Published in final edited form as:

J Speech Lang Hear Res. 2011 April ; 54(2): 632–657. doi:10.1044/1092-4388(2010/09-0148).

Assessing multimodal spoken word-in-sentence recognition in children with normal hearing and children with cochlear implants

Rachael Frush Holt,

Department of Speech and Hearing Sciences, Indiana University

Karen Iler Kirk, and

Department of Speech, Language, and Hearing Sciences, Purdue University and DeVault Otologic Research Laboratory, Indiana University School of Medicine

Marcia Hay-McCutcheon

DeVault Otologic Research Laboratory, Indiana University School of Medicine

Abstract

Purpose—To examine multimodal spoken word-in-sentence recognition in children.

Method—Two experiments were undertaken. In Experiment I, the youngest age with which the multimodal sentence recognition materials could be used was evaluated. In Experiment II, lexical difficulty and presentation modality effects were examined, along with test-retest reliability and validity in normal-hearing children and those with cochlear implants.

Results—Normal-hearing children as young as 3.25 years and those with cochlear implants just under 4 years who have used their device for at least 1 year were able to complete the multimodal sentence testing. Both groups identified lexically easy words in sentences more accurately than lexically hard words across modalities, although the largest effects occurred in the auditory-only modality. Both groups displayed audiovisual integration with the highest scores achieved in the audiovisual modality, followed sequentially by auditory-only and visual-only modalities. Recognition of words in sentences was correlated with recognition of words in isolation. Preliminary results suggest fair to good test-retest reliability.

Conclusions—The results suggest that children's audiovisual word-in-sentence recognition can be assessed using the materials developed for this investigation. With further development, the materials hold promise for becoming a test of multimodal sentence recognition for children with hearing loss.

Introduction

The Multimodal Nature of Speech Perception

Multisensory integration of speech is a natural and automatic process in individuals of all ages, even infants (Desjardins, Rogers, & Werker, 1997; Erber, 1969; Kuhl & Meltzoff, 1982; McGurk & MacDonald, 1976; Sumby & Pollack, 1954). Children and adults with

Corresponding Author and Address: Rachael Frush Holt, Ph.D., Indiana University, Department of Speech and Hearing Science, 200 South Jordan Avenue, Bloomington, IN 47405. Office: 812-855-6659, Fax: 812-855-5531, raholt@indiana.edu.

Karen Iler Kirk currently is affiliated with the University of Iowa, Department of Communication Sciences and Disorders, Wendell Johnson Speech and Hearing Center, Iowa City, IA 52242. Marcia Hay-McCutcheon currently is affiliated with The University of Alabama, Department of Communicative Disorders, Tuscaloosa, AL 35487.

normal hearing and those with hearing loss benefit from combining auditory and visual speechreading cues in speech recognition (Bergeson, Pisoni, & Davis, 2003; Erber, 1972, 1975; Geers, Brenner, & Davidson, 2003; Kaiser, Kirk, Lachs, & Pisoni, 2003; Sumby & Pollack; Tyler et al., 1997). The ability to integrate visual and auditory cues is particularly useful when listening conditions are compromised, such as in interfering background noise (Erber, 1972, 1975; Massaro & Cohen, 1995; Sumby & Pollack; Summerfield, 1987).

Sumby and Pollack (1954) were among the first investigators to recognize that speech perception is multimodal in nature. They demonstrated that the addition of visual cues to the auditory signal improved word recognition, particularly under adverse listening conditions, such as in background noise. The absolute magnitude of improvement increased with higher levels of background noise; however, the ratio of obtained improvement to possible available improvement varied little across the tested signal-to-noise ratios (SNRs). In other words, although the visual contribution varied in magnitude across the noise levels, the actual visual contribution relative to the possible maximum contribution was constant across SNRs. Later, McGurk and McDonald (1976) demonstrated that vision does not simply enhance speech perception; it can alter it, as well. When presented with conflicting visual and auditory speech cues, sometimes participants perceive a syllable that was not present in either the visual or auditory stimuli themselves. For example, some participants will fuse an auditory/ba/ and a visual/ga/, resulting in a perceived/da/(known as the “McGurk Effect”). Despite knowing that speech perception is a multimodal process, most research in the speech and hearing sciences, like other sensory fields (Calvert & Lewis, 2004), has exclusively focused on a single modality: in our case, the auditory perception of speech.

From a clinical perspective, the majority of speech perception testing conducted for diagnostic purposes and for tracking performance over time with sensory aids has been conducted in an auditory-only (A-only) modality for good reason. Clinicians and researchers were interested in examining the effects of sensory aid use on auditory skills. The influence of visual information primarily was thought of as a supplement to auditory perception via speechreading (e.g., Erber, 1972, 1975; Sumby & Pollack, 1954). However, recent research indicates that the ability to integrate speech information from both modalities, auditory and visual, may have important implications for determining cochlear implant (CI) candidacy: speechreading and audiovisual (AV) speech perception appear to be reliable pre-implantation predictors of post-implantation success and benefit in prelingually deaf children (Bergeson & Pisoni, 2004; Bergeson, Pisoni, & Davis, 2003, 2005; Lachs, Pisoni, & Kirk, 2001). Further, audiovisual speech recognition is correlated strongly with other outcome measures of speech and language development. These relationships suggest a common source of underlying linguistic knowledge. AV speech perception also better reflects the demands of everyday communication than A-only (Robbins, Renshaw, & Osberger, 1995). Therefore, a picture of the speech perception development and/or current skill level of a sensory aid user is not complete without testing her/his speech perception in all the perceptual modes in which speech is perceived. Thus, it is reasonable to suggest that multimodal spoken word recognition testing be a component of the test battery used to assess speech perception development in children with sensory aids.

A Need for Audiovisual Speech Perception Assessments

To evaluate multimodal spoken word recognition testing in children with sensory aids, we looked to available tests as a logical first step. However, despite the importance of multimodal speech perception and Erber’s observation made over 30 years ago that there was a “need for standardization and validation of auditory-visual stimulus items” (1975, p. 489), few standard clinical tests of audiovisual word recognition are available. Two audiovisual tests have been used with pediatric sensory aid users: the Children’s Audio-Visual Enhancement Test (CAVET) (Tye-Murray & Geers, 2001) and the Audiovisual

Feature Test (Tyler, Fryauf- Bertschy, & Kelsay, 1991). The CAVET (published by the Central Institute for the Deaf) is an audiovisually recorded open-set word recognition test that consists of three lists of 20 individual words that are reported to be familiar to hearing-impaired children in the fourth grade. Each list is evenly divided into low-visibility and high-visibility items, which were determined based on data from young, normal-hearing adults (Tye-Murray, Sommers, & Spehar, 2007). The test is scored by calculating the gain provided by the addition of auditory cues to the visual cues. Spehar, Tye-Murray, and Sommers (2004) reported that the three lists were equivalent in the V-only modality. The Audiovisual Feature Test is a closed-set test of consonant feature recognition. A recorded version of the test is used by the test developers (Tyler et al.), but is not commercially available. The stimulus items consist of seven letters (all consonants) and three words that are judged to be familiar to young children. Due to the closed-set nature of the task, the same stimulus items can be used in consecutive test administrations to compare performance in multiple modalities.

In addition to these AV speech recognition tests, two A -only tests have been modified for multimodal presentation to pediatric sensory aid users: the Common Phrases Test (Robbins et al., 1995) and the Pediatric Speech Intelligibility Test (PSI) (Jerger, Lewis, Hawkins, & Jerger, 1980). The Common Phrases Test is an open-set test of spoken sentence comprehension of phrases used in everyday situations. The test is administered live-voice. The PSI is a closed-set test of word and sentence recognition that originally was presented in competing speech to evaluate central auditory processing in children. Similar to the Common Phrases Test, it typically is administered live-voice (although Eisenberg and Dirks [1995] used a recorded A- only version in competing noise). In the word recognition section, the tester says a word and the child is to respond by pointing to the corresponding item depicted pictorially on one of five cards. The response task in the sentence portion is the same as in the word recognition task, but the stimulus is a sentence describing an action portrayed in one of the picture cards. Both the PSI and the Common Phrases tests were adapted by researchers and clinicians in our lab for multimodal use in outcomes research on pediatric sensory aid users (e.g., Lachs, Pisoni, and Kirk [2001] used the Common Phrases Test and Bergeson et al. [2003] used the PSI in their respective audiovisual word recognition work).

Although these tests are valuable for various purposes, we saw a need for a multimodal test that was recorded, used connected speech, and was based on a theory of spoken word recognition. Of the AV tests currently in use, only two tests (the CAVET and the Audiovisual Feature Test) were used in an audiovisually recorded format. Live-voice presentation does not allow for control over level, speaking rate, inflection, and vocal clarity across speakers and across stimuli within a single speaker. These factors are known to influence performance on speech perception tests (Brandy, 1966; Penrod, 1979). In addition, it would be useful to know how visual cues influence perception of connected speech. Both of the available audiovisually recorded tests use either isolated words or consonants. Finally, the manner in which the available tests typically are used results in them primarily being descriptive in nature. Descriptive tests provide a general estimate of an individual's speech understanding, but they do not provide information about the underlying encoding, storage, and retrieval of speech by listeners. One way to use these measures to examine how listeners process and identify speech audiovisually, is to modify the way in which the tests are used. For example, Tye-Murray et al. (2007) recently used the CAVET to examine how adults perceive audiovisual speech by examining lexical properties of the words used in the CAVET from both acoustic and visual perspectives. Lexical properties of speech (such as the frequency with which a word occurs in a language and the number of words that can be confused with the target word) have been shown to influence the accuracy and speed with which individuals identify spoken words (see Luce and Pisoni [1998] for a review).

