, 71*, 491-504.

Shepard, R. N. 1964. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1,* 54-87.

Wilson, C. 2003. Analytic bias in artificial phonology learning: consonant harmony vs. random alternation. Paper presented at the Workshop on Markedness and the Lexicon, Massachusetts Institute of Technology, January 25, 2003.

Xu, F., & J. B. Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological Review, 114,* 245-272.

Appendix: Instructions for the experiments

Experiment I:

You will hear made-up words over the headphones.
Your task is to identify the FIRST vowel in each word.

Press the correct button as soon as you know which it is.

If the FIRST vowel is the same as the vowel in 'BAT’, then press the LEFT button.

If the FIRST vowel is the same as the vowel in 'BEAT’, then press the RIGHT button.

Experiment II:

Participants who categorized velars:

You will hear made-up words over the headphones.
Your task is to identify the SECOND CONSONANT in each word.

Press the correct button as soon as you know which it is.

If the second consonant is ‘k’ as in 'rocky’, then press the LEFT button.

If the second consonant is ‘g’ as in 'rugged’, then press the RIGHT button.

Participants who categorized alveolars:

You will hear made-up words over the headphones.
Your task is to identify the SECOND CONSONANT in each word.

Press the correct button as soon as you know which it is.

If the second consonant is ‘t’, then press the LEFT button.

If the second consonant is ‘d’, then press the RIGHT button.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Visual Recognition Memory in 5- and 8-Month-Old Infants and its Relation to Vocabulary Development[1]

Swapna Musunuru, Derek M. Houston and Sarah Pope[2]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by a research grant awarded to Derek Houston (R01DC006235) and a training grant to David Pisoni (T32DC00012). We would like to acknowledge the staff of the Infant Language Lab, especially coordinator Kabreea York, for their assistance in recruiting and testing infants. We are also very grateful to the infants and their parents who participated in the research.

[2] Department of Otolaryngology – Head & Neck Surgery; Indiana University School of Medicine, Indianapolis, IN

Visual Recognition Memory in 5- and 8-Month-Old Infants and its Relation to Vocabulary Development

Abstract. Early infancy measures that may predict language and cognitive outcomes are important diagnostic tools when considering early intervention strategies. These infancy measures are especially of value to hearing impaired populations whose language development is often significantly delayed. These populations can benefit tremendously from early intervention. Visual recognition memory is one such infancy measure that relies on measuring short-term memory capacity to predict later cognitive outcomes. In this study, we used a variant of a visual recognition memory task designed by Rose et al. (2001) to test normal hearing and hearing impaired infants' visual recognition memory. Results suggested that 8.5-month-olds recognized the third object in the largest span length, though 5-month-olds did not demonstrate any notable preferences. A correlation between looking scores and the MacArthur Bates CDI also indicated that infant ability to recognize objects is related to their ability to their early gesturing skills.

Introduction

Finding early predictors for language and cognitive outcomes is important for identifying infants at risk for later deficits, especially since early identification is the first step towards implementing an intervention. Outcome predictors are especially valuable to obtain for hearing impaired populations as several studies have demonstrated the benefits of early intervention for children with hearing impairment. Typically, these populations experience significant delays in language development and academic achievement (Yoshinaga-Itano et al., 1998). Early identification of hearing loss in children prior to six months of age immediately followed by an appropriate intervention has been shown to improve language, speech, and social-emotional development in comparison to children identified later in life (Yoshinaga-Itano, 1999; Downs & Yoshinaga-Itano, 1999; Yoshinaga-Itano et al., 1998). In particular, those with hearing losses identified before six months of age had higher receptive and expressive language quotients, larger productive vocabulary lexicons, increased speech intelligibility, better academic achievement, and less developmental delay compared to those identified at a later age (Yoshinaga-Itano, 1999; Downs & Yoshinaga-Itano, 1999; Yoshinaga-Itano et al., 1998).

Since language abilities are frequently impaired in populations with hearing loss, determining which abilities predict language development is especially important for children with hearing impairments (Miyamoto et al., 2003). In normal hearing infants, later language development may be predicted through expressive and receptive language performances, early phonetic speech perception, speech segmentation ability, as well as various speech processing tasks (Newman et al., 2006; Tsao, Liu, & Kuhl 2004; Hohm et al., 2007). However, these auditory-based measures cannot be used to predict language outcomes in deaf infants; more appropriate measures may be a range of non-auditory cognitive tasks.

Since attention is an important component of learning in infancy, understanding the connection between attention and language outcomes could provide insight regarding which early abilities are predictive of later cognition. Tasks that examine language and cognitive predictors but do not depend on auditory experience have been conducted with normal hearing children, though only a few have examined predictors in hearing impaired populations. Among studies that have examined pre-cochlear implant predictors, Bergeson and colleagues (2001) demonstrated that pre-implantation Pediatric Sentence Intelligibility scores are strongly correlated with vocabulary, receptive and expressive language, and speech intelligibility scores obtained two years post-implantation. These findings imply that the pre-implantation lip-reading and audiovisual speech perception scores measured by Pediatric Sentence Intelligibility can be used to predict speech and language skills after several years of implant use. Audiovisual speech perception measures may also provide reliable behavioral markers that can be used to predict and identify which children will benefit the most from their cochlear implants (Bergeson et al., 2001). Other studies conducted by Horn and colleagues (2004) have demonstrated the link between motor development and perceptual-motor functions with post implant language measures. Individual differences in visuomotor integration skills in prelingually deaf children are predictive of post-implant performance on open-set speech perception, auditory sentence comprehension, and speech intelligibility skills when measured over three years of cochlear implant experience (Horn et al., 2004). Motor development measures taken from behavioral functioning assessments (specifically the Vineland Adaptive Behavioral Scales) have also been predictive of performance on spoken-word recognition, receptive and expressive language, and vocabulary knowledge tests measured three years post-implantation (Horn et al., 2005). It is important to note that the previously mentioned measures are all indicated for older CI children, since early identification and intervention often does not take place in infancy. As of yet there are not many studies examining early abilities in infancy for children with hearing impairment.

In contrast, a number of studies have looked at non-auditory cognitive predictors for language and cognitive outcomes in children with normal hearing. For instance, multiple components of information processing such as memory, processing speed, attention, representational competence and cross modal transfer have all been shown to predict later differences in IQ and the Mental Development Index (Rose et al., 2005; Rose and Feldman, 1995; Rose & Feldman, 1997). Information processing in infancy is even linked to adolescent intelligence scores (Sigman & Beckwith, 1997) and young adult IQ and academic achievement (Fagan, Holland, & Wheeler, 2007). Studies on infant habituation and novelty preference have demonstrated a link between attention and cognitive outcomes, such that shorter looking times were indicative of better outcomes in childhood (McCall & Carriger, 1993; Colombo, Shaddy, Richman, Maikranz, & Blaga, 2004). In addition, language impairments, such as specific language impairment (SLI), are associated with impaired cognitive processing (Benasich & Tallal, 2002) and impaired auditory processing (Choudbury, Leppanen, Leeyers & Benasich, 2007). In typically developing populations, visual novelty preferences in infancy predict not only IQ, but also language abilities independent of IQ level (Thompson, Fagan & Fulker, 1991).

One particular component of information processing, visual recognition memory (VRM), is highly related to outcomes (Colombo, Shaddy, Richman, Maikranz & Blaga, 2004; Fagan & Krahe McGrath, 1981). Specifically, better information processing is correlated with better visual recognition memory and consequently better cognitive outcomes (Rose et al., 2001; Rose & Feldman, 1997). Visual recognition memory is considered a particularly strong predictive measure for later specific cognitive outcomes and demonstrates that early cognitive processes are important for later language acquisition (Rose et al., 1992; Rose et al., 1991). VRM requires learning to attend to some features and to ignore others and form perceptual categories. Formation of these categories is important for early language acquisition. Children's abstraction of perceptual features forms the basis for concepts of objects and these concepts need to be in place before language may be acquired (Rose et al., 1991). Specific language outcomes that have been measured through VRM include receptive and expressive language at 2.5, 3, 4, and 6 years, vocabulary scores at 4, 7, and 11 years, IQ at 3, 4, 5, 6, and 11 years and language proficiency at 3 years (Rose et al., 2004; Rose et al., 1991).

The goal of the present study was to extend studies conducted on normal hearing infants to hearing impaired infants and toddlers in hopes of finding pre-implant predictors in infancy and early childhood. Employing a similar span-task paired-comparison paradigm as used in Rose et al. (2001), we tested both normal hearing and hearing impaired infants on their VRM. A series of images was shown for familiarization, followed by the same images paired with new images for the test phase. The target images were in either a series of two or three (spans) and the percentage of time looking at the new image was calculated. This preference for looking at the new image indicates short term memory, or their visual recognition memory. We will then observe how the individual differences in infant VRM scores predict later language outcomes, as measured by the MacArthur Bates Cognitive Development Index and other language measures. It is hoped that by assessing VRM for these infants, we can predict the nature and severity of later language deficit they may incur and consequently implement appropriate early intervention measures.

Method

Participants

Twenty 5-month-old infants (mean age: 5.0, range: 4.3-5.6) and fifteen 8.5-month-olds (mean age: 8.6, range: 8.1-9.0) with normal hearing participated in this study. All participants were recruited from the greater Indianapolis metropolitan area and passed their newborn hearing screenings.

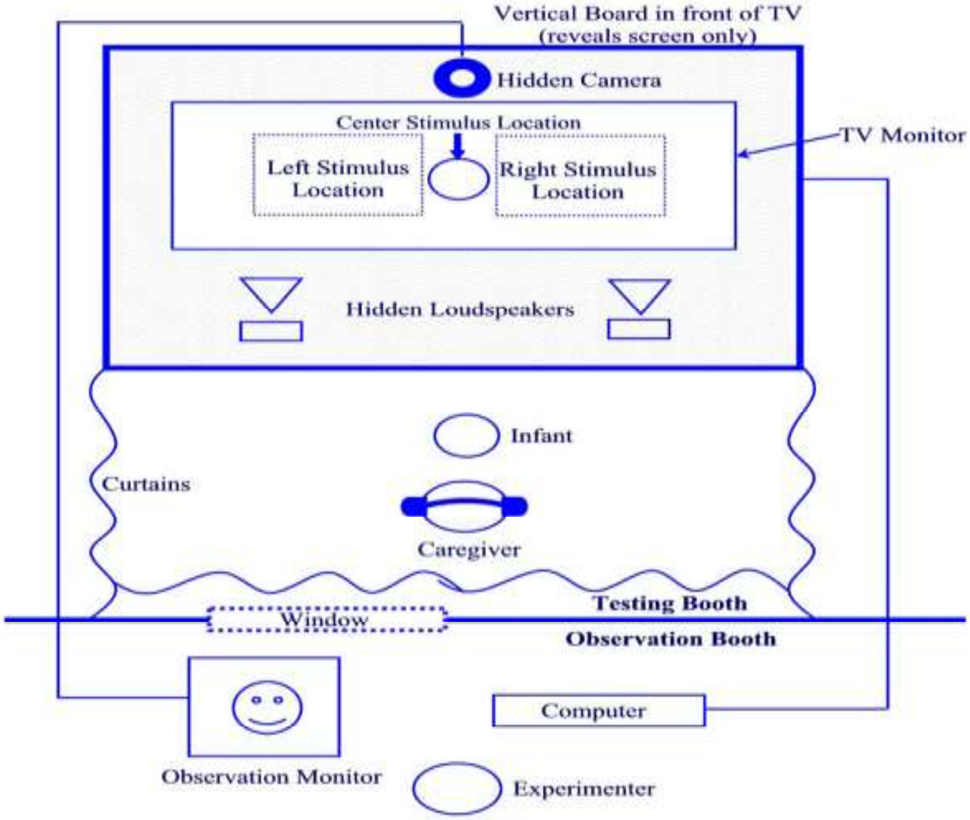

Figure 1. During the span tasks, the caregiver sits in a neutral position looking forward and holds the infant in his or her lap. The visual stimuli appear at the left and right stimulus locations. The “attention getter” appears at the center stimulus location.

Apparatus

Testing was conducted in a custom-made double-walled IAC sound booth (see Figure 1). Infants sat on their caregivers' laps approximately five feet in front of a 55 inch wide-aspect television monitor. Visual stimuli were displayed on the monitor at approximately eye level to the infants. The experimenter observed the infant from a separate room via a hidden, closed-circuit digital camera and controlled the experiment using the Habit software package (Cohen et al. 2004) running on a Macintosh G5 desktop computer.

Stimuli

The stimuli consisted of 24 images of colorful objects. Images were found using an image search on the Internet and were selected if it was unlikely that the infant would already be familiar with the image (e.g. an image of a spoon would not have been selected since it is likely that the infant has seen a spoon before, but an image of a unique candleholder would have been selected). Images were then organized into 12 pairs: 2 pairs for familiarizing the infant to the testing procedure and 10 pairs for the experiment. Image pairs were designed to be easily discriminable from each other yet equal in attractiveness – no single image in a paired set was significantly more intricate or colorful than its corresponding paired image (See Figure 2). In order to create the paired image slide, Photoshop was used to create an initial 12x8 inch blue background template slide with two equally sized white boxes placed side by side on top of the blue background. Next, the individual images from each pair were made an equal size (0.75x0.75 inches) and pasted within the white boxes. This process was repeated for each paired set of images to yield a collection of slides (12 total) that were identical to one another except for the unique image in each white box. Two additional familiarization slides were then made for each corresponding paired image slide. These familiarization slides consisted of the same 12x8 inch blue background slide with a single centrally located white box which was equivalent in size to the paired image slide boxes. Each familiarization slide corresponding to a given paired image slide consisted of one of the two images in the pair. For example, if a paired image slide consisted of images A and A', two familiarization slides were created - one with image A in the centrally located box, and the other with image A' in the centrally located box. This was done for counterbalancing purposes so that any bias for one image over the other would be canceled out since half of the infants would be familiarized with image A while the other half was familiarized with A'. All of these images were again resized to 0.75x0.75 inches before they were pasted into the centrally located box. This process was repeated for every paired image slide of the experiment phase and for only one of the two images in each of the two familiarization phase slides, thus yielding a total of 22 familiarization slides (20 for the experiment phase and 2 for the familiarization phase).

Once pairs were created, pairs were further organized into two sets of spans of two and three. Pairs within each span were designed to be highly discriminable from other pairs in the span. Extra care was also taken to make sure that images that appeared in the left boxes were all extremely different from one another, and images that appeared in the right boxes were extremely different from one another.

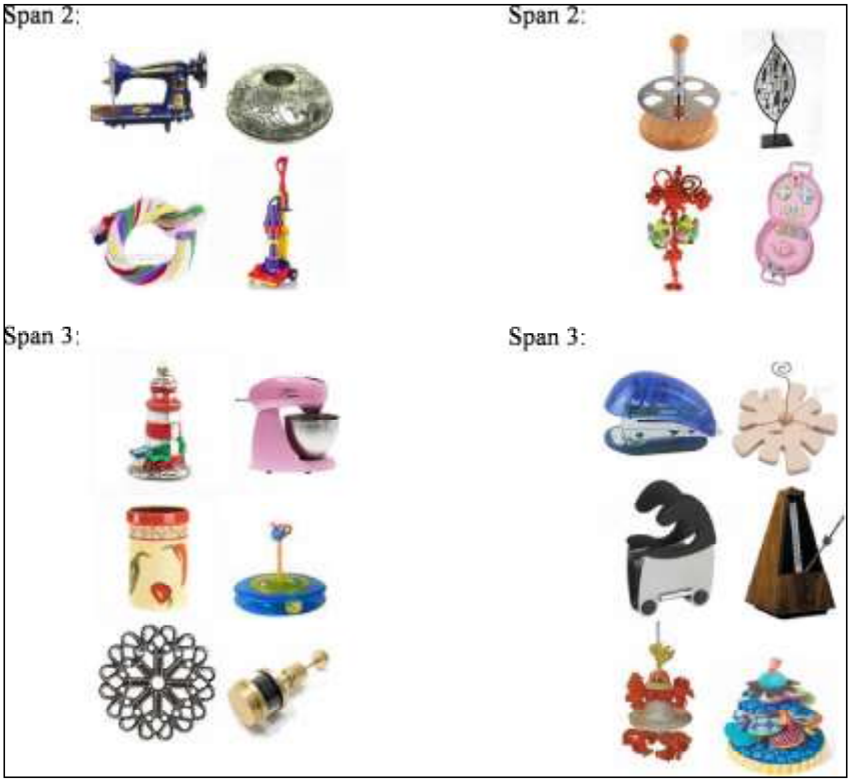

Figure 2. Images were placed in pairs of equal attractiveness, and then further grouped into two groups of span 2 problems and span 3 problems.

Aside from the 24 stimuli, an "attention-getter" video clip was also created. This clip consisted of a black screen with an animation of a baby laughing located at the center of the screen. This was used to redirect the infants' attention back to the screen in between trials.

Procedure

The basic design of the experiment consisted of 2 sets of 5 problems each, in spans of two or three items. The procedure for each of these spans followed a paired-comparison paradigm. For each span, the infant was familiarized to two or three images in succession depending on span length, and then given a series of test trials with each successive familiar image now paired with a new image. For example, for a span of two, the infant would first be familiarized with images "A" and "B" in succession, and then tested with A vs. A' and B vs. B' (the prime mark denotes the new image in the pair). All spans were presented in ascending order (from span 2 to span 3). Previous studies have demonstrated that ascending versus descending order of span length does not affect outcome of results (Rose et al., 2001).
Two initial pre-test problems were given first to familiarize the infant with the testing procedure. Each problem consisted of a single familiarization slide followed by its corresponding paired image slide, with the novel image being on the right side for pre-test test trial 1 and on the left side for pre-test test trial 2. Each slide was shown for a total of 10 seconds and in between slides the brief "attention-getter" clip was shown to redirect the infant's attention to the screen.

For the test phase of the experiment, infants were presented with two sets of the five problems mentioned earlier. There was no break between the two sets. Each of the familiarization slides were shown for a total of 10 seconds, followed immediately by the paired image slides, also each shown for 10 seconds. Familiarization slides were selected at random to ensure that there would be an equal number of paired image slides with the novel stimuli on both the right and left sides to control for side preference. In addition, each infant had a different randomized set of familiarization images, which also helped control error due to side preference and image preference. In between each slide, the infant's attention was always redirected to the screen by the "attention-getter" clip.

The infant's looks during testing were recorded by a digital video camera operated by a research assistant who adjusted the camera throughout testing to maintain focus on the infant's eyes. The digital video recordings were then coded for eye-movements offline, via a frame-by-frame analysis. During coding, the researcher was aware of when the infant was looking at a paired versus a familiarization slide, but was not aware of which side the novel stimuli was located on in the paired slides. During familiarization trials, looks were coded as either "center" or "away." During test trials, looks were coded as either "left," "right," or "away." The beginning of each trial was determined as the moment that the infant's face became bright as consequence of the TV monitor shining light from the light-colored study slides into the dark testing room.

Once raw frame-by-frame data was obtained from coding the videos, it was entered into a spreadsheet which then computed the total looking time and longest looks in each direction and to target versus non-target via use of a macro designed to compute these factors. These looking times were then entered into SPSS for further analysis.

Results

Span Three

Because the number of conditions differed between the span 2 and span 3 strings, the data from each span were analyzed separately. For Span 3, we performed a three-way, repeated measures analysis of variance (ANOVA) with age (5 months, 8.5 months) as a between-subjects factor, and stimulus condition (novel, old), serial position (one, two, three) and phase (one, two) as within-subject factors. The analysis demonstrated a main effect of phase, F(1,33) = 6.38, p = .02, and an interaction between target, serial position and age group, F(2,33) = 5.88, p = .007. Due to this interaction we decided to perform two-way repeated measures ANOVAs combined across phases by age group and serial position. Among 5-month-olds in span 3 there were no main effects or interactions. However, in the 8.5-month-old population, there was a main effect of phase in serial position 3, F(1,14) = 12.9, p = .003, as well as an effect of target that neared significance, F(1, 14) = 4.30, p = .057. Since there was no significant interaction between target and phase, we grouped across phase and used t-tests to assess looking time differences at each serial position. We found no significant difference in any of the serial positions for 5-month-olds, but there was a significant difference in looking times in serial position 3 for the 8.5-month-olds, t(14) = 3.59, p = .003, suggesting 8.5-month-olds recognized the final objects in the span 3 problems.

Span Two

We decided to analyze the data from span 2 in the same way that we analyzed the data in span 3, in order to compare the phases to each other. The only difference between the analyses was that the serial position factor had three levels (one, two, three) rather than two. In the 5-month-old age group, there was a significant interaction between phase and target in Span 2, serial position 2, F(1,19) = 4.69, p = .043. Due to this, we analyzed differences in looking time by both phase and serial position for the 5-month-olds in span 2. However, there were no significant differences in any of the serial positions, regardless of phase. There were also no significant differences in the 8.5-month-olds.

MacArthur Bates CDI

Out of the twenty 5-month-olds, 13 had completed the MacArthur Bates CDI (MCDI) in time for analysis. Out of fifteen 8.5-month-olds, 12 subjects completed the MCDI. Correlations between performance on each phase of the recognition memory task and infants' vocabulary as measured by the MCDI were assessed. Infants' overall performance in phase 2 correlated significantly with their early gestures (r = .59, p = .002), suggesting that infants' ability to recognize objects is related to their ability to recognize early gestures.

Discussion

For the span 2 task we found no statistically reliable evidence of object recognition for either age. These findings are inconsistent with the Rose et al. (2001) study, where a novelty response was demonstrated in the span 2 task. One possible reason for our failure to replicate the findings is that we did not have enough statistical power. Consistent with that possibility, infants' mean looking times were in the direction of a novelty preference, though they were not statistically significant. Further testing with additional subjects may lead to the trend being statistically reliable.

