

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

**Talker Discrimination by Prelingually Deaf Children with Cochlear Implants:
Some Preliminary Results¹**

Miranda Cleary and David B. Pisoni²

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

¹ This work was supported by NIDCD Research Grant DC00111 and NIDCD Training Grant DC00012 to Indiana University. We gratefully acknowledge the kind cooperation of Dr. Ann Geers and C. Brenner in making this project possible. We also thank T. Wood and C. Dillon for their help during data collection.

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

Talker Discrimination by Prelingually Deaf Children with Cochlear Implants: Some Preliminary Results

Abstract. Forty-four school-age children (ages 8 and 9) each of whom had used their Nucleus-22 device for at least four years were tested to assess their ability to discriminate differences between recorded pairs of female voices uttering sentences. Children were asked to respond “same voice” or “different voice” for each trial. The correct answer was “same” for half of the trials, and “different” on the remaining trials. Two conditions were tested. In one condition, the linguistic content of the sentence was always held constant and only the voice of the talker was permitted to vary from trial to trial. In another condition, the linguistic content of the utterance was also varied so that to correctly respond “same” the child needed to recognize that two different sentences were spoken by the same talker. Data from a group of 21 normal-hearing five-year-old children with the same stimulus materials were used to establish that these tasks were well within the capabilities of children without hearing impairment (mean proportion correct on the varied sentence condition = 89%). For the CI children tested, in the “fixed sentence condition” the mean proportion correct was 67% which although significantly different from the score expected by chance of 50%, suggests that the CI children found this discrimination task rather difficult. In the “varied sentence condition,” however, the mean proportion correct was only 57%. Although this value is statistically above chance, the results suggest that the CI users were essentially unable to recognize an unfamiliar talker’s voice when the content of the paired sentences differed. These findings are discussed in terms of how cochlear implants transmit the speech signal, the contribution of various acoustic cues to talker identity, and known interactions between the perception of linguistic and indexical properties of spoken words. Correlations between performance on the “fixed sentence” version of the task and other processing measures are also reported for the CI group.

Introduction

A large body of research has shown that normal-hearing listeners are sensitive to properties in the acoustic speech signal that provide information about the speech producer. These properties are sometimes referred to as “indexical” properties of the signal and can convey, though imperfectly, information about the talker’s gender, age, regional background, emotional state of mind, etc. (Kramer, 1963; McGehee, 1937; Ptacek & Sanders, 1966). Indexical information is usually conceptualized as contrasting with “linguistic” information about the intended pattern of phonemes/phonemic contrasts (Pisoni, 1997). Since linguistic and indexical information are both simultaneously encoded in the same physical acoustic energy, the primary question of interest to speech researchers is how the parallel extraction of these two types of information takes place, and the degree to which these processes interact with each other.

The ability to use indexical information to perceptually discriminate between the speech of different talkers is often taken for granted in communicative situations. In order for a listener to interpret what is being said in the larger context of a spoken conversation, it is usually important to know who is speaking. In situations involving discussion among a large number of people, a listener must keep track of the current speaker and register a change of speaker when it occurs. The difficulty of this task is increased when associated visual cues are unavailable such as in the case of communicating via telephone, listening to the radio, or if one has momentarily turned away from the speaker’s face.

As most normal-hearing persons can attest to, even under ideal listening conditions, confusions between talkers sometimes occur. For a hearing-impaired listener using a cochlear implant, these problems are compounded. Research has shown that perceptual errors and communicative breakdowns are more likely to occur even for generally successful users of cochlear implants when they are faced with having to rapidly decode the speech of multiple talkers (Sommers, Kirk, & Pisoni, 1997). Even though generally identified as a problem, relatively little research has investigated how cochlear implants convey the acoustic properties associated with indexical information or how cochlear implant users perceive and interpret this information. Instead, understandably, the focus has been primarily on the perception of acoustic properties known to contribute towards making phonemic distinctions.

In the current study, we asked pediatric users of cochlear implants to make judgments about whether or not pairs of recorded sentences were spoken by the same talker. Two conditions were tested. In one condition, the linguistic content of the paired utterances was identical (referred to henceforth as the “fixed sentence condition”). In the other condition, the linguistic content of the two sentences always differed (the “varied sentence condition”). In this latter condition it was necessary for the listener to be able to identify two separate utterances as spoken by either the same talker or by two different talkers. Since the talkers used in this study were previously unfamiliar, we reasoned that in order to perform the varied sentence condition it would be necessary for the listeners to form a representation, or expectation, of what the speaker of the first utterance in each pair would sound like in a subsequent, linguistically different utterance. Upon hearing the second utterance, the listener would then be able to make a judgment about whether the two sentences were, in fact, spoken by the same talker or by two different talkers.

We attempted to minimize the potential difficulty of this second task condition through several methodological simplifications. In addition to similarities between talkers, key factors in determining the difficulty of a talker discrimination task are the amount of information provided per talker, and the number of different talkers among which the listener is asked to discriminate (Murray & Cort, 1971; Pollack, Pickett, & Sumby, 1954). We therefore used relatively long sentence-length stimuli and only a very small set of three female talkers.