However, most clinicians and researchers use the available test materials in the manner in which they were intended to be used and none of the current AV testing materials were intended to assess the underlying encoding, storage, and retrieval of speech by listeners. Evaluating these underlying perceptual processes is important because they might contribute to the enormous variability in outcomes observed in the pediatric CI population (Krull, Choi, Kirk, Prusick, & French, 2010). Therefore, we saw a need for theoretically based, audiovisually recorded tools that can assess the multimodal nature of speech perception over a wide range of listening abilities to supplement the available measures.

Currently, there are two word recognition tests administered in the A-only format that are theoretically motivated tests of speech perception: the Lexical Neighborhood Test (LNT) and the Multisyllabic Lexical Neighborhood Test (MLNT) (Kirk, Pisoni, & Osberger, 1995). Both tests evaluate open-set word recognition in children with sensory aids. Kirk et al. selected the words from a subset of items compiled by Logan (1992) from the CHILDES database (MacWhinney & Snow, 1985). The subset includes words produced by children with normal hearing between the ages of 3 and 5 years and whose lexical characteristics are well documented. According to the Neighborhood Activation Model of spoken word recognition (Luce & Pisoni, 1998), recognition of spoken words is influenced by three lexical factors: the frequency of occurrence of a given word in a language; how phonemically similar the target word is to other words in the lexicon (typically calculated by counting the number of words, or lexical neighbors, that can be created from the target by substitution, addition, or deletion of a single phoneme); and how often the neighbors occur in the lexicon. Two of these characteristics were used to define a given word's lexical difficulty on the LNT and MLNT: lexically easy words are those that occur frequently in a language and have few neighbors; alternatively, lexically hard words are used infrequently in a language and have many neighbors with which they can be phonemically confused. Because there is less competition for lexical selection, easy words tend to be identified more accurately and quickly than hard words (Luce & Pisoni). Using the LNT and MLNT, Kirk et al. reported that children with CIs were sensitive to the lexical properties of the stimulus words, similar to what has been reported in listeners with normal hearing. Their results suggest that despite having limited auditory experience and language skills and receiving a degraded signal, children with CIs organize, store and access items in memory like children with normal hearing. Later studies have reported similar word recognition results in both clinical and nonclinical populations (e.g., Carter & Wilson, 2001; Dirks, Takayanagi, Moshfegh, Noffsinger, & Fausti, 2001; Kirk, Pisoni, & Miyamoto, 1997; Luce, 1986; Luce & Pisoni), and Bell and Wilson (2001) found similar results for sentence recognition in normal-hearing adults.

Expanding on Bell and Wilson's (2001) work, Eisenberg, Martinez, Holowecky, and Pogorelsky (2002) created lexically controlled word-in-sentence materials in the A-only presentation modality using the same database employed by Kirk et al. (1995). The sentences are five to seven words in length, of which three are scored key words. The key words in a given sentence are all lexically easy or all lexically hard; the remaining words are lexically "neutral." Lexically neutral words were those selected from the database that are not classified as "easy" or "hard." Half of the sentences contain lexically easy key words and half are comprised of lexically hard key words. The sentences, along with the key words presented in isolation, were administered to 12 children with normal hearing (NH) and 12 with CIs who were oral communicators. Both groups of children ranged in age from 5 to 14 years. NH children were presented with spectrally degraded stimuli (the signal was reduced to 4 spectral bands) using a noise-band processing procedure described by Shannon, Zeng, Wygonski, Kamath, and Ekelid (1995). Both groups of children demonstrated more accurate identification of easy words than hard words, both when presented in sentences and in isolation. Eisenberg et al.'s results suggest that even over a wide range of CI benefit,

children with CIs organize, store and retrieve items in their mental lexicons similarly to children with normal hearing.

To evaluate the multimodal spoken word recognition abilities of children, we chose to video record the sentences developed by Eisenberg et al. (2002), so that they could be presented multimodally. We selected Eisenberg et al.'s sentences for several reasons. First, their development was based in a model of spoken word recognition, the NAM (Luce & Pisoni, 1998), thereby allowing examination not just of percent correct, but of how the children organize, store and access lexical items in memory. Although the NAM certainly is not the only model of spoken word recognition, it is one that has been well tested and whose tenets have been demonstrated across many investigations with many different clinical and non-clinical populations (e.g., Carter & Wilson, 2001; Dirks et al., 2001; Kirk et al., 1997; Luce, 1986; Luce & Pisoni). By including both lexically easy and hard words, one can evaluate potential underlying perceptual processes (such as encoding, storage and retrieval) that might influence the widely variable outcomes in pediatric cochlear implantation. If recognition scores of lexically easy and hard words do not differ from each other, it would suggest that the listener with hearing loss might not organize and/or access her/his mental lexicon as typical listeners do. Second, the materials have been used with young NH children and those with hearing loss who use CIs (Eisenberg et al.). Third, they are composed of sentences, which tend to be more reflective of everyday communication demands than are tests using individual words or phonemes, because communication occurs in context. This investigation is the first step toward a long-term goal of developing these materials into a test of audiovisual spoken word recognition.

The purpose of the current investigation was three-fold: 1) to examine presentation modality and lexical difficulty effects in NH children and those with CIs; 2) to obtain preliminary data from which to identify the lower age limit with which the multimodal spoken word-in-sentence recognition materials can be used; and 3) to obtain preliminary validity and test-retest reliability data on the multimodal sentence materials. This investigation reports on the first steps in a series in which we are developing the multimodal sentence recognition materials. Two experiments were carried out to address the goals of the current investigation. In Experiment I, we sought to determine the youngest age with which the multimodal sentences could be administered and to examine lexical difficulty and modality effects. In Experiment II, we examined the same effects in noise that were evaluated in Experiment I, and evaluated the materials' validity and test-retest reliability in children with normal hearing and those with CIs. In addition to the primary experiments, a Preliminary Experiment was conducted to generate lists of sentences that were equally difficult within each testing modality for use in Experiments I and II.

Preliminary Experiment

To use the sentence materials across multiple modalities without redundancy, we conducted a preliminary study to create equivalent lists of sentences within each presentation modality. Phase I of the Preliminary Experiment was conducted to generate lists of equivalent sentences and Phase II was carried out to verify the lists were indeed equivalent.

Method

Participants—Two different groups of 30 young, normal-hearing adults with normal or normal- corrected vision who spoke English as their primary language participated in the Preliminary Experiment (Phase I mean age = 23.0 years; Phase II mean age = 24.2 years). As a first step in developing these materials, adults were selected as participants because they could undergo more testing than children. Further, other investigators have used adults

in generating AV testing materials (e.g., Tye-Murray et al. [2007] used NH adults in categorizing low- and high- visibility items on the CAVET).

Materials—The multimodal sentences were drawn from the 50-sentence corpus developed by Eisenberg et al. (2002). The sentences are composed of single- and two-syllable words selected from the same subset of words from the CHILDES database (MacWhinney & Snow, 1985) used by Kirk et al. (1995) in the LNT and MLNT. The database is composed of 994 words. Using two of the three parameters for lexical difficulty outlined by the NAM, 148 are classified as “easy” and 232 as “hard” words. The remaining 614 words are classified as neither easy nor hard and thus are considered “neutral” in lexical difficulty. Each sentence composed by Eisenberg et al. consists of five to seven words, of which three are key words that are scored. Half of the sentences consist of lexically easy key words and half of lexically hard key words. The remaining words in each sentence are lexically neutral and are unscored. Each sentence is “syntactically correct but semantically neutral (low in word predictability)” (Eisenberg et al., p. 453). The sentences composed of easy key words and those of hard key words are similar with respect to the occurrence of various speech features, such as voiced and voiceless fricatives, diphthongs, etc. Further information regarding sentence development can be found in Eisenberg et al.

A professional video recording team digitally video recorded a professional Caucasian female announcer who spoke a General American dialect of English. To reduce any visual distractions, the speaker wore plain clothing, no jewelry, and minimal make-up. The recording took place in a single session in a sound booth and the speaker was instructed to use a natural speaking style. Each sentence was recorded three or more times in order to select a good “take.” A good “take” was one that was fluent, had a similar prosody to the other sentences, and did not include visual distractions, such as excessive eyeblinks. Only one “take” was selected for each sentence. The selected token of each sentence was spliced into individual digital movie files with a 16 bit, 44.1 kHz audio sampling rate, 29.97 frames per second, 720 × 480 frame size, 24-bit color depth, in DV/DVCPRO video format. The audio and video streams were separated in Final Cut Pro HD v.4.5 (Apple Computer Inc., 2004) and the audio portion of each stimulus was adjusted to equalize the total root mean square amplitude (RMS) of each sentence in Audition v. 2.0 (Adobe, 2005). Then the video and equalized audio streams were merged back together for multimodal presentation in Final Cut Pro. This procedure does not interfere with the temporal aspects of the stimuli.

Procedures—Within each phase, participants were randomly assigned to one of three presentation modalities (A-only, V-only and AV), such that 10 participants were assigned to each modality. In Phase I, all 50 sentences were presented in random order; in Phase II, the materials were randomly blocked by list and the sentences were randomized within each list. Only the key words were scored. In the A-only and AV conditions, the speech was presented at an average RMS of 65 dB A and a SNR of -7 dB. The SNR was selected because preliminary data suggested it achieved average performance of approximately 70% correct in the AV modality. The noise was a white noise shaped to match the long-term spectrum of the sentences as a whole. See Experiment II for a full description of the generation of the speech-shaped noise.

