Published in final edited form as:

Audiol Med. 2007 December 1; 5(4): 250–261. doi:10.1080/16513860701673892.

Audiovisual Spoken Word Recognition by Children with Cochlear Implants

Karen Iler Kirk^{1,2}, Marcia J. Hay-McCutcheon², Rachael Frush Holt³, Sujuan Gao⁴, Rong Qi⁴, and Bethany L. Gehrlein²

¹ Dept. of Speech, Language and Hearing Sciences, Purdue University, West Lafayette, Indiana USA

² Dept. of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine Indianapolis, Indiana USA

³ Dept. of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana USA

⁴ Dept. of Medicine, Section of Biostatistics, Indiana University School of Medicine Indianapolis, Indiana USA

Abstract

This study examined how prelingually deafened children with cochlear implants combine visual information from lipreading with auditory cues in an open-set speech perception task. A secondary aim was to examine lexical effects on the recognition of words in isolation and in sentences. Fifteen children with cochlear implants served as participants in this study. Participants were administered two tests of spoken word recognition. The LNT assessed isolated word recognition in an auditory-only format. The AV-LNST assessed recognition of key words in sentences in a visual-only, auditory-only and audiovisual presentation format. On each test, lexical characteristics of the stimulus items were controlled to assess the effects of lexical competition. The children also were administered a test of receptive vocabulary knowledge. The results revealed that recognition of key words was significantly influenced by presentation format. Audiovisual speech perception was best, followed by auditory-only and visual-only presentation, respectively. Lexical effects on spoken word recognition were evident for isolated words, but not when words were presented in sentences. Finally, there was a significant relationship between auditory-only and audiovisual word recognition and language knowledge. The results demonstrate that children with cochlear implants obtain significant benefit from audiovisual speech integration, and suggest such tests should be included in test batteries intended to evaluate cochlear implant outcomes.

Keywords

hearing impairment; speech perception; assessment; cochlear implant

INTRODUCTION

Cochlear implants (CIs) are surgically implanted medical devices designed for individuals with hearing loss who do not receive adequate benefit from conventional hearing aids. CIs bypass damaged or missing hair cells in the cochlea and directly stimulate surviving neural elements in the auditory nerve. In 1990, CIs were approved by the United States Food and Drug

Administration (FDA) for use in children with profound hearing loss aged 2–17 years (1). Since then, there have been substantial technological advances in CI systems, with a concomitant broadening of pediatric candidacy criteria. Current FDA guidelines now permit the implantation of children with profound hearing loss as young as 12 months of age, and children with severe-to-profound hearing loss as young as 24 months of age. Over the last decade, advances in the design and implementation of CI systems and changes in candidacy criteria have led to large improvements in speech perception and spoken word recognition obtained by children following cochlear implantation. Today the expectations for speech and language development in children who receive a CI are higher than ever. However, enormous variability in outcomes still exists. Whereas many children with CIs can use auditory input from their devices to acquire age-appropriate communication skills, others demonstrate very limited auditory and language skills.

Fortunately, everyday speech communication is not limited to input from only one sensory modality. Speech has both an auditory component, the acoustic waveform, and a visual component, the visible articulatory gestures generated during speech production (e.g., lip rounding). Visual cues are particularly important because they help to specify place of articulation, a speech feature that often is not accessible to many individuals with significant hearing loss (2). Sumbly and Pollack (3) were the first researchers to document that the addition of visual speech cues to the auditory speech signal yielded substantial improvements in speech perception for listeners with normal hearing, especially in adverse listening conditions such as noise. Similar results have been obtained by other investigators for listeners with normal hearing (4) (5) and for listeners with hearing loss (6) (7).

Auditory-Visual Speech Integration

The cognitive processes by which individuals combine and integrate auditory and visual speech information with lexical and syntactic knowledge have become an important area of research in the field of speech perception. It has been hypothesized that visual cues enhance speech understanding because they provide segmental and suprasegmental information that is complementary (i.e., place of articulation) to the acoustic speech cues, and because they reduce the attentional demands placed on the auditory signal (8). Audiovisual spoken word recognition therefore appears to be more than the simple addition of auditory and visual information (9). A well-known example of the robustness of audiovisual spoken word recognition is the “McGurk” effect (10). When presented with an auditory stimulus/ba/ and a visual stimulus/ga/, many listeners report perceiving/da/. Thus, information from separate sensory modalities can be combined to produce percepts that differ predictably from either the auditory or visual percept alone. However, substantial individual variability is noted in the ability to integrate the two types of speech cues (11). One factor that appears to influence auditory-visual, or audiovisual integration is degree of hearing loss. Individuals with lesser degrees of hearing loss obtain better auditory-only and audiovisual performance than individuals with greater degrees of hearing loss (12). The nature of a listener’s early linguistic experience also influences audiovisual speech perception. Individuals with an earlier onset of deafness demonstrate better visual-only spoken word recognition than listeners with a late onset of deafness (13). The ability to combine and integrate auditory and visual speech information has been found to be an important predictor of speech perception benefit with a sensory aid (14–16).