Differences between these results and Rose et al. may also be due to differences in the nature of the stimuli. Two-dimensional images on a TV screen were used in this study while 3-D objects placed in a tray were used in the Rose study. While 3-D objects are more complex than 2-D images, the 2-D images may have been less stimulating or they may have been more complex to process since they appeared on a TV screen, a presentation mode that the infant may not be used to. This may have contributed to increased task difficulty, and the subsequent familiarity response.

The span 3 task, in contrast to the span 2, demonstrated interesting serial position effects. The 8.5-month age group demonstrated a recency effect, an indication that the images were stored in their short-term memory (as opposed to moving to their long-term memory). Just as demonstrated by Rose et al. (2001), they had a significant novelty response to novel over familiar in serial position 3 of span 3. The 5-month old age group, on the other hand, did not show object recognition at any serial position.

Aside from measuring performances on the span task, another goal for this study was to correlate results from the span task to language development via use of the MacArthur Cognitive Development Index. We found that performance on the span task did correlate with early gestures. This suggests that while our version of the span task may be more difficult for infants that the Rose et al. task, it is possible that our version may serve as a tool for predicting language acquisition in deaf children before implantation. We are currently testing that possibility.

References

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predictslater language impairment**.** *Behavioural Brain Research, 136,* 31 – 49.

Bergeson, T.R., Pisoni, D.B., & Davis, R.A.O. (2001). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review, 103(4),* 347/370.

Choudbury, N., Leppanen, P., Leevers, H., & Benasich, A. (2007). Infant information processing and family history of specific language impairment: Converging evidence of RAP deficits from two paradigms. *Developmental Science, 10*, 213-236.

Cohen LB, Atkinson DJ, Chaput HH. Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas 2004.

Colombo J., McCardle, P., & Freund, L. (Eds.). (2009). *Infant pathways to language: Methods, models, and research directions*. Mahwah, NJ: Erlbaum.

Colombo, J., Shaddy, D. J., Richman,W. A., Maikranz, J. M., & Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. *Infancy, 5*, 1−38.

Downs, M.P., Yoshinaga-Itano, C. (1999). The efficacy of early identification and intervention for children with hearing impairment. *Pediatric Clinics of North America, 46*, 79-87.

Fagan, J., Holland, C., & Wheeler, K. (2007). The prediction, from infancy, of adult IQ and achievement. *Intelligence, 35*, 225-231.

Fagan, Joseph F., & Krahe McGrath, Susan (1981), Infant recognition memory and later intelligence. *Intelligence, 5*, 121-130.

Hohm, E., Jennen-Steinmetz, C., Schmidt, M.H., & Laucht, M. (2007). Language development at ten months: Predictive of language outcome and school achievement ten years later? *Eur Child Adolesc Psychiatry, 16*, 149-156.

Horn, D.L., Davis, R.A.O., Pisoni, D.B., & Miyamoto, R.T. (2004). Visuomotor integration ability of pre-lingually deaf children predicts audiological outcome with a cochlear implant: A first report. *International Congress Series, 1273,* 356-359.

Horn, D.L, Pisoni, D.B., Sanders, M.S., & Miyamoto, R.T. (2005). Behavioral assessment of prelingually deaf children before cochlear implantation. *The Laryngoscope, 115*, 1603-1611.

Hunter, M.A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5,* 69-95.

Knoll, T., Steetharam, N., Coven, A., Kmoch, J., Byer, S., et al. (2007). Adobe Photoshop CS3 (Version 10.0): Adobe Systems Incorporated.

McCall, Robert B., & Carriger, Michael S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, *64*, 57−79.

Miyamoto, R.T., Houston D.M., Kirk, K.I., Perdew A.E., & Svirsky, M.A. (2003). Language development in deaf infants following cochlear implantation. *Acta Otolaryngol*, *123*, 241-244.

Newman, R., Ratner, N.B., Jusczyk, A.M, Jusczyk, P.W., & Dow, K.A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology, 42*, 643-655.

Rose, S.A., & Feldman, J.F. (1995). Prediction of IQ and specific cognitive abilities at 11 years from infancy measures. *Developmental Psychology, 31*, 685-696.

Rose, S.A., & Feldman, J.F. (1997). Memory and speed: Their role in the relation of infant information processing to later IQ. *Child Development, 68*, 630-641.

Rose, S.A., Feldman, J.F., & Jankowski J.J. (2001). Attention and recognition memory in the 1st year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology, 37(1)*, 135-151.

Rose, S.A., Feldman, J.F., & Jankowski, J.J. (2001). Visual short-term memory in the first year of life: capacity and recency effects. *Developmental Psychology, 37(4),* 539-549.

Rose, S.A., Feldman, J.F., & Jankowski, J.J. (2004). Infant visual recognition memory. *Developmental Review, 24*. 74-100.

Rose, S.A., Feldman, J.F., Jankowski, J.J., & Van Rossem, R. (2005). Pathways from prematurity and infant abilities to later cognition. *Child Development*, *76*, 1172-1184.

Rose, S.A., Feldman, J.F., & Wallace, I.F. (1991). Language: A partial link between infant attention and later intelligence. *Developmental Psychology*, *27*, 798-805.

Rose, S.A., Feldman, J.F., & Wallace, I.F. (1992). Infant information processing in relation to six-year cognitive outcomes. *Child Development*, *63*, 1126-1141.

Rose, S.A., Gottfried, A.W., Melloy-Carminar, P., & Bridger, W.H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*, 704-713.

Sigman, M., Cohen, S. E., & Beckwith, L. (1997). Why does infant attention predict adolescent intelligence? *Infant Behavior and Development*, 20, 133–140.

Thompson, L., Fagan, J., & Fulker, D. (1991). Longitudinal prediction of specific cognitive abilities from infant novelty preference. *Child Development*, *67,* 530-538.

Tsao, F., Liu, H., & Kuhl, P.K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*, 1067-1084.

Yoshinaga-Itano, C. (1999). Benefits of early intervention for children with hearing loss. *Otolaryngologic Clinics of North America, 32*, 1089-1102.

Yoshinaga-Itano, C., Sedey A.L., Coulter, D.K., & 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. 29 (2008)
Indiana University

Infant Dialect Discrimination[1]

Jennifer Phan[2] and Derek M. Houston[2]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This research was supported by the NIH-NIDCD Research Grant (R01DC006235), a grant from the IUPUI Undergraduate Research Opportunities Program, a grant from the IUPUI Summer Research Opportunities Program, and an Educational Enhancement Grant from the IU Graduate Student Organization.

[2] Indiana University School of Medicine, Indianapolis, IN

Infant Dialect Discrimination

Abstract. In order to understand speech, infants must differentiate between phonetic changes that are linguistically contrastive and those that are not. Research has shown that infants are very sensitive to fine-grained differences in speech sounds that differentiate words in their own or another language. However, little is known about infants' ability to discriminate phonetic differences associated with different dialects of their native language. Using a Visual Habituation procedure, 7-, 11-, 18-, 24-, and 30-month-olds were tested on their ability to discriminate two linguistically equivalent variants of the diphthong (/aI/) - one produced in their native dialect (North Midland American English) versus one produced in a nonnative dialect (Southern American English). Seven-month-olds discriminated the variants but 11-month-olds did not. Infants from 18 to 30 months of age did not demonstrate statistically significant discrimination, but they did show a trend toward discrimination with increasing age. The findings suggest that dialect discrimination follows a U-shaped course of development. Because 11-month-olds demonstrated the poorest dialect discrimination performance, we are currently assessing their ability to discriminate linguistically different speech sounds varying in degree of acoustic similarity. Preliminary findings suggest that both language experience and acoustic differences may influence infants' discrimination of phonetic contrasts in the native language.

Introduction

In order to understand speech and learn language, infants must be able to distinguish between many different speech sounds. Furthermore, infants must be able to discriminate between phonetic contrasts that are linguistically relevant and those that are not. Cross-language speech perception studies demonstrate that the ability to distinguish and categorize phonetic differences is significantly influenced by the phonological qualities of the native language. Young infants are especially sensitive to differences in a wide variety of speech sounds; however, during the second half of the first year of life, linguistic experience begins to modify infants' ability to distinguish many speech contrasts so that they are generally less attuned to nonnative phonemic distinctions. This perceptual trend usually persists into adulthood (Miyawaki, Strange et al. 1975; Trehub 1976; Werker, Gilbert et al. 1981; Flege and Eefting 1987; Best and Strange 1992). Consequently, most adults have difficulty discriminating nonnative contrasts, even with rigorous laboratory training or early exposure to speech sounds of the foreign language (Lively, Logan et al. 1993; Lively, Pisoni et al. 1994; Pallier, Bosch et al. 1997). The course of early perceptual reorganization is influenced by native language experience and likely influences later language development and speech perception.

The progression from a language-general to a language-specific pattern of speech discrimination has been described for both consonants and vowels. With respect to consonants, investigators have found that younger infants are able to discriminate a wide variety of consonant contrasts, both native and nonnative (Eimas, Siqueland et al. 1971; Holmberg, Morgan et al. 1977; Aslin, Pisoni et al. 1983; Bertoncini, Bijeljac-Babic et al. 1987). However, infants' sensitivities to nonnative consonant differences begin to decline around 10-12 months of age (Werker and Tees 1983; Werker and Tees 1984; Werker and Lalonde 1988; Best 1993; Best, McRoberts et al. 1995). Infants' discrimination of vowel contrasts is also evident early in development (Trehub 1973; Trehub 1976; Bertoncini, Bijeljac-Babic et al. 1987; Kuhl, Williams et al. 1992; Polka and Werker 1994; Polka and Bohn 1996). In early studies, English-learning infants aged one to four months were able to discriminate /a/-/i/ and /a/-/u/ contrasts in their native language (Trehub 1973), and they were also able to differentiate between the nonnative [pa]-[pã] which demonstrate the oral/nasal vowel contrast that occurs only in French and Polish (Trehub 1976). Two-month-olds were able to distinguish subtle contrasts like /i/-/I/ in a continuous (versus categorical) manner (Swoboda, Morse et al. 1976). In the native language, six-month-olds were able to distinguish between the vowel distinctions /a/-/e/ (Kuhl 1983). Six-month-olds (Kuhl 1979) as well as younger 2- and 3-month-olds (Marean, Werner et al. 1992) were also able to discriminate the contrasts /a/-/i/ even when listening to talkers varying across age and gender (Kuhl 1979; Clarkson, Eimas et al. 1989; Marean, Werner et al. 1992).

Modification of consonant perception to adapt to language-specific constraints occurs around ten to twelve months of age; however, this change is thought to take place earlier for vowels (at six to eight months of age). When investigators tested English- and Swedish-learning 6-month-olds on their ability to distinguish vowel archetypes from prototypical and nonprototypical examples using the English and Swedish vowels /i/ and /y/, they discovered that discrimination was maintained around nonprototypical exemplars whereas increased generalization was observed around prototypical exemplars. However, this finding, termed the "perceptual magnet effect," was not found in the younger 4-month-old age group (Kuhl, Williams et al. 1992). Polka and Werker (1994) demonstrated that sensitivity to nonnative vowel distinctions declines earlier in comparison to consonant differences (Polka and Werker 1994). The investigators showed that, among the age groups tested, English-learning 4-month-old infants were better able to discriminate two German vowel contrasts embedded in minimal pairs, /dYt/-/dUt/ and /dyt/-/dut/. English-learning 6- to 8-month-olds , while still showing better discrimination of these foreign contrasts than the 10- to 12-month-olds, were already beginning to show a decline in their ability to discriminate nonnative vowel distinctions (Polka and Werker 1994). Thus, it appears that alterations in the perceptual organization of vowels begin around six to eight months of age.

Regional dialect variation poses an interesting situation in the perception of speech sounds. Dialect differences can occur through variation in the pronunciations of vowels within the native language, though these differences are not linguistically contrastive. Unlike some nonnative vowel contrasts, many dialect-based vowel differences within the native language remain perceivable by adults (Clopper and Pisoni 2001-2002; Clopper and Pisoni 2004a; Clopper and Pisoni 2004b). This may be because, while not linguistically relevant, dialect-based vowel differences are at least meaningful to listeners exposed to more than one dialect. Researchers have shown that adults can distinguish between different dialects of their native language (Clopper and Pisoni 2001-2002; Clopper and Pisoni 2004a; Clopper and Pisoni 2004b). They demonstrated that American English-speaking adults are able to use phonetic properties of different dialects to categorize various talkers into at least three main American English dialect groups (Clopper and Pisoni 2004b). Their research also suggests that adults who have had early exposure to different dialects are more accurate at identifying an unfamiliar talker's native region, indicating that early exposure to linguistic variation affects dialect discrimination (Clopper and Pisoni 2004a; Clopper and Pisoni 2004b). In another study, perception of differences among six Swedish dialects in native and nonnative speakers was examined. Many nonnative speakers were as proficient as native speakers in dialect discrimination, though native speakers were significantly better than nonnative speakers in naming dialect (Cunningham-Andersson 1996). To date, there have been a few studies investigating the initial development of dialect perception, mainly using fluent speech stimuli. For instance, Nazzi et al. (2000) found that five-month-old infants were able to discriminate American English versus British English when listening to fluent speech (Nazzi, Jusczyk et al. 2000). Another study observed that Australian and American 6-month-olds and Australian 3-month-olds preferred listening to Australian English sentences; this finding that infants are able to generalize across two dialects was attributed to language experience. In addition, the researchers proposed that with age, infants sort out extraneous phonetic information and cluster American and Australian dialects into the same group (Kitamura, Panneton et al. 2006).

Despite this evidence that dialect variation plays a role in speech perception, little is known about the development of dialect discrimination. In particular, little is known about how infants discriminate specific contrasts that vary with regard to dialect. Therefore, the present study seeks to investigate several questions: Do infants discriminate speech sounds that differ by dialect at the same age as when they can discriminate other contrasts that are linguistically relevant in the ambient language? Is dialect discrimination maintained throughout development? Or, does the ability to discriminate dialect differences develop in a similar way as other phonetic differences (that is, do individuals lose the ability to discriminate contrasts that are not relevant to them)?

Experiment 1: Dialect Discrimination in Younger Infants

Experiment 1 was conducted in order to investigate 7- and 11-month-old infants' ability to discriminate the North Midland American English and Southern American English dialect pronunciations of the word "pine," which differ primarily in the vowel sound /aI/. The 7-and 11-month-old age groups were selected because several studies have shown that infants at about 6-8 months of age are able to discriminate most contrasts but by 10 months of age have declined in their ability to discriminate many nonnative contrasts. The diphthong /aI/ was used to test the infants because it is one of the most prominent differences between the North Midland and Southern American English dialects (Wolfram and Schilling-Estes 1998; Clopper 2000). For instance, Southern talkers tend to produce less diphthongization of the /aI/ sound than talkers of other dialects (Wolfram and Schilling-Estes 1998; Clopper and Pisoni 2004b). Consequently, listeners generally use the /aI/ diphthong in order to identify and distinguish between talkers of the North Midland and Southern American English dialects (Clopper and Pisoni 2004b).

Methods

Participants

Twenty American 7-month-olds (12 males and 8 females) and twenty American 11-month-olds (8 males and 12 females) served as participants. The infants were all from monolingual American English-speaking homes in central Indiana. The mean age for the 7-month-olds was 6.76 months (SD = 0.55, range = 6.02 months – 7.76 months), and the mean age for the 11-month-olds was 11.21 months (SD = 0.52, range = 10.43 months – 11.97 months).

Stimuli

The auditory stimuli consisted of repetitions of different tokens of the word "pine" in the North Midland American English and Southern American English dialects. The repetitions were spoken by a female speech-language pathologist who was originally from North Carolina and had lived in Indiana for the past four years. She was capable of speaking in the North Midland American and Southern American English dialects. The visual stimuli consisted of the same checkerboard pattern displayed concurrently with all auditory stimuli.

Acoustic analyses were performed in order to determine that the /aI/ dipthong was different between the North Midland and Southern dialects. In order to analyze the auditory stimuli and to characterize dipthongization of the vowels, the computer program Praat (version 4.1.28) was used. All of the analyses were done using Praat's standard settings. Two tokens of "pine" in the North Midland dialect and two tokens of "pine" in the Southern dialect were used. The durations of the tokens of "pine" in the North Midland dialect were 706 ms (token 1) and 691 ms (token 2), and the durations for the tokens of "pine" in the Southern dialect were 653 ms (token 1) and 725 ms (token 2). The durations of the vowel sound /aI/ of "pine" in the North Midland dialect were 346 ms (token 1) and 339 ms (token 2). The durations of the vowel sound /aI/ of "pine" in the Southern dialect were 334 ms (token 1) and 394 ms (token 2).

To characterize the dipthongization of the vowels, the second formant (F2) of the vowel was measured. The change in F2 has been found to most clearly reflect the difference in diphthongization between the North Midland and Southern dialects (Clopper and Pisoni 2004b). The second formant of the vowel was measured at two temporal points, one-third and two-thirds into the vowel. Then, the difference between the formant measurements at these two points was determined (i.e. F2(2/3) – F2(1/3)). For the vowels in the North Midland dialect, ΔF2 was 410 Hz (token 1) and 494 Hz (token 2). For the vowels in the Southern dialect, ΔF2 was 45 Hz (token 1) and 45 Hz (token 2). These analyses show a greater ΔF2 in the North Midland dialect than the Southern dialect, which is consistent with other investigators' measures of ΔF2 in these dialects (Clopper and Pisoni 2004b).

Apparatus

The experiment was conducted in a sound booth. A 55'' wide-screen television screen was located inside a panel at the front of the sound booth. There was a small hole above the panel through which a video camera was placed in order to watch and record the infants' movements. A G4 Macintosh computer running Habit software, which contained the experiment files used to test the infants, was located in a separate room from the sound booth (Cohen, Atkinson et al. 2004). The experimenter could view the subject from a closed circuit television from this room. The computer allowed the experimenter to start and stop the visual and sound stimuli and record and store the information on looking times in a data file.

Procedure

The experiment was conducted using a version of the Hybrid Visual Habituation (VH) procedure (Houston and Horn 2007). Infants were seated on their caregiver's lap in front of a TV monitor. At the beginning of each trial, a video of a baby, which served as the attention-getter, was presented in the center of the screen until the infant oriented to the center. Then, the attention-getter turned off, and a checkerboard pattern appeared concurrently with the auditory stimuli. Each trial continued until the infant looked away for 1 second or until the infant looked for a maximum of 20 seconds. The amount of time the infant looked at the checkerboard while the stimuli were presented was recorded for each trial. The experiment consisted of two phases, the habituation phase and the test phase. During the habituation phase, half of the infants were presented with repetitions of a single token of the word "pine" in the North Midland American English dialect, and half of the infants were presented with repetitions of a single token of the word "pine" in the Southern American English dialect. The habituation phase continued until there was a 50% decrease in looking time over three trials compared to the first three trials. After the habituation criterion was reached, the test phase began. During the test phase, infants were presented with ten "old" trials and four "novel" trials. The old trials consisted of repetitions of the token of "pine" presented during habituation alternating with another token of the word "pine" spoken in the same dialect. The novel trials consisted of repetitions of the old token of "pine" presented during habituation alternating with repetitions of a token of the word "pine" spoken in novel dialect. The first two test trials consisted of a novel trial and an old trial, and the order was counterbalanced across subjects. The remaining twelve test trials occurred pseudorandomly so that the novel dialect set was never presented on two consecutive trials. After the experiment was completed, the infant's looking times between the "novel" versus "old" dialects were measured and compared.

Results and Discussion

Paired t-test analyses were performed to compare the mean looking times to the novel stimulus versus the old stimulus. The 7-month-old infants showed significantly longer looking times to the novel dialect versus the old dialect, $t\ (19) = 2.19$, $p \leq 0.02$ (one-tailed). The mean looking time to the old dialect was 3.63 s (SD = 1.38), and the mean looking time to the novel dialect was 4.55 s (SD = 2.18). In contrast, the 11-month-old infants did not demonstrate significant differences in looking times to the novel dialect versus the old dialect, $t\ (19) = -0.08$, $p \leq 0.47$ (one-tailed). The mean looking time to the old dialect was 4.47 s (SD = 1.75), and the mean looking time to the novel dialect was 4.42 s (SD = 2.30). Fig. 1 displays the results for the 7- and 11-month-olds in Experiment 1.

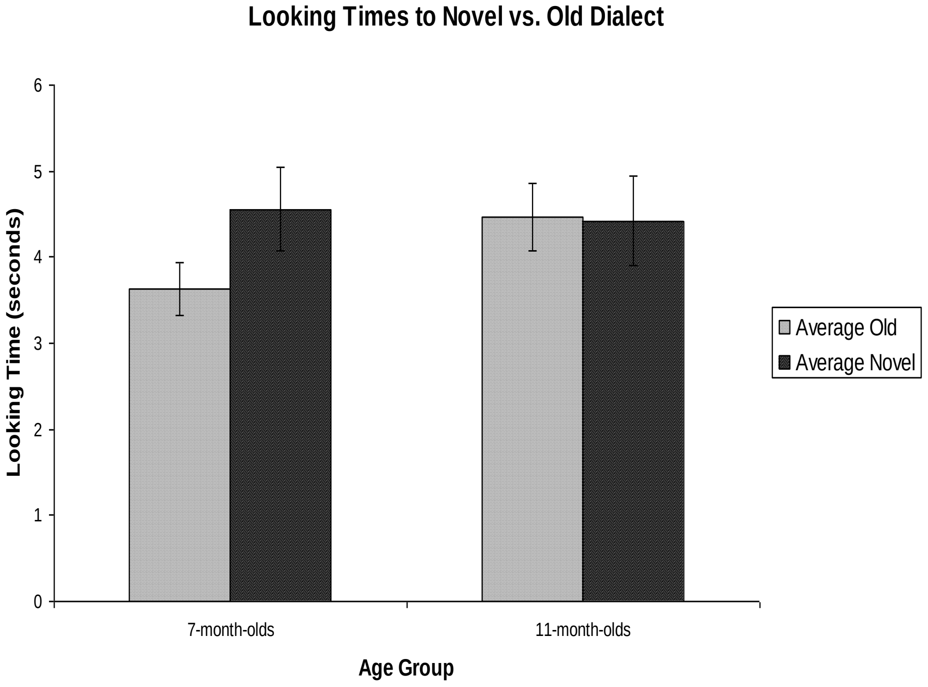

Figure 1. Mean looking times to the novel versus the old dialect in 7- and 11- month-olds.