Results

The results of the Preliminary Experiment appear in Appendix A. The average percent key words correct on each individual sentence from Phase I is displayed in Figure A1. The intelligibility data from Phase I were then used to organize the sentences into six 8-sentence lists that were intended to be equally difficult within each modality. The stars displayed in Figure A1 indicate the sentences (E13 and H23) that were excluded from the final sentence

lists due to reduced intelligibility across all three testing modalities. Each list contained four sentences composed of easy key words and four containing hard key words. Although 8-sentence lists may seem short relative to other spoken word recognition materials, three key words are scored in each sentence resulting in 24 test items scored per list. Thornton and Raffin (1978) and Gelfand (1993) have demonstrated that confidence limits and standard deviations increase with smaller numbers of scored responses. In other words, reliability is compromised with smaller numbers of test items. Typical word recognition tests use between 10 test items (Isophonemic Word Lists [Boothroyd, 1984; Boothroyd & Nitttrouer, 1988]) and 50 test items (NU-6 [Tillman & Carhart, 1966]), however, most audiologists usually only present 25 of the items when using the NU-6. Although a direct comparison to these individual word recognition tests cannot be made because the words in the sentence materials are not independent of one another due to context effects, the 24 test items scored on each of the sentence lists are about the average number of items scored on a typical speech recognition test, while leaving enough lists for testing across multiple modalities. Table A1 displays the mean intelligibility of each sentence and each list as a whole in all three modalities from Phase I of the Preliminary Experiment. The results of a one-way ANOVA of list difficulty suggest that there were no significant differences in difficulty across the lists within each modality (V-only: $F(5, 59) = .219, p = 0.953$; A-only: $F(5, 59) = .255, p = 0.935$; AV: $F(5, 59) = .153, p = 0.978$).

Figure A2 displays mean percent key words correct for each sentence list in all three modalities from Phase II of the Preliminary Experiment, which was carried out to confirm that the lists were indeed equivalent. Within each modality, average performance across sentence lists varied by 4 to 8 percentage points. The results of a one-way ANOVA of list difficulty suggest that the lists were not significantly different from one another within each modality (V-only: $F(5, 59) = .356, p = 0.876$; A-only: $F(5, 59) = .759, p = 0.583$; AV: $F(5, 59) = .217, p = 0.954$). Although the variability in performance across the six lists within each modality varied slightly (see list 6 in the AV modality), the average performance was shown to not be significantly different across the lists within a given modality in two separate groups of participants in Phases I and II of the preliminary study. Therefore, for the purposes of the current investigation, any list or lists could be used in any of the three presentation modalities to evaluate multimodal perception of speech. Further work will evaluate whether variability across lists could be improved (perhaps with longer lists). In addition, future investigations will evaluate list equivalency in children, because their speech perception is still developing, which could potentially influence item difficulty and thus, list equivalency.

It should be noted that two different versions of the stimuli were used in Phases I and II of the preliminary study. The second version came about after “color-correcting” the stimuli (adjusting the color spectrum across the image), a technique used to correct color discrepancies resulting from lighting artifacts that occur during recording. The result is a color spectrum that replicates that of the subject being recorded, thereby improving the perceived color conformity of the recording to the original subject (which in this case was the speaker). In other words, color correcting improves the naturalness of the color spectrum. The original version was used in Phase I and the color-corrected version in Phase II. Whereas no significant differences in word recognition were found between the two versions of materials in either the AV or A-only presentation modalities, a one-way ANOVA revealed that performance on the color-corrected version (from Phase II) in the V-only presentation modality was significantly better than on the original version of the materials (from Phase I), $F(1, 19) = 12.583, p = .002$. The difference between the mean scores in V-only was 11 percentage points. Importantly however, the study design used in Phases I and II of the Preliminary Experiment did not allow for determination of whether differences in performance on the two versions of the stimuli were due to color correction or

to differences between the two groups of participants used in Phases I and II. Therefore, a third phase of the preliminary study was undertaken using a repeated measures design to specifically examine the effects of color correction.

In Phase III, a different group of 20 young, normal-hearing adults with normal or normal-corrected vision (mean age = 23.05 years) was tested on their recognition of the sentences. All of the participants completed five presentation conditions, each composed of different sentences (half of the sentences in each condition were composed of easy key words and half of hard key words): V-only color-corrected, V-only original, A-only, AV color-corrected and AV original. The sentence order used across conditions was randomized and the sentences assigned to each condition were randomized across participants. For the A-only and AV conditions, the SNR used was -7 dB, with the speech presented at an average RMS of 65 dB A.

Average percent key words correct in each condition from Phase III of the preliminary experiment is displayed in Figure A3 of the Appendix. Two paired-samples t-tests revealed that there were no significant differences between the color-corrected and the original versions of the materials in the V-only condition, $t(19) = .068$, $p = .947$, or the AV condition, $t(19) = .678$, $p = .506$. This suggests that any differences observed in scores in Phases I and II were due to group differences and not to differences in the minor effects of color correction.

Experiment I

Method

Participants—All of the participants in Experiment I spoke English as their first language and their parents reported no speech or language problems. Further, they were required to have normal visual acuity or normal corrected visual acuity, if glasses or contacts were worn (Snellen acuity of 20/25), and pass a hearing screening bilaterally at 20 dB HL at .5, 1, 2, 4, and 8 kHz and 25 dB HL at .25 kHz (re: ANSI, 1989). Data for Experiment I were collected at two different sites: the Indiana University School of Medicine (IUSOM) and Purdue University. Institutional review board approval was obtained from both institutions. A total of 61 participants were recruited. Three participants did not pass either the hearing or the vision screening (at Purdue) and were not tested further. One child was the first 6-year-old recruited (at the IUSOM) and based on the performance of the 4- and 5-year-old children, no more 6-year-olds were recruited. Therefore, this single 6-year-old's data were not included in any analyses. This left 57 participants who were administered the protocol (35 from IUSOM and 22 from Purdue). The children ranged in age from 3;0 to 5;11 (years; months) and were separated into three groups based on chronological age. There were 23 3-year-olds, 20 4-year-olds and 14 5-year-olds. Participant demographics for Experiment I are displayed in the top half of Table 1. Six of the 57 children (four from IUSOM and two from Purdue) were not administered one or more tests because of experimenter error, being unable to complete the entire testing protocol due to fatigue, or the parent not being able to return for the second visit.

Materials

Multimodal Sentences: Recall from the preliminary study that the video-recorded multimodal sentences were drawn from the 50-sentence corpus developed by Eisenberg et al. (2002). Each sentence composed by Eisenberg et al. consists of five to seven single- and two-syllable words, of which three are key words that are scored. Half of the sentences consist of three lexically easy key words and half of three lexically hard key words. The remaining words in each sentence are lexically neutral and are unscored. In the Preliminary

Experiment, six 8-sentence lists were developed that were equally difficult in each testing modality (A-only, V-only, and AV). These sentence lists were used in both Experiments I and II.

Goldman-Fristoe Test of Articulation-2: The Goldman Fristoe Test of Articulation - Second Edition (GFTA-2; Goldman & Fristoe, 2000) is a standardized articulation test for individuals aged 2–21 years. It consists of a sounds-in-words section, sounds-in-sentences section, and a stimulability section, in order to assess an individual's articulation of consonants found in Standard American English. The sounds-in-words section of the GFTA-2 utilizes colored pictures on an easel to elicit 53 words from the child, targeting consonants in the initial, medial and final positions, as well as consonant clusters in the initial position. Only the sounds-in-words section was administered in the current investigation to identify any gross articulation deficits. In other words, the GFTA-2 primarily was used to verify that the child being tested could be understood by the tester and thus responses to the multimodal sentence materials could be scored accurately.

Expressive Vocabulary Test: The Expressive Vocabulary Test (EVT; Williams, 1997) is a standardized measure of expressive vocabulary for children and adults ages 2.5 – 90 years. The test uses a broad sample of American English vocabulary. It is administered by presenting colored pictures located on an easel and asking the examinee about each picture. The first 38 items are labeling items in which the examiner points to the picture to be named. The remaining 152 items are synonym items in which the examiner points to the picture while stating a label for the pictured item along with a carrier phrase. The examinee is to respond with a synonym for the item. Age- appropriate start items and basal and ceiling rules allow for administration of only those items that closely approximate the examinee's vocabulary level. Responses are recorded on a standard form and then scored to yield a raw score, standard score, normal curve equivalent, percentile rank, and test-age equivalent score. The EVT was used to ensure that participants had sufficient vocabulary knowledge to complete the testing. No child's standard score was less than 90, indicating that all of the children had appropriate vocabulary knowledge for their chronological age.

Procedures—The multimodal sentence materials were presented in quiet at an average RMS of 65 dB A in the sound field. Participants were tested in a sound booth while seated in front of a 19- inch computer monitor with speakers placed at ± 45 degrees azimuth relative to the listener. The speaker's image consumed approximately one-third of the horizontal dimension and over one-half of the vertical dimension of the monitor. The lighting in the sound booth was rather dim, as it is in most sound booths. The materials were presented using dedicated software via a Pentium computer under the control of the tester. Participants were instructed to repeat what they perceived and responses were scored on-line by the tester outside of the booth. Only the three key words in each sentence were scored. A second tester sat inside the booth with the child to provide reinforcement through verbal praise and to help keep the child attending to the task and watching the computer monitor. The tester in the booth also helped the child keep track of the number of trials completed and remaining by stamping pictures on sheets of paper. The number of pictures corresponded to the number of trials in that condition. After a specific number of pages were completed, the child was given a break during which she/he could play a game with the tester, color, and/or talk with the tester. The children were provided a longer snack break half-way through the entire session.

At the IUSOM, children visited the lab twice. On the first day, children completed the hearing and vision screenings, the GFTA-2, and the multimodal sentences. On the second testing day, which occurred between 1 and 3 weeks after the first visit, children completed the multimodal sentences again and the EVT. The order of test administration was

randomized each day, with the exception that the hearing and vision screenings were always completed first. Each child was presented two multimodal sentence lists in each of three presentation modalities (three sets of two lists): A-only, V-only and AV. The presentation modality and the lists presented in each modality were randomized, as were the order of the lists within each modality and the order of the sentences within each list.