Our manipulation of the linguistic content of the utterances was motivated in part by recent findings on the interaction in perception between the linguistic and indexical properties of speech. Research has shown that the speed of identification of linguistic information (i.e., the phonemic content of the message) is influenced by the presence of indexical variability, and that the perception of certain types of indexical information is, in turn, influenced by linguistic variability (e.g., Miller, 1978; Mullennix & Pisoni, 1990). It is this second relationship that is being explored in the present study, that is, the effect of linguistic variability on judgments about the indexical properties of speech.

A number of related studies involving both normal-hearing and hearing-impaired children have been conducted. In a series of published papers, Jerger and her colleagues investigated whether children demonstrate the same degree of interaction between linguistic and indexical processing as shown by adults. In one of the earlier studies in this series, Jerger et al. (1993) asked normal-hearing children three to six years of age to decide whether a spondee (e.g., “ice cream”) was spoken in a male or a female voice. The experimenters then varied whether the judgment was made under the condition of particular words being *consistently associated* with either the male voice or the female voice, or with no predictable association present. They also included a control condition in which only a single word was used for all male/female voice judgment trials. Jerger et al. found that the presence of unpredictable variability in the association between a particular voice and particular word had a significant effect on reaction times in the task, slowing the decision speed by about 95 ms relative to the control condition. The predictable

variation condition was also slower on average than the control condition, but just barely so, on the order of about 30 ms.

In a subsequent study, Jerger, Martin, Pearson, & Dihn (1995) used a similar task with 40 school-age children diagnosed with mild to severe hearing impairments. All of the children used conventional hearing aids and 90% of the group were believed to have acquired their hearing impairment before age 2. According to Jerger et al., these children were able, when using their hearing aids, to identify the two talkers used in the study as either male or female with very high accuracy. The reaction time results from this study suggested that the hearing aid users, particularly the younger children, found it easier to ignore linguistic information while making judgments about indexical information than did a comparison group of normal-hearing children. That is to say, the hearing-impaired children's speeded judgments regarding talker gender showed less interference from unpredictable variation in the linguistic dimension than did the judgments of normal-hearing children.

Although the methodology used in the present study differs considerably from Jerger et al.'s, the theoretical issues involved are similar. Specifically, we are interested in assessing how well experienced pediatric users of cochlear implants are able to ignore (or generalize beyond) linguistic variability when asked to discriminate between the voices of different talkers. In addition to this theoretical issue which deals with processing strategies, we are also interested more generally in the perception of similarity between talkers, and in determining which acoustic properties cochlear implant users are best able to use to differentiate between talkers. The present study begins to address some of these larger issues.

It has long been known that even given relatively short samples of speech, listeners can readily form impressions (though not necessarily accurate impressions) of the talker's gender, age, emotional state, and linguistic background (Kreiman, 1997). An early study by Mann, Diamond and Carey (1979), reported that the ability to recognize briefly studied unfamiliar voices continues to improve throughout the school-age years and plateaus around adult levels by adolescence. Despite much research in this area, however, the perceptual factors that children and adults use to distinguish between different talkers are far from completely understood. Speech researchers have identified a number of measurable acoustic dimensions along which talkers differ, including fundamental frequency (range, average value, and irregularities in) and absolute formant frequencies, but the degree to which listeners "perceptually weigh" the importance of each dimension has not yet been resolved (see Kreiman, 1997 for discussion).

Our use of the term "indexical" in this report admittedly encompasses a great many aspects of the speech signal, not all of which necessarily relate to cross-talker variability. An important consideration in research on voice perception is that listeners may use a variety of encoding and processing strategies and/or resources in order to accomplish the task of talker discrimination (Kreiman, 1997). Thus, the acoustic cues that yield indexical information are, to a degree, functionally defined according to whether or not they can be shown to have been used by a perceiver for the particular task. In a laboratory situation, the experimenter's selection of the stimulus set largely constrains what possible strategies for discriminating between talkers can be utilized.

Because of the range of strategies listeners may use to discriminate between talkers, it is not straightforward to determine which prior research on the perceptual skills of cochlear implant users is relevant to the present task. In particular, the degree to which earlier studies of cochlear implant users' perception of "supra-segmental" properties of speech such as intonation are relevant to our results is unclear. For example, the prosodic contrasts involved in word stress and intonation are, in fact, linguistic in nature, and depend more on perception of relative f_0 values (among other factors) than on absolute f_0 levels such as might provide cues to the identity of the speaker. On the other hand, absolute f_0 is fairly well established as a very basic perceptual dimension along which listeners tend to discriminate different

talkers, assuming such variation is present. As such, perception of fundamental frequency even as a linguistic entity may be an important factor to consider. In general, the existing literature indicates that for children with cochlear implants, the large acoustic differences in f_0 that distinguish declarative versus WH-question intonation, and male from female speech are fairly easily discriminated, even before phonemic distinctions are readily made (e.g., Osberger et al., 1991). Although much of the available research on f_0 perception was carried out with the early implant designs, many of these findings should be generalizable to the current generation of cochlear implants as well.