The results suggest that 7-month-olds discriminate the North Midland and Southern American dialectical variations of /aI/ in "pine" but that 11-month-olds do not.

Experiment 2: Dialect Discrimination in Older Infants and Toddlers

Adults are able to differentiate between many different dialects of American English. This suggests that the ability to discriminate dialectical variants is regained at some point in development. We explored this possibility in Experiment 2 by investigating 18-, 24-, and 30-month-old infants' ability to discriminate the North Midland American English and Southern American English dialect pronunciations of the word "pine." In addition, because older infants and toddlers have more language experience, we examined whether exposure to the Southern American English dialect influenced their discrimination of the dialectical variants.

Methods

Participants

Thirty American 18-month-olds (15 males and 15 females), thirty American 24-month-olds (19 males and 11 females), and thirty American 30-month-olds (17 males and 13 females) served as participants. The infants were all from monolingual American English-speaking homes in central Indiana. The mean age for the 18-month-olds was 17.80 months (SD = 0.63, range = 16.61 months – 18.95 months), the mean age for the 24-month-olds was 23.97 months (SD = 0.64, range = 23.03 months – 25.00 months), and the mean age for the 30-month-olds was 29.99 months (SD = 0.74, range = 28.98 months – 30.99 months).

Stimuli

These were identical to Experiment 1.

Apparatus

This was identical to Experiment 1.

Procedure

This was identical to Experiment 1. Additionally, the subjects' parents and other caregivers were asked to complete a questionnaire. The survey asked the caregivers to estimate the number of hours per week an infant was exposed to the Southern American English dialect by a live speaker and various forms of media (i.e. television, radio, etc.) as well as other demographic information.

Results and Discussion

Paired t-test analyses were performed to compare the mean looking times to the novel stimulus versus the old stimulus. The 18-, 24-, and 30-month-olds did not demonstrate significant differences in looking times to the novel dialect versus the old dialect; however, discrimination of the two dialectical variants appeared to approach significance with increasing age. For the 18-month-olds, the mean looking time to the old dialect was 4.53 s (SD = 1.69), and the mean looking time to the novel dialect was 4.81 s (SD = 2.98), t (29) = 0.55, $p \leq 0.30$ (one-tailed). For the 24-month-olds, the mean looking time to the old dialect was 3.93 s (SD = 1.69), and the mean looking time to the novel dialect was 4.34 s (SD = 1.87), t (29) = 1.07, $p \leq 0.15$ (one-tailed). For the 30-month-olds, the mean looking time to the old dialect was 3.75 s (SD = 2.21), and the mean looking time to the novel dialect was 4.59 s (SD = 3.82), t (29) = 1.48, $p \leq 0.06$ (one-tailed). Fig. 2 displays the difference in mean looking times to the novel versus old dialect for the 7-, 11-, 18-, 24-, and 30-month-olds.

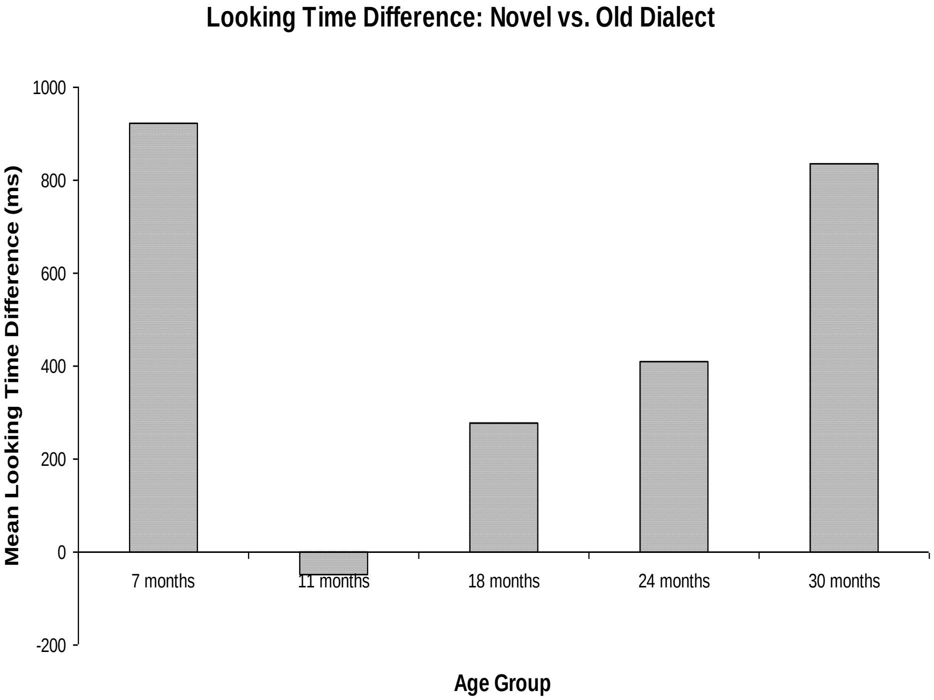

Figure 2. Mean looking time differences to the novel versus the old dialect in 7-, 11-, 18-, 24-, and 30-month-olds.

The results suggest that infants from 11 to 30 months of age do not discriminate between the North Midland American and Southern American English dialect pronunciations of the word "pine." However, there is a trend toward discrimination with increasing age, which may be due to infants' maturity and/or exposure to the Southern dialect.

In order to help determine whether dialect discrimination is influenced by exposure to the Southern dialect, analyses of variance (ANOVAs) were performed on the 18- to 30-month-olds. The data for one 30-month-old subject tested later, which was not included in the original paired t-test analyses, was included in these analyses because the participant was exposed to the Southern dialect. Infants who did not return completed questionnaires were not included in the ANOVA evaluations. The results revealed no significant effect of exposure to the Southern dialect on discrimination of dialect, $F(1, 73) = 0.122$, $p \leq 0.36$ (one-tailed). Furthermore, there was no significant effect of exposure time to the Southern dialect on the older infants' discrimination abilities, $F(1, 73) = 0.129$, $p \leq 0.36$. These findings suggest that exposure and exposure time to the nonnative dialect does not necessarily facilitate discrimination of dialects from 18 month to 30 months of age.

Experiment 3: Linguistic and Acoustic Variability

Infants at 11 months of age demonstrated the poorest dialect discrimination performance among the age groups tested. We propose two possibilities for this finding: (1) 11-month-olds do not show discrimination of dialect because the dialectical variants are linguistically equivalent or (2) infants at this age are generally less sensitive to acoustic differences in speech sounds. In Experiment 3, we explored these possibilities by assessing 11-month-old infants' ability to discriminate linguistically different speech sounds varying in degree of acoustic similarity.

Methods

Participants

Thirty American 11-month-old infants (14 males and 16 females) were tested on the words “pine” in the Southern dialect and “pawn” in the North Midland dialect (Group 1). The mean age for these infants was months 11.00 months (SD = 0.55, range = 9.97 months – 11.81 months). To date, twelve American 11-month-old infants (5 males and 7 females) have been tested on the words “pine” and “pawn” both in the North Midland dialect (Group 2). The mean age for these infants was 11.35 months (SD = 0.45, range = 10.63 months – 12.07 months). Twenty American 11-month-old infants (9 males and 11 females) were tested on the words “pun” and “pawn” both in the North Midland dialect (Group 3). The mean age for these infants was 10.85 months (SD = 0.58, range = 10.03 months – 12.01 months). Twenty American 11-month-old infants (11 males and 9 females) were tested on the words “pine” and “soil” both in the North Midland dialect (Group 4). The mean age for these infants was 10.99 months (SD = 0.54, range = 10.03 months – 11.65 months). The infants were all from families in central Indiana.

Stimuli

The visual stimulus was identical to Experiment 1.

Each group of infants was tested on two different word pairs. Table 1 shows the participant group and their corresponding testing condition.

	Condition
Group 1	pine (S) vs. pawn (NM)
Group 2	pawn (NM) vs. pun (NM)
Group 3	pine (NM) vs. pawn (NM)
Group 4	pine (NM) vs. soil (NM)

Table 1. Groups of participants and the corresponding word pairs on which they were tested, where S = Southern American English dialect and NM = North Midland American English dialect.

Acoustic analyses were performed in order to compare the acoustic characteristics of the words and vowels. In order to analyze the auditory stimuli, the computer program Praat (version 4.1.28) was used. All of the analyses were done using Praat’s standard settings. The auditory stimuli consisted of repetitions of different tokens of the words “pine” in the Southern dialect, “pine” in the North Midland dialect, “pawn” in the North Midland dialect, “pun” in the North Midland dialect, and “soil” in the North Midland dialect. The tokens of the word “pine” in the North Midland and Southern dialects were identical to the ones used in Experiment 1. The durations of the tokens of “pawn” were 786 ms (token 1) and 782 (token 2), and the durations of the vowel sound /a:/ in “pawn” were 371 ms (token 1) and 341 (token 2). The durations of the tokens of “pun” were 644 ms (token 1) and 678 ms (token 2), and the durations of the vowel sound /ʌ/ in “pun” were 218 ms (token 1) and 227 ms (token 2). The word “soil” was selected because of its markedly different acoustic qualities from the other words described above, which allow it to be easily differentiated. The durations of the tokens of “pine” were 706 ms and 691 ms, and the durations for the tokens of “soil” were 747 ms and 744 ms. The visual stimulus was identical to Experiment 1.

To characterize the vowels, the first (F1) and second formants (F2) of the vowels were measured. The formants of the vowel were measured at two temporal points, one-third and two-thirds into the vowel. Then, the difference between the formant measurements at these two points was determined (i.e. F1(2/3) – F1(1/3) or F2(2/3) – F2(2/3)). Table 2 summarizes the formant measurements for the vowel sound /aI/ the Southern dialect and /aI/, /a:/, and /ʌ/ in the North Midland dialect.

	F1(1/3) (Hz)	**F1(2/3) (Hz)**	**Percent ΔF1**	**F2(1/3) (Hz)**	**F2(2/3) (Hz)**	**Percent ΔF2**
/aI/ **Pine (S)**	989	921	-7%	1707	1752	2%
/aI/ **Pine (NM)**	719	607	-16%	1617	2069	28%
/a:/ **Pawn (NM)**	966	921	-5%	1438	1461	2%
/ʌ/ **Pun (NM)**	1011	989	-2%	1685	1797	7%

Table 2. Mean formant measurements for the vowel sounds /aI/ in "pine" in the Southern (S) dialect, /aI/ in "pine" in the North Midland (NM) dialect, /a:/ in "pawn" in the NM dialect, and /ʌ/ in "pun" in the NM dialect.

Apparatus

This was identical to Experiment 1.

Procedure

This was identical to Experiment 1.

We initially tested 11-month-old infants (Group 1) using the words "pine" in the Southern dialect and "pawn" in the North Midland dialect. While the two words are linguistically different, the onset of the diphthong /aI/ and the vowel /a:/ are similar acoustically. We then tested Group 2 on the words "pawn" and "pun" in the North Midland dialect. The /a:/ in "pawn" and /ʌ/ in "pun" are acoustically similar and are both back vowels that have an average F1 difference of 57 Hz and an average F2 difference of 292 Hz. Next, we tested Group 3 using the words "pine" and "pawn" in the North Midland dialect to determine the 11-month-olds' discrimination ability of a diphthong versus a non-diphthong. Finally, we tested Group 4 using "pine" and "soil" in the North Midland dialect. This was presumed to be the least challenging discrimination task, since this word pair demonstrated distinctly different words both linguistically and phonologically.

Results and Discussion

The 11-month-old infants in Group 1 (pineS-pawnNM) did not show a significant difference between the average looking times to the novel stimulus versus the old stimulus t (29) = 0.95, $p \leq 0.18$ (one-tailed). The mean looking time to the old stimulus was 4.39 s (SD = 2.32), and the mean looking time to the novel stimulus was 4.82 s (SD = 3.39). Group 2 (pawnNM-punNM) did not show a significant difference between the average looking times to the novel versus old stimulus, t (19) = -0.15, p ≤ 0.44 (one-tailed). The mean looking time to the old stimulus was 5.20 s (SD = 2.80), and the mean looking time to the novel stimulus was 5.14 s (SD = 3.41). Group 3 (pineNM-pawnNM) did not show a significant difference between the average looking times to the novel versus old stimulus, t(11) = -0.89, p ≤ 0.196 (one-tailed). The mean looking time to the old stimulus was 4.99 s (SD = 2.18), and the mean looking time to the novel stimulus was 4.49 s (SD = 1.99). The infants in Group 4 (pineNM-soilNM) showed a significant difference between the average looking times to the novel stimulus versus the old stimulus, t (19) = 3.20, p ≤ 0.003 (one-tailed). The mean looking time to the old stimulus was 3.05 s (SD = 0.89), and the mean looking time to the novel stimulus was 5.21 s (SD = 2.96). Figure 3 shows the results of Experiment 3 and also includes the results of the 11-month-olds from Experiment 1.

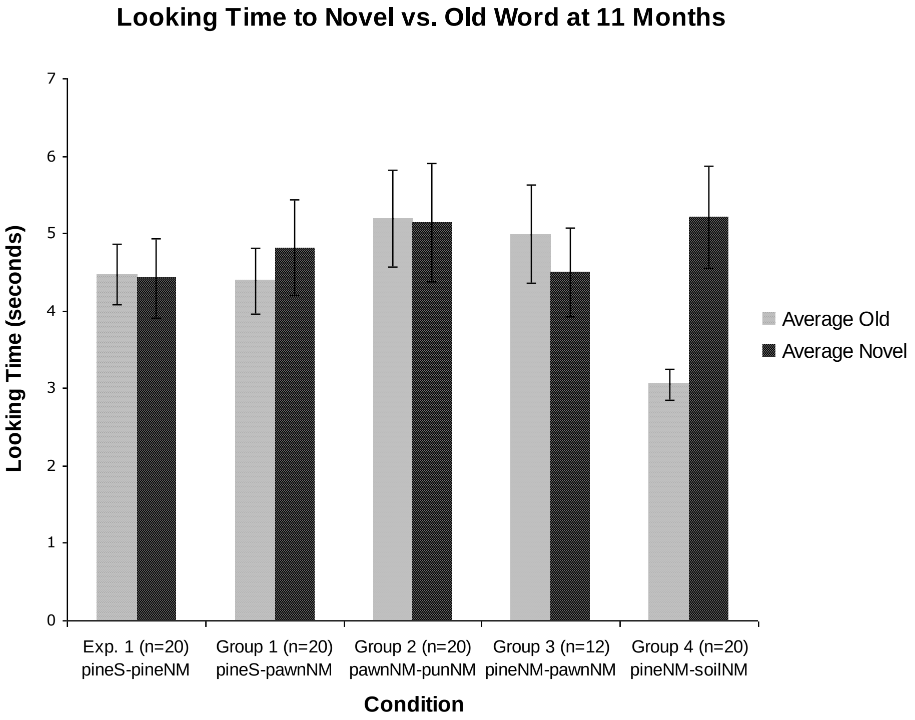

Figure 3. Mean looking times to the novel versus the old stimulus in 11-month-olds of Experiment 1 and in Groups 1, 2, 3, and 4 of Experiment 3.

These findings suggest that 11-month-olds have difficulty discriminating varying degrees of acoustically similar vowel sounds, regardless of dialect differences, even when these differences affect the linguistic meaning of the words. It seems that the vowel discrimination task still remains the most challenging when the vowels are acoustically similar and linguistically equivalent, as occurs in the case of the words "pine" in the Southern dialect and "pine" in the North Midland dialect. However, 11-month-olds are able to distinguish speech sounds markedly different in their acoustic and linguistic qualities, as in "pine" and "soil."

General Discussion

The results provide evidence that infants at 7 months of age discriminate dialectical differences of linguistically equivalent vowel sounds but that infants at 11 months to 30 months of age do not. Similar to Polka and Werker's (1994) study using nonnative vowel contrasts, younger infants discriminated vowel contrasts better than older infants. Unlike discrimination of vowel contrasts, however, findings from the present study may suggest that not all non-linguistically relevant vowel contrasts decline in perceptibility at the same age or earlier than consonants. Although the 11- to 30-month-olds did not show discrimination of the dialect-based vowel variants, there was a trend toward discrimination with increasing age. Interestingly, the present study found that experience with the nonnative dialect in 18-, 24-, and 30-month-olds did not facilitate discrimination of dialect. This finding is similar to the results of studies on perception of vowel contrasts in monolingual and bilingual infants. Bosch and Sebastian-Galles (2003) found that Spanish monolingual, Catalan monolingual, and Spanish/Catalan bilingual 4-month-olds were able to discriminate the Catalan contrasts /e/-/ɛ/ while both monolingual Spanish and bilingual infants at 8 months of age showed decreased discrimination of the contrasts. The 8-month-old bilingual subjects demonstrated a decline in sensitivity to the vowel contrasts despite daily exposure to both languages. Discrimination for bilingual infants was ultimately recovered at 12 months of age (Bosch and Sebastian-Galles 2003). Taken together, the results from this and the present study indicate that it is possible that the speech perception processes in the first few years of life are not the mere outcome of exposure to a specific language or dialects within a language.

The perception and discrimination of both native and nonnative vowels may be influenced by factors other than language-specific constraints. This idea is supported by our findings that 11-month-old infants, who showed the poorest discrimination ability in this study, failed to discriminate linguistically different vowels sounds of similar acoustic quality, whether or not dialect differences were present. Eleven-month-olds were, however, able to distinguish speech sounds markedly different in their acoustic and linguistic qualities, as in “pine” and “soil.” Despite the influence of acoustic variation on discrimination, the discrimination task still seems to be the most challenging for 11-month-olds when the vowels are both acoustically similar and linguistically equivalent, as occurs in the case of the words “pine” in the Southern dialect and “pine” in the North Midland dialect; after 11 months, discrimination of these contrasts does seem to gradually improve with increasing age. These findings suggest that acoustic variability and general developmental maturity may play more significant roles than language experience in the discrimination of vowel sounds than was previously thought.

These findings raise several important developmental questions. First, how general is the developmental pattern found in the present study? Can 7-month-olds discriminate more subtle dialectical contrasts? Can 11-month-olds discriminate any dialectical contrasts? Further studies on other dialectical contrasts would provide valuable information about the generality of this developmental pattern and about the role of phonemically-irrelevant acoustic similarity on infants’ discrimination of dialectical contrasts.

Another developmental question that these findings raise is how and at what age do listeners recover their sensitivity to the dialect contrasts that they lose during infancy? Unlike many nonnative phonemic contrasts, dialect contrasts are perceptible to adult listeners even though they are not linguistically relevant (Clopper and Pisoni 2001-2002; Clopper and Pisoni 2004a; Clopper and Pisoni 2004b). Also, how and at what age do listeners recover their sensitivity to linguistically different yet acoustically similar phonemic contrasts? While discrimination of many nonnative phonemic contrasts declines with age, some studies have demonstrated that listeners eventually become better at discriminating certain nonnative contrasts. Polka and Werker (1994) found that adults, both native and nonnative speakers, were able to discriminate two German vowel contrasts (/Y/-/U/ and /y/-/u/) better than 10- to 12- month-old infants. These findings suggest that discrimination of dialect and of other vowel contrasts are recovered at some point during development and may follow a similar pattern of development as discrimination of at least some nonnative contrasts. Further, how acoustically diverse must linguistically different vowel sounds be in order for 11-month-olds to show discrimination? To address this question, we plan to test 11-month-olds using the words “pine” and “poin” in the North Midland dialect. We know that 11-month-olds can discriminate the words “pine” and “soil,” but will they be able to discriminate the vowel sounds /aI/ from /oI/ when the beginning and ending consonants are similar?

Finally, do listeners ever learn to discriminate dialects through exposure to them? The findings in the current study do not support the idea that exposure to different dialects aids in dialect discrimination, at least for infants and toddlers. However, the research described by Clopper and Pisoni (2004a, 2004b) suggests that experience with different dialects plays at least some role in dialect discrimination. Further studies with older toddlers and children may help to delineate the developmental time course and role of dialect exposure in becoming sensitive to dialect contrasts.

References

Aslin, R. N., D. B. Pisoni, et al. (1983). Auditory development and speech perception in infancy. Infancy and the biology of development. M. M. Haith and J. J. Campos. New York, Wiley. **2:** 573-687.

Bertoncini, J., R. Bijeljac-Babic, et al. (1987). "Discrimination in neonates of very short CVs." Journal of the Acoustical Society of America **82**: 31-37.

Best, C. T. (1993). Emergence of language-specific constraints in perception of native and non-native speech: A window on early phonological development. Developmental neurocognition: Speech and face processing in the first year of life. B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage and J. Morton. Dordrecht, Kluwer**:** 289-304.

Best, C. T., G. W. McRoberts, et al. (1995). "Divergent developmental patterns for infants' perception of two nonnative consonant contrasts." Infant Behavior & Development **18**: 339-350.

Best, C. T. and W. Strange (1992). "Effects of phonological and phonetic factors on cross-language perception of approximants." Journal of Phonetics **20**(3): 305-330.