At Purdue, the children visited the lab once. They were administered the same test battery used at the IUSOM, with the exception that they were not administered the multimodal sentences twice. Therefore, they completed the multimodal sentences, EVT, and screenings at a single visit. The test order was randomized similarly to the method used at the IUSOM.

Participants were told to repeat what they perceived and to guess when necessary. Although they were verbally praised for each response, they were not given feedback regarding their accuracy. The entire response was recorded on-line, but only the three key words were scored for the multimodal sentences.

Although two testing sites were used, most influences on the task were consistent across testing sites: computer monitor size, distance from the monitor, computer program that ran the experiment, and incentives and instructions to participants. Variables that differed included: the sound booths and the testers themselves; the number of visits and what was completed during each visit – children at Purdue were tested once, whereas those at the IUSOM were tested over the course of two visits; and the original version of the sentence materials were used at the IUSOM, whereas the color-corrected materials were used at Purdue. Although Phase III of the Preliminary Experiment suggested that there was no difference in performance in the V-only and AV modalities for the two versions of the sentence stimuli, we compared the pediatric data from two sites to evaluate any potential differences. Mean scores for each presentation modality (V-only, A-only, and AV) for each age group (3-, 4-, and 5- year-olds) for those subjects who were presented the original multimodal sentences materials (at the IUSOM) and those administered the color-corrected multimodal sentences materials (at Purdue) are presented in Table 2. A two-way ANOVA with repeated measures revealed that, just as with adults, color-correction did not significantly affect performance in any modality ($p = .67$), nor was there a significant interaction between color correction and presentation modality ($p = .65$). These consistent negative findings for color correction supported our decision to collapse the pediatric data collected across both centers in the following analyses.

Results

Figure 1 displays the individual results on the multimodal sentences. The data are organized by chronological age with results for the A-only presentation modality appearing in the top panel and those for the AV presentation modality in the bottom panel. The V-only scores are not displayed because this condition was very difficult for the children: only three children (two 4-year-olds and one 5-year-old) each got a single word correct. Although most of the scores, particularly for the 4- and 5-year-olds, were at or approaching ceiling in the A-only and AV presentation modalities, there was a trend for older children to score higher than younger children. In A-only none of the 5-year-olds scored below 90% correct, whereas 4 of the 20 4-year-olds (20%) and 11 of the 23 3-year-olds (49%) scored below 90% and 5 of them scored below 65%. Similar trends were found in the AV condition. In general, by the time children were 3½-years-old, most were performing at or near ceiling in both the A-only and AV presentation modalities (although there are three notable exceptions: two 4-year-olds and one 3-year-old).

Figure 2 displays average data on the multimodal sentences plus 1 standard deviation for each age group as a function of lexical difficulty. The top panel displays the V-only

presentation modality. Again, this presentation modality proved to be extraordinarily difficult. The middle and bottom panels display the A-only and AV presentation modalities, respectively. Although there was a trend for the 4- and 5-year-olds to demonstrate better word- in-sentence recognition in A-only and AV than the 3-year-olds, all of the groups scored above 80% on average regardless of the presentation modality or the lexical nature of the key words in the sentences. A three-way ANOVA with repeated measures (within-subject factors: modality, lexical difficulty; between-subject factor: age) revealed that presentation modality, but not lexical difficulty, significantly influenced performance on the multimodal sentences, $F(2,108) = 2287.389, p < .0001$. Further, there were significant performance differences across age groups, $F(2, 54) = 4.818, p = .012$. A planned post-hoc t-test using a conservative Bonferroni correction revealed that the main effect of age group was due to the differences between the 3-year-olds and the 5-year-olds ($p = .019$). The performance differences between the 3-year-olds and 4-year-olds was marginally significant ($p = .069$). The two oldest age groups performed similarly. There was an interaction between modality and age group, $F(4,108) = 3.515, p = .01$, perhaps reflecting that collapsed across lexical difficulty, the highest scores were achieved by 4-year-olds in the AV modality, whereas 3- and 5-year-olds' highest scores were obtained in the A-only modality. Still, the difference scores between the modalities were minimal: .10 percentage points for the 3-year-olds, .15 percentage points for the 4-year-olds and .30 percentage points for the 5-year-olds. The interaction also could perhaps reflect that no 3-year-olds could identify any words in V-only. The 4-year-olds showed the expected pattern of results with a trend for lexically easy words in sentences to be identified more accurately than lexically hard words, whereas the other age groups did not show this trend, $F(2,108) = 3.249, p = .047$. Still, the differences in performance between lexically easy and hard words is extremely small, likely because the words were presented in quiet and children were performing near ceiling in A-only and AV. Because there was no main effect of lexical difficulty or an interaction between lexical difficulty and presentation modality, the data were collapsed across lexical difficulty level and subjected to paired-samples t-tests to examine presentation modality effects. Not surprisingly, the results revealed that performance in V-only was significantly poorer than A-only, $t(56) = -53.344, p < .0001$, and AV, $t(56) = -46.066, p < .0001$. Performance in A-only and AV were not significantly different from one another.

Finally, test-retest reliability was evaluated on the children tested at the IUSOM, although the interpretation is limited due to floor effects for V-only and ceiling effects for A-only and AV from testing in quiet. If a test is reliable, scores at test and retest should be correlated (Anastasi, 1961). Scores at test and retest were correlated in the A-only and AV presentation modalities (A-only sentences with hard key words, $r = .799, p < .0001$; A-only sentences with easy key words, $r = .865, p < .0001$; AV sentences with hard key words, $r = .638, p < .0001$; and AV sentences with easy key words, $r = .561, p = .001$). Floor effects in V-only precluded any meaningful analysis of test-retest reliability in that presentation modality.

Discussion

The results of Experiment I suggest that the multimodal sentence materials can be used to test word-in-sentence recognition in multiple presentation modalities at least down to age 3;25 years in children with normal hearing. Although it is unlikely that NH children between 3;0 and 5;11 will demonstrate V-only recognition of more than one or two words, the overwhelming majority of children will perform with good accuracy on A-only and AV in quiet. These results can begin to be used to develop test-age guidelines for use with children with hearing loss in that if children with normal hearing older than 3.25 years can be administered the materials in quiet, it is likely that children with hearing loss at this age, or perhaps older, will be able to be administered the materials, as well.

That there were slight differences between the number of visits and the number of tests completed at the visits between the two testing sites potentially could be a limitation of the study. Participants at Purdue completed more testing at their visit than did those at the IUSOM and thus, might have experienced more fatigue. Because test order and items presented within each test were randomized, any effects of fatigue would be spread across the tests equally. Furthermore, children were given breaks, a snack, and time to play games with the tester at both sites in an effort to reduce effects of fatigue. Although procedural differences existed across testing sites, some differences can be valuable to collecting preliminary normative data on these new materials. Data collected at multiple sites are more likely to have broader applicability than those collected at one center.

The primary aim of Experiment I was to assess the youngest age with which the multimodal sentence materials can be used and for that reason the materials were presented in quiet. An obvious limitation of this approach is that although several factors significantly influenced children's word recognition accuracy (e.g., presentation modality and chronological age), the analyses on lexical difficulty were not meaningful because of ceiling effects in A-only and AV. Further, the effect of presentation modality was due to the children having so much difficulty in the V-only presentation modality. Therefore, a new group of NH children was recruited in Experiment II to repeat the protocol from Experiment I in noise to eliminate ceiling effects. In addition to more carefully examining these effects, a group of deaf children with CIs, a target population for use of the multimodal sentence materials, completed the protocol as well.

Experiment II

Method

Participants—Two groups of children participated in Experiment II. The first group consisted of 29 typically developing children between the ages of 3;4 and 5;11 with normal hearing and vision. They met the same participant selection criteria outlined in Experiment I. NH participant demographic information is provided in the bottom half of Table 1. The NH participants were tested at IU-Bloomington. The second group of participants consisted of children with prelingual hearing loss who received CIs by age 3.25 years, had no additional disabilities beyond sequelae related to hearing loss (e.g., speech and language delays), and used current CI devices and speech sound processors. Demographic information for the children with CIs appears in Table 3. Eight of the children with CIs (participant numbers 1 – 8) were tested at IU-Bloomington. This group of children completed the entire protocol outlined in the Procedures section of Experiment II. Because this is a low-incidence population and we wanted to increase power, an additional group of CI users (participant numbers 9 – 19) was tested on the multimodal sentences as part of an ongoing longitudinal study on cochlear implantation outcomes at the IUSOM. Institutional review board approval was obtained from both institutions. Because these children were participating in the IUSOM protocol, they could not be administered the identical protocol as those at IU-Bloomington due to the restrictions required of the ongoing longitudinal study at this site. Specific details regarding procedural differences will be discussed in the Procedures section. The average age at testing of the children with CIs was 6.20 years and the average length of device use was 4.23 years. All but two of the children had air-conduction pure-tone averages (at .5, 1, and 2 kHz) greater than 80 dB HL in their better-hearing ears. All but three of the children used a monaural CI exclusively; two were binaural CI users (who received their CIs sequentially) of just over .6 years and one child used a hearing aid on the contralateral ear. Finally, five children used Total communication as their primary mode of communication, in which signing in English word order is combined with spoken English. The remaining 14 children used oral communication, in which no signing is used.

Materials—Some of the same materials that were used in Experiment I were used in Experiment II: the color-corrected multimodal sentence materials, the GFTA-2, and the EVT. In addition, color-corrected audiovisual versions of the LNT and MLNT (AV-LNT and AV-MLNT) were used. The LNT and MLNT are monosyllabic and multisyllabic tests of spoken word recognition, respectively. The LNT consists of two lists of 50 words and the MLNT is comprised of two lists of 24 words. Half of the words on each list are lexically easy and half are lexically hard. The LNT and MLNT words were audiovisually recorded during the same session that the multimodal sentence materials were recorded, and the same RMS equalization and color-correction used with the sentence materials was carried out with the AV-LNT and AV-MLNT words.

The Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) was used to measure vocabulary in the group of CI users tested at the IUSOM because this is the vocabulary measure used in their longitudinal outcomes study on cochlear implantation. The PPVT is a receptive vocabulary test in which the child is presented with individual words and is asked to identify each item from a closed-set of four pictured items. The EVT, an expressive vocabulary test, was administered to the children at IU-Bloomington because some of these children were enrolled in the longitudinal study at the IUSOM and administering the PPVT at intervals between their typical testing intervals (which occur every 6 months) would have invalidated the results. Although it was expected that some of the hearing-impaired children would have gaps or delays in their lexical development (Boothroyd, 1993; Carney et al., 1993; Moeller, Osberger, & Eccarius, 1986), the vocabulary measures were used to verify that the children had adequate vocabulary knowledge to complete the spoken word recognition testing. Standard scores on the EVT and PPVT varied widely among the CI children, but none resulted in vocabulary levels lower than a 3-year-old level.

Speech-shaped Noise: A limitation of Experiment I was that ceiling effects were encountered, particularly for the 4- and 5-year-old children. Therefore, to eliminate these effects, the NH children in Experiment II were presented with the speech recognition materials (the multimodal sentences, AV-LNT, and AV-MLNT) in white noise shaped to match the long-term average spectrum of the sentences as a whole. The white noise was generated using the Kay Elemetrics Computerized Speech Laboratory (CSL) software at a 16 bit, 44.1 kHz sampling rate. It was then shaped in third-octave bands from .25 – 10 kHz to match the long-term average spectrum of the sentences as a whole.

Procedures—The NH children visited the lab twice. On the first day, they completed the hearing and vision screenings, the GFTA-2, the multimodal sentences, and either the AV-LNT or the AV-MLNT. On the second day, which occurred within one to three weeks of the first visit, they were retested on the multimodal sentence materials, the EVT, and either the AV-LNT or the AV-MLNT (whichever was not completed at the first visit). The speech recognition materials in the A-only and AV modalities were presented in speech-shaped noise at a SNR of -2 dB, because pilot testing suggested that this SNR was sufficient for reducing performance in AV to approximately 70%. The speech materials were presented at an average total RMS of 65 dB A.

The CI users who were tested at IU-Bloomington (participant numbers 1 – 8) completed the same protocol as the NH children, except that they were administered the entire speech recognition protocol in quiet. The CI users tested at the IUSOM (participant numbers 9 – 19) were administered the multimodal sentence materials in quiet once as part of their regular longitudinal testing that occurs every 6 months. These children were not administered the AV-LNT or AV-MLNT. Further, as was stated earlier, they completed the PPVT rather than the EVT. Again, the multimodal sentence data from these children were included to

increase the power of some of the statistical analyses and to provide more data from which to make recommendations regarding the multimodal sentence materials' use.

For the multimodal sentences, the modality and list order were randomized, as were the order of sentences within each list. Two different sentence lists were completed in each modality. Because there are only two AV-LNT and AV-MLNT lists and three presentation modalities, one list was presented twice – once in V-only and once in A-only. The list assigned to the AV condition and that assigned to the A-only and V-only conditions were randomized across subjects. The order of presentation modality was randomized with the caveat that the V-only condition always was presented before (but not necessarily immediately before) the A-only condition. This control was used because it was expected that V-only performance would be much poorer than A-only and because no visual information is available in the A-only mode that could cue the participants that they had been given the same list twice.

Results

Normal-hearing children—Figure 3 displays the individual multimodal sentence data for the A-only modality (top panel) and the AV modality (bottom panel). In A-only, only one of the 5-year-olds scored below 65% correct, whereas seven of the 4-year-olds and all but one of the 3-year-olds scored below 65%. Similar trends were noted for the AV presentation modality. As in Experiment I, the V-only data are not displayed because this presentation modality proved to be extraordinarily difficult for the NH children. Figure 4 displays average multimodal sentence data plus 1 standard deviation for each age group as a function of lexical difficulty. The top panel displays the V-only data. One 3-year-old identified a single word correct and two 5-year-olds each identified two words correctly; all of the other children were unable to identify any key words correctly in V-only. The middle and bottom panels display the A-only and AV data, respectively. In general, there is a trend for the older children to achieve higher accuracy, regardless of presentation modality or lexical difficulty, than the younger children.

The data were subjected to a three-way ANOVA with repeated measures (within-subject factors: modality, lexical difficulty; between-subject factor: age). In contrast to Experiment I, all of the main effects were significant when ceiling effects were eliminated. First, presentation modality significantly influenced performance, $F(2,52) = 531.448, p < .0001$, with post-hoc paired comparison t-tests indicating that the AV presentation modality resulted in significantly better performance than A-only or V-only ($p < .0001$) and in turn, A-only resulted in significantly better performance than V-only ($p < .0001$). Second, lexically easy key words were identified more accurately than lexically hard ones, $F(1,52) = 47.879, p < .0001$. And third, older children performed significantly better than younger children, $F(2,26) = 7.578, p = .003$. Planned post-hoc t-tests using a conservative Bonferroni correction revealed that the effects of age were due to the 5-year-olds significantly outperforming both the 4-year-olds ($p = .018$) and the 3-year-olds ($p = .004$). The 3- and 4-year-olds performed similarly. There was a significant interaction between modality and age group, $F(4,52) = 3.713, p = .01$. This result primarily appears to be due to some age groups (e.g., 3-year-olds and 5-year-olds) capitalizing more on AV gain than other groups (e.g., 4-year-olds). Finally, the interaction between modality and lexical difficulty was significant, $F(2,52) = 17.814, p < .0001$, suggesting that lexical difficulty effects are greater in the A-only modality than in the AV modality (and are virtually nonexistent in V-only because of floor effects).

The AV-LNT and AV-MLNT were used to examine the validity of the multimodal sentence materials. Because no audiovisually recorded sentence materials exist to which to compare these new materials, we correlated word-in-sentence recognition in each modality with

isolated word recognition performance in each modality on the AV-LNT and AV-MLNT. AV word-in-sentence recognition was significantly correlated with AV isolated word recognition (LNT words, $r = .461$, $p = .012$ and MLNT words, $r = .367$, $p = .050$). A-only word-in-sentence recognition and A-only isolated word recognition was significantly correlated for the MLNT words, $r = .497$, $p = .006$, but not the LNT words, $r = .345$, $p = .067$. The lack of variability in the V-only scores precluded a meaningful analysis in this modality.

Finally, because the interpretation of test-retest reliability was limited in Experiment I (due to floor and ceiling effects), reliability was evaluated again for the new set of NH children tested in Experiment II. Mean scores increased by 11.7 and 7.9 percentage points going from test to re-test in A-only and AV, respectively. Scores at test and retest were significantly correlated in the A-only and AV modalities (A-only, $r = .587$, $p = .001$; AV, $r = .577$, $p = .001$). With virtually no variability in V-only scores at test or retest, no meaningful evaluation of test-retest reliability in that modality could be carried out. Therefore, although some learning took place when the materials were presented twice in a short period of time, test-retest reliability was fair to good.

Children with cochlear implants—Individual scores from the CI users on the multimodal sentences as a function of presentation modality are displayed in Figure 5. The data are ordered by length of device use. Note that the participant who had used her/his device for 3.6 years was not administered the multimodal sentence materials in V-only due to the inability to maintain attention during the entire testing session. As is typical in this clinical population, there was considerable variability in performance in all three modalities across participants. In contrast to the NH children, just over half of the CI users demonstrated measurable, albeit low, performance in the V-only condition. Eleven of the CI users scored higher than 50% and just four had scores below 20% correct in the A-only presentation modality, indicating that many children had reasonably good A-only speech recognition of the multimodal sentence materials. Furthermore, all but three children showed evidence of AV gain by achieving their best performance in the AV presentation modality. Although there was a wide range of ages included in the sample, no significant effects of either chronological or hearing age were found.

Figure 6 displays mean CI data plus 1 standard deviation on the lexically easy and lexically hard key words on the multimodal sentences as a function of presentation modality. The results of a two-way ANOVA with repeated measures (within-subject factors: modality, lexical difficulty) revealed that easy key words were identified more accurately than hard key words, $F(1,17) = 10.120$, $p = .005$. Further, there was a significant main effect of presentation modality, $F(2,34) = 59.089$, $p < .0001$. Planned post-hoc paired-samples t-tests revealed that all the presentation modalities were significantly different from each other ($p < .0001$ for all paired comparisons).

To preliminarily evaluate the validity of the sentence materials with a target population, we again correlated performance on the sentence materials with the isolated word recognition measures (the AV-LNT and the AV-MLNT) on the subset of children with CIs who completed the entire protocol. The results suggest that word-in-sentence recognition was strongly correlated with isolated word recognition in A-only ($r = .932$, $p = .001$ for LNT words and $r = .964$, $p < .0001$ for MLNT words). In the AV modality, word-in-sentence recognition was strongly correlated with isolated word recognition on the LNT words, $r = .881$, $p = .004$, but not the MLNT words, $r = .552$, $p = .184$. Neither of the V-only correlations was significant.

The eight participants who were administered the entire protocol completed the multimodal sentences twice to evaluate test-retest reliability. Table 4 displays individual and mean test and retest scores collapsed by lexical difficulty in each modality. Mean scores increased between 1.6 and 2.3 percentage points from test to retest across the three testing modalities. Pearson correlations across presentation modalities revealed that test and retest scores were highly correlated in all three modalities: V-only, $r = .746, p = .033$; A-only, $r = .986, p < .0001$; and AV, $r = .992, p < .0001$.