Lexical Effects on Spoken Word Recognition and Audiovisual Speech Integration

Speech perception is a complex process in which information from separate auditory and/or visual sensory modalities is combined with prior linguistic knowledge stored in long-term memory. The Neighborhood Activation Model (NAM) of Luce and Pisoni (17) offers a two stage account of how the sound pattern of words in memory contribute to the perception of

spoken words. According to NAM, a stimulus input activates a set of similar acoustic-phonetic patterns in memory in a multi-dimensional acoustic-phonetic space, with activation levels proportional to the degree of similarity to the target word. This initial activation stage is followed by “lexical selection” among a large number of potential candidates that are consistent with the acoustic-phonetic input. Word frequency is assumed to act as a biasing factor in this model by multiplicatively adjusting the activation levels of the acoustic-phonetic patterns. In lexical selection, the activation levels are then summed and the probability of choosing each pattern is computed based on the overall activation level. Word recognition occurs when a given acoustic-phonetic representation is chosen based on the computed probabilities. Luce and his colleagues (18) showed that word frequency (i.e., how often words occur in the language) (19) and neighborhood density (i.e., the number of similar sounding words or “lexical neighbors” for the target word) influence auditory-only word recognition. Lexically “easy” words (high frequency words from sparse neighborhoods) were recognized with greater accuracy and speed than lexically “hard” words (low frequency words from dense neighborhoods). These results demonstrate that the structure and organization of sound patterns in memory affects spoken word recognition performance by listeners with normal hearing.

Over the last decade, Kirk and colleagues have investigated lexical effects on auditory-only spoken word recognition by children with CIs using two tests created specifically for this purpose. Based on computational analyses, word lists were constructed to allow systematic examination of the effects of word frequency, lexical density and word length. The resulting Lexical Neighborhood Test (20) consists of two lists of 50 monosyllabic words, whereas the Multisyllabic Lexical Neighborhood Test (MLNT) consists of two lists of 24 two- or three-syllable words. On each list, half of the words are lexically easy and half are lexically hard. In the first of a series of studies, Kirk et al. (20) found that children with cochlear implants recognized lexically easy words with significantly greater accuracy than lexically hard words. These early results demonstrated that pediatric CI users are sensitive to the acoustic-phonetic similarity among words, that they organize words into similarity neighborhoods in long-term memory, and that they use this structural information in recognizing words in an open-set response format.

In a follow-up study, Eisenberg and colleagues (21) created lexically controlled sentences from the subset of words used to develop the LNT and MLNT. One list contained sentences with lexically easy key words; the other list contained key words that were lexically hard. Eisenberg et al. (21) then used these sentences to investigate the effects of word frequency and lexical density on children’s recognition of isolated words and the same words produced in a sentence context. Children also completed a test of vocabulary recognition. Participants included children with normal hearing and children with CIs. The study revealed that auditory-only word recognition was influenced both by the lexical characteristics of the stimulus items and by the linguistic context in which they were presented. Sentences containing lexically easy key words were recognized better than sentences containing lexically hard key words. Children with normal hearing and high-performing children with CIs were better at recognizing key words in a sentence context than in isolation. Pediatric CI users with poor auditory-only recognition skills showed the opposite pattern. That is, word recognition scores were higher for isolated words than for key words presented in sentences. For children with CIs, the relationship between language abilities and spoken word recognition was strong and significant. This result demonstrates the influence of linguistic knowledge on phonological processing of words.