Expectations regarding the children's performance on our talker discrimination task were based in part on the children's history of CI use and the design of the device itself. The eight- and nine-year old children who participated in this study had all used a Nucleus 22 multi-channel cochlear implant for at least four years, and the majority of children at time of testing were using a coding strategy that is capable of representing fairly detailed spectral information about the speech signal. The Nucleus 22 device and its associated spectral peak coding strategy (SPEAK) use a filter bank of 20 filters with center frequencies ranging between 250 Hz to 10,000 Hz (with variable bandwidths between filters) to continually process the incoming waveform (Loizou, 1998). Depending on the distribution of the spectral information, 5 to 10 of these filters (those that best correspond to amplitude peaks in the spectrum) are selected to pass along information to the internal part of the implant. The selection of the active filters is engineered so that vowel sounds retain more spectral detail, while sounds such as fricatives use a smaller number of spectral peaks. According to Loizou (1998), approximately six maxima are used on average. The rate at which pulses are sent via the individual electrodes is dependent on the number of maxima being conveyed and parameters of the individual patient's mapping, but tends to range between 180-300 cycles per second. There is a tradeoff based on available current such that if more maxima are being conveyed, the rate of stimulation is reduced. Fewer maxima are associated with faster stimulation rates and thus better time resolution of the original acoustic information. The SPEAK strategy uses interleaved pulses, such that about every 4 ms, the selected subset of electrodes are stimulated in descending amplitude order. The amplitude of each pulse is governed by the amplitude envelope of the signal issuing from the particular bandpass filter with which it is associated. Unlike previous coding strategies which tried to provide an independent temporal cue to fundamental frequency via stimulation rate, the SPEAK coding strategy, like most other "state of the art" coding strategies, does not use stimulation rate to code f_0 but instead leaves fundamental frequency to be decoded by the listener from patterns in the waveform and spectra (Jones, McDermott, Seligman, & Millar, 1995; Seligman & McDermott, 1995).

Given the design of the device and the children's history of use we judged that at least some of the pediatric CI users would be able to make the simple discriminations presented under the "fixed sentence" condition. Because the linguistic content of the sentence was held constant across all comparisons, inter-talker differences should constitute the primary source of any perceived acoustic variation between sentences. For the "varied sentence" condition it was anticipated that the task would prove more difficult, since a generalizable representation of each talker's voice is presumably necessary to accomplish the task. However, if the children were able to ignore the linguistic variability as directed, for the purpose of the task at hand, we judged that the signal provided by the implant should be sufficient to permit some children to form the necessary representations of the different voices. Data from younger normal-hearing children would help us to judge the overall difficulty of this "varied sentence" condition.

Method

Participants

Normal-hearing (NH) Preschoolers. Twenty-one normal-hearing preschoolers were tested as part of a larger project being conducted at the Indiana University Speech Research Laboratory. Thirteen

female and eight male children participated. The children ranged in age from 5;3 to 5;8, mean age = 5;6 (SD = 0;2 months). The mean PPVT receptive vocabulary standardized score for the group was 115.6 (range = 97-138, SD = 14). This average score is one standard deviation above the expected mean for this age. The results reported here were gathered from 22 consecutively recruited children, with data from one child eliminated from the final analysis due to experimenter error.

Pediatric Cochlear Implant Users. Forty-five hearing-impaired pediatric users of cochlear implants participated in this study. All children were participants in a larger study currently being conducted at the Central Institute for the Deaf (see Geers et al., 1999, for details). One child in this group only completed one of the two conditions and all data from this child were subsequently dropped from the analysis, reducing the sample size to 44. The remaining children ranged in age from 7;11 to 9;11, mean age = 8;9 years (SD = 0;6). As can be seen in Table 1, all pediatric cochlear implant users in this study had lost their hearing before age three, with the majority reported as congenitally deaf. The duration of deafness prior to implantation averaged approximately three years and every child had used his/her implant for at least four years prior to the present testing. The group included children who use auditory/oral language as their primary means of communication as well as some children who use total communication, i.e., who rely on manual signs to supplement spoken language.

N=44	Mean	Minimum	Maximum	Std Deviation
Age at Onset of Deafness, in Months	2.52	0	36	7
Duration of Deafness in Years	2.94	.58	5.17	1.11
Duration of CI Use in Years	5.60	4.09	6.87	.66
Number of Active Electrodes	18.20	8	22	2.82

Table 1. Participant characteristics for the pediatric cochlear implant users.

Although we report results below for both the NH and CI users described above, we do not wish to suggest that a direct comparison is appropriate. In actuality, the NH children whose data are reported here completed this discrimination task as part of another study completed prior to any testing of the CI users. That is to say, the NH children were not recruited as a direct comparison group. Nevertheless, we feel that reporting the NH children's performance here is useful at this time to establish that the procedure used was well within the perceptual and cognitive abilities of normally developing children three to four years younger than the CI users in this study.