To further evaluate the precision of the multimodal sentence materials, the 95% binomial confidence interval is plotted around the A-only mean score in Figure 5. This analysis examines what constitutes a significant difference in scores across modalities on these materials. Five children's AV scores were clearly outside the A-only confidence interval and four were approaching the edge of the confidence interval. Although this is an imperfect approach, because the words are not independent of each other due to context effects, it, along with the test-retest reliability data, serves as a first approximation of test precision. Together, these preliminary analyses suggest that the materials have fair to good reliability in the NH pediatric population and high reliability for these eight children with CIs. Although further work is needed on test-retest reliability, these findings look hopeful for the larger hearing3 impaired population.

Discussion

The results from Experiment II suggest that children with CIs just under 4 years of age who have used their devices for at least 1 year can be administered the multimodal sentence materials. Considerable variability is expected across all modalities in this target population of sensory aid users. Like their NH counterparts who were tested in noise, children with CIs are sensitive to lexical difficulty effects of both isolated words and those in sentences across multiple modalities. This lends more support to the idea that children with CIs organize their mental lexicons in ways similar to children with normal hearing.

Although we could not run statistical analyses comparing data from the NH children who were tested in quiet to data from the CI children (who also were tested in quiet) due to ceiling and floor effects for the NH children, there was a trend for most of the children with CIs to display some limited ability to identify words by lipreading alone, an ability that NH children did not seem to possess. These V-only trends in performance might be artifact due to floor effects in that modality for the NH children, however, they are consistent with the results of Jerger, Tye-Murray, and Abdi (2009), who reported similar findings for a group of children with hearing loss and a group with normal hearing on a word recognition task that combined the Word Identification by Picture Identification test (Ross & Lerman, 1971) and the CAVET. NH children appear to be poor lipreaders, whereas the exposure children with CIs have to using lipreading cues either by necessity or by explicitly being taught to use them for perceptual and production purposes in aural (re)habilitation seems to assist them in identifying words in V-only.

The fact that many children were unable to identify words in the V-only condition need not limit the utility of the materials. Although one way of evaluating AV benefit is to compare performance in AV to V-only, how to go about calculating gain due to additional visual input has been a matter of debate for some time (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). One method that is widely applied in the speech perception literature is that proposed by Sumby and Pollack (1954) in which the improvement going from A-only to AV is evaluated against the total possible improvement (Ross et al.). Sumby and Pollack's method allows evaluation of gain from visual speech separate from V-only performance. Therefore, although it is expected that children, particularly those with normal hearing, will struggle in the V-only modality on this open-set word-in-sentence recognition task, this need not limit

the utility of the materials themselves. Difficulty in the V-only condition also need not preclude testing in that modality. Although the V-only scores for the CI children were low, many of them were able to identify some words through only visual speech cues. In fact, there was one child with a CI (one of the children who had her/his device for 6.2 years) who primarily relied on visual cues. This kind of information would be lost if the materials were not presented in V-only. This suggests that it is worthwhile to use the materials in all three modalities with children with CIs.

There were two instances where sentence recognition was not correlated with isolated word recognition (the NH children in A-only when comparing the sentences with the AV-LNT and the CI children in AV when comparing the sentences with the AV-MLNT), but the remaining six significant correlations between performance on these multimodal sentence materials and the AV-LNT and AV-MLNT suggest, at least preliminarily, that the sentence materials are a valid measure of spoken language. Both the sentence materials and the isolated word recognition tests appear to measure similar underlying processes. Further development of the materials, for example increasing the number of scored items, might serve to further improve the materials' validity.

Both Experiments show promise for the materials' ability to reliably measure multimodal sentence recognition. However, the current investigation leaves open some areas for further test development. For example, although the test-retest correlations were significant for NH children, they were lower than is desirable for new test development (which typically is approximately .70 to .80 [Nunnally & Bernstein, 1994]). The reliability coefficients for the children with CIs were at this level or higher, but this is based on just eight children's data. In addition, NH children's test-retest scores increased in A-only and AV by approximately 11 and 8 percentage points, respectively, suggesting some learning effects. The experimental design used did not allow for examination of when these effects plateau. However, data from the eight children with CIs preliminarily suggest very limited learning effects in that population (mean scores increased by approximately 2 percentage points across modalities). In addition to testing more children with cochlear implants, future investigations into the development of these multimodal materials will further evaluate approaches that might improve test-retest reliability. One method is to develop practice materials that could be used to first achieve stable performance before testing commences. Another method is to increase list length and thus, the number of scored items per list. However, it is possible that the psychometric properties that could be improved with longer lists might be restricted by children's limited attention spans. Further, multimodal assessments involve 2–3 times more testing than typical A-only evaluations require. And because most children with CIs undergo a battery of tests, the additional testing time required of multimodal testing might not be trivial. Therefore, there likely is a trade-off between potential improved reliability and possible attention effects that would be evaluated in future development of the materials.

Another area that is important for further test development is evaluating list equivalency in children. In the current investigation, list equivalency was tested on adults. It is possible that differences between children's and adults' speech perception might result in differences in list equivalency in children. Future investigations will examine this possibility.

With further work in developing these materials into a test of spoken sentence recognition, in particular examining test-retest reliability and list equivalency in children, these new materials have promise for tracking performance over time, a critical component of any test used to follow speech and language development following cochlear implantation.

Conclusion

This investigation was the first step toward a long-term goal of developing these multimodal sentence materials into a test of audiovisual spoken word recognition, the Audiovisual – Lexical Neighborhood Sentence Test (AV-LNST). In doing so, we sought: to identify a lower age limit for which the materials can be used reliably; to examine presentation modality and lexical difficulty effects of the new materials; and to preliminarily evaluate the materials' validity and test-retest reliability. Both NH children and those with CIs were administered the materials, along with multimodal tests of isolated word recognition.

NH children as young as 3.25 years were able to complete the multimodal sentence testing with little difficulty in the A-only and AV modalities. All of the children, even the oldest 5-year-olds with normal hearing, struggled in the V-only modality, suggesting that young NH children are not able to use lipreading cues when they are not accompanied by corroborating acoustic speech cues. In contrast, there was a trend for children with cochlear implants, regardless of their communication modality (OC or TC), to be able to use lipreading cues more effectively than NH children, albeit to a limited extent. These results are consistent with those of Jerger et al. (2009). The children with cochlear implants ranged in age from 3.8 to 9.6 years of age, a much wider age range than the NH children, and had used their devices for between 1 and 7.75 years. However, there were no age-related (chronological or hearing age) performance trends for the CI users. All but four of the CI users had A-only scores greater than 20% correct and nearly all showed some degree of AV benefit. These data suggest that children with CIs just under 4 years of age who have had their devices for at least 1 year can be administered the multimodal sentence materials, although performance varies considerably across children, similar to other spoken word recognition tests used with pediatric CI users. It certainly is possible even younger children with CIs could be administered the multimodal sentences; we, however, did not test any children with CIs younger than 3.86 years.

All participant groups were sensitive to presentation modality effects on the multimodal sentences. In general, the highest scores were achieved when listeners had access to both the auditory and visual speech cues simultaneously, demonstrating their ability to integrate cues from both modalities to improve performance over one modality alone. Intermediate scores were obtained when participants had access to the auditory speech cues only and the lowest scores were recorded when the participants only had access to the visual speech cues. NH children were particularly poor at identifying any words in the V-only modality, whereas the CI users showed some limited ability to identify words using visual speech cues exclusively.

Previous investigations have reported lexical difficulty effects for words in sentences presented in the A-only modality in children with normal hearing and those with CIs (Eisenberg et al., 2002) and in NH adults (Bell & Wilson, 2001). Kaiser et al. (2003) reported lexical difficulty effects for multimodally presented isolated words spoken by multiple talkers in NH adults and those with CIs and in our preliminary study, we found lexical difficulty effects for words in sentences presented multimodally in NH adults. The results of the current investigation extend the results of these previous investigations: lexically easy words in sentences were identified with higher accuracy than lexically hard words in sentences across modalities by children with normal hearing and those with CIs. However, the magnitude of the lexical effects varied with presentation modality. Scores were so low in the visual modality that examination of lexical difficulty effects by NH participants was impossible. CI users, on the other hand, did have better recognition of lexically easy words than lexically hard words, even in the V-only modality. For both the NH children and those with CIs, lexical difficulty effects were greater for the A-only modality than the AV modality, consistent with the expectation that the addition of visual

speech cues reduces the competition for lexical selection. The finding of lexical effects across modalities also is consistent with results of other investigators who have suggested that lexical difficulty effects on speech perception, at least for individual phonemes, are likely modality independent (e.g., Brancazio, 2004). That lexical difficulty effects were found in NH children and children with CIs across modalities also suggests that the multimodal sentence recognition materials are sensitive to differences in how listeners organize and/or access their mental lexicons and thus, could be used to evaluate these abilities in children with hearing loss.

The significance of this work lies in its potential to fill the need for carefully designed test materials and methods to investigate auditory and visual cue integration for speech, particularly the influence of audiovisual perception on the organization of and access to words stored in memory. Furthermore, this work provides a new tool to examine AV integration skills of sensory aid users, a neglected area of research, particularly in children. This will be particularly useful clinically in light of recent findings suggesting that pre-implant V-only and AV speech perception performance reliably predicts post-implantation success and benefit in prelingually deaf children (Bergeson & Pisoni, 2004; Bergeson et al., 2003; Lachs et al., 2001). Although additional work is needed to fully evaluate test-retest reliability, validity, and list equivalency, the results of this investigation are hopeful that the materials could serve as a valid and reliable measure of multimodal sentence recognition in children with hearing loss. A long-term goal of this work is to develop benchmarks that could aid clinicians in making evidence-based decisions regarding sensory aids and aural (re-)habilitation. Finally, work on AV speech perception is important because speech is a multimodal signal that is processed by multiple systems in the brain – not just auditory centers, but also visual/motor centers, as well (Calvert & Lewis, 2004). Therefore, materials that can be used to further investigate the elaborate system humans possess for processing speech are critical for moving this area of research forward.