At present, little is known about the contribution of the lexicon and knowledge of sound patterns of words to audiovisual speech integration by individuals with cochlear implants. Kaiser et al. (22) examined lexical effects on isolated word recognition in adult CI users with postlingual deafness. Their performance was assessed using lexically controlled word lists presented under

auditory-only, visual-only, and combined audiovisual presentation formats. Kaiser et al. (22) found that word recognition performance was highest for audiovisual presentation followed by auditory-only and then visual-only formats, respectively. Word recognition performance was better for lexically easy words than for lexically hard words, regardless of presentation format. The pattern of results suggested that information from the auditory and visual modalities is used to access common, multimodal lexical representations in memory. However, participants in the Kaiser et al. (22) study acquired language with an intact auditory system and lost their hearing later in life. In contrast, the majority of children with cochlear implants are congenitally or prelingually deafened; they must use the degraded auditory information provided by a CI to learn the spoken language of their surrounding communities and to encode and access multimodal lexical representations in memory.

Purpose of the Present Investigation

The purpose of the present investigation was to examine audiovisual integration in early implanted children with prelingual deafness. A secondary goal was to evaluate the effects of lexical difficulty on audiovisual speech integration in this same population of participants.

MATERIALS AND METHODS

Participants

This was an analysis of data collected as part of a prospective longitudinal study of pediatric cochlear implant outcomes being conducted at the Indiana University School of Medicine. The procedures followed were in accordance with the ethical standards of the Indiana University Purdue University Indianapolis Institutional Review Board (IUPUI IRB) and the Helsinki Declaration of 1975, as revised in 1983. The longitudinal study has been approved by the IUPUI IRB (Protocol No. 8606-11). The participants in this investigation were administered a battery of speech and language tests prior to cochlear implantation, and again at 6- or 12-month intervals postimplantation. In order for their data to be included in the current study, participants had to meet the following inclusion criteria: 1) congenital or prelingual onset of hearing loss; 2) three-frequency pure tone average threshold (PTA) in the severe-to-profound range in the ear to be implanted; 3) implantation prior to age 3 years; 4) user of current CI technology; 5) native speaker of English; 6) no disabilities in addition to hearing loss; and 7) completion of a test of an audiovisual word recognition test that recently was added to the longitudinal test battery. Fifteen children met these criteria.

Participant characteristics are presented in Table 1. Best PTA refers to thresholds in the better hearing ear. Four children had thresholds in the moderate or severe range in their non-implanted ear. The remaining children had a bilateral, profound hearing loss. The mean age at which the children received their CI processor was 1.91 years (range: 0.86 to 2.98 years). The mean age at time of testing was 5.89 years (range: 3.8 to 9.0 years). Thus, the mean duration of CI use at the time of testing was 3.97 years (range: 1.0 to 6.9 years). Device condition refers to the sensory aid configuration employed by the children in their daily activities. Thirteen of the children were fitted with a monaural CI. One child had bilateral CIs and one child used a hearing aid on the non-implanted ear in conjunction with a monaural CI. Finally, 14 of the children used Oral Communication and one child used Total Communication (the combined use of signed and spoken English).

Stimulus Materials and Testing Procedures

Data from three different tests were analyzed in this investigation of audiovisual speech integration and lexical influences on spoken word recognition. Participants were tested using the sensory aid configuration they employed in daily activities.

Speech Perception Tests—Two tests of spoken word recognition were included. As noted above, the Lexical Neighborhood Test is an open-set test of auditory-only word recognition consisting of two lists of 50 monosyllabic words. On each list, half of the words are lexically easy and half are lexically hard. An audio-recorded, five-talker version of the test was used. Each of the talkers spoke a General American dialect of English.

The Audiovisual-Lexical Neighborhood Sentence Test (AV-LNST) (23) is an audiovisually recorded test derived from the Lexical Sentence Test developed by Eisenberg et al. (21). The AV-LNST sentences were produced by a professional Caucasian female announcer who spoke a General American dialect of English in a conversational speaking style. The sentences developed by Eisenberg et al. (21) are syntactically correct but semantically neutral (low in word predictability). The AV-LNST contains six lists of eight sentences that can be administered in three different presentation formats: visual-only (V-only), auditory-only (A-only), and audiovisual (AV). The sentence lists are equally difficult within each presentation format. Sentences are five-to-seven words in length and contain three key words with controlled lexical characteristics. In each sentence list of eight sentences, four sentences contain lexically easy key words and four sentences contain lexically hard key words. This results in 24 scored items per list. The remaining unscored words in each sentence are lexically neutral. Both spoken word recognition tests were administered with the participants seated inside a sound-treated booth approximately one meter from a computer monitor at zero degrees azimuth. Stimulus presentation was via computer control. Auditory stimuli were presented via loudspeaker; due to differences in methods over time, the LNT was administered at approximately 70 dB SPL whereas AV-LSNT stimuli were presented at approximately 65 dB SPL. During A-only testing, the computer monitor was black. During V-only testing, participants saw the talker on the computer monitor but heard no sound. During AV testing, participants saw the talker and heard the auditory stimuli. All testing was open-set. Participants responded by repeating the word they heard. The participant who used Total Communication also was allowed to sign the word. Responses were scored on-line by an examiner seated in the booth with the child. If the examiner was unsure of the child's intended response, the child was prompted to provide more information (e.g., "What is that?" or "Can you tell me more?"). If a response could not be clarified in this way, it was scored as incorrect. Responses were scored as the percent of words correct as a function of lexical difficulty.