Stimulus Materials

The stimuli were selected from the Indiana Multi-Talker Sentence Database (Karl & Pisoni, 1994; Bradow, Torretta, & Pisoni, 1996), a CDROM containing digital recordings of 21 talkers each uttering 100 sentences selected from the Harvard Sentence lists (Egan, 1948; IEEE, 1969). All sound files were sampled at 20 kHz with 16-bit amplitude quantization and normalized such that the average RMS values for all files were equated. For detailed description of the recording procedures see Karl & Pisoni (1994). Eight sentences were used for the practice trials and another twenty-four sentences were selected for use during the test trials. (See Appendix.) Speaking rate #02 (medium rate) from the CDROM was used for all stimuli. The sentences were selected to have roughly similar construction and were all between 1.61 and 2.16 seconds in duration (8 to 11 syllables in length). An effort was made to not select sentences containing vocabulary the children would be unfamiliar with, however, due to the nature of the available database, there remain some words that are probably unfamiliar to hearing-impaired children (e.g., "colt")

“brim” and “reef”). For related future studies we are currently making a new set of recordings of the HINT-C, the sentences of which contain more appropriate vocabulary.

For this preliminary study, tokens from two male talkers were selected for the practice trials and tokens from three female talkers were selected for the test trials. The male talkers used for the practice stimuli were talkers #01 (gravely), and #21 (deeper). (These talkers are referred to as m1 and m9 in Bradlow, Torretta, & Pisoni, 1996.) The three female talkers used for the test stimuli were talkers #06 (smooth, deeper, bit older), #07 (gravely, young, unpleasant), #23 (higher, young, sweet) (talkers f2, f3, and f10 in Bradlow, Torretta, & Pisoni, 1996). The three female talkers were judged by the experimenter to differ, at least impressionistically, along the dimensions of age, and roughness of voice. Thus, although better-controlled examination of particular indexical dimensions is something we are working towards, the type of variation represented by the three talkers represented here is itself multidimensional. The recordings from talkers #06, #07, and #23 are, however, similar in that all are clearly produced by female adults with similar speaking rates, similar regional accents, and no marked emotional quality.

Among the reasons for selecting the female talkers over the male talkers as the test stimuli was the fact that the female talkers in this particular database have generally higher speech intelligibility scores than the male talkers (Bradlow, Torretta, & Pisoni, 1996). We reasoned that use of less intelligible stimuli could possibly distract listeners from the primary talker discrimination task in spite of the fact that participants were aware that the linguistic content of the test tokens was irrelevant. Of the ten available female talkers, talkers #06, #07, and #23 all had speech intelligibility scores above the mean for the group (>89.5%)(Bradlow, Torretta, & Pisoni, 1996). In addition, the three talkers were selected such that 1) their recorded tokens were quite close in overall duration for each sentence, on average, 2) there was some separation between the talkers’ mean f0 values and 3) the talkers were not strongly idiosyncratic relative to the other talkers (i.e., no extremely strongly evident age or regional dialect separation among the three, unlike the two remaining highly intelligible female talkers). As reported by Bradlow, Torretta, and Pisoni (1996), mean f0 for the talkers over the full set of 100 sentences contained in the original database were as follows: #06 = ~168 Hz, #07 = ~179 Hz, #23 = ~237 Hz. Our impression was that even normal-hearing persons might occasionally confuse the three selected talkers if close attention was not paid, thus limiting the possibility of ceiling effects in the simple accuracy measure.

Because a same/different discrimination task was to be used, six trials representing every possible ordered pairing of the three voices were employed for the “different voice” trials. For the six “same voice” trials, each of the three voices was paired with itself twice. This was the case in both the fixed sentence and varied sentence conditions. Within each pair, a one-second silent interval was inserted between the offset of the first sentence and the onset of the second sentence.

Procedure

Normal-hearing Children. Each of the 21 normal-hearing children passed a hearing screening at 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz at a level of 20 dB HL using a portable Maico Hearing Instruments pure tone audiometer (MA27) and TDH-39P headphones. A response at 25 dB HL was accepted for 250 Hz due to ambient room noise. Left and right ears were tested separately. Before the discrimination task was introduced, the children were tested on their understanding of the terms “same” and “different” using picture cards. All of the five-year-olds in this group easily identified a pair of pictures that were the same, and a pair that was different, indicating that they understood the concepts of same and different.

This group of children only received only one condition of the talker discrimination task, namely, the “varied sentence” condition. The discrimination trials were administered using a PC computer using a

control program written in C. After instructing the child about the basic nature of the task, four practice trials were administered via a tabletop loud speaker. All children received the same ordering of practice trials (same, different, same, different) using stimuli from two male talkers. On the first two practice trials the experimenter modeled the task by giving the correct answer after the pair of sentences was played. The child was encouraged to do the last two practice trials on his/her own and feedback was provided. During the practice period, the experimenter explained that if the child wasn't sure about the correct answer, he or she could ask for the (same pair of) sentences to be presented again and this option was demonstrated. Repetition could occur up to two additional times. This option was available for both the practice and test trials. The 12 test trials were presented via headphones (Beyerdynamic, DT100), with the examiner being unable to hear the current trial as it was played. The child was asked to verbally report whether the two talkers were the "Same" or if they were "Different" and was shown how the experimenter would circle the child's answer on a response sheet. Assignment of the 24 different sentences to the 12 test trial pairs, and the order of presentation of the test trials were pseudo-randomized by the computer.