Bosch, L. and N. Sebastian-Galles (2003). "Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life." Language and Speech **46**(2-3): 217-243.

Clarkson, R. L., P. D. Eimas, et al. (1989). "Speech perception in children with histories of recurrent otitis media." Journal of the Acoustical Society of America **85**(2): 926-933.

Clopper, C. G. (2000). Some acoustic cues for categorizing American English regional dialects: an initial report on dialect variation in production and perception. Research on Spoken Language Processing, Progress Report**:** 43-65.

Clopper, C. G. and D. B. Pisoni (2001-2002). Perception of dialect variation: some implications for current research and theory in speech perception. Research on Spoken Language Processing, Progress Report 270-290.

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

Cohen, L. B., D. J. Atkinson, et al. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin, University of Texas.

Cunningham-Andersson, U. (1996). "Learning to interpret sociodialectal cues." Quarterly Progress and Status Report - Royal Institute of Technology, Department of Speech, Music, and Hearing **2**: 155-158.

Eimas, P. D., E. R. Siqueland, et al. (1971). "Speech perception in infants." Science **171**(968): 303-6.

Flege, J. and W. Eefting (1987). "The production and perception of English stops by Spanish speakers of English." Journal of Phonetics **15**: 67-83.

Holmberg, T. L., K. A. Morgan, et al. (1977). Speech perception in early infancy: Discrimination of fricative consonants. Meeting of the Acoustical Society of America, Miami Beach, FL.

Houston, D. M. and D. L. Horn (2007). "Assessing Speech Discrimination in Individual Infants." Infancy **12**(2): 119-145.

Kitamura, C., R. Panneton, et al. (2006). Attuning to the native dialect: when more means less. 11th Australian International Conference on Speech and Science Technology, University of Auckland, New Zealand.

Kuhl, P. K. (1979). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories." Journal of the Acoustical Society of America **66**: 1668-1679.

Kuhl, P. K. (1983). "Perception of auditory equivalence classes for speech in early infancy." Infant Behavior and Development **6**: 263-285.

Kuhl, P. K., K. A. Williams, et al. (1992). "Linguistic experiences alter phonetic perception in infants by 6 months of age." Science **255**: 606-608.

Lively, S. E., J. S. Logan, et al. (1993). "Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories." Journal of the Acoustical Society of America **94**(3, Pt 1): 1242-1255.

Lively, S. E., D. B. Pisoni, et al. (1994). "Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories." Journal of the Acoustical Society of America **96**(4): 2076-2087.

Marean, C. G., L. Werner, et al. (1992). "Vowel categorization in very young infants." Developmental Psychology **28**: 396-405.

Miyawaki, K., W. Strange, et al. (1975). "An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English." Perception & Psychophysics **18**: 331-340.

Nazzi, T., P. W. Jusczyk, et al. (2000). "Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity." Journal of Memory & Language **43**(1): 1-19.

Pallier, C., L. Bosch, et al. (1997). "A limit on behavioral plasticity in speech perception." Cognition **64**(3): B9-B17.

Polka, L. and O. S. Bohn (1996). "A cross-language comparison of vowel perception in English-learning and German learning infants." Journal of the Acoustical Society of America **100**(1): 577-592.

Polka, L. and J. F. Werker (1994). "Developmental changes in perception of non-native vowel contrasts." Journal of Experimental Psychology: Human Perception and Performance **20**: 421-435.

Swoboda, P., P. A. Morse, et al. (1976). "Continuous vowel discrimination in normal and at-risk infants." Child Development **47**: 459-465.

Trehub, S. E. (1973). "Infants' sensitivity to vowel and tonal contrasts." Developmental Psychology **9**: 91-96.

Trehub, S. E. (1976). "The discrimination of foreign speech contrasts by infants and adults." Child Development **47**: 466-472.

Werker, J. F., J. H. Gilbert, et al. (1981). "Developmental aspects of cross-language speech perception." Child Development **52**: 349-355.

Werker, J. F. and C. E. Lalonde (1988). "Cross-language speech perception: Initial capabilities and developmental change." Developmental Psychology **24**: 672-683.

Werker, J. F. and R. C. Tees (1983). "Developmental changes across childhood in the perception of non-native speech sounds." Canadian Journal of Psychology **37**: 278-286.

Werker, J. F. and R. C. Tees (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life." Infant Behavior and Development **7**: 49-63.

Wolfram, W. and N. Schilling-Estes (1998). American English. Malden, Blackwell.

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Visual Sequence Learning in Normal Hearing and Hearing-Impaired Infants: Finding an Early Predictor of Language[1]

Suzanne L. Field[2], Carissa L. Shafto[3], Derek M. Houston[2,4], Christopher M. Conway[5], and Sara Tinter[4]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by a research grant awarded to Derek Houston (R01DC006235) and a training grant to David Pisoni (T32DC00012). We would like to acknowledge the staff of the Infant Language Lab, especially coordinator Kabreea York, for their assistance in recruiting and testing infants. We are also very grateful to the infants and their parents who participated in the research.

[2] Indiana University School of Medicine
[3] University of Louisville
[4] DeVault Otologic Research Laboratory, Riley Hospital
[5] St. Louis University.

Visual Sequence Learning in Normal Hearing and Hearing-Impaired Infants: Finding an Early Predictor of Language

Abstract. Despite the success of cochlear implantation in many deaf and hearing-impaired children, there remains considerable variability in individual language outcomes. Research suggests that implicit learning abilities might explain some of this variability. For example, Conway, Pisoni, Anaya, Karpicke, & Henning (under revision) found that implicit visual sequence learning abilities in normal hearing (NH) children and deaf children with cochlear implants (CIs) were related to individual differences in spoken language processing and development. Deaf children with CIs also had atypical sequence-learning abilities when compared to their NH peers, suggesting that in addition to normal language development, other cognitive functions may also be negatively affected by auditory deprivation. The current study investigated visual sequence learning as a possible early predictor of vocabulary development in NH and deaf infants. Two groups of NH infants- ten 8.5-months-olds and six 5-month-olds- as well as two cochlear implanted infants- a 23- and 30-month-old- were presented with a three-location spatiotemporal sequence of multi-colored geometric shapes. Early language skills were assessed using the MacArthur-Bates CDI. Analyses of children's reaction times to the stimuli suggest that children learned the spatiotemporal sequence and their increased reaction times were significantly correlated with vocabulary comprehension abilities. Clinical and theoretical implications are discussed.

Introduction

Hearing-loss and deafness have long been associated with delayed spoken language development. For example, hearing-impaired children perform below average on verbal abstract reasoning tests when compared with hearing peers (Moeller, 2000). One approach to reducing the effects of hearing loss on language development is by early intervention programs. The benefit of early identification and intervention in deafness for the development of language skills has been well established in recent years. There is both neural-cognitive and linguistic evidence to support the theory that early exposure to auditory input is advantageous to language development. Proponents of early intervention often reference early plasticity of the brain and the importance of auditory input for the normal development of the central auditory system (Sininger, Doyle, & Moore, 1999, Harrison, 2001) and maintenance of the auditory nerve (Sly *et al.*, 2007, Manrique, Cervera-Paz, Huarte, & Molina, 2004). Significantly better receptive and expressive language outcomes have been found in infants whose deafness or hearing loss was identified before six months of age when compared to infants whose hearing loss was identified after six months of age (Yoshinaga-Itano, Sedey, Coulter, & Mehl, 1998). In addition, improved speech intelligibility has been found in earlier-identified hearing-impaired children versus later-identified hearing-impaired children regardless of the degree of hearing loss (Levitt, McGarr, & Geffner, 1987). The strength of the evidence presented for early intervention and its effect on better language outcomes has led to mandatory newborn hearing screening in 42 states (Hearing Loss Association of America, 1999, U.S. Preventive Services Task Force Recommendation, 2008).

For children with severe-to-profound hearing loss, the evidence that early intervention leads to improved language outcomes has led to earlier cochlear implantation. A cochlear implant is an auditory prosthesis that is often recommended in cases where hearing loss is too great to benefit from a conventional hearing aid. The U.S. FDA recently approved cochlear implantation for ages as young as 12-months old, and some surgeons are providing CIs to infants as young as 6 months of age. Early implantation is motivated by the research reviewed above on the benefits of early intervention and also by other studies that have shown that early cochlear implantation leads to better language outcomes. One study comparing the receptive and expressive language outcomes of a child implanted at six months of age to a group of children implanted at older ages found that at 18 months post-implantation, the early implanted child exceeded age equivalency on the Reynell Developmental Language Scales (Miyamoto, Houston, Kirk, Perdew, & Svirsky, 2003). The early implanted child outperformed even those children who were implanted before three years of age. Another study found that children who were implanted at less than two years of age had significantly better speech intelligibility than children implanted after two years of age (Svirsky, Chin, & Jester, 2007). In an analysis of the CI children's performance as a function of age at implantation and language skills, their findings also suggested that the ability of children to learn language decreases with age (Svirsky *et al*., 2007). Further evidence has been found that cochlear implantation at or before one year of age may facilitate better pre-word learning skills important in language development such as audio-visual mapping (Houston, Ying, Pisoni, & Kirk, 2003). The goal of early cochlear implantation is to provide auditory input during the critical period for language development in order to lessen the gap between chronological age and auditory age in hearing-impaired and deaf children. Thus, if implanted at a very young age, deaf children will have a better chance of developing age-appropriate speech and language abilities.

Despite the success of early cochlear implantation in many deaf and hearing-impaired children, there remains considerable variability in individual language outcomes. Previous studies have focused on identifying the individual factors that predict later language outcomes after cochlear implantation. Finding an early predictor for later language development in both NH and CI children would allow clinicians to assess the type of language problems these children are likely to have or develop. This knowledge would then allow clinicians to tailor therapy and treatment programs to the specific needs of individual patients in order to achieve better speech and language outcomes. A pre-implant predictor would be especially useful because clinicians could provide effective, focused therapy during the period at which language development is most influenced by treatment.

Some recent studies have investigated the relationship between performance on cognitive and linguistic tasks before implantation and language outcomes after implantation. For example, Horn, Davis, Pisoni, and Miyamoto (2004) examined visual-motor integration skills in deaf children before implantation via the Beery Developmental Assessment of Visual-motor Integration (VMI) (Beery, 1989). Performance on the VMI was found to account for individual differences in outcome and to accurately predict later speech perception, sentence comprehension, and speech intelligibility skills at one, two, and three years post-implantation. Another study of pre-implant predictors found that lip reading and AV speech perception scores on the Pediatric Sentence Intelligibility (PSI) test were strongly associated with vocabulary, receptive and expressive language, and speech intelligibility scores two or more years after implantation (Bergeson, Pisoni, & Davis, 2001). Although some correlations between pre-CI performance on cognitive tasks and later language outcomes have been found, most of the variance in language outcomes is still unexplained.

The findings concerning the variability of language abilities in CI children have led some researchers to suggest that these differences in language outcomes should be attributed to cognitive variables; namely, perception, attention, learning, and memory (Pisoni *et al*., 2000). Several studies have compared CI children's performance on language tasks to their performance on cognitive-processing tasks, such as forward and backward digit spans. In one such study, Cleary, Pisoni, and Geers (2001) found that when presented with a working memory task designed to evaluate short-term storage of auditory and visual spatial sequences, CI children had shorter span scores on average than NH children. The results suggested that CI children display atypical working memory development, which could perhaps be attributed to differences in verbal/auditory encoding or rehearsal strategies. Another study comparing the working memories of NH and CI children via the digit span task (Burkholder & Pisoni, 2002). By correlating digit span lengths to the speaking rates during digit span recall they found that CI children had longer pauses during recall and shorter digit spans overall. These findings suggest that lack of early auditory input and processing in deaf infants affects the sub-vocal rehearsal speed and memory scanning processes that usually act as part of working memory to store and process linguistic information. Pisoni and Cleary (2003) found that spoken word recognition scores of CI deaf children correlated with individual working memory capacity and with differences in cognitive processes that encode, store, and retrieve linguistic information from working memory. This same study also found that the speaking rate of CI children was related to both forward and backward digit spans, a phenomenon that has been similarly reported in the NH population. Although these findings strongly suggest that intrinsic cognitive abilities such as working memory contribute to language outcomes, there has been very little work investigating the role played by fundamental learning processes, and no such work in infants. In this study, we will investigate implicit sequence learning as a possible cognitive process that can be assessed during infancy and may account for some of the variability in language outcome currently observed in pre-lingually deaf CI children.

Implicit sequence learning and language development

Implicit sequence learning is the ability to acquire knowledge about complex sequential stimulus patterns, usually occurring under conditions without conscious intent or awareness (Cleeremans and McClelland, 1991; Berry and Dienes, 1993). Previous research on sequence learning has established that it is a type of nondeclarative or procedural memory that both infants and adults demonstrate (Clegg, DiGirolamo, & Keele, 1998). Current theories suggest that implicit sequence learning may be related to language acquisition because it is an unconscious developmental process (Cleeremans, Destrebecqz, & Boyer, 1998). Because people generally use language without an explicit understanding of the rules of grammar dictating its structure, it is likely that much knowledge of language is gained through implicit learning mechanisms such as sequence learning (Cleeremans *et al*., 1998). One study of sequence learning and language (Saffran, Aslin, & Newport, 1996) tested word segmentation in 8-month-old infants by exposing them to a continuous stream of a nonsense language where the only clue to word boundaries was that the transitional probabilities between syllable pairs were higher within words than across word boundaries. After only two minutes of exposure to streaming nonsense speech, the infants were able to distinguish the "words" from the non-words. Saffran and colleagues performed a second experiment on a new group of 8-month-old infants to find out if they could also differentiate the syllable sequences of nonsense words from syllable sequences spanning word boundaries. The infants listened to streaming speech containing three-syllable nonsense words and were then tested on the nonsense words and "part-words," created by joining the final syllable of one word to the first two syllables of another word. In spite of the increased difficulty of this task and the short time of exposure (again two minutes), the infants were able to discriminate between the words and part-words. The results from both these experiments support the conclusion that some statistical learning mechanism must be responsible for word segmentation in first-language acquisition. Similar researched has been carried out with groups of first-grade children and adults (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). In that study both age groups were able to identify individual words from a nonsense language when only the transitional probabilities of the syllables indicated word boundaries. These results are even more impressive considering that the subjects in this experiment were not told they were listening to a language, nor were they told to listen at all. The subjects were simply instructed to create designs on a computer program while the nonsense language was streamed in the background. The fact that older subjects were still able to acquire information about the new language without conscious effort supports the theory that statistical learning plays a role in language learning at all ages.

Recent research has found that implicit learning abilities in normal hearing adults were significantly correlated with individual performance on a sentence perception task which relied on context for perception (Conway, Bauernschmidt, Huang, & Pisoni, under revision). More recently, Conway, Pisoni, Anaya, Karpicke, and Henning (under revision) further explored implicit learning in normal hearing children and deaf children with cochlear implants. They found that implicit visual sequence learning abilities were related to individual differences in spoken language processing and development in both normal hearing and deaf children. In addition, deaf children with CIs had atypical sequence learning abilities as compared to the age-matched normal hearing children. These findings suggest that aside from the already established disadvantage of loss of normal language development in deafness, other cognitive functions may also be affected by auditory deprivation (Conway *et al*., under revision).

So far, there has not been any investigation of visual sequence learning abilities in young deaf infants. In regards to normal-hearing infants, previous work has shown that infants as young as 3-months old can learn relatively simple visual sequential patterns (Wentworth, Haith, Hood, 2002; Haith, Hazan, Goodman, 1988). In one of the first studies on infant sequence learning, Haith, Hazan, and Goodman (1988) found that infants could form expectations for left-right visual patterns when expectation was measured by shorter reaction times to upcoming events. Clohessy, Posner, and Rothbart (2001) confirmed that young infants can learn a visual sequence and further investigated age differences in infant sequence learning abilities. Three different subject populations aged 4-, 10-, and 18-months-old were shown both unambiguous and context dependent visual sequences. Four-month-olds could learn the unambiguous sequences but only 18-month-olds demonstrated context dependent sequence learning (Clohessy *et al*., 2001). Wentworth, Haith, and Hood (2002) found that 3-month-old infants were able to anticipate spatiotemporal regularity in a three-location event sequence where the central picture predicted the location of the following stimulus.

In the present study, we will investigate visual sequence learning and its connection to language development in both normal hearing and cochlear-implanted infants. We have created four simple geometric spatiotemporal sequences where different shapes in various colors appear on three different screens one at a time. We showed these sequences to a group of normal hearing and hearing-impaired or CI deaf infants and video recorded their eye movements to gauge sequence learning. Each infant was then given a MacArthur Communicative Development Inventory (CDI) to determine his or her language skills. We hypothesized that normal hearing infants will have better sequence learning abilities when compared to age-matched CI deaf children because NH infants will have had considerably more exposure to auditory sequences and, therefore, more experience in encoding and learning sequential information.

Methods

Participants

The participants were sixteen NH infants[6] and two CI infants recruited from the greater Indianapolis metropolitan area. Six of the normal hearing infants were approximately 5 months old (M=5.01 months, range= 4.34 to 5.69, 3 female) and ten were 8.5 months old (M=8.66 months, range= 8.06 to 9.05, 5 female). All normal hearing infants had passed their newborn hearing screening. The chronological ages of the two CI infants were 22.99- and 29.93-months-old, and their hearing ages were 8.52 and 17.99 months respectively.

[6] 21 additional infants participated (13 in the 5-month-old group and 8 in the 8.5-month-old group), but were excluded from analyses for failing to complete the experiment.

Apparatus

The experiment was conducted within a custom-built double-walled IAC sound booth approximately six feet in width. Infants were tested while seated on a caregiver’s lap directly in front of a 55-inch wide-aspect TV monitor with two smaller 19-inch Dell computer monitors on either sidewall (see Figure 1). The infant sat on the caregiver’s lap and watched three screens at approximately eye level. Experimental sessions were recorded via a hidden camera and the experimenter (unable to see which stimulus was being presented) observed the session on a monitor that displayed the live-action video of the infant and controlled the stimulus presentation from outside the sound booth. Infants were approximately five feet from the center screen and three feet from the side monitors. Visual images were displayed on all three screens at approximately infant eye level. Infants viewed the images on the side monitors at an angle of 57°. The experimenter observed the session from a separate room via hidden, closed circuit TV camera, and was blind to which stimuli were presented. The experiment was controlled by the Habit software package (Cohen, Atkinson, & Chaput, 2004) run on a Macintosh G4 desktop computer.

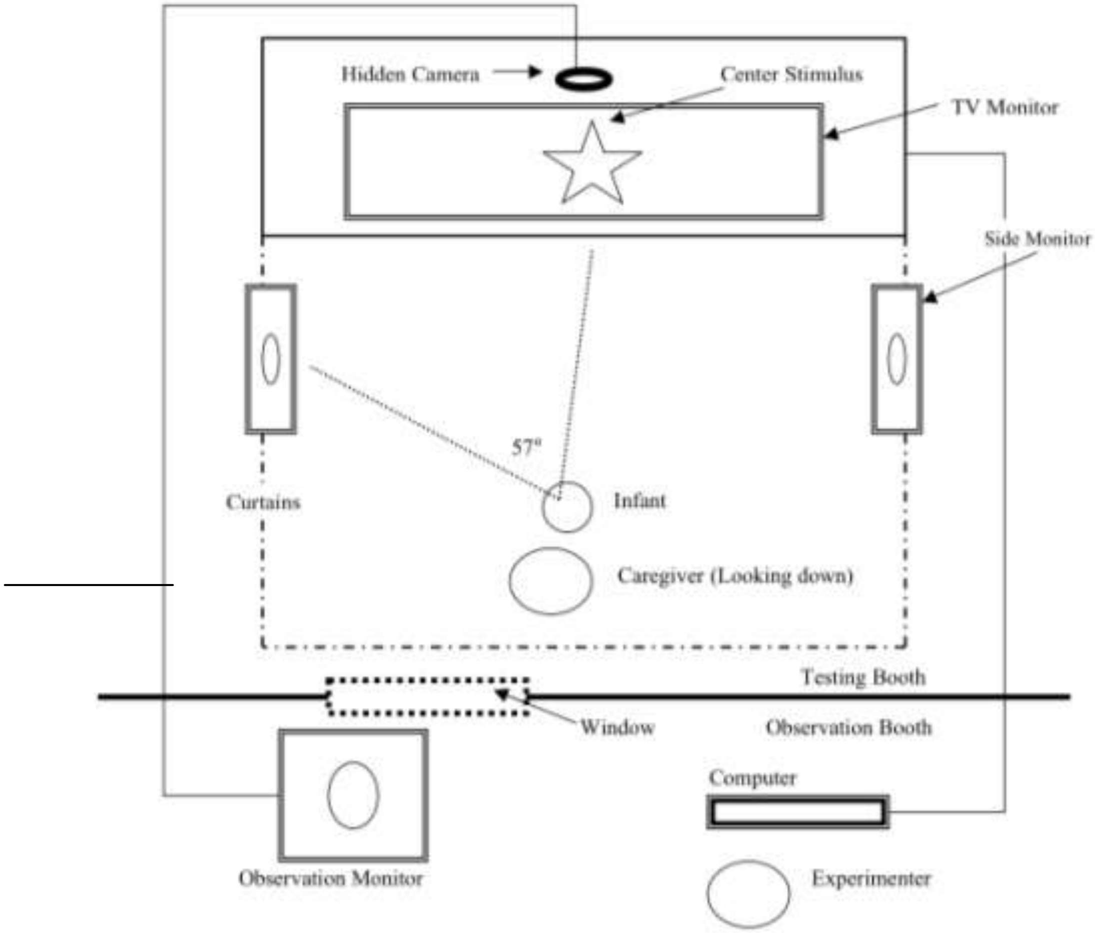

Figure 1. Sound Booth. The infant sat on the caregiver¹s lap and watched three screens at approximately eye level. Experimental sessions were recorded via the hidden camera and the experimenter (unable to see which stimulus was being presented) observed the session on a monitor that displayed the live-action video of the infant and controlled the stimulus presentation from outside the sound booth.