Receptive Vocabulary Test—Previous work has demonstrated that vocabulary knowledge influences spoken word recognition by children with cochlear implants (21) (24). Therefore, data from the Peabody Picture Vocabulary Test, a measure of receptive vocabulary knowledge, were included in our analyses. The test consists of a series of four-picture plates. The examiner says one word per page and the child is asked to point to the picture that best represents the word. Children who used Oral Communication were allowed to see and hear the examiner. The child who used Total Communication was provided with the sign representing the word in addition to receiving auditory and visual speech cues from the examiner. Testing was carried out in a quiet therapy room with both the child and the examiner seated at a small table. The examiner sat facing the child so that the examiner's face was clearly visible during testing.

RESULTS

Fourteen of the children included in our retrospective analyses were administered the AV-LNST at one postimplant interval. The remaining child was administered the AV-LNST at two successive postimplant intervals (approximately 4.0 and 4.5 years postimplant). For this child, only data from the later interval were included in the analyses (4.5 years postimplant). Descriptive statistics were calculated for all variables of interest.

Auditory-Visual Speech Integration

Figure 1 illustrates the participant's mean performance on the new audiovisual speech recognition task, the AV-LNST. In this figure, the percent of lexically easy key words correctly identified in a sentence context is shown on the left. The right side of the figure shows mean recognition of lexically hard key words in sentences. For the lexically easy words, the mean percent correct scores were 5%, 60% and 67%, respectively, for the V-only, A-only and AV presentation formats. Performance was slightly poorer for lexically hard words, with mean scores of 1%, 53% and 66%, respectively, for the V-only, A-only and AV presentation formats.

Figure 2 presents individual data for recognition of words in sentences. The top panel represents recognition of lexically easy key words and the bottom panel depicts performance on the lexically hard key words. Recognition of key words in the V-only presentation format was very poor, with scores ranging from 0–33% for easy words and from 0–8% correct for hard words. In fact, very few participants achieved scores greater than 0% correct using visual speech information alone. Only four participants achieved V-only scores greater than 0% correct for lexically words, and just two participants scored greater than 0% correct on V-only lexically hard words. When the participants were provided with auditory speech cues in the A-only presentation format, their performance improved. There was a great deal of individual variability in the A-only presentation format, with scores ranging from 0–100% correct for both lexically easy and hard words. A majority of the participants recognized key words with greater than 60% accuracy using auditory speech cues alone. The highest word recognition scores were obtained when children were provided with both auditory and visual speech cues, although large individual variability was evident. Scores for key word recognition in the AV presentation format ranged from 8% to 100%. The participants appeared to benefit from the integration of auditory and visual speech information. For the lexically easy key words, nine of 12 participants achieved higher scores in the AV presentation format compared to the A-only presentation format. (Three achieved scores of 100% correct in the A-only condition.) Of the 13 participants who did not recognize the lexically hard words with 100% accuracy in the A-only presentation format, nine achieved higher scores when they had access to both auditory and visual speech cues.

A mixed model was used to examine the effects of presentation format and lexical difficulty on the AV-LNST. Fixed factors were presentation format (V-only, A-only and AV), lexical difficulty (Easy, Hard), the interaction of presentation format and lexical difficulty, age at CI fitting, age at testing and pure tone average thresholds. Participant served as a random factor in this model. The mixed model did not detect a significant interaction between presentation format and lexical difficulty ($p = .78$). After controlling for age at CI fitting, age at testing and pure tone average threshold, there was no significant effect of lexical difficulty on identification of key words on the AV-LNST. However, there was a significant main effect of presentation format ($p < .0001$). Scores were highest in the AV presentation format, followed by the A-only format and the V-only format. Performance in the V-only format did not differ significantly from 0% correct. The adjusted mean scores were 67%, 57% and 0% for the AV, A-only and V-only presentation formats, respectively. Post hoc analyses revealed that scores in the AV condition were significantly higher than scores in either the A-only ($p < .05$) or the V-only presentation format ($p < .0001$). Furthermore, scores in the A-only format were significantly higher than scores in the V-only presentation format ($p < .0001$).