Pediatric Cochlear Implant Users. The pediatric cochlear implant users were tested in a manner very similar to the NH children except that the discrimination task involved an additional condition. The fixed sentence condition was administered first, followed by the varied sentence condition. In the fixed sentence condition, the child heard only one sentence across all twelve trials, as spoken by the three different talkers. Each child heard one of the twenty-four sentences selected for use in the varied sentences condition, and this assignment was balanced such that each sentence used in the varied sentence condition was heard in the fixed sentence condition by approximately two children. A practice block of four trials preceded the running of the test trials. All children received the same four practice trials using a single sentence and two different male voices. In all other aspects, the administration of the practice trials was the same as described for the normal-hearing group

After completing twelve test trials in the fixed sentence condition, the child was given the revised instructions for the varied sentence condition. A practice block of four trials again preceded the running of the test trials. All children received the same four practice trials using eight different sentences and two different male voices. Twelve test trials using the three female talkers and 24 different sentences were then administered.

The pediatric cochlear implant users were tested using a Macintosh portable laptop computer using a Psyscope script written to mimic the C program used with the normal-hearing children. Stimuli were presented via a loudspeaker (Advent AV280, 10 Watts amplifier output power, THD < 1%, frequency response 70 Hz-20 kHz) at approximately 70 dB SPL as judged by a sound level meter placed near the location of the child's head. In some cases the level was adjusted upwards at the request of the child. Presentation of all stimuli was audible to the examiner. Although the practice trials were repeated for a few children in order to get the child on task, no test pairs were repeated. Although this is different from the methodology used with the NH children, the impact of this change is probably small because very few of the NH preschoolers requested any repetitions of the test trials.

Four different pseudo-random assignments of the 24 different sentences to the twelve available test pairs were generated prior to testing and nearly equal numbers of children were tested with each randomization. Presentation order of the same talker/different talker test trials was pseudo-randomized by the computer.

The procedures followed with the cochlear implant users were administered by a clinician experienced in working with hearing-impaired children. This clinician was trained in the task administration by the researcher responsible for gathering the data from the normal-hearing children.

Results and Discussion

Normal-hearing Preschoolers. The normal-hearing children had very little difficulty with the varied sentence condition on which they were tested, scoring 89% correct on average as a group. Most children scored either 12/12 or 11/12 correct. The distribution of scores obtained from the children in the NH group is shown on the far right hand panel of Figure 1. The scores of the children as a group differed significantly from chance performance of 50% ($t(20) = 15.28, p < .001$). The few errors that were observed primarily involved children incorrectly responding “same” for different voice pairs involving comparisons between talkers #06 and #07. Very few other errors were obtained.

Pediatric Cochlear Implant Users. The score distributions obtained from the CI users for both the “Fixed” and “Varied conditions are shown in the left and center panels of Figure 1. The mean accuracy for the group in the fixed sentence condition was 67% which is significantly above chance performance of 50% (one-sample t-test, $t(43) = 7.13, p < .001$). The mean accuracy for the group in the varied sentence condition was 57% correct, which, although significantly above chance (one sample t-test, $t(43) = 3.10, p = .003$), indicates that the pediatric CI users encountered considerable difficulty with this task. A paired-samples t-test between scores in the two conditions showed a significant drop in scores for the varied sentence task over the fixed sentence task ($t(43) = 3.66, p = .001$) with an average drop of about 11% (or 1.33 trials) across the two conditions. Scores in the two conditions showed a weak but significant positive correlation ($r = +.30, p = .049$).