Stimuli

The stimuli consisted of twelve 2D visual images of colorful geometric shapes organized into four object sets (A-D). Each object set consisted of three unique geometric shapes made with the Adobe custom shape tool in Adobe Photoshop CS3. The Photoshop .png files were then animated to look like they were looming in and out using Final Cut Express HD. The looming images were saved as Quicktime movies.

Object sets were designed to include maximally dissimilar shapes, all of which were set to the same size and animation speed. Because of the size difference between the center screen and side monitors, the Photoshop .png files were scaled down so that they would appear as the same size on all screens. The background color for the Quicktime movies was white, but screens that were not displaying an image during a trial were set to black to keep from distracting the infant. The shapes in each set were all different colors, selected according to a specific pattern from the Adobe Photoshop color palette such that no colors repeated within or between sequences. All shapes in the object sets were also unique and did not repeat within or between sets. See Figure 2 for images of all four object sets. For example, Object Set A consisted of an ellipse, a triangle, and a flower. All stimuli loomed from small to large and back to small within 2.66 seconds, and each stimulus loomed up to five times within the course of one trial or presentation. The maximum size for each shape was either 31 cm or 34 cm depending upon whether the shape appeared on the center or side monitors respectively. No infant saw the same shape on both the side and center monitors, so this slight difference in size was not problematic.

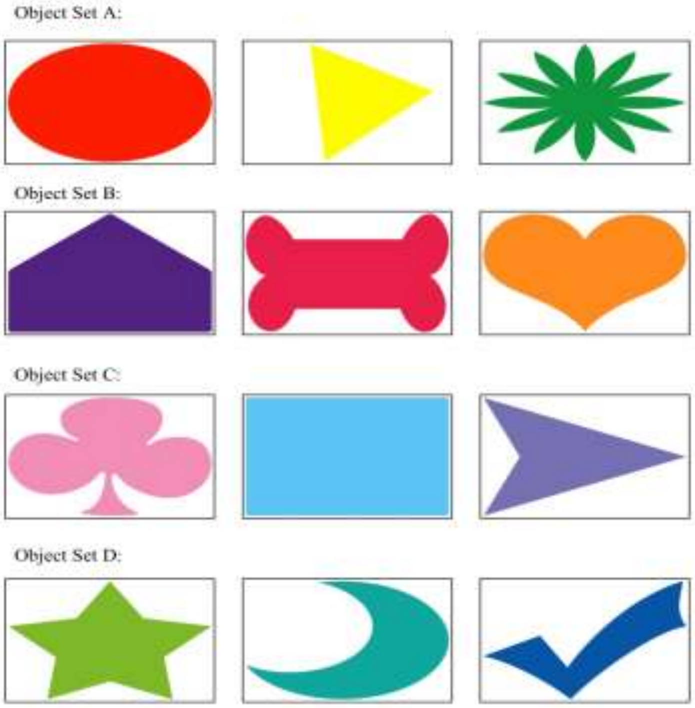

Figure 2. Object Sets. Set A: pure red ellipse, pure yellow triangle, pure green starburst, Set B: pure violet pentagon, pure magenta red dog bone, pure yellow orange heart, Set C: pastel cyan rectangle, pastel blue violet arrowhead, pastel magenta club, and Set D: pure pea green star, pure green cyan crescent, pure blue checkmark.

Procedure

The experiment consisted of three pre-test familiarization trials and four phases, all of which were completed within a single experimental session. The object sets were presented in two different spatiotemporal sequences or patterns on the lab monitors: Left, Center, Right or Right, Center, Left (L-C-R or R-C-L). Each infant viewed only one of the two spatial patterns during the experimental phases. Therefore, the object sets presented in each phase changed, but the spatial organization of the sequence did not. Each object set was presented four times during a given phase for a total of twelve trials per phase. The spatiotemporal sequence and object set viewed during each phase was determined by random assignment and counterbalanced across the experiment to control for any variation in learning different colors or shapes. One infant may have viewed Object Set A in Phase 1 and then continued through the phases and object sets alphabetically (Phase 2: set B, Phase 3: set C, etc.). Another infant may have begun Phase 1 with object set B and then also continued through the rest of the phases in alphabetical order (ending with Phase 4: set A).

Shapes within each object set were always presented in the same location, even when the spatial pattern or phase was different. For example, if one infant observed Object Set A in the L-C-R pattern and another observed Object Set A in the R-C-L pattern, both infants saw an ellipse on the left, a triangle on the center screen, and a flower on the right screen. All that changed between participants was the temporal order in which these images appeared (L-C-R or R-C-L). In all phases, the presentation of the stimuli was contingent upon the infant looking at the screen (infant controlled). Each trial began with the appearance of the stimulus and ended 700 milliseconds after the infant looked at the stimulus. Stimuli within each sequence were separated by an inter-stimulus interval (ISI) of 1100 milliseconds.

The maximum playtime for one experimental session including all four phases (3 pre-test trials plus 48 test trials, 53 trials total) lasted approximately 14.4 minutes with each phase lasting approximately 3.6 minutes. However, the actual length of the trials and phases varied depending upon when the infant looked at the screen. The average length of a testing session for the entire group of nineteen infants was 6 minutes, 8 seconds. All phases from the pre-test stimuli till the end of Phase 4 were presented to the infant without breaks or pauses. The parents or caregivers holding the infants were instructed to look down and keep their eyes closed or wear a visor to limit their influence on the infant looking at the screens. Experimental sessions were recorded using the closed circuit TV camera, and infants' eye movements as well as head movements were later analyzed to determine how quickly they reacted to the location of the next stimulus.

Familiarization Phase: To orient the infant to the task, warm-up stimuli were displayed on each of the screens one at a time. The presentation pattern of the familiarization sequence depended upon the spatiotemporal sequence that followed. For example, infants began the experiment by viewing the familiarization trials in one of two orders: C-L-R or R-L-C. If the infant saw C-L-R, then the spatiotemporal sequence of the following phases was L-C-R. If the infant saw R-L-C during the warm-up phase, then the spatiotemporal sequence that followed was R-C-L. Two different warm-up sequences were used to prevent the first trial of Phase 1 from appearing on the same screen as the last trial of the familiarization phase.

All three screens displayed the same stimulus, a looming blue lightning bolt on a white background. The warm-up stimulus served as an "attention getter" and was created in the same way that the other training/testing stimuli were made. The presentation of the warm-up stimulus was contingent upon the infant looking, identical to the presentation of the phase stimuli.

Phase 1:Training The infant was presented with one of the object sets A-D in one of the two spatiotemporal patterns (L-C-R or R-C-L). Each object was presented four times on each of the screens for a total of twelve trials.

Phase 2: Testing In this phase, the infant was tested for its ability to predict the location of the next stimulus based upon the spatial pattern used in Phase 1. A new set of objects was used but they were presented in the same spatiotemporal sequence as Phase 1 (L-C-R or R-C-L). Infant eye movements were recorded during all phases, and later coded for reaction time (RT).

Phase 3: Quasi-random Next the infant was presented with a quasi-random spatiotemporal sequence of another new object set. Two sequences were created using the www.random.org list randomizer function. Each infant only viewed one or the other quasi-random sequence during Phase 3.
Both quasi-random sequences were controlled for the total number of appearances of stimuli on each screen, and no stimulus repeated on the same screen two times consecutively. Also, each object was always presented in the same location, irrespective of the spatial organization of the phase. The quasi-random phase allowed for the measurement of the infant's baseline RT to stimuli presented without a L-C-R or R-C-L sequence. RTs during this phase were therefore predicted to be slower than those of Phases 1 and 2. Slower RTs were expected during this phase because there was no spatiotemporal pattern for the infants to learn.

Phase 4: Testing Another new object set was presented to the infant during this phase but this time it was in the same spatiotemporal sequence used in Phases 1 and 2. Faster or equal RTs during this last phase of the experiment compared to those of Phase 3 signified that the infant had learned the sequence from Phases 1 and 2. Phase 4 also served as a control for possible decreased looking time by the infants due to fatigue.

Data Collection

The video recordings of the experimental sessions were coded offline using Supercoder (Hollich, 2005) for right, left, and center looks. The coded files were then run through an Excel Macros program along with the spatiotemporal sequence of the stimuli for the testing session. The Macro calculated the RTs for each trial. RTs were generally defined as follows. The RT for trial X was the time between the onset of the first correct look for trial X-1 and the onset of the first correct look for trial X. Thus, all RTs were positive even if they were anticipatory. An anticipatory look was any look that immediately followed the correct look to the previous stimulus and occurred before 200 ms after the onset of the subsequent stimulus. A look was counted as anticipatory even if it ended before the onset of the stimulus. When the first correct look for a trial (e.g., trial X) was anticipatory, the RT start point for the next trial (e.g., trial X+1) began the moment when there was a correct look following the onset of the stimulus. For example, if there was an anticipatory look for trial X, the start point for trial X+1 RT began either at the onset of the stimulus of trial X, if the infant's correct anticipatory look continued to that point, or at the onset of the first post-stimulus-onset correct look, if the anticipatory look ended before the onset of the stimulus of trial X. All anticipatory looks were classified as either correct or incorrect depending upon whether the infant looked to the location of the next stimulus or not. The median RT and standard deviation were calculated for each phase.

Results

Due to the small number of participants with CIs tested thus far, only the NH participants were included in the group analyses. A 4 x 2 x 2 repeated measures ANOVA (Phase x Age group x Gender) was conducted on the median RTs for each of the four phases, with phase as the repeating factor. There was a main effect of phase ($F_{(1,12)}$=6.952, *p*<.025)[7], which suggests that the stimuli sequences presented in the different phases of the experiment had a significant effect on how quickly participants reacted to the stimuli. Specifically, when children were shown stimuli in a consistent spatiotemporal sequence (*e.g*. left, right, center) they were much faster at reacting to the next stimulus presented than when they were presented with stimuli in a quasi-random spatiotemporal sequence. This suggests that participants were learning the sequence. There was no significant main effect of age ($F_{(1,12)}$=.02, *p*=.89) or gender ($F_{(1,12)}$=.951, *p*=.35) and no significant interactions.

CDI Results

CDIs were given to all 8.5-month-old infants who completed the study and are being distributed to the 5-month-old infants three months after they participate in the study. At the time of this writing, all of the 8.5-month-olds and two of the 5-month-olds had completed CDIs. We conducted correlation analyses between median RT and scores on the CDI, focusing on the number of *phrases understood*, *vocabulary comprehension* score, and *total gestures* score. Percentages on these sections were calculated by dividing the number reported by the total number on that section of the CDI. The change in RT from Phase 1 to 2 (a measure of learning) was positively correlated with vocabulary comprehension score (r^2=.593, *p*<.05), but not significantly correlated with *phrases understood* or *total gestures* score (r^2=.445, *p*=.15 and r^2=.232, *p*=.47 respectively). This suggests children's success at learning the spatiotemporal sequence was positively related to their vocabulary comprehension ability at 8 months of age (see Table 1). With a larger data set we expect this link between early nonverbal cognitive abilities (*e.g*. visual sequence learning) and later language development (*e.g*. vocabulary) will persist. A significant correlation was also found between the Phase 2 and 3 RT difference and Phase 3 and 4 RT difference (r^2=.735, *p*<.01), reflecting that infants who had faster RTs during Phase 2 also had faster RTs during Phase 4, relative to the quasi-random sequence (Phase 3).

Measure	Phase 1 RT minus Phase 2 RT	Phase 2 RT minus Phase 3 RT	Phase 4 RT minus Phase 3 RT	CDI Phrases Understood	CDI Vocab Comprehension	CDI Total Gestures
Phase 1 RT minus Phase 2 RT	---					
Phase 2 RT minus Phase 3 RT	.44	---				
Phase 4 RT minus Phase 3 RT	.35	.74**	---			
CDI Phrases Understood	.45	-.09	.07	---		
CDI Vocab Comprehension	.59*	.15	.24	.50	---	
CDI Total Gestures	.23	-.31	-.01	.71**	.44	---

Table 1. Correlations between MacArthur CDI measures & sequence learning, *p<.05, **p<.01

[7] Using an alpha level of .05.

Summary

The results from the NH infants' average RTs for each phase suggest that 5- and 8.5-month-old NH infants can learn a three location spatiotemporal sequence. The differences between RT in the different phases suggest that infants were learning the sequence during the testing session. In this manner, the data followed the hypothesis that RTs would get faster during Phase 2, be slower during the random Phase 3, and finally be faster again during Phase 4 when the stimuli returned to the sequential pattern from Phases 1 and 2. There was also a significant correlation between vocabulary comprehension abilities and sequence learning from Phase 1 to 2, which indicates that sequence learning may have an important relationship to vocabulary development in NH infants.

Discussion

The goals of this study were to test sequence-learning skills in infants and to determine if sequence learning is an early predictor of later language development. Other studies have shown that in addition to age at implantation, several other demographic variables have been identified as predictors for language outcomes after cochlear implantation, including amount of residual hearing, length of deafness before implantation, and communication mode (oral vs. total communication) (Pisoni, Cleary, Geers, & Tobey, 2000). Thus, patients with more residual hearing, a shorter period of deafness and who use oral communication are more likely to be successful at developing language (Pisoni *et al*., 2000).

The first goal of this study was to determine if sequence learning could also be used as a predictor of language outcome in both NH and deaf infants with CIs. Due to the limited number of 5-month-old participants who were old enough to complete the CDI, the relationship between sequence learning ability and vocabulary abilities should be further investigated with more subjects. However, we have some preliminary data to suggest that visual sequence learning in NH infants significantly correlates with vocabulary comprehension.

The second goal of this study was to further our understanding of how the lack of auditory input in cases of deafness affects general cognitive processes in regards to sequence learning. There is evidence to suggest that when compared to age-matched normal hearing children, pre-lingually deaf children demonstrate atypical development in some aspects of executive function, such as working memory (Horn, Conway, Henning, Pisoni & Kronenberger, 2008). Thus, with this study we hoped to investigate differences in cognitive development for NH and deaf infants through visual sequence learning abilities. Because NH infants would have had more exposure to encoding and learning sequential information in the form of auditory language input, we hypothesized that there may some differences between visual sequence learning skills in NH versus CI infants. Thus far we were only able to enroll a moderate number of participants to adequately research this phenomenon. However, our results do encourage further investigation into sequence learning in NH and CI infants.

Recruitment for this study is ongoing with the goal of gathering additional data to further establish the evidence for sequence learning and its correlation with language development found here. Additional NH and CI participants are currently being recruited, as well as pre-implant deaf children, in order to assess visual sequence learning prior to implantation. We expect deaf infants may show lower sequence-learning abilities compared to NH infants and that their sequence-learning abilities may help predict later language outcomes.

References

Beery, K. (1989). *The VMI Developmental Test of visual-motor integration* (Vol. 3rd Revision ed.). Cleveland: Modern Curriculum Press.

Bergeson, T. R., Pisoni, D.B., & Davis, R.A.O. (2001). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review*, *103*(4), 23.

Berry, D. C., & Dienes, Z. (1993). *Implicit Learning: Theoretical and empirical issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Christiansen, M. H., Allen, J., & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 47.

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(5), 16.

Cleeremans, A. & J. L. McClelland. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120, 18.

Clohessy, A. B., Posner, M.I., & Rothbart, M.K. (2008). Development of the functional visual field. *Acta Psychologica*, 106(2), 17.

Cohen, L. B., Atkinson, D.J., & Chaput, H.H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas.

Conway, C. M., Bauernschmidt, A., Huang, S.S., & Pisoni, D.B. (under revision). Implicit statistical learning in language processing: Word predictability is the key. *Cognition.*

Conway, C. M., Karpicke, J., & Pisoni, D.B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Oxford University Press*.

Conway, C. M., Pisoni, D.B., Anaya, E.M., Karpicke, J. & Henning, S.C. (under revision). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science.*

Downs, M. P. & Yoshinaga-Itano, C. (1999). The efficacy of early identification and intervention for children with hearing impairment. *Pediatric Clinics of North America, 46*(1).

Haith, M. H., Hazan, C. & Goodman, G.S. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development*, *59,* 12.

Harrison, R. V. (2001). Age-related tonotopic map plasticity in the central auditory pathways. *Scandinavian Audiology*, *3*(53), 6.

Hollich, G. (2005). SuperCoder (Version 1.5).

Horn, D. L., Conway, C.M., Henning, S.C., Pisoni, D.B., & Kronenberger, W. (2008). Behavioral assessment of executive function in pre-lingually deaf children with cochlear implants. Indiana University School of Medicine: Department of Otolaryngology--Head & Neck Surgery.

Horn, D. L., Davis, R.A.O., Pisoni, D.B., & Miyamoto, R.T. (2004). Visuomotor integration ability of pre-lingually deaf children predicts audiological outcome with a cochlear implant: a first report. *International Congress Series*, 1273, 3.

Kirkham, N. Z., Slemmer, J.A., Richardson, D.C., & Johnson, S.P. (2007). Location, location, location: Developing spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 13.

Knoll, T., Steetharam, N., Coven, A., Kmoch, J., Byer, S., *et al*. (2007). Adobe Photoshop CS3 (Version 10.0): Adobe Systems Incorporated.

Levitt, H., McGarr, N.S. & Geffner, D. (1987). *Development of language and communication skills in hearing-Impaired children.* Washington, D.C.: American Speech-Language-Hearing Association.

Manrique, M., Cervera-Paz, F.J., Huarte, A., & Molina, M. (2004). Prospective long-term auditory results of cochlear implantation in prelinguistically deafened children: The importance of early implantation. *Acta Otolaryngologica, Suppl. 552,* 8.

Melnik, V., Malyarenko, A., Purdenko, A., Kysil, I., Alexeyenko, K. *et al*. (2006). SPSS 13 for Mac OS X. Chicago: SPSS Inc.

Miyamoto, R. T., Houston, D.M., Kirk, K.I., Perdew, A.E., & Svirsky M.A. (2003). Language development in deaf infants following cochlear implantation, *Acta Otolaryngologica, 123*, 3.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics, 106*(3).

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. *The Volta Review*, *101,* 53.

Pisoni, D. B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear & Hearing,24(*1S), 14.

Saffran, J. R., Aslin, R.N., and Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 3.

Saffran, J. R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 4.

Sarant J, B. P., Dowell R, Clark G, Gibson W. (2001). Variation in speech perception scores among children with cochlear implants. *Ear & Hearing, 22,* 10.

Sininger, Y. S., Doyle, K.J., & Moore, J.K. (1999). The case for early identification of hearing loss in children. Auditory system development, experimental auditory deprivation, and development of speech perception and hearing. *Pediatric Clinics of North America, 46*(4), 14.

Sly, D. J., Heffer, L.F., White, M.W., Shepard, R.K., Birch, M.G.J., Minter, R.L., Nelson, N.E., Wise, A.K., & O'Leary, S.J. (2007). Deafness alters auditory nerve fibre responses to cochlear implant stimulation. *European Journal of Neuroscience*, 26(2), 12.

Svirsky, M. A., Chin, S.B., & Jester, A. (2007). The effects of age at implantation on speech intelligibility in pediatric cochlear implant users: Clinical outcomes and sensitive periods. *Audiological Medicine, 5*(4), 13.

Universal Screening for Hearing Loss in Newborns: U.S. Preventive Services Task Force Recommendation Statement. (2008). *Pediatrics, 122*(1), 3.

Wentworth, N., Haith, M.M., & Hood, R. (2001). Spatiotemporal regularity and interevent contingencies as information for infants' visual expectations. *Infancy, 3*(3), 19.

Yoshinaga-Itano, C., Sedey, A.L., Coulter, D.K., & Mehl, A.L. (1998). Language of early- and later-identified children with hearing loss. *Pediatrics, 102*, 10.

Yoshinaga-Itano, C. (1999). Benefits of early intervention for children with hearing loss. *Otolaryngologic Clinics of North America, 32*(6).

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

Perceptual Similarity of Unfamiliar Regional Dialects: Some Preliminary Findings[1]

Terrin N. Tamati[2]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant T32-DC00012 and NIH-NIDCD Research Grant R01-DC000111. I would like to thank Robert Nosofsky for his help with data analysis. I would also like to thank Luis Hernandez for his helpful advice and assistance with this project.

[2] Also, Department of Linguistics, Indiana University, Bloomington, IN

Perceptual Similarity of Unfamiliar Regional Dialects: Some Preliminary Findings

Abstract. Linguistic experience has been shown to affect the perception of regional dialect variation. This pilot study examined the effect of familiarity on the perceived similarity of regional dialects. Native speakers of American English completed a paired comparison perceptual similarity rating task with a group of unfamiliar regional dialects. Speakers of American English were asked to make explicit judgments about the similarity of a set of unfamiliar talkers from the United Kingdom and Ireland on the basis of regional dialect. Results showed that listeners judged pairs of talkers from the same dialect region as more similar than pairs of talkers from different dialect regions. A multidimensional scaling analysis revealed two dimensions of perceptual dialect similarity, which reflected the geographic locations of the cities of origin of the talkers. These findings suggest that, despite being unfamiliar with the regional dialects in the study, listeners were able to use dialect-specific differences in the acoustic signal to make judgments of the perceptual similarity of talkers based on their regional dialect.