Auditory-only Isolated Word Recognition

Figure 3 depicts the participant's performance on the isolated test of spoken word recognition, the LNT. Recall that this test is given in an auditory-only presentation format. Mean percent correct word recognition scores were 66% for lexically easy words and 52% for lexically hard words. Individual scores for the LNT ranged from 8–92% correct for the lexically easy words,

and from 0–92 % correct for the lexically hard words. A mixed model was used to examine the effects of lexical difficulty on the children's performance on the LNT. Fixed factors were lexical difficulty (Easy, Hard), age at CI fitting, age at testing, and pure tone average threshold. Participant served as a random factor. The mixed model detected a significant association between auditory-only word recognition on the LNT and the lexical difficulty of the test items. After controlling for age at CI fitting, age at testing and pure tone average threshold, easy words were recognized with significantly greater accuracy than hard words ($p < .0001$). The adjusted mean LNT scores for Easy and Hard words were 65.7% and 52.4%, respectively.

Word Recognition in Isolation and in Sentence Contexts

Overall, the average scores for auditory-only recognition were similar for words presented in isolation and for words in a sentence context. However, large individual differences were noted in the children's performance on each test. Following Eisenberg et al (21), we elected to examine the data separately for individuals with good and poor auditory-only word recognition. The median auditory-only word recognition score on the AV-LNST (67% for both lexically easy and hard key words) was used to separate the participants into groups of high and low performers. Figure 4 presents the mean auditory-only word recognition scores for words presented in isolation and in a sentence context for the two groups of participants. The high performing participants obtained higher scores on key words presented in sentences than on isolated words. The high performers obtained average word and sentence scores of 78% and 86%, respectively, for easy words, and 67% and 78% for hard words. In contrast, the poorer performers showed the opposite pattern. Their word recognition performance was better for isolated words than for words presented in sentences. Their mean word and sentence scores were 50% and 31%, respectively, for easy words and 36% and 24%, respectively, for hard words.

Pearson Correlation Coefficients were calculated to measure the correlation between performance on the isolated word recognition test (the LNT) and the recognition of key words in sentences (the AV-LNST). Table 2 presents the results of the correlational analysis between the participants' performance on the AV-LNST and the LNT. After adjusting for age at CI fitting, age at testing and pure tone average threshold, it was found that visual-only recognition of key words in sentences was not correlated with recognition of isolated words presented in an A-only format. However, the participants' performance on the Auditory-only and the AV presentation formats of the AV-LNST was strongly correlated with recognition of isolated words through listening alone ($r > 0.74$; $p < .0001$). This was true for both easy and hard words.

Receptive Vocabulary Knowledge and Spoken Word Recognition

Figure 5 shows the relationship between the participants' receptive vocabulary knowledge and their recognition of key words on the AV-LNST in the A-only and AV presentation formats. Data obtained from the V-only condition are not illustrated because of floor effects. Figure 6 shows the relationship between vocabulary knowledge and the auditory-only recognition of isolated words on the LNT. In both of the figures, performance on the Easy and Hard words is plotted separately. Partial correlations were used to test the association between receptive vocabulary skills and spoken word recognition while controlling for other factors known to affect spoken word recognition (age fit with a CI, age at testing, and pure tone average threshold). V-only scores on the AV-LNST were included in the analyses. After controlling for the other factors, partial correlations revealed a significant association between receptive vocabulary knowledge, as measured by the PPVT, and recognition of easy AV-LNST words in the A-only ($r = 0.67$; $p = .018$) and AV ($r = 0.74$; $p = .006$) presentation formats. Similarly, there was a significant correlation between PPVT scores and recognition of hard words on the AV-LNST in the A-only ($r = .71$; $p = .009$) and in the AV ($r = 0.64$; $p = .026$) presentation formats. There was no significant correlation between PPVT scores and performance on the

AV-LNST In the V-only presentation format. For the LNT, recognition of hard words through listening alone was significantly associated with receptive vocabulary knowledge ($r = 0.43$; $p = .03$), but recognition of easy words was not.