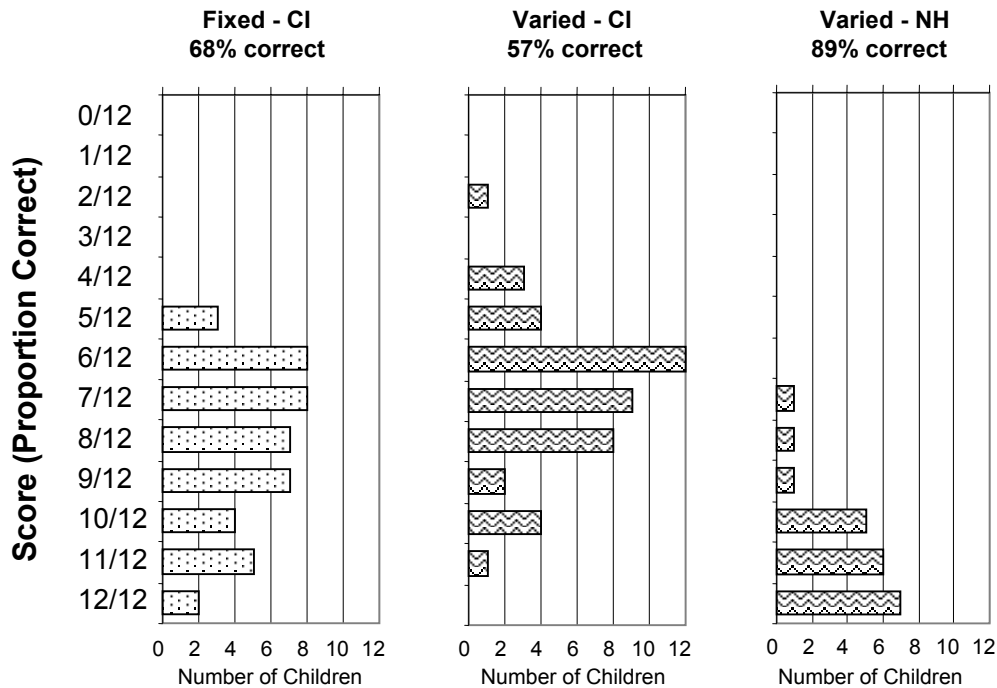


Figure 1. Distribution of scores obtained by the pediatric CI users ($N = 44$) in left and center panels, and normal-hearing five-year olds ($N = 21$), rightmost panel.

The cochlear implant users clearly had much greater difficulty with the varied sentence condition of the talker discrimination task than did the normal-hearing children on the same task. Although we did not test the normal-hearing children on the fixed sentence condition, it is likely that they would have done very well, probably better than the 89% they scored on the varied sentence condition.

Table 2 illustrates the distribution of the two possible error types in each condition for the CI users. One pattern evident in Table 2 is a bias for more often incorrectly responding “different” rather than “same” for pairs tested in the varied sentence condition. No such response bias was observed in the NH children.

Proportion of Possible Errors	Fixed Sentence Condition		Varied Sentence Condition	
	Said “Same” when Different	Said “Different” when Same	Said “Same” when Different	Said “Different” when Same
	# of children	# of children	# of children	# of children
6/6		1		2
5/6		1	4	8
4/6	3	5	4	9
3/6	9	10	6	9
2/6	14	13	11	8
1/6	12	5	13	6
0/6	6	9	6	2
Mean Proportion of Errors	.30	.35	.34	.52

Table 2. Distribution of errors for the children with cochlear implants. Errors made in the Fixed Sentence condition are shown on the left, errors made in the Varied Sentence condition are shown on the right.

Despite the fact that the talker discrimination scores obtained were not very continuously distributed due to the small number of trials, the variability present in the obtained scores allowed us to calculate correlations between talker discrimination scores and other measures available for these children. Because the pediatric CI users were nearly at chance in the varied sentence condition, meaningful correlations obtained with this measure are unlikely, and although shown in tables below, must be interpreted cautiously.

The results shown in Table 3 indicate that within this sample of cochlear implant users, talker discrimination performance was not significantly correlated in either condition with age at onset of deafness, duration of deafness, or duration of device use. Recall, however, that this group of CI users is relatively homogenous with respect to these factors. In particular, the children were pre-selected to demonstrate relatively little variability along these dimensions, unlike many prior studies involving samples of CI children with greater variation in these traditional predictor variables. For number of active electrodes and degree of exposure to an oral-only communication environment, there is only a slight indication that a greater number of active electrodes and more exposure to oral-only environment are

positively associated with better talker discrimination scores in both conditions. Exposure to oral-only communication was quantified using the communication mode scoring procedure described in Geers et al. (1999), which takes into account the type of communication environment experienced by the child in the year just prior to implantation, each year over the first three years of CI use, and then in the year just prior to the current testing. Table 3 also shows that the two-year spread in the chronological ages of the children was not a factor in performance.

	Proportion Correct Fixed Sentence Condition	Proportion Correct Varied Sentence Condition
Age at Onset of Deafness in Months	.19	.16
Duration of Deafness in Months	-.12	.06
Duration of CI use in Years	-.06	-.19
Number of Active Electrodes	.26	.19
Degree of Exposure to an Oral-Only Communication Environment	.27	.32*
Age in Years	-.13	.06

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 3. Correlations between talker discrimination performance and demographic variables.

	Fixed Sentence Condition	Varied Sentence Condition
BKB – Open Set Sentence Test Key Word Identification	.44**	.16
LNTE – Open Set Spoken Word Identification	.48**	.32*

* $p < 0.05$ (2-tailed), ** $p < .01$

Table 4. Correlations between talker discrimination performance and word recognition measures.

As shown in Table 4, performance on the talker discrimination task was positively correlated with two measures of spoken word recognition. These correlations were moderately large and statistically significant in the case of the fixed sentence condition. In the varied sentence condition, as might be surmised from the proximity of the group mean to chance performance, smaller correlations were observed, with only the correlation with LNTE open-set word recognition reaching statistical significance.

One question that arose in the design of this study was the question of whether the pair-wise comparison task draws on individual differences in short-term memory ability. That is, do the processing demands of the talker discrimination task make use of memory resources given that one stimulus must be kept in working memory for at least one second before the next sentence is played out?