Introduction

Studies on the perception of regional dialect variation have shown that naïve listeners can perceive dialect-specific differences and can use stored knowledge of such differences to make reliable judgments about the regional background of talkers. Van Bezooijen and Gooskens (1999) and Van Bezooijen and Ytsma (1999) examined the categorization of regional varieties of Dutch in a multi-level forced-choice categorization task. They found that naïve listeners, when presented with audio recordings of talkers from different regions of the Netherlands and Belgium, were able to categorize talkers by regional dialect. Van Bezooijen and Gooskens (1999) also observed similar results for regional varieties of English spoken in the United Kingdom.

More recently, Clopper and Pisoni (2004b) used a forced-choice categorization task to investigate the perceptual categorization of regional varieties of American English. In their study, naïve listeners were able to place talkers into dialect categories. An analysis of the acoustic properties of the speech samples used in the study, the regional dialect of the talkers, and the responses of listeners in the categorization task showed that listeners used dialect-specific acoustic-phonetic properties in making their judgments. In addition, the listeners' response errors in the categorization task were consistent, revealing patterns of perceptual similarity among dialects.

Other studies have observed similar findings. Clopper and Pisoni (2007) asked naïve American English listeners to group unfamiliar talkers by region of origin in an auditory free classification task. Listeners were able to use dialect-specific phonetic differences to group talkers into regional dialect categories. Using a paired comparison perceptual similarity rating task, Clopper, Levi, and Pisoni (2006) also found that naïve listeners can make explicit judgments about the similarity of talkers' voices based on regional origin. In both studies, listeners' judgments reflected the perceived similarity of regional dialects of American English.

Linguistic experience and a listener's past developmental history has been shown to influence the perception of regional dialects. Williams, Garrett, and Coupland (1999) looked at the categorization of six regional varieties of English in Wales by adolescent males. The listeners were from the same six regions under study. In a forced-choice categorization task, overall accuracy was fairly low (30%).

However, a more detailed analysis of the various listener groups revealed that listeners were more accurate at identifying the region of origin of talkers from the same region (45%) than talkers that were from other regions (24%). Similarly, Baker, Eddington, and Nay (2009) investigated the effects of region of origin and amount of time spent in Utah on listeners' ability to identify talkers from Utah. In their study, listeners who were from Utah were better at identifying Utah talkers than listeners from other regions of the United States and listeners from the Western United States were more accurate at identifying Utah talkers than listeners from other regions. Experience also influenced performance, as listeners who had spent more time in Utah were better able to identify the Utah talkers. In addition, region of origin, but not experience, determined the features the listeners used to identify the talkers in the task.

Other studies have investigated how residential history influences the perceived similarity among dialects in further detail. Analyzing the error patterns in the perceptual categorization task, Clopper and Pisoni (2004b) found significant differences in the perceived similarity of dialects among three listeners groups (Northern Indiana, Southern Indiana, and Out-of-State). In another study, Clopper & Pisoni (2004a) examined the effect of geographic mobility on dialect categorization in American English. In that study, a group of mobile listeners, who had lived in at least three different states, and a group of non-mobile listeners, who had only lived in Indiana, performed a six-alternative forced-choice categorization task. Mobile listeners were more accurate than the non-mobile listeners in the categorization task and mobile listeners who had lived in a particular dialect region were better at categorizing talkers from that region than mobile listeners who had not lived in that specific region. Clustering analyses on the categorization responses also revealed that the perceived similarity of the regional dialects differed for the mobile and non-mobile listeners.

In another study, Clopper and Pisoni (2006) explicitly looked at the effect of region of origin on dialect categorization. Four groups of listeners who differed with respect to region of origin (North or Midland dialect regions) and geographic mobility (mobile or non-mobile) performed a forced-choice categorization task. While listeners did not differ in accuracy, geographic mobility and region of origin affected the perceived similarity of the regional dialects. Clopper and Pisoni (2007) found similar listener background effects using an auditory free classification task.

Clopper and Bradlow (2007) examined how native language status (native or non-native) affects the perceived similarity of regional dialects. In their study, despite limited experience with American English regional dialects, non-native listeners were able to use variation in the acoustic signal to classify talkers by regional dialect of American English in an auditory free classification task. The non-native and native listeners formed a similar number of talker groups and showed a similar perceptual similarity space. However, the type and amount of experience the listeners had with English influenced some aspects of perceptual dialect classification. Non-native listeners were less accurate and less consistent than native listeners in dialect categorization and, while they had a similar dialect similarity structure, the talker groups were more clearly defined in the native similarity space. Furthermore, the non-native listeners relied on dialect-specific phonetic features in grouping the talkers to a greater extent than the native listeners, who were able to use their knowledge of regional dialect variation in American English in completing the task.

Taken together, these earlier studies establish an important role for linguistic experience in the perception of regional dialect variation. The findings suggest that the more experience one has with a particular regional variety, the better one will be at identifying the regional variety of unknown talkers. In addition, these studies show that experience with regional varieties affects the perceived similarity among regional varieties. In other words, the more familiar the listener is with a given variety, the more distinct that variety will seem, perhaps resulting in easier identification and greater perceived distinctiveness among regional dialects.

Based on the previous studies, then, it would be expected that listeners would have difficulty in differentiating unfamiliar regional dialects. This prediction seems to be supported by recent findings from Ikeno and Hansen (2007). In their study, they looked at the effect of listener background on the comprehension and identification of regional dialects of English in the United Kingdom. Native speakers of British English, native speakers of American English, and non-native speakers of English living in the United States completed a forced-choice dialect categorization task. These groups represented three levels of familiarity with the regional dialects: familiar with both English and the regional dialects (British English group), familiar with the language but not the regional dialects (American English group), and limited familiarity with both English and the regional dialects (non-native group). Listeners classified talkers as having one of three different types of accents, which were introduced in a pre-test familiarization period. Talkers were from Belfast (Northern Ireland), Cambridge (England), and Cardiff (Wales).

The British English listeners were much more accurate (83%) than the American English listeners (56%) and the non-native listeners (45%). The listener groups also showed different confusion patterns in their responses, suggesting that the perceived similarity of the dialects was different for each listener group. Thus, the groups who were not familiar with the three dialects in the study (the American English listeners and the non-native listeners in the United States) had more difficulty in categorizing unfamiliar talkers by regional dialect. However, these groups did perform quite well with 56% correct for the American English listeners and 45% for the non-native listeners, suggesting that they were able to reliably detect differences between the three unfamiliar dialects.

One of the limitations of the Ikeno and Hansen (2007) study is that they used a three-alternative forced-choice categorization task. Providing the dialect regions and labels imposes response constraints on the listeners in completing the task (Clopper & Pisoni, 2007). Given that the listeners were unfamiliar with the regional dialects in the study, the forced-choice categorization task would likely have been very difficult. The auditory free classification task (Clopper & Pisoni, 2007) and the paired comparison similarity rating task (Clopper *et al.*, 2006) have recently been used to examine the perceived similarity of regional dialects of American English. Since both the auditory free classification task and the paired comparison similarity rating task allow listeners to directly compare regional varieties without imposing experimenter-based categories, they might be better suited for studies on how familiarity affects the perception of regional dialect variation.

The purpose of the current pilot study was to evaluate, as a basis for future studies on the effect of familiarity of the perception of regional dialect variation, the perceived similarity of a group of unfamiliar regional dialects using the direct two-interval dialect similarity rating task adopted from Clopper *et al.* (2006). Native speakers of American English were presented with a set of talkers from five different regions of the United Kingdom and Ireland and asked to rate the similarity of the talkers based on regional dialect. Based on previous studies, it was expected that listeners would have a great deal of difficulty in distinguishing the regional dialects. It was unclear if listeners would be able to perceive talkers from the same dialect region as more similar than talkers from different dialect regions or if similarity judgments would reflect the regional background of the talkers.

Method

Listeners

Ten native speakers of American English participated in the study. Listeners had diverse residential histories but none reported having lived in the United Kingdom, Ireland, or any other English-speaking country outside of the United States. Three listeners reported being fluent in at least one other language. All of the nine participants were between the ages of 22 and 27 at the time of testing. One participant reported a history of a speech or hearing disorder.

Talkers

Twenty talkers were selected from the IVie Corpus (Grabe, Post, & Nolan, 2001). The IVie Corpus consists of audio recordings of twelve to fourteen talkers for each of nine cities in the British Isles. Half of the talkers for each city were female and half were male. Cities represented in the corpus include Belfast, Bradford, Cambridge, Cardiff, Dublin (Malahide), Leeds, Liverpool (Scouse), London, and Newcastle upon Tyne. All talkers were around the age of sixteen and enrolled in urban high schools at the time of recording. Talkers from Bradford, Cardiff, and London were ethnic minorities. Talkers from Bradford were bilingual Punjabi-English speakers, talkers from Cardiff were bilingual Welsh-English speakers, and talkers from London were of Jamaican descent. Materials for each talker include a conversation, a map task, a story told from memory, read sentences, and a read passage. This corpus was designed for studies on variation in the prosodic systems of different regional dialects.

The twenty talkers used in the current study were females around the age of sixteen at the time of recording. Four talkers were chosen from each of the following five cities: Belfast, Dublin (Malahide), Cambridge, Leeds, and Newcastle upon Tyne. These five cities were chosen for their geographic locations, as they were, geographically, distant from one another. Talkers from Bradford, Cardiff, and London were not used in the study in order to avoid introducing additional within-dialect variation from varying degrees of bilingualism or contact with other varieties of English, such as English spoken in India or Jamaica. The talkers from each city were selected based on the quality of the recording available for the read passage and the fluency of their speech in the target phrase.

Stimulus Materials

Stimulus materials consisted of a single short phrase for each talker. The phrase was selected from the read passage in the IVie Corpus (Grabe *et al.*, 2001). An attempt was made to select a phrase that would include a variety of speech sounds that could be used to distinguish the regional dialects in the study. The phrase used throughout this study is reported in (1):

(1) the beauty who had stolen his heart

For each talker, the target phrase was extracted from the read passage and saved in an individual sound file. When necessary, the sound file was edited so that there would be no noticeable speech errors.

Procedure

Listeners were seated in individual booths in a quiet testing room. Each sat in front of a computer equipped with a keyboard and headphones. On each trial, listeners heard the target phrase produced by two different talkers separated by 500 ms of silence. The phrase also appeared visually on the computer screen. Listeners were asked to judge the similarity of the accents of the two talkers on a scale from 1 ("very different") to 7 ("very similar"). This way, they were asked to make an explicit judgment on the similarity of the unfamiliar dialects, with higher ratings reflecting greater dialect similarity than lower ratings. For each trial, listeners had seven seconds from the onset of the second production of the target phrase to respond. Listeners responded by pressing the button on the keyboard that corresponded to the similarity rating for a particular talker pair. If the listener failed to give a response in that time, the next trial would begin. They received no feedback during the experiment.

Each listener heard all twenty talkers paired with all other talkers two times throughout the study, for a total of 380 trials. Of the total number of trials, 60 trials included two talkers from the same city ("same-dialect" trials) and 320 included two talkers from different cities ("different-dialect" trials). For each talker pair, both possible orders of presentation of the talkers (AB and BA) were included. The order of presentation of the trials was determined randomly. The experiment was divided into two halves, with a short break provided after the first 190 trials. It took approximately 45 minutes to complete. The similarity ratings and response times for the ratings were collected. All trials for which no response was recorded were removed (*N* = 20). Only the similarity ratings will be discussed here.

Results

To see whether listeners were able to make explicit judgments on dialect similarity for unfamiliar dialects, responses for same-dialect and different-dialect pairs were examined. Table 1 shows the mean similarity ratings for the same-dialect and different-dialect pairs.

Trial	Response
Same-dialect	5.40 (1.50)
Different-dialect	3.85 (1.76)

Table 1: Mean and standard deviation of responses for same-dialect and different-dialect trials.

Overall, listeners showed higher similarity ratings for same-dialect pairs than different-dialect pairs. To assess the difference between ratings for same-dialect and different-dialect pairs, a paired *t*-test was carried out on similarity ratings with dialect match (whether it was a "same-dialect" or "different-dialect" trial) as the factor. The difference in responses for same-dialect and different-dialect trials was highly significant (*t*(9) = 9.064, p < .001). Thus, same-dialect pairs were rated significantly higher than different-dialect pairs.

A multidimensional scaling analysis was then carried out to obtain measures of the underlying similarity space. A 20 x 20 talker similarity matrix was constructed from the similarity ratings given to a particular talker pair averaged across all listeners. An ALSCAL analysis was performed on the similarity matrix. One-, two-, three-, and four-dimensional solutions were obtained. Mean stress values for each solution were .296, .162, .103, and .085, respectively. These values suggest an "elbow" at the two-dimensional solution, as stress was greatly reduced from the one- to two-dimensional solution and the addition of the third dimension did not result in a large decrease in stress. Given these results and the high interpretability of the two-dimensional solution, the two-dimensional solution was selected for discussion.

The resulting two-dimensional multidimensional scaling solution from the ALSCAL analysis was plotted. The first dimension appeared to separate the talkers from England on the left and the talkers from Ireland on the right and the second dimension appeared to separate the talkers from the North of each country from talkers from the South of each country. Since the dimensions seemed to roughly correspond to the geographic locations of the cities, in order to better visualize the results, the plot was rotated, with the geographic coordinates of the cities as the normative solution, using the ROTAT program. The resulting plot after rotation is shown in Figure 1. Each talker is represented by the first letter of her city. The four *B*s represent the four talkers from Belfast, Northern Ireland, the four *C*s represent the four talkers from Cambridge, England, the four *D*s represent the four talkers from Dublin, Ireland, the four *L*s represent the four talkers from Leeds, England, and the four *N*s represent the four talkers from Newcastle upon Tyne, England.

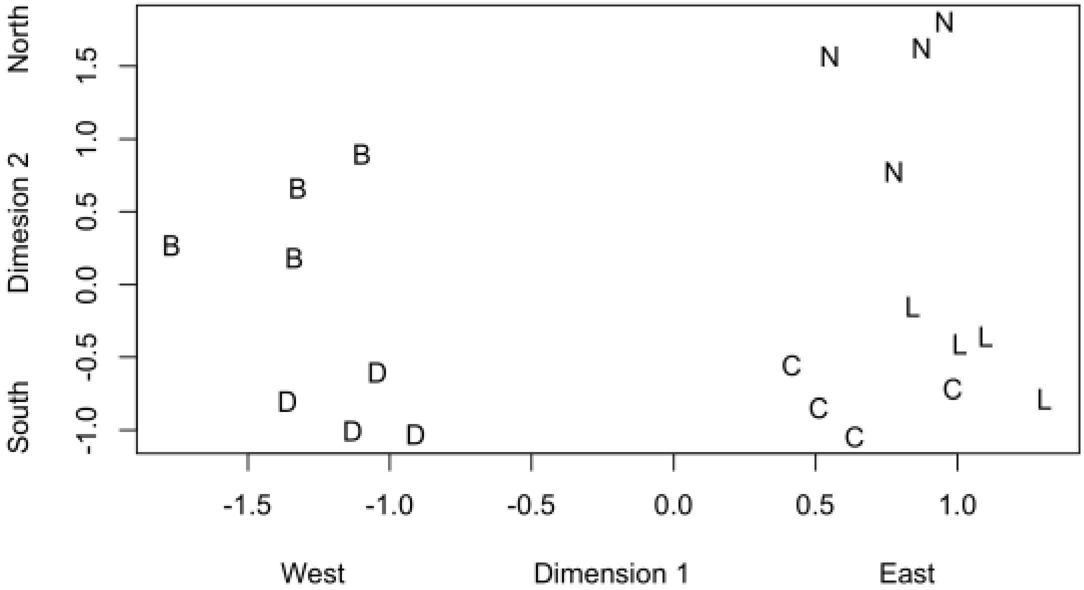

Figure 1. Rotated two-dimensional multidimensional scaling solution. Talkers are represented by the first letter of their home city: *B* for Belfast, *C* for Cambridge, *D* for Dublin, *L* for Leeds, and *N* for Newcastle upon Tyne.

This plot shows that the talkers from Dublin and Belfast are separated from the talkers from Cambridge, Leeds, and Newcastle along the first dimension and that the talkers from Dublin, Cambridge, and Leeds are separated from the talkers from Belfast and Newcastle along the second dimension. In other words, the first dimension distinguishes the talkers from the West from the talkers from the East (or Ireland from England) and the second dimension distinguishes the talkers from the North of each country from the talkers from the South of each country. These two underlying perceptual dimensions can be interpreted, then, in terms of geography: West versus East and North versus South.

Discussion

The results of the two-interval similarity rating task show that, given the phrase *the beauty who had stolen his heart*, listeners judged the similarity of the accents of two talkers based on their regional background. The analysis of the same dialect and different dialect ratings showed that listeners rated pairs of talkers from the same city as more similar than pairs of talkers from different cities.

In addition, the multidimensional scaling analysis of the similarity ratings revealed two dimensions of perceptual dialect similarity. These dimensions reflected the geographic location of the cities; the first dimension was interpreted as West versus East and the second dimension was interpreted as North versus South. These perceptual dimensions produced a space that roughly corresponded to the geography of England and Ireland. As shown in Figure 1, the Belfast talkers are in the Northwest quadrant (top left in Figure 1), the Dublin talkers are in the Southwest quadrant (bottom left in Figure 1), the Newcastle talkers are in the Northeast quadrant (top right in Figure 1), and the Leeds and Cambridge talkers are in the Southeast quadrant (bottom right in Figure 1), with the Leeds talkers located slightly above the Cambridge talkers.

The listeners' ability to judge the similarity of the talkers' accents based on regional background is consistent with the earlier results of Clopper *et al.* (2006). In that study, native speakers American English made explicit judgments about the dialect of talkers, who were also native speakers of American English, based on their regional background. However, unlike Clopper *et al.* (2006), where listeners would have had varying degrees of experience with the different American English regional dialects, in the current study, listeners were unfamiliar with the regional dialects that they rated. That is, listeners had very little experience or familiarity with the English and Irish dialects.

As mentioned above, none of the listeners lived in England, Ireland, or any other English-speaking country outside the United States. Some of them, however, had acquaintances from England and Ireland, and, given the linguistics background of some of the participants, it is also likely that a few of the listeners were familiar with studies on regional dialects in England and Ireland. In a post-test questionnaire, most listeners reported that they recognized the talkers as being from England or Ireland, but none of the listeners were able to identify explicitly the exact city of origin of any of the talkers. In a few instances, listeners erroneously thought that some of the talkers were speakers of American English or non-native speakers. Only two of the listeners were able to identify the broad regions of origin of some of the talkers. One listener indicated that some of the talkers were from Northern Ireland and another indicated that some of the talkers were from Southern England.

Thus, despite having little direct experience with the regional dialects in the study, listeners were able to make explicit judgments about the perceptual similarity of unfamiliar talkers' voices that reflected regional dialects of Ireland and the United Kingdom. Since the results of this pilot study cannot be compared to perception data from listeners familiar with all, or some, of the regional dialects, they cannot yet be interpreted in terms of the effects of familiarity on the perceptual similarity of regional dialects. Still, the performance of the speakers of American English raises the issue of how familiarity with regional dialects influences the perceived similarity of those dialects.

Previous studies suggest that, because they were unfamiliar with the regional dialects, listeners should have had difficulty judging the similarity between the regional accents. Since they have had such little experience with the dialects, they would not have knowledge of the phonetic differences that are important for distinguishing the dialects nor would they have knowledge of variation irrelevant to distinguishing the dialects. Nevertheless, listeners were able to identify dialect-specific phonetic features in the speech of the unfamiliar talkers and use these features, while disregarding other irrelevant variation, to make judgments about the similarity of talker pairs based on regional dialect. The multidimensional scaling solution obtained from the similarity data shows that talkers were clearly grouped based on their city of origin along geographic dimensions.

So, given how well the American English listeners performed, could a listener familiar with the regional dialects in the study do any better? A listener who is familiar with the regional dialects would be influenced by their linguistic experience, including their own dialect, mobility, and social background. This linguistic experience would shape the way that they hear the differences between the dialects. However, as the results of the current study suggest, the "perceptual warping" of the dialect similarity space resulting from linguistic experience would not necessarily result in dialects being heard as more distinct from one another. Experience may actually lead to regional dialects being perceived as more similar, rather than different. This interpretation is consistent with previous studies that have shown that non-mobile American English listeners from the Northern dialect region actually perceive their own dialect as more similar to the Midland dialect or a supraregional standard variety (Clopper & Pisoni, 2006; Clopper & Pisoni, 2007; Niedzielski, 1999). To examine this in further detail, another study should also be carried out with listeners with different degrees and types of familiarity with the regional dialects.

Conclusions

In this study, a paired comparison perceptual similarity rating task was used to examine the role of familiarity on the perceptual similarity space of regional dialects. Speakers of American English made judgments of the similarity of the accents of unfamiliar talkers from various regions of the United Kingdom and Ireland. Results showed that listeners were generally successful at judging the similarity of the accents of two talkers based on their regional background. Listeners rated pairs of talkers from the same city as more similar than pairs of talkers from different cities. Furthermore, the multidimensional scaling analysis of the similarity ratings revealed two underlying dimensions of perceptual dialect similarity, which reflected the geographic location of the cities. These findings differ from previous studies on the effect of familiarity on regional dialect identification or classification, which predicted that listeners with little or no experience with a set of regional dialects should be very poor at differentiating such dialects. The results of this study, however, showed that listeners were successful at distinguishing the British and Irish regional dialects, suggesting that experience with regional dialect variation may not necessarily result in greater perceived differences between regional dialects.

References

Baker, W., Eddington, D., & Nay, L. (2009). Dialect identification: The effects of region of origin and amount of experience. *American Speech , 84*, 48-71.

Clopper, C. G., & Bradlow, A. R. (2007). Native and Non-native Perceptual Dialect Similarity Spaces. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. 665-668). Saarbrücken, Germany.