DISCUSSION

Auditory-Visual Speech Integration

The present study revealed that access to both auditory and visual speech cues yields substantial perceptual benefit to children with CIs. When administered a recorded, open-set test of sentence recognition, the participants were best at identifying key words in the AV presentation format, followed by the A-only and V-only presentation formats, respectively. This finding is consistent with previous studies concerning audiovisual spoken word recognition in adult CI users tested with isolated words (22) and in pediatric CI recipients tested with sentences presented via live voice (16). Mean scores for the AV and A-only presentation formats were generally quite high. However, there was large amount of individual variability in performance in each condition. Some participants with relatively poor speech perception abilities through listening alone showed great improvement when both auditory and visual speech cues were available. This suggests that auditory-only word recognition tests may underestimate a child's communication abilities in the real world. Auditory-only word recognition tests serve as the gold standard for evaluating cochlear implant outcomes and they should remain an important part of the test battery. However, the present results suggest that spoken word recognition also should be routinely tested in an audiovisual format which may better represent daily listening demands.

Lexical Effects on Spoken Word Recognition

The lexical characteristics of the stimulus items significantly influenced the recognition of isolated words through listening alone. That is, the CI recipients in this study were more accurate at recognizing easy words than hard words. This finding is in agreement with previous studies (20) (22). The present results support the notion that pediatric CI users are sensitive to the acoustic-phonetic similarity among words, that they organize words into similarity neighborhoods in long-term memory, and that they use this structural information in recognizing isolated words in an open-set response format.

Although there was a trend for the easy words on the AV-LNST to be recognized more accurately than hard words, this trend was not significant. Furthermore, there was no interaction between presentation format and lexical difficulty of the AV-LNST stimulus items. The lack of a lexical effect for key words in a sentence conflicts with the results of Eisenberg et al (21), who examined lexical effects on the identification of key words using listening alone. Our results differed despite the fact that the current study used the sentences created by Eisenberg et al. (21). There are several possible reasons for the divergent results. One potential explanation is that different talkers produced the sentences in each study. Talker characteristics have been shown to influence spoken word recognition in listeners with hearing loss (25). However, this explanation is unlikely, given that the current CI recipients demonstrated similar auditory-only key word recognition as the children in the previous study. The differences in findings also may be due to the fact that the two investigations employed different data analysis techniques. Eisenberg and her colleagues used analysis of variance to examine the effects of presentation format and lexical difficulty. In contrast, the present investigation employed a mixed model that controlled for other factors known to influence word recognition performance before examining the effects of presentation format and lexical difficulty. Finally, the divergent results may be accounted for by the way in which the stimuli were presented in each investigation. Whereas Eisenberg et al. (21) administered the sentences in an A-only format alone, the present investigation employed three different presentation formats. It may be that

floor effects in the V-only and ceiling effects in the AV presentation formats reduced the overall effects of lexical competition on word recognition. Future studies using competing noise when presenting auditory or audiovisual stimuli may help to clarify the discrepancies in findings between our results and those of the earlier investigators.

Our results were similar to those of Eisenberg et al. (21) in at least one respect. Like those investigators, we found that high-performing children could use contextual cues to improve word recognition. That is, the percent of words correctly recognized was higher in sentence contexts than when words were presented in isolation. However, both studies revealed the opposite pattern for CI recipients with poor recognition scores. This finding suggests that the low-performing children may process connected speech as strings of unrelated words. As Eisenberg et al. (21) pointed out, such a perceptual processing strategy would place increased demands on short-term memory and may result in reduced performance.

The Relationship Between Spoken Word Recognition and Language Skills

Overall, our results suggest that language knowledge is significantly related to auditory-only and audiovisual open-set word recognition in this population of children with CIs. When listeners with normal hearing are presented with an auditory speech signal, they can be expected to repeat words correctly, even if they do not know them. When the speech signal is degraded, as with cochlear implant speech processing, listeners must recognize words with limited auditory cues and may not be certain of the response. If a word is not in a child's mental lexicon, he or she will not be able to identify it from limited speech information. However, a child may be able to "fill in the blanks" to correctly recognize words that are in their mental lexicon. Vocabulary knowledge relates most strongly to the identification of lexically difficult words. In this condition, a listener must be able to process relatively fine acoustic cues to identify the target from many similar sounding words in memory. It may be that vocabulary knowledge helps children to focus on potential target words and suppress non-word competitors.