	Fixed Sentence Condition	Varied Sentence Condition
WISC Digit Span Forward Points (lip-reading permitted)	.18	.11
Memory Game Span Auditory-Only	.38*	.05
Memory Game Span Auditory Plus Lights	.14	.21
Memory Game Span Lights-Only	~.00	.05

* Correlation is significant at the 0.05 level (2-tailed).

Table 5. Correlations between talker discrimination scores and memory measures obtained from the same sample in a separate study.

As shown in Table 5, performance on the fixed sentence condition of the talker discrimination task was positively correlated with memory span for stimuli presented only in the auditory modality. With the memory span measures that included visual information, however, there was little or no correlation observed with talker discrimination in the fixed sentence condition. This pattern of correlations suggests that either auditory memory plays some small role in the pair-wise discrimination task or that performance on the auditory-only memory span task may in some part reflect basic individual differences in auditory discrimination ability.

General Discussion

The talker discrimination study reported in this paper is very preliminary and we are currently in the process of expanding the scope of this research in a number of directions. Our results do, however, confirm the expectation that hearing-impaired prelingually-deafened children who have acquired language via a multi-channel cochlear implant have more difficulty discriminating between similar-sounding talkers than do normal-hearing children, particularly under conditions where the linguistic content of the message is varied.

There are several aspects of this study that suggest that the results should be interpreted somewhat cautiously. For the pediatric CI users who completed both versions of the task, the “fixed” and “varied” sentence conditions were not counterbalanced in their order of administration. Therefore it is

possible that some type of fatigue effect or order effect related to the task instructions could be responsible for the drop in the cochlear implant users' performance on the "varied sentence" condition. Given the relative brevity of the tasks, however, this is probably unlikely.

Other limitations of the present study include the fact that we do not report data from normal-hearing children using the same-sentence task. As stated previously, we assume they would do very well on this task, however we have not actually tested this. One of the reasons that we initially were hesitant to run a "fixed sentence" version of this task with normal-hearing children was that there are limitations on what we can conclude from any set of results generated using the stimuli described in our method section. The primary problem is that since the Indiana Multi-Talker Sentence Database only contains one recorded token of each talker saying each sentence, the "same talker" trials of the "fixed sentence" condition involve comparing a token with itself. It is not clear that a correct response under these conditions must necessarily reflect encoding of indexical attributes using a representation of a talker's voice. In our future studies we will be including conditions in which different recorded tokens of the same sentence are used in "fixed sentence" same-talker trials.

It may be of interest to note that we have recently tested one post-lingually deafened adult user of a Clarion 8-channel cochlear implant on the same tasks as used with the children discussed in this paper. This individual, "Mr. S", was previously identified as an extremely successful user of his cochlear implant (e.g., see Goh, Pisoni, Kirk, & Remez, 1999; Herman & Clopper, 1999). For this gentleman, in addition to the fixed sentence and varied sentence conditions, we also included an additional "fixed sentence condition" involving 24 additional trials with two male and two female talkers, such that, in 6 of the 12 trials in which the correct response was "different", the voices of the talkers differed by gender. This adult user made no errors at all in this condition, or in the original fixed sentence condition using only female talkers. He responded correctly in 9 of 12 test trials in the "varied sentence" condition, making three incorrect responses of "different" when the talkers were, in fact, the same. Thus, this cochlear implant user's performance resembled the performance of the pediatric CI users on the "varied sentence" condition. On the fixed sentence condition, however, he performed much more consistently than most of the children with cochlear implants.

The limited indexical variability present in our stimulus set required a more difficult discrimination than is typically used in the few studies that have asked hearing-impaired children to make judgments about indexical properties. This selection was intended to help us look at how small differences between talkers are perceived. Since the cochlear implants of today do, in fact, convey a fairly detailed spectral representation of the speech signal, this is not an unreasonable goal.

In future research we intend to further test whether hearing-impaired children find it easier to ignore linguistic information while making judgments about indexical information than do normal-hearing children. This research is related to a previous line of research in which we have found evidence suggesting that children with cochlear implants do not automatically use semantically redundant linguistic information which can potentially help them perform a multi-modal working memory task, despite the fact that these children can be shown to be able to identify the linguistic auditory information when played in isolation (Cleary, Pisoni, & Geers, in press). We plan to further test the hypothesis that hearing-impaired children, due to their atypical history of spoken language acquisition, are less likely than normally-developing children to try to automatically integrate semantically reinforcing/redundant or semantically conflicting information. It is this development of automaticity that makes normal-hearing children susceptible to interference in auditory Stroop tasks and the related Garner tasks as reported on by Jerger and colleagues.

In our new projects we will need to have recorded sentences that are more suitable for use with young hearing-impaired children so as to make the gathering of linguistic judgments possible alongside indexical judgments. As noted earlier, we plan to record multiple tokens from twelve female talkers of the sentences in the HINT-C Sentence Test as well as generate re-synthesized tokens of these recordings. Re-synthesized versions of these recordings with adjustment of acoustic parameters associated with perception of pitch and breathiness will permit us to have more precise experimental control over the acoustic similarities between talkers.