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change, 18,* 193-221.

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics, 35,* 421-438.

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.

Clopper, C. G., Levi, S. V., & Pisoni, D. B. (2006). Perceptual Similarity of Regional Dialects of American English. *Journal of the Acoustical Society of America, 119*, 566-574.

Grabe, E., Post, B., & Nolan, F. (2001). *The IViE corpus.* Department of Linguistics, University of Cambidge. http://www.phon.ox.ac.uk/IViE/.

Ikeno, A., & Hansen, J. H. (2007). The Effect of Listener Accent Background on Accent Perception and Comprehension. *EURASIP Journal on Audio, Speech, and Music Processing, 3,* 1-8.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of the Acoustical Society of America*, *18,* 62-85.

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, *18,* 31-48.

Van Bezooijen, R., & Ytsma, J. (1999). Accents of Dutch: Personality impression, divergence, and identifiability. *Belgian Journal of Linguistics*, *13,* 105-129.

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

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 29 (2008)
Indiana University

A Software-Based System for Synchronizing and Preprocessing Eye Movement Data in Preparation for Analysis[1]

Mohammad B. Afaneh, Visal Kith, and Tonya R. Bergeson[2]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Research Grant R21DC06682 and Training Grant T32DC00012. We thank Luis Hernández for his insightful comments on this project.

[2] Babytalk Research Laboratory, Department of Otolaryngology – Head and Neck Surgery, Indiana University School of Medicine. Correspondence concerning this article should be addressed to Tonya R. Bergeson, Ph.D., Department of Otolaryngology – Head and Neck Surgery, Indiana University School of Medicine, 699 West Drive – RR044, Indianapolis, IN

A Software-Based System for Synchronizing and Preprocessing Eye Movement Data in Preparation for Analysis

Abstract. When upgrading or adding a new component or system to an existing system in a research lab, the problem of incompatibility often arises. In this paper, we present a software-based solution for integrating and synchronizing an eye-tracking system with another software system used for stimulus presentation in an infant speech research lab. The algorithms developed and implemented in this system originate from different tracks of images and data processing algorithms. The solution presented is specific to our particular system set up. Nevertheless, it can be easily applied to any other similar setup using an eye-tracker system with stimuli presentation software.

Introduction

Traditionally, researchers of visual attention and perception have made use of techniques such as monitoring a live view of the subject's head and face to get a rough idea of their gaze direction (e.g., looking right, left, or center). With recent advances in eye tracking technology, however, eye tracking systems have been introduced and integrated in visual perception laboratories where both high accuracy and high resolution are necessary to investigate looking behavior towards a variety of visual scenes.

The use of eye trackers has several advantages over the traditional methods. The first and most important advantage is the increased accuracy over traditional methods (up to 0.5 degrees visual angle). Second is the measurement of other types of useful information in addition to the direction of gaze (e.g., pupil diameter, blinks, head position, and pupil position). Another very important advantage is the ability to have external software analyzing eye tracker output data. By using such software, the process of detecting blinks, fixations, dwell times and saccades can be done automatically.

A typical eye tracker consists of a camera, which is encircled by a ring of infrared LEDs, and a control unit in which the image captured is processed before being transferred to a PC or other interface device. The ring of LEDs illuminates the eye, and when placed on the axis of the camera lens it produces an interesting effect on the pupil. That is, the subject’s pupil will appear as a bright object in the captured image, similar to the “red-eye” effect in photography. The infrared light also causes a reflection off the cornea. By computing the vector between the corneal reflection and the pupil center, the system can compute the direction of the subject’s gaze.

In many laboratories, software specific to an operating system is used for presenting stimuli in experiments. For example, Habit software, which runs on Mac OS, is used in several infant laboratories (Cohen, Atkinson, & Chaput, 2004). Experiments run with Habit allow the duration of trials to be set on-line according to predefined criteria regarding the attention and behavior of the subject. To do this, Habit monitors the operator's key strokes on a computer keyboard, which tell the software where the subject is looking (left, right, or center), and according to these strokes Habit determines the duration of the trial on the screen and moves on to the next trial. The output file of the program contains the direction of looks and cumulative look times in each trial.

Problem Statement

In the infant laboratory setup described above, the Mac will send both an output video and audio stimulus to a TV screen or a monitor to be viewed by the subject. When an eye tracker is introduced into the system, the video output first has to pass through a "scan converter" device before continuing on its way to the TV screen. This device splits the video signal into two signals: one which goes into the eye tracker control unit, and a second which passes to the monitor to be displayed (see Figure 1).

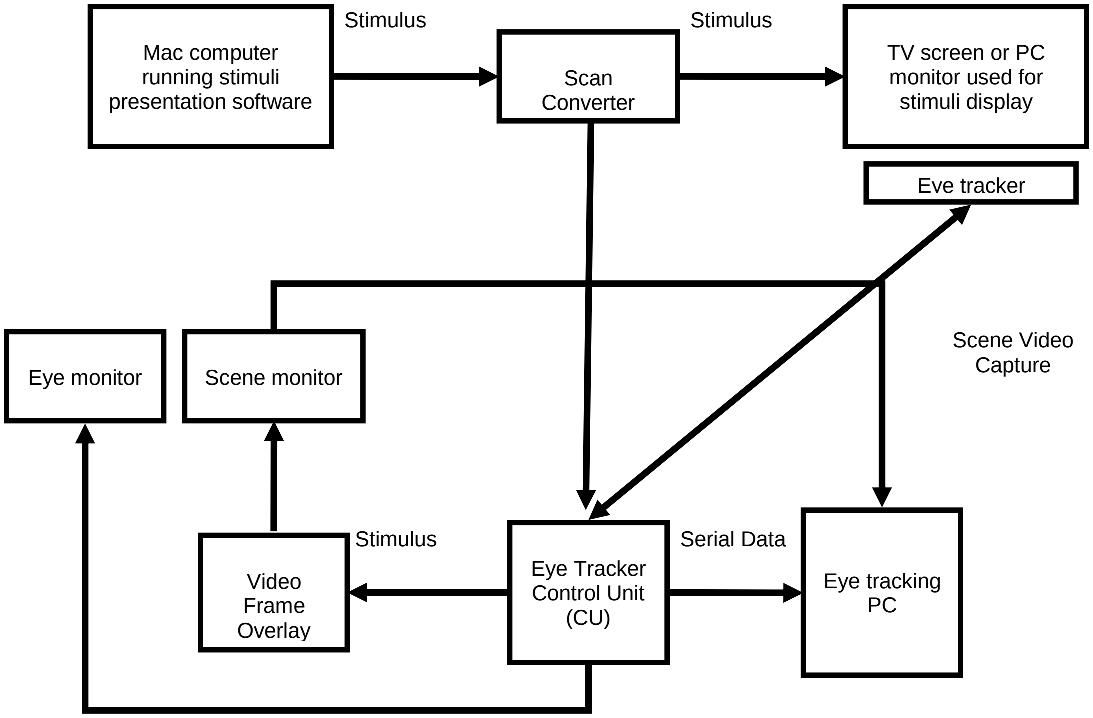

Figure 1. System model represents the interaction between a Mac, a PC and an Eye-Tracker system. The Mac sends video and audio output to the TV by pass through a scan converter. The scan converter then splits the video signal into two: one goes to the TV and one goes to the eye-tracker control unit. The eye-tracker control unit superimposes crosshairs on the frames indicating the gaze target of the subject and then sends to the scene monitor to be captured by the PC. At the same time the eye-tracker control unit also receives video input from an eye-tracker camera which to be displayed on the eye monitor.

The eye tracker obtains the video of the stimulus and then superimposes crosshairs on the frames indicating the gaze target of the subject. In such a setup the eye tracker gets an input of the video, but does not know the start or end frame of a trial within an experiment. Also, the output file consists of rows of eye movement data, each of which corresponds to one frame (or 1/60th of a second) for the entire experiment. That is, the output file contains no separation between trials within each experiment. The separation between trials is very important since statistical analyses are performed using data from each trial within an experiment. Finally, each video is time stamped; we will take advantage of this timestamp as part of our solution.

Proposed Solution

The solution can be approached in two different ways: hardware or software. Each has its advantages and disadvantages. The advantage of software over hardware is usually cost, and the disadvantage is usually speed. Another important advantage for software-based solutions is portability. To avoid adding new hardware components to the system and reduce cost, we chose the software approach.

The video output presented on the TV screen by Habit consists of pretest, habituation, and/or test phases of stimuli, each of which is called a trial. In between the different trials, a short video (the "attention getter") is repeated to draw the attention of the infant. The general idea behind our solution is to make use of the captured "attention getter" video to detect the start and end frames of the different trials. Then the program can extract this information, detect the rows in the output data file corresponding to these trials, and mark them as separate trials for the purpose of later analyses.

Even though eye movement data usually exists for the majority of the trials, the output of the Habit software can still serve as a backup of the gaze direction results in cases where the eye tracker loses data or does not work properly. It also serves to mark the trials in the output file with their specific descriptions.

Finally, our hope was to minimize the requirement of user interaction and make the use of the software as easy as possible. We chose to develop a program with a simple Graphical User Interface (GUI) with this in mind. Figure 2 shows the design and appearance of the GUI.

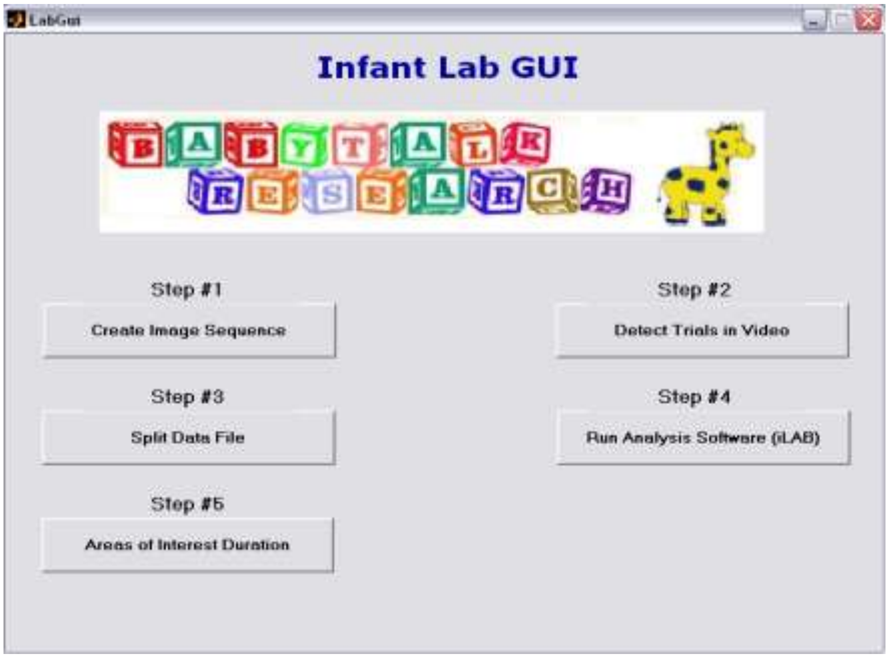

Figure 2. The design of GUI is to make the use of the software as easy as possible. It contains step by step instruction for the user to follow to achieve the data analysis.

Algorithm Discussion

There are two phases in the proposed algorithm. The first phase consists of the image processing and object recognition processing, which help in detection of the different trials in the video. The second phase is the data processing phase which includes reading, preprocessing, and then marking of the rows of the eye movement data file.

Phase I

Figure 3 shows a flow chart of the basic steps in the first phase of our system solution. To make the process of the algorithm clearer we will describe each step in detail.

The first step in this phase is to split the video file into a sequence of jpeg image files with each file name containing the sequence number of that image in the original video (frame0001.jpg, frame0002.jpg, etc.). This video decomposition will allow us to process and analyze each frame in its original sequential order. This step is also beneficial for programs that deal with images, rather than extracting individual frames from the original video file format.

The program then loops over all the image files in order, starting with the first frame (**frame_number** = 1). To be able to tell the software to look for a trial start or end frame in the entire series of frames, we define a variable which is set to 1 when looking for a start frame of a trial (**start_flag** = 1), and set to zero when looking for a trial end frame (**start_flag** = 0).

If **start_flag** is set to 1 then the algorithm will first find the correlation of the current frame with a reference image (defined prior to entering the loop). The reference image should be an image which represents the image displayed between the trials (see Figure 4). To find the start and end of trials, we calculate the autocorrelation between the current frame in the video and the reference frame. However, in our case a movie file was presented between trials. This meant that we could not rely on just one reference image in detecting the *separation movie*. The more practical solution was to choose the correlation coefficient to be 0.8 instead of 1 (chosen by trial and error). This technique required only one reference image which is relatively similar to the frames of the separation movie. The choice of a correlation coefficient of less than 1 also accounted for any artifacts present in the frames. If the correlation is indeed less than 0.8 then we have detected the start frame of the first trial.

Next, the algorithm calls another routine which recognizes and extracts the sequence number present in the current frame and stores it in the start frames array.

The next step in the algorithm sets the **start_flag** equal to 0 until we find the end frame of the trial. If **start_flag** equals 0 then the algorithm will find the correlation between the current frame and the reference frame. If the correlation is larger than or equal to 0.8, then it will decide that the image is indeed the start of the separation movie. In this case, the algorithm calls the digit recognition routine to extract the sequence number present in the **previous** frame and **not** the current frame because the previous one was the last frame of the trial and the current frame is not included in the trial. The algorithm then stores the digit recognition result in the end frame array.

Afterwards, **start_flag** is set to 1 so that the program searches for the start frame of the next trial and so on until it detects the end frame of the last trial. Finally, the algorithm outputs the two arrays (start frame array and end frame array), each in a separate line to an output file chosen by the user as an argument to the algorithm. Table 1 is an example of the output. The first row represents the start frames and the second represents the end frames of the different trials in order. In this case there were three trials in the experiment.

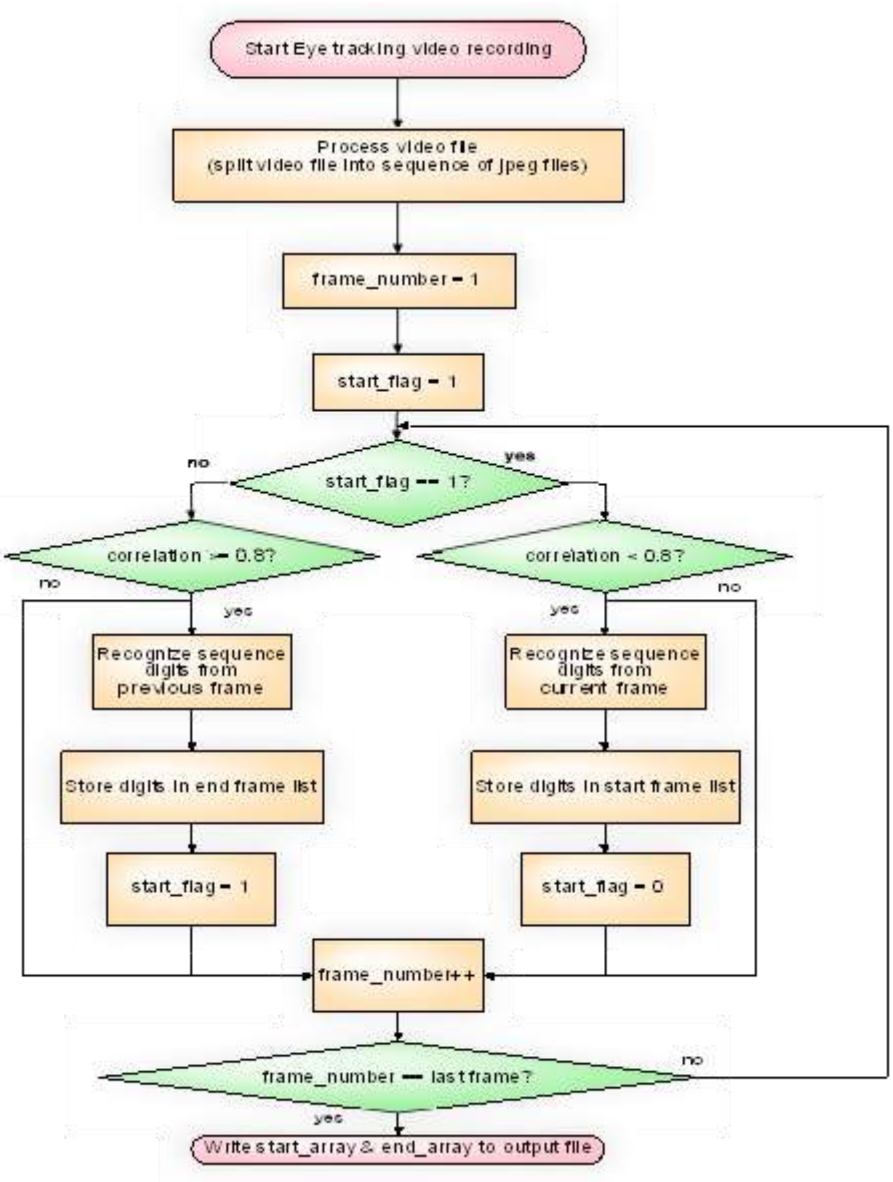
Start Eye tracking video recording
Process video file
(split video file into sequence of jpeg files)
frame_number = 1
start_flag = 1
no
yes
start_flag == 1?
correlation >= 0.8?
correlation < 0.8?
no
no
yes
yes
Recognize sequence digits from previous frame
Recognize sequence digits from current frame
Store digits in end frame list
Store digits in start frame list
start_flag = 1
start_flag = 0
frame_number++
no
frame_number == last frame?
yes
Write start_array & end_array to output file

Figure 3. Phase I Flow Chart shows the basic steps in detecting the starting and ending of each trial. First, the algorithm checks if the variable start_flag is equal to 1 or not. If it is, then it searches for a start frame; otherwise it searches for an end frame. After detecting the starting or ending frame, the algorithm recognizes sequence digits and stores those digits in an array.

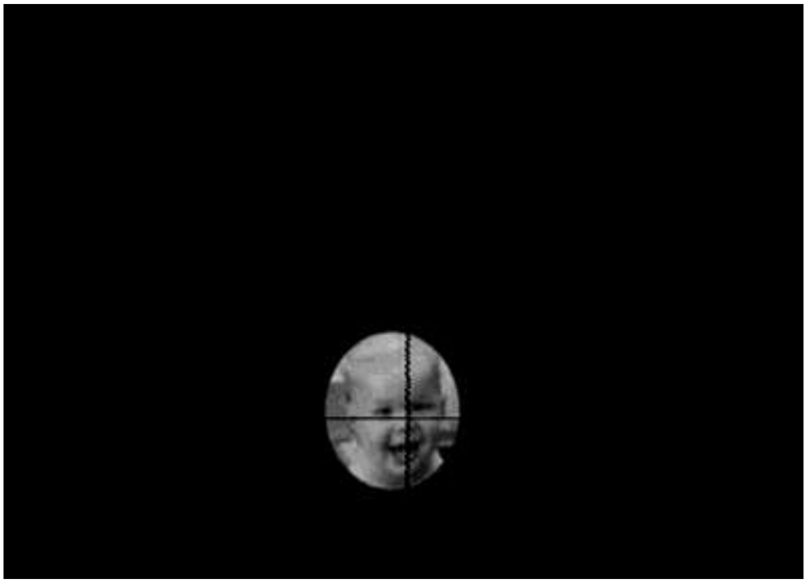

Figure 4. The reference image is an image which represents the image displayed in between the trials. We are able to find the start and end frame for each trial by calculating the correlation between the current frame and this reference image. If the correlation is less then 0.8 then we have detected the start frame.

	Trial 1	Trial 2	Trial 3
Start Frames	12056	14657	17569
End Frames	13503	15798	20033

Table 1. Output of Start and End Frames for 3 Trials

Digit recognition

Digit detection and extraction from a particular frame are included in Figure 3. In our system, we had 5 digits superimposed on each video frame, each of which corresponds to a row of data in the eye movement data file. Those frame numbers are superimposed by an external hardware device (called *video frame overlay*) which takes in two inputs, the video output and the data stream from the eye tracking system; and one output, the video with the superimposed frame numbers.

The idea behind using such a device is to synchronize the video output with the eye movement data. By reading the frame number from the video frame one can know the corresponding row of data in the eye movement file. In our solution, we use image processing techniques to detect the different trials in the video, and then go back to the data file and mark the lines corresponding to each trial in order to perform the analysis on each trial separately.

Figure 5 shows an example of a sub-image taken from a video frame containing the superimposed frame numbers. We chose this frame to illustrate a serious problem: two frame numbers are overlapped on the same frame. This is caused by the down-sampling of the video due to the mismatch between the frame grabber capturing video at 30 Hz and the eye tracker outputting frames at 60 Hz. Because the video is interlaced, the odd rows of the image are captured at the first run of the image capture while the even rows are captured at the second run of the capturing. One digit comes from the odd pass while the other comes from the even pass. We have to note here, however, that this is a worst-case scenario since in most images only the last digit (far right) of the sequence will be an overlap of two consecutive digits. The second to last digit will appear as an overlap every 5 frames and the third to last every 25 and so on.

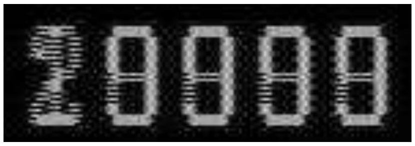

Figure 5. Superimposed frame numbers on a frame are output frame numbers from eye-tracker which are superimposed on a frame. This figure shows the worst-case scenario where two frame numbers are overlapped on the same frame.

To solve this issue, for each frame we extract the odd rows and extract the digits from this sub-image. There will still be an error detecting the exact row which represents either the start or end frame of that trial, but the maximum error would be 2 frames in each trial, which is negligible, compared to the total number of frames in each trial. Figure 6 shows the even and odd rows extracted from the sub-image in Figure 5.