The present results did not reveal any relationship between knowledge of the language and the recognition of speech through visual cues alone. This finding differs from those of Bergeson et al. (16) who found a significant relationship between visual-only word recognition and performance on measures of language processing. The contrasting results probably are due to differences in the participant characteristics between our study and that of Bergeson et al. (16). They longitudinally examined audiovisual word recognition in a group of children with profound deafness who received a CI before 9 years of age. Participants were divided into four groups as a function of Communication Method (Oral vs. Total Communication) and age at time of implantation (Early vs. Late). The mean age at implantation was approximately 3 years for the Early groups and 6 years for the Late groups. Children were administered a live voice test of audiovisual word recognition, the Common Phrases Test in V-Only, A-only and AV presentation formats. They also carried out an analysis of scores on other speech and language tests administered as part of the longitudinal battery used in the Indiana University School of Medicine study. Their results revealed significant main effects of duration of cochlear implant use, communication method, age at implantation and presentation format; they also found a number of significant interactions among their main effects. Of particular interest to the present study is the interaction between age at implantation and presentation format. Children who were implanted at an earlier age demonstrated better A-only skills than V-only skills. Conversely, children in the Late group demonstrated better V-only speech recognition than A-only recognition. In our study, nearly all of the children performed poorly in the V-only condition; all were implanted prior to age 3 years and had early access to spoken language. The lack of a relationship between language processing and visual recognition in our study is probably due to the floor effects noted in the V-only condition.

In summary, our results suggest that many children with cochlear implants obtain very high levels of word recognition through listening alone. However, large individual differences in performance still remain. Our results further demonstrate that audiovisual speech integration contributes substantially to word recognition in children with cochlear implants. As such, measures of this skill should be a standard part of any clinical evaluation battery. Finally, results from lexically controlled word and sentence tests help us to understand sensory, linguistic and cognitive processes that support spoken word recognition in this clinical population.

Acknowledgments

This work was supported by National Institutes of Health/National Institute on Deafness and Other Communication Disorders grants R01 DC00064 and R01 DC008875. We thank Shirley C. Henning, M.S., for assistance in data collection, Terri Kerr, B.A., and Randy Richter, B. Comp. Sci. for assistance with data analysis and Robyn Britt for clerical assistance. We also gratefully acknowledge the children who participated in this investigation and their parents.

References

1. Clark GM. The cochlear implant: A search for answers. *Cochlear Implants International* 2000;1(1):1–15. [PubMed: 18791989]
2. Walden BE, Grant KW, Cord MT. Effects of amplification and speechreading on consonant recognition by persons with impaired hearing. *Ear and Hearing* 2001;22(4):333–41. [PubMed: 11527039]
3. Sumby WH, Pollack I. Visual contribution of speech intelligibility in noise. *Journal of the Acoustical Society of America* 1954;26:212–5.
4. Demorest M, Bernstein L, DeHaven G. Generalizability of speechreading performance on nonsense syllables, words, and sentences: Subjects with normal hearing. *Journal of Speech and Hearing Research* 1996;39:697–713. [PubMed: 8844551]
5. Sommers MS, Tye-Murray N, Spehar B. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing* 2005;26:263–75. [PubMed: 15937408]
6. Erber NP. Auditory and audiovisual reception of words in low-frequency noise by children with normal hearing and by children with impaired hearing. *Journal of Speech and Hearing Research* 1971;14:496–512. [PubMed: 5163883]
7. Moody-Antonio S, Takayanagi S, Masuda A, Auer ET, Fisher L, Bernstein LE. Improved speech perception in adult congenitally deafened cochlear implant recipients. *Otology and Neurology* 2005;26:649–54.
8. Grant KW, Seitz PF. The use of visible speech cues for improving auditory detection of specific words. *Journal of the Acoustical Society of America* 2000;108:1197–208. [PubMed: 11008820]
9. Bernstein LE, Demorest ME, Tucker PE. Speech perception without hearing. *Perception & Psychophysics* 2000;62:233–52. [PubMed: 10723205]
10. McGurk H, MacDonald JW. Hearing lips and seeing voices. *Nature* 1976;264:746–8. [PubMed: 1012311]
11. Demorest ME, Bernstein LE. Sources of variability in speechreading sentences: A generalizability analysis. *Journal of Speech and Hearing Research* 1992;35:876–91. [PubMed: 1405543]
12. Seewald RC, Ross M, Giolas TG, Yonovitz A. Primary modality for speech perception in children with normal and impaired hearing. *Journal of Speech and Hearing Research* 1985;28:36–46. [PubMed: 3981996]
13. Bergeson, TR.; Pisoni, DB. Audiovisual speech perception in deaf adults and children following cochlear implantation. In: Calvert, GA.; Spence, C.; Stein, BE., editors. *The Handbook of Multisensory Perception*. Cambridge, MA: MIT Press; 2004. p. 749–71.
14. Bergeson TR, Pisoni DB, Davis RAO. A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review* 2003;103:347–70.
15. Pisoni, DB.; Svirsky, MA.; Kirk, KI.; Miyamoto, RT. Looking at the “Stars”: A first report on the intercorrelations among measures of speech perception, intelligibility and language development in pediatric cochlear implant users. Bloomington, IN: Indiana University; 1997.