Acknowledgments

Funding Sources: NIH-NIDCD (T32 00012, R01 DC00064, 5R01 DC008875), the American Speech-Language-Hearing Foundation, and the American Hearing Research Foundation

We appreciate the contributions of: Amy Martinez for making the stimulus-shaped noise; Elicia Roos, Meredith Newman, Lisa Burkhartzmeyer, Lindsay Mitchell Fletcher, Stacy Howell, Bea Smith, Andrea Moore, Katie Wallace, Charis Fong, Tricia Ray, Spencer Smith, Kathy Carbonell, Shirley Henning, and Bethany Gehrlein for assisting in data collection; and Tessa Bent and Luis Hernandez for their contributions to the color correction portion of the investigation. Portions of this work were presented at the fall meeting of the Acoustical Society of America, Minneapolis, MN (October, 2005), the annual meeting of the American-Speech- Language-Hearing Association, San Diego, CA (November, 2005), the 11th International Conference on Cochlear Implants in Children, Charlotte, NC (April, 2007), and the 2008 Newborn Hearing Screening Conference, Cernobbio, Italy (June 2008). This work was funded by the American Speech-Language-Hearing Foundation, the American Hearing Research Foundation and the NIH-NIDCD (T32 DC00012, R01 DC00064, and 5R01 DC008875).

References

- Adobe. Adobe Audition. San Jose, CA: Adobe Systems, Inc; 2005.
- American National Standards Institute. Specification for audiometers (ANSI S3.6–1989). New York: Acoustical Society of America; 1989.
- Anastasi, A. Psychological Testing. New York: The Macmillan Company; 1961. Test reliability; p. 105-134.
- Apple Computer. Final Cut Pro. Cupertino, CA: Apple Computer, Inc; 2004.
- Bell TS, Wilson RH. Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology*. 2001; 12:514–522. [PubMed: 11791938]

- Bergeson, TR.; Pisoni, DB. Audiovisual speech perception in deaf adults and children following cochlear implantation. In: Calvert, G.; Spence, C.; Stein, BE., editors. The handbook of multisensory processes. Cambridge, MA: MIT Press; 2004. p. 749-771.
- 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; 103:347–370. monograph.
- Bergeson TR, Pisoni DB, Davis RAO. Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. Ear and Hearing. 2005; 26:149–164. [PubMed: 15809542]
- Boothroyd A. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. Journal of Speech and Hearing Research. 1984; 27:134–144. [PubMed: 6716999]
- Boothroyd, A. Profound deafness. In: Tyler, RS., editor. Cochlear Implants: Audiological Foundations. San Diego, CA: Singular Publishing Group, Inc; 1993. p. 1-33.
- Boothroyd A, Nitttrouer S. Mathematical treatment of context effects in phoneme and word recognition. Journal of the Acoustical Society of America. 1988; 84:101–114. [PubMed: 3411038]
- Brancazio L. Lexical influences in audiovisual speech perception. Journal of Experimental Psychology: Human Perception and Performance. 2004; 30:445–463. [PubMed: 15161378]
- Brandy WT. Reliability of voice tests in speech discrimination. Journal of Speech and Hearing Research. 1966; 9:461–465.
- Calvert, GA.; Lewis, JW. Hemodynamic studies of audiovisual interactions. In: Calvert, G.; Spence, C.; Stein, BE., editors. The handbook of multisensory processes. Cambridge, MA: MIT Press; 2004. p. 483-502.
- Carney AE, Osberger MJ, Carney E, Robbins AM, Renshaw J, Miyamoto RT. A comparison of speech discrimination with cochlear implants and tactile aids. Journal of the Acoustical Society of America. 1993; 94:2036–2049. [PubMed: 8227746]
- Carter AS, Wilson RH. Lexical effects on dichotic word recognition in young and elderly listeners. Journal of the American Academy of Audiology. 2001; 12:86–100. [PubMed: 11261462]
- Desjardins RN, Rogers J, Werker JF. An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. Journal of Experimental Child Psychology. 1997; 66:85–110. [PubMed: 9226935]
- Dirks DD, Takayanagi S, Moshfegh A, Noffsinger PD, Fausti SA. Examination of the neighborhood activation theory in normal and hearing-impaired listeners. Ear and Hearing. 2001; 22:1–17. [PubMed: 11271971]
- Dunn, ML.; Dunn, ML. Peabody Picture Vocabulary Test. Pearson Assessments; Bloomington, MN: 1997.
- Eisenberg LS, Dirks DD. Reliability and sensitivity of paired comparisons and category rating in children. Journal of Speech and Hearing Research. 1995; 38:1157–1167. [PubMed: 8558884]
- 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; 23:450–462. [PubMed: 12411778]
- Erber NP. Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech and Hearing Research. 1969; 12:423–425. [PubMed: 5808871]
- Erber NP. Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. Journal of Speech and Hearing Disorders. 1972; 15:413–422.
- Erber NP. Auditory-visual perception of speech. Journal of Speech and Hearing Disorders. 1975; 40:481–492. [PubMed: 1234963]
- Geers A, Brenner C, Davidson L. Factors associated with development of speech perception skills in children implanted by age five. Ear and Hearing. 2003; 24:24S–35S. [PubMed: 12612478]
- Gelfand, SA. A clinical speech recognition method to optimize reliability and efficiency. Paper presented at the Convention of American Academy of Audiology; Phoenix, AZ. 1993.
- Goldman, R.; Fristoe, M. Goldman-Fristoe Test of Articulation – 2. Pearson Assessments; Bloomington, MN: 2000.

- Jerger S, Tye-Murray N, Abdi H. Role of visual speech in phonological processing by children with hearing loss. *Journal of Speech, Language and Hearing Research*. 2009; 52:412–434.
- Jerger S, Lewis S, Hawkins J, Jerger J. Pediatric speech intelligibility test, I. Generation of test materials. *International Journal of Pediatric Otorhinolaryngology*. 1980; 2:217–230. [PubMed: 7188063]
- Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech and Hearing Research*. 2003; 46:390–404.
- Kirk, KI.; Pisoni, DB.; Miyamoto, RT. Lexical discrimination by children with cochlear implants: Effects of age at implantation and communication mode. In: Waltzman, S.; Cohen, N., editors. *Proceedings of the Vth International Cochlear Implant conference*. New York: Thieme Medical Publishers; 1997.
- Kirk KI, Pisoni DB, Osberger MJ. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing*. 1995; 16:470–481. [PubMed: 8654902]
- Krull V, Choi S, Kirk KI, Prusick L, French B. Lexical effects on spoken- word recognition in children with normal hearing. *Ear and Hearing*. 2010; 31:102–114. [PubMed: 19701087]
- Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. *Science*. 1982; 218:1138–1141. [PubMed: 7146899]
- Lachs L, Pisoni DB, Kirk KI. Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing*. 2001; 22:236–251. [PubMed: 11409859]
- Logan, JS. *Research on Spoken Language Processing Technical Report N. 8*. Bloomington, IN: Indiana University; 1992. A computational analysis of young children's lexicons.
- Luce, PA. *Research on Speech Perception Technical Report No. 6*. Bloomington, IN: Indiana University; 1986. Neighborhoods of words in the mental lexicon.
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*. 1998; 19:1–36. [PubMed: 9504270]
- MacWhinney B, Snow C. The child language data exchange system. *Journal of Child Language*. 1985; 12:271–296. [PubMed: 4019604]
- Massaro DW, Cohen MM. Perceiving talking faces. *Current Directions in Psychological Science*. 1995; 4:104–109.
- McGurk H, McDonald J. Hearing lips and seeing voices. *Nature*. 1976; 264:746–748. [PubMed: 1012311]
- Moeller, MP.; Osberger, MJ.; Eccarius, M. Receptive language skills. In: Osberger, MJ., editor. *Language and Learning Skills of Hearing-Impaired Students*, ASHA Monogram. Vol. 23. 1986. p. 41-54.
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. 3. New York: McGraw-Hill; 1994.
- Penrod J. Talker effects on word discrimination scores of adults with sensorineural hearing impairment. *Journal of Speech and Hearing Disorders*. 1979; 44:340–349. [PubMed: 480938]
- Robbins, AM.; Renshaw, JJ.; Osberger, MJ. *Common Phrases Test*. Indianapolis, IN: Indiana University School of Medicine; 1995.
- Ross, M.; Lerman, J. *Word intelligibility by picture identification*. Pittsburgh, PA: Stanwix House; 1971.
- Ross LA, Saint-Amour D, Leavitt V, Javitt D, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech. *Cerebral Cortex*. 2007; 17:1147–1153. [PubMed: 16785256]
- Shannon RV, Zang FG, Wygonski J, Kamath V, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. [PubMed: 7569981]
- Spehar R, Tye-Murray N, Sommers M. Time-compressed visual speech and age: a first report. *Ear and Hearing*. 2004; 25:565–572. [PubMed: 15604917]
- Sumby WH, Pollack I. Visual contribution of speech intelligibility in noise. *Journal of the Acoustical Society of America*. 1954; 26:212–215.
- Summerfield, Q. Some preliminaries to a comprehensive account of audio-visual speech perception. In: Dodd, B.; Campbell, R., editors. *Hearing by Eye: The Psychology of Lip-Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987. p. 3-51.