(a) Odd Row

(b) Even Row

Figure 6. Odd and even row images extracted from the image in Figure 5. (a) A frame number in odd row. (b) A frame number in even row.

In order to recognize all digits in a certain frame, we propose an algorithm which uses template-matching theory to recognize one digit at a time and output an array of five elements representing the frame sequence (see Figures 7 and 8). Since the digits in the frames are not located in the exact position in each experiment (possibly because of jpeg artifacts or random errors in the image capture process), the algorithm first searches for the area in which each digits is located.

Figure 7. Template images used in the template matching algorithm to recognize each digit of a frame number.

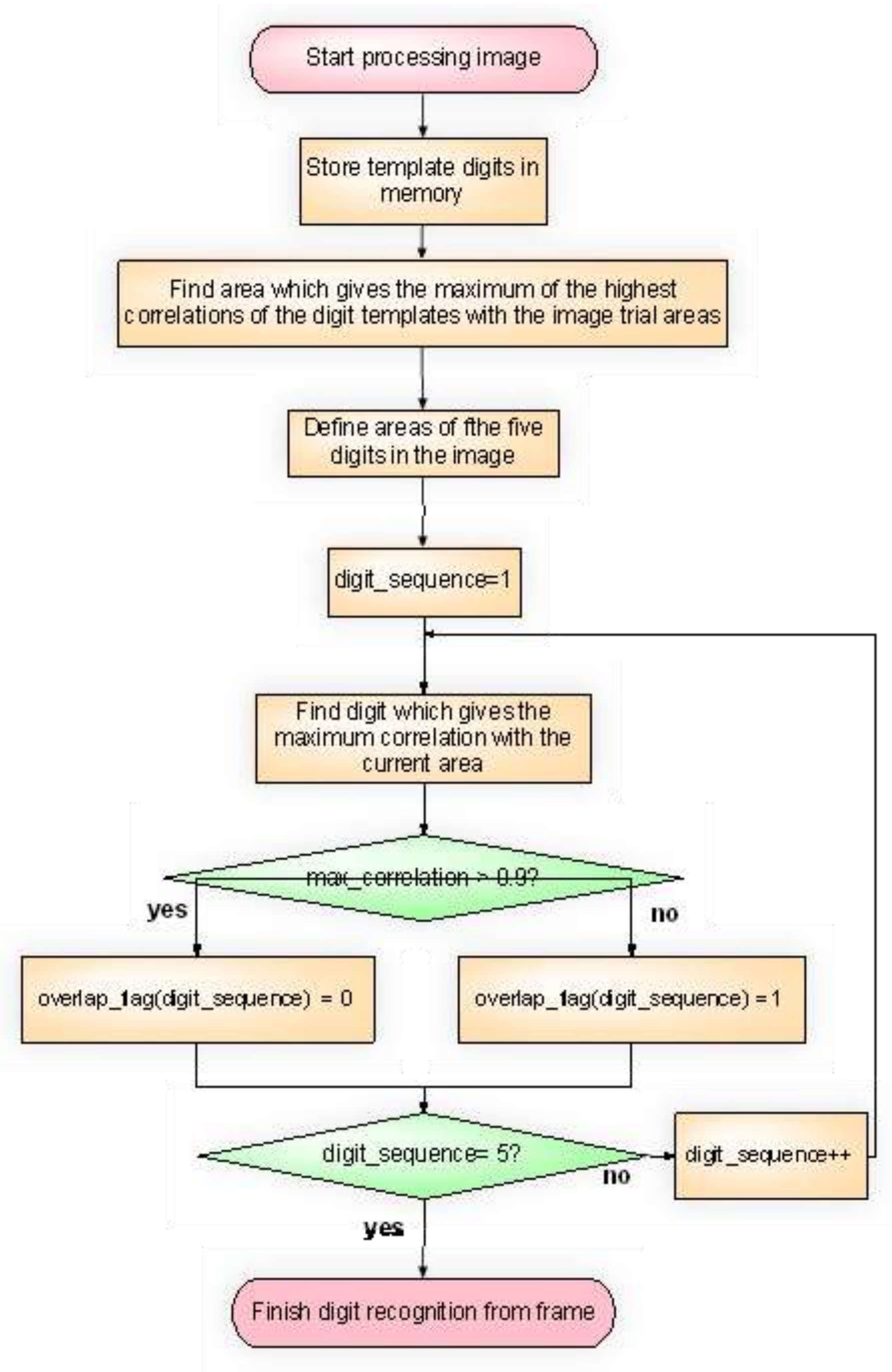
Start processing image
Store template digits in memory
Find area which gives the maximum of the highest correlations of the digit templates with the image trial areas
Define areas of fthe five digits in the image
digit_sequence=1
Find digit which gives the maximum correlation with the current area
max_correlation > 0.9?
yes
no
overlap_flag(digit_sequence) = 0
overlap_flag(digit_sequence) = 1
digit_sequence= 5?
no
digit_sequence++
yes
Finish digit recognition from frame

Figure 8. The Digit Recognition Algorithm is used to recognize each digit of a frame number. First, this algorithm searches for the area in which each digit is located. Then the variable digit_sequence specifies how many digits need to recognize. Note that for optimization, not all digits will be recognized all the time.

The method approached was to search for the area that had the highest correlation with each of the digit template images. That is, the correlation between each area and the image templates is calculated and the highest correlation with the digits is stored for each candidate region. Across all the areas, the one with the highest of these correlations is chosen to be the correct digit. This is done for each of the five digits in order to find all the exact regions in each frame.

This solution was very accurate in recognizing all five digits. However, the processing time turned out to be much longer than we expected, so to optimize the algorithm we made the program only extract all five digits in cases where it was needed.

As we mentioned above, each frame contained an overlap of two sequence numbers. So if a frame was overlapped by the numbers 00001 and 00002, then the following frame will most likely be stamped with an overlap of 00003 and 00004 unless there was loss of one or more frames during the frame capture process. Thus, we could use an approximation method to calculate the number on each frame. In other words, the program could be modified to extract the digits from one frame and then use an offset to determine the frame sequence number on any subsequent frame.

However, we found that this approximation method might not always get the exact number on the frame due to frames dropped during capture. For example, sometimes the approximation was 00100 but the exact number was 00114. The difference between these two numbers might vary depending on how many frames dropped during video capture. If 7 frames were dropped, then the difference would be 14.To account for the loss in frames captured and still be able to reduce the running time of the algorithm, we check for the difference between the approximation and the recognition of the last two digits. If the error is less than **α,** the approximation is added to the recognition of the first frame in that trial. If the error is greater than or equal to **α**, then we perform recognition for all five digits by using algorithm in Figure 8, where **α** is defined as the maximum number of frames loss in each trail.

This is true assuming that the maximum number of lost frames is 25, thus **α** = 25 * 2 = 50 (each frame overlap by two numbers). If the difference between the approximation and the recognition of these two digits is less than 50, that leaves no chance for the other digits to be different. An example in which the difference could be greater than 50 is where the approximation is 1196 and the recognition is 04. Thus 96 – 04 = 92 (compare only the last two digits), so after performing analyses of all five digits the exact digit would be 1204.

This modification would mostly use the Digits Recognition Algorithm described in Figure 8 to recognize only two digits instead of five digits all the time. Thus, it would improve the performance speed by up to 60%.

Data Alignment and Analysis

The next step after detecting the start and end frames of each trial would be to mark the rows in the eye movement data file according to the trial number. After this step, the data are ready for analysis using any eye tracking analysis program which is capable of handling text files as data input. An example of such analysis software is ILAB (Gitelman, 2002).

After obtaining the fixation output file, we use it to calculate the fixations on predefined areas of interest (AOI). The idea is simple: detect the coordinates of each fixation, and then determine in which area of interest it exists. This is done per trial because we are interested in analyzing each trial separately.

Future Directions

In the system solution proposed, we used MATLAB, which is known to be simple and powerful though relatively slow in comparison to other programming languages. Because of this, we believe that migrating from MATLAB to a faster programming language (such as C/C++) would be an important improvement and would save time in executing the steps.

Another important future development would be to convert the software to a real-time application so that it detects the trials online. This could also lead to development of software programs that could control the stimulus according to the eye movements of the subject.

References

Cohen, L.B., Atkinson, D.J., and Chaput, H. H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas.

Gitelman, D. R., (2002). ILAB: A program for post experimental eye movement analysis. Behavior Research Methods, Instruments and Computers, 34(4), 605-612.

Gonzalez, R. C., and Woods, R. E. (2000). Digital Image Processing. 2nd Edition, Prentice Hall.

Jacob, R. J. K., (1991). The use of eye movements in human – computer interaction techniques: what you look at is what you get. ACM Transactions on Information Systems 9 (3), 152–169.

Jacob, R. J. K., (1995). Eye tracking in advanced interface design. In Barøeld, W., & Furness, T. (Eds.), Advanced Interface Design and Virtual Environments, 258-288. Oxford: Oxford University Press.

Pelz, B. J., Canosa, L. R., Kucharczyk, D., Babcock, J., Silver, A. and Konno, D., (2000). Portable Eyetracking: A Study of Natural Eye Movements. Proceedings of the SPIE, Human Vision and Electronic Imaging, San Jose, CA: SPIE.

Young, L., and Sheena, D., (1975). Survey of eye movement recording methods. Behavior Research Methods and Instrumentation 7, 397–429.

IV. Publications

JOURNAL ARTICLES PUBLISHED:

Altieri, N., Gruenenfelder, T., & Pisoni, D.B. (2010). Clustering coefficients of lexical neighborhoods: Does neighborhood structure matter in spoken word recognition? *Mental Lexicon* 5(1), 1-21.

Bradlow, A. R. and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*. 106 (2), 707-729.

Beer, J., Pisoni, D.B., & Kronenberger, W. (2009). Executive Function in Children with Cochlear Implants: The Role of Organizational-Integrative Processes. *Volta Voices* 16(3), 18-21.

Bent, T., Buchwald, A., & Pisoni. D.B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *Journal of the Acoustical Society of America*, 126 (5), 2660–2669.

Bent, T., Bradlow, A. R., and Smith, B.L. (2008). Production and Perception of Temporal Patterns in Native and Non-Native Speech. *Phonetica*. 65 (3), 131-147.

Bradlow, A. R. and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*. 106(2), 707-729.

Buchwald, A. (2009). Minimizing and optimizing structure in phonology: Evidence from aphasia. *Lingua*, 119(10), 1380-1395.

Buchwald, A., Winters, S.J., & Pisoni, D.B. (2009). Visual speech primes open-set recognition of auditory words. *Language and Cognitive Processes*, *24(4),* 580-610.

Buchwald, A., Rapp, B., and Stone, M. (2007). Insertion of discrete phonological units: An ultrasound investigation of aphasic speech. *Language and Cognitive Processes*, 22(6), 910-948.

Casserly, E. & Pisoni, D.B. (2010). Speech perception & production. In L. Nadel (Ed.), Wiley Interdisciplinary Reviews: Cognitive Science. *John Wiley & Sons*, 1(5), 629-647.

Conway, C. M., Bauernschmidt, A., Huang, S.S. & Pisoni, D.B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356-371.

Conway, C.M. & Christiansen, M.H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, 21, 561-580.

Conway, C.M. & Pisoni, D.B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, 1145, 113-131.

Conway, C.M., Pisoni, D.B., & Kronenberger, W.G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science*, 18(5), 275-279.

Fagan, M. K., & Pisoni, D.B. (2009). Perspectives on multisensory experience and cognitive development in infants with cochlear implants. *Scandinavian Journal of Psychology*, 50(5), 457-462.

Fagan, M. K. & Pisoni, D. B. (2010). Hearing experience and receptive vocabulary development in deaf children with cochlear implants. *Journal of Deaf Studies and Deaf Education*, 15(2), 149-161.

Felty, R., Buchwald, A. & Pisoni, D.B. (2009). Adaptation to frozen babble in spoken word recognition. *Journal of the Acoustical Society of America – Express Letters*, 125(3), EL93-EL97.

Gruenenfelder, T.M., & Pisoni, D.B. (2009). The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons. *Journal of Speech, Language, and Hearing Research*, 52(3), 596-609

Harnsberger, J.D., Pisoni, D.B. & Wright, R. (2008). A new method for eliciting three speaking styles in the laboratory, *Speech Communication*, 50(4), 323-336.

Hay-McCutcheon, M. J., Pisoni, D.B., & Hunt, K.K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. *International Journal of Audiology*, 48(6), 321-333.

Hayes-Harb, R., Smith, B. L., Bent, T., and Bradlow, A. R. (2008). Production and Perception of Final Voiced and Voiceless Consonants by Native English and Native Mandarin Speakers: Implications Regarding the Interlanguage Speech Intelligibility Benefit. *Journal of Phonetics*, 36(4), 664-679.

Kapatsinski, V. (2008a.) Constituents can exhibit partial overlap: Experimental evidence for an exemplar approach to the mental lexicon. *Chicago Linguistic Society 41: The Panels*, 227-242.

Kapatsinski, V. (2008). Rethinking rule reliability: Why an exceptionless rule can fail. *Chicago Linguistic Society,* 44 (2), 277-291.

Kapatsinski, V. (2009). Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language*, 85(2), 248-77

Kapatsinski, V. (2009). Adversative conjunction choice in Russian: Semantic and syntactic influences on lexical selection. *Language Variation and Change*, 21(2), 157-73

Kapatsinski, V., & Radicke. J.(2009). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley, eds. *Formulaic Language. Vol. II: Acquisition, loss, psychological reality, functional explanations*. Amsterdam: John Benjamins. (Typological Studies in Language 83).

Kapatsinski, V. (2010). Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Language and Speech*, 53(1), 71-105.

Loebach, J. L., Bent, T. and Pisoni, D. B. (2008). Multiple routes to perceptual learning. *Journal of the Acoustical Society of America*, 124 (1), 552-561.

Loebach, J.L. & Pisoni, D.B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *Journal of the Acoustical Society of America*, 123(2), 1126-1139.

Loebach, J. L., Pisoni, D.B., & Svirsky, M.A. (2009). Transfer of Auditory Perceptual Learning with Spectrally Reduced Speech to Speech and Nonspeech Tasks: Implications for Cochlear Implants. *Ear & Hearing*, 30(6), 662-674.

Loebach, J.L., Pisoni, D.B. & Svirsky, M.A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 224-234.

Hay-McCutcheon, M. J., Pisoni, D.B., & Hunt, K.K. (2009). Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis. *International Journal of Audiology*, 48(6), 321-333.

Peterson, N.R., Pisoni, D.B., & Miyamoto, R.T. (2010). Cochlear Implants and Spoken Language Processing Abilities: Review and Assessment of the Literature. *Restorative Neurology and Neuroscience*, 28(2), 237-250.

Ronquest, R.E., Levi, S.V., & Pisoni, D.B. (2010). Language Identification from Visual-only Speech Signals. *Attention, Perception, & Psychophysics*, 72(6), 1601-1613.

Stevenson, R.A., Altieri, N.A., Kim, S., Pisoni, D.B. & James, T.W. (2010). Neural processing of asynchronous audiovisual speech perception. *Neuroimage*, 49(4), 3308-3318.

Tierney, A.T., Bergeson-Dana, T., & Pisoni, D.B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical Musicology Review*, 3(4), 217-186.

Tierney, A.T., Bergeson, T.R., & Pisoni, D.B. (2009). General intelligence and modality-specific differences in performance: a response to Schellenbert (2008). *Empirical Musicology Review*, Vol. 4, 217-186, 37-39.

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America*, 123(6), 4524-4538.

BOOK CHAPTERS PUBLISHED:

Conway, C.M., Loebach, J.L., & Pisoni, D.B. (2009). Speech perception. In B. Goldstein (Ed.), Encyclopedia of Perception, Los Angeles, CA: *SAGE Publications, Inc*., 918-923.

Loebach, J.L., Conway, C.M., & Pisoni, D.B. (2009). Audition: Cognitive influences. In B. Goldstein (Ed.), Encyclopedia of Perception, Los Angeles, CA: *SAGE Publications, Inc*., 138-141.

Pisoni, D.B., Conway, C.M., Kronenberger, W., Horn, D.L., Karpicke, J. & Henning, S. (2008). Efficacy and effectiveness of cochlear implants in deaf children. In M. Marschark & P. Hauser (Eds.), *Deaf Cognition: Foundations and Outcomes*. New York: Oxford University Press, 52-10.

Pisoni, D.B., Conway, C.M., Kronenberger, W., Henning, S. & Anaya, E. (2010). Executive function, cognitive control and sequence learning in deaf children with cochlear implants. In M. Marschark & P. Spencer (Eds), *Oxford Handbook of Deaf Studies, Language, and Education*, 439-457.

PROCEEDINGS PUBLISHED:

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2008). A cross-language familiar talker advantage? *Acoustics 08-Paris: Proceedings of the Acoustical Society of America meeting/Euronoise*, Paris. 2435-2439.

Taler, V., Grove, L.M., Aaron, G.A., & Pisoni, D.B. (2009). Lexical competition and spoken word recognition in aging and MCI: preliminary findings. *TENNET*, Montreal, Quebec, Canada.

Taler, V., Saykin, A.J., Wishart, H., Pisoni, D.B., Flashman, L.A., Rabin, L.A., Nutter-Upham, K.E., & Pare, N. (2008). Neuropsychological correlates of fluency measures in older adults with cognitive complaints and amnestic MCI. *46th Annual Meeting of the Academy of Aphasia*. Turku, Finland.

MANUSCRIPTS ACCEPTED FOR PUBLICATION (IN PRESS):

Bent, T. Loebach, J.L. Phillips, L., & Pisoni, D.B. (In Press). Perceptual adaptation to sinewave-vocoded speech across languages. *Attention Perception and Psychophysics*.

Conway, C.M., Karpicke, J., Anaya, E.M., Henning, S.C., Kronenberger, W.G., & Pisoni, D.B. (In Press). Nonverbal cognition in deaf children following cochlear implantation: Motor sequencing disturbances mediate language delays. *Developmental Neuropsychology*.

Conway, C.M., Pisoni, D.B., Anaya, E.M. Karpicke, J., & Henning, S.C. (In Press). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science.*

Geers, A.E., Strube, M.J., Tobey, E.A., Pisoni, D.B. & Moog, J.S. (In Press). Epilogue: Factors Contributing to Long-Term Outcomes of Cochlear Implantation in Early Childhood. *Ear & Hearing*

Kapatsinski, V. (In Press). Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology.*

Pisoni, D.B., Kronenberger, W.G., Roman, A.S. & Geers, A.E. (In Press). Measures of Digit Span and Verbal Rehearsal Speed in Deaf Children After More than 10 Years of Cochlear Implant Use. *Ear & Hearing.*

Taler, V., Aaron, G.P., Steinmets, L.G., & Pisoni, D.B. (In Press). Lexical Neighborhood Density Effects on Spoken Word Recognition and Production in Healthy Again. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*.

MANUSCRIPTS SUBMITTED:

Altieri, N., Pisoni, D.B., & Townsend, J. (Submitted). Some normative data on speech-reading skills. *Journal of the Acoustical Society of America*.

Bent, T. Loebach, J.L. Phillips, L., & Pisoni, D.B. (Under Revision). Perceptual adaptation to sinewave-vocoded speech across languages. *Attention Perception and Psychophysics*.

Conway, C.M., Karpicke, J., Anaya, E.M., Henning, S.C., Kronenberger, W.G., & Pisoni, D.B. (Submitted). Nonverbal cognition in deaf children following cochlear implantation: Motor sequencing disturbances mediate language delays. *Developmental Neuropsychology*.

Dillon, C.M., deJong, K. & Pisoni, D.B. (Submitted). Phonological awareness, reading skills and vocabulary knowledge in deaf children who use cochlear implants. *Journal of Deaf Studies and Deaf Education*.

Dillon, C.M., deJong, K. & Pisoni, D.B. (Submitted). Phonological processing and reading in deaf children with cochlear implants. *Cognition.*

Dillon, C.M., Pisoni, D.B. & deJong, K. (Submitted). Nonword factors and vocabulary effects in nonword repetition and reading skills of children with cochlear implants. *Applied Psycholinguistics*.

Geers, A.E., Strube, M.J., Tobey, E.A., Pisoni, D.B. & Moog, J.S. (Submitted). Epilogue: Factors Contributing to Long-Term Outcomes of Cochlear Implantation in Early Childhood. *Ear & Hearing.*

Kitano, K., Nichols, T.M., Britton, R.A., Pisoni, D.B. & Koceja, D.M. (Submitted). Executive function and postural control in the elderly. *Current Gerontology and Geriatrics Research*.

Kronenberger, W.G., Pisoni, D.B., Henning, S.C., Colson, B.G. & Hazzard, L.M. (Submitted). Working Memory Training Improves Memory Capacity and Sentence Repetition Skills in Deaf Children with Cochlear Implants: A Pilot Study.

Pisoni, D.B., Kronenberger, W.G., Roman, A.S. & Geers, A.E. (Submitted). Measures of Digit Span and Verbal Rehearsal Speed in Deaf Children After More than 10 Years of Cochlear Implant Use.

Radicke, J.L., Levi, S.V., Loebach, J.L. & Pisoni, D.B. (Submitted). Audiovisual phonological fusion. *Perception & Psychophysics*.

Taler, V., Pisoni, D.B., Farlow, M.R., Hake, A.M., Kareken, D.A. & Unverzagt, F.W. (Submitted). Reduced cluster switching in category fluency reveals cognitive decline: A longitudinal study. *Journal of the International Neuropsychological Society*.

Taler, V. & Pisoni, D.B. (Submitted). Effects of talker-specific encoding on recognition memory for spoken sentences. *Memory*.