16. Bergeson TR, Pisoni DB, Davis RAO. Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear and Hearing* 2005;26:149–64. [PubMed: 15809542]
17. Luce PA, Pisoni DB. Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing* 1998;16:470–81.
18. Cluff MS, Luce PA. Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance* 1990;16(3):551–63. [PubMed: 2144570]
19. Kucera, F.; Francis, W. *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press; 1967.
20. Kirk KI, Pisoni DB, Osberger MJ. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing* 1995;16:470–81. [PubMed: 8654902]
21. Eisenberg LS, Martinez AS, Holowecky SR, Pogorelsky S. Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. *Ear and Hearing* 2002;23:450–62. [PubMed: 12411778]
22. Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research* 2003;46:390–404.
23. Holt, RF.; Kirk, KI.; Pisoni, DB.; Burckhartzmeyer, L.; Lin, A. Lexical and context effects in children's audiovisual speech recognition. 150th Meeting of the Acoustical Society of America; 2005; Minneapolis, MN. 2005.
24. Kirk KI, Sehgal ST, Hay-McCutcheon M. Comparison of children's familiarity with tokens on the PBK, LNT, and MLNT. *Annals of Otology, Rhinology, and Laryngology* 2000;185:63–4.
25. Kirk KI, Pisoni DB, Miyamoto RT. Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research* 1997;40:1395–405.

Figure 1. Mean percent of words in sentences correctly recognized on the AV-LNST as a function of lexical difficulty and presentation format.

Figure 2.

Individual data for the percent of key words correctly recognized in sentences on the AV-LNST as a function of presentation format. The top panel shows scores for the lexically easy key words and the bottom panel shows scores for the lexically hard key words.

Figure 3. Mean percent of auditory-only isolated words correctly recognized on the LNT as a function of lexical difficulty.

Figure 4. A comparison of auditory-only word recognition in isolated word and sentence contexts as a function of lexical difficulty. The data are plotted separately for high and low performing participants.

Figure 5.

The association between vocabulary recognition performance (PPVT standard scores) and the percent correct scores for words in sentences presented in an A-only or an AV presentation format. The top panel plots scores for the lexically easy key words and the bottom panel plots scores for the lexically hard key words on the AV-LNST.

Figure 6.

The association between vocabulary recognition performance (PPVT standard scores) and the percent correct scores for isolated words presented in an A-only format. The top panel plots scores for the lexically easy key words and the bottom panel plots scores for the lexically hard key words on the LN

Table 1

Participant Characteristics

Participant ID	Best PTA (dB HL)	Age Fit with CI (yrs)	Age at Testing (yrs)	Communication Mode	Device Condition
496	67	0.86	4.75	Oral	CI + HA
497	113	0.99	5.03	Oral	Monaural CI
406	103	1.48	8.41	Oral	Binaural CI
547	118	1.89	7.87	Oral	Monaural CI
346	113	1.59	7.75	Oral	Monaural CI
587	95	1.79	3.86	Oral	Monaural CI
482	102	1.15	5.70	Oral	Monaural CI
507	118	1.40	4.96	Oral	Monaural CI
465	82	2.18	7.16	Oral	Monaural CI
563	92	2.80	3.79	Oral	Monaural CI
515	118	2.30	5.79	TC*	Monaural CI
558	85	2.54	4.47	Oral	Monaural CI
344	118	2.98	8.96	Oral	Monaural CI
562	113	2.41	3.88	Oral	Monaural CI
525	58	2.39	5.55	Oral	Monaural CI
Mean	99.81	1.92	5.89	-	-

* Total Communication

Table 2

Pearson Correlation Coefficients for performance on the two spoken word recognition measures.

AV-LNST		LNT Easy Words	LNT Hard Words
V-only	Easy words	0.29	0.22
	Hard Words	0.40	.31
A-only	Easy Words	0.82**	0.88**
	Hard Words	0.81**	0.74*
AV	Easy Words	0.79**	0.77**
	Hard Words	0.77**	0.76**

*
p < .01;

**
p < .001