- Thornton A, Raffin MJM. Speech discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research*. 1978; 21:507–518. [PubMed: 713519]
- Tillman, TW.; Carhart, R. Northwestern University Auditory Test No. 6. Brooks AFB, TX: USAF School of Aerospace Med; 1966. An expanded test for speech discrimination utilizing CNC monosyllabic words. Tech. Report SAM-TR-66-55
- Tye-Murray, N.; Geers, A. Children's Audio-Visual Enhancement Test. St. Louis, MO: Central Institute for the Deaf; 2001.
- Tye-Murray N, Sommers M, Spehar B. Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*. 2007; 11:233–241. [PubMed: 18003867]
- Tyler, RS.; Fryauf-Bertschy, H.; Kelsay, D. Audiovisual Feature Test for Young Children. Iowa City: The University of Iowa; 1991.
- Tyler RS, Fryauf-Bertschy H, Kelsay DMR, Gantz BJ, Woodworth GP, Parkinson A. Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology – Head and Neck Surgery*. 1997; 117:180–187. [PubMed: 9334763]
- Williams, KT. Expressive Vocabulary Test. Pearson Assessments; Bloomington, MN: 1997.

Appendix A

Table A1

Mean proportion key words correct for each sentence and each entire list in all three presentation modalities (A-only, V-only, and AV) from Phase I of the Preliminary Experiment.

List	Sentence	A-only	V-only	AV
1	E1	0.40	0.20	0.83
	E2	0.43	0.00	0.77
	E12	0.60	0.37	0.73
	E14	0.40	0.47	0.97
	H3	0.20	0.00	0.37
	H10	0.17	0.13	0.60
	H18	0.23	0.20	0.77
	H20	0.73	0.00	0.73
	Mean	0.40	0.17	0.72
2	E4	0.47	0.13	0.57
	E9	0.87	0.40	1.00
	E19	0.53	0.13	0.90
	E20	0.33	0.10	0.70
	H1	0.13	0.07	0.63
	H4	0.20	0.10	0.50
	H11	0.37	0.03	0.73
	H15	0.57	0.33	0.83
	Mean	0.43	0.16	0.73
3	E8	0.57	0.17	0.80
	E15	0.50	0.03	0.87
	E18	0.60	0.13	0.90
	E24	0.20	0.20	0.50
	H2	0.23	0.13	0.70
	H6	0.23	0.07	0.57

List	Sentence	A-only	V-only	AV
4	H8	0.67	0.30	0.77
	H17	0.07	0.17	0.60
	Mean	0.38	0.15	0.71
	E3	0.67	0.13	0.87
	E5	0.43	0.00	0.73
	E10	0.87	0.43	1.00
	E16	0.30	0.50	0.87
	H9	0.37	0.00	0.57
	H12	0.30	0.07	0.53
	H21	0.37	0.00	0.57
	H25	0.17	0.37	0.70
	Mean	0.43	0.19	0.72
5	E6	0.40	0.23	0.80
	E7	0.37	0.03	0.50
	E22	0.90	0.53	1.00
	E25	0.23	0.03	1.00
	H5	0.40	0.10	0.37
	H7	0.10	0.30	0.77
	H19	0.47	0.00	0.80
	H22	0.20	0.17	0.33
	Mean	0.38	0.18	0.70
6	E11	0.70	0.27	0.80
	E17	0.13	0.13	0.53
	E21	0.40	0.30	0.83
	E23	0.57	0.23	0.93
	H13	0.47	0.10	0.60
	H14	0.27	0.00	0.67
	H16	0.30	0.13	0.40
	H24	0.37	0.10	0.80
	Mean	0.40	0.16	0.70

* V-only = visual-only; A-only = auditory-only; AV = audiovisual

Figure A1.

Mean percent key words correct (+1 standard deviation) for individual sentences in each of three modalities (V-only: top panels; A-only: middle panels; and AV: bottom panels) from Phase I of the Preliminary Experiment. Sentences with easy key words and those with hard key words are displayed in the left- and right-hand columns, respectively. Stars indicate sentences that were excluded from the final word lists.

Figure A2.

Mean percent key words correct (+1 standard deviation) for each sentence list from the multimodal sentence materials from Phase II of the Preliminary Experiment. The top, middle, and bottom panels display data from V-only, A-only and AV, respectively.

Figure A3.

Mean percent key words correct (+1 standard deviation) on the multimodal sentence materials for each presentation modality and color -corrected condition from Phase III of the Preliminary Experiment. The unfilled bars represent the original stimuli and the filled bars the color-corrected stimuli.

Figure 1. Individual percent key words correct on the multimodal sentence materials for normal-hearing children from Experiment I in auditory-only (top panel) and audiovisual (bottom panel) presentation modalities. Data are ordered from youngest age to oldest age at test. Three-year-olds' data are displayed with gray bars, 4-year-olds' with unfilled bars, and 5-year-olds' with black bars.

Figure 2.

Mean percent key words correct (+1 standard deviation) on the multimodal sentence materials for normal-hearing children from Experiment I for each age group as a function of lexical difficulty in visual-only (top panel), auditory-only (middle panel), and audiovisual (bottom panel) presentation modalities.

Figure 3. Individual percent key words correct on the multimodal sentence materials for normal-hearing children from Experiment II in auditory-only (top panel) and audiovisual (bottom panel) presentation modalities. Data are ordered from youngest to oldest age at test. Three-year-olds' data are displayed with gray bars, 4-year-olds' with unfilled bars, and 5-year-olds' with black bars.

Figure 4. Mean percent key words correct (+1 standard deviation) on the multimodal sentence materials for normal-hearing children from Experiment II for each age group as a function of lexical difficulty in visual-only (top panel), auditory-only (middle panel), and audiovisual (bottom panel) presentation modalities.

Figure 5.

Individual percent key words correct on the multimodal sentence materials for children with cochlear implants from Experiment II as a function of presentation modality (gray bars reflect the visual-only presentation modality, unfilled bars denote the auditory-only presentation modality, and black bars reflect the audiovisual presentation modality). Data are ordered from fewest to greatest number of years using a cochlear implant. Dashed lines represent the 95% confidence interval around the A-only mean.

Figure 6.

Mean percent key words correct (+1 standard deviation) on the multimodal sentence materials for children with cochlear implants from Experiment II for lexically easy and lexically hard key words as a function of presentation modality (gray bars reflect the visual-only presentation modality, unfilled bars denote the auditory-only presentation modality, and black bars reflect the audiovisual presentation modality).

Table 1

Normal-hearing participant demographics

	n	Age Range(mos)	Mean Age (mos)
Experiment I			
3-year-olds	23	36–47	41.3
4-year-olds	20	48–59	53.1
5-year-olds	14	60–71	64.5
Experiment II			
3-year-olds	9	40–46	43.9
4-year-olds	10	49–59	53.5
5-year-olds	10	60–71	66.1

*
mos = months

Table 2

Experiment I: mean percent key words correct for the original and color-corrected multimodal sentence recognition materials

	V-only	A-only	AV
3-year-olds			
Original (n=16)	0.00	82.29	84.25
Color-corrected(n=7)	0.00	88.10	84.82
4-year-olds			
Original (n=11)	0.00	93.37	93.53
Color-corrected(n=9)	0.46	93.06	92.59
5-year-olds			
Original (n=8)	0.26	95.05	95.57
Color-corrected(n=6)	0.00	97.92	96.53

* V-only = visual-only; A-only = auditory-only; AV = audiovisual

Table 3

Experiment II CI participant demographics

Subject ID	Age at Implant (Years)	Age at Testing (Years)	Length of CI Use (Years)	Best PTA(dB HL)	Comm Mode	Device Condition	Other
1	1.42	5.00	2.67	118	TC	Mon. CI	
2	1.83	6.42	4.25	82	OC	Mon. CI	
3	3.25	7.67	4.83	118	TC	Mon. CI	
4	0.92	7.08	5.25	118	OC	Mon. CI	
5	2.17	6.33	5.42	118	TC	Bin. CI	LBCI = 0.67 yrs
6	2.33	9.42	6.17	100	TC	Mon. CI	
7	2.83	7.58	6.17	103	OC	Bin. CI	LBCI = 0.67 yrs
8	1.83	9.58	7.75	107	TC	Mon. CI	
9	0.86	4.75	3.89	67	OC	CI + HA	
10	0.99	5.41	4.43	113	OC	Mon. CI	
11	1.59	7.75	6.16	113	OC	Mon. CI	
12	1.79	3.86	2.07	95	OC	Mon. CI	
13	1.15	5.70	4.55	102	OC	Mon. CI	
14	1.40	4.96	3.56	118	OC	Mon. CI	
15	2.80	3.79	0.99	92	OC	Mon. CI	
16	2.54	4.47	1.93	85	OC	Mon. CI	
17	2.98	8.96	5.98	118	OC	Mon. CI	
18	2.41	3.88	1.47	113	OC	Mon. CI	
19	2.39	5.55	3.16	58	OC	Mon. CI	
Mean	1.97	6.20	4.23	102.07			
SD	0.74	1.88	1.85	18.25			

* CI = cochlear implant; PTA = pure-tone average; Comm = Communication; TC = Total communication; OC = oral communication; HA = hearing aid; Mon. = monaural; Bin. = binaural; LBCI = length of bilateral cochlear implant use; yrs = years; SD = standard deviation

Table 4

CI individual and mean scores at test and retest collapsed by lexical difficulty in each modality

Subject ID	V-only Test	V-only Retest	A-only Test	A-only Retest	AV Test	AV Retest
1	4%	13%	85%	94%	96%	98%
2	6%	6%	69%	73%	73%	75%
3	6%	6%	90%	90%	98%	100%
4	4%	0%	15%	10%	23%	15%
5	8%	10%	88%	98%	96%	100%
6	4%	13%	23%	6%	29%	38%
7	19%	23%	2%	4%	19%	23%
8	6%	6%	65%	73%	94%	96%
Mean	7.3%	9.6%	54.4%	56.0%	65.9%	68.0%