Once the appropriate stimulus materials are available, in addition to testing additional normal-hearing children on talker discrimination judgments under conditions of “listening in the clear,” we plan to examine the perceptual judgments of normal-hearing children under simulated conditions of hearing-loss such as those that mimic the signal processing which takes place using the Nucleus 22 device with the SPEAK processing strategy (e.g., Eisenberg et al., 2000). This research should help shed further light on how pediatric cochlear implant users encode and process the indexical variability that is present in the spoken language they encounter in their everyday lives.

References

- Bradlow, A.R., Torretta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.
- Cleary, M., Pisoni, D.B., & Geers, A. (in press). Some measures of verbal and spatial working memory in eight- and nine-year old hearing-impaired children with cochlear implants. *Ear and Hearing*.
- Egan, J.P. (1948). Articulation testing methods. *Laryngoscope, 58*, 955-991.
- Eisenberg, L.S., Shannon, R.V., Martinez, A.S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America, 107*, 2704-2710.
- Geers, A.E., Nicholas, J., Tye-Murray, N., Uchanski, R., Brenner, C., Crosson, J., Davidson, L.S., Spehar, B., Torretta, G., Tobey, E.A., Sedey, A., & Strube, M. (1999). Center for Childhood Deafness and Adult Aural Rehabilitation, Current research projects: Cochlear implants and education of the deaf child, second-year results. In *Central Institute for the Deaf Research Periodic Progress Report No. 35* (pp. 5-20). St. Louis, MO: Central Institute for the Deaf.
- Goh, W.D., Pisoni, D.B., Kirk, K.I., & Remez, R.E. (1999). Audio-visual perception of sinewave speech in an adult cochlear implant user: A case study. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 201-210). Bloomington, IN: Indiana University.
- Herman, R., & Clopper, C. (1999). Perception and production of intonational contrasts in an adult cochlear implant user. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 301-321). Bloomington, IN: Indiana University.
- IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Report No. 297*.
- Jerger, S., Martin, R., Pearson, D.A., & Dihn, T. (1995). Childhood hearing impairment: Auditory and linguistic interactions during multidimensional speech processing. *Journal of Speech and Hearing Research, 38*, 930-948.
- Jerger, S., Pirozzolo, F., Jerger, J., Elizondo, R., Desai, S., Wright, E., & Reynosa, R. (1993). Developmental trends in the interaction between auditory and linguistic processing. *Perception and Psychophysics, 54*, 310-320.
- Jones, P.A., McDermott, H.J., Seligman, P.M., & Millar, J.B. (1995). Coding of voice source information in the Nucleus cochlear implant system. *Annals of Otolaryngology, Rhinology, & Otolaryngology-Supplement, 166*, 363-365.

- Karl, J.R., & Pisoni, D.B. (1994). Effects of stimulus variability on recall of spoken sentences: A first report. In *Research on Spoken Language Processing Progress Report No. 19* (pp. 145-193). Bloomington, IN: Indiana University.
- Kramer, E. (1963). Judgement of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin*, *60*, 408-420.
- Kreiman, J. (1997). Listening to voices: Theory and practice in voice perception research. In K. Johnson & J.W. Mullennix (Eds.) *Talker Variability in Speech Processing* (pp. 85-108). San Diego: Academic Press.
- Loizou, P.C. (1998). Introduction to cochlear implants. *IEEE Signal Processing Magazine*, September, 101-130.
- Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*, 153-165.
- McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology*, *17*, 269-271.
- Miller, J.L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics*, *24*, 175-180.
- Mullennix, J., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *47*, 379-390.
- Murray, T., & Cort, S. (1971). Aural identification of children's voices. *Journal of Auditory Research*, *11*, 260-262.
- Osberger, M.J., Miyamoto, R.T., Zimmerman-Phillips, S., Kemink, J.L., Stroer, B.S., Firszt, J.B., & Novak, M.A. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant. *Ear and Hearing*, *12* (Supplement), 66S-80S.
- Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.) *Talker Variability in Speech Processing* (pp. 9-32). San Diego: Academic Press.
- Pollack, I., Pickett, J., & Sumbly, W. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America*, *26*, 403-406.
- Ptacek, P., & Sanders, E. (1966). Age recognition from voice. *Journal of Speech and Hearing Research*, *9*, 273-277.
- Seligman, P., & McDermott, H. (1995). Architecture of the Spectra 22 speech processor. *Annals of Otology, Rhinology, and Laryngology, Supplement*, *166*, 139-141.
- Sommers, M.S., Kirk, K.I., & Pisoni, D.B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*, *18*, 89-99.

Appendix

The stimuli were selected from the Indiana Multi-Talker Database, a CDROM containing recordings of 100 of the Harvard Sentences as spoken by 21 talkers.

Practice stimuli: Speaking rate #02, Male talkers #01 (gravely), #21(deeper) (~2 seconds long)

1. Glue the sheet to the dark blue background.(02)
2. Kick the ball straight and follow through.(14)
3. Help the woman get back to her feet.(15)
4. Take the winding path to reach the lake.(32)
5. Mend the coat before you go out. (35)
6. March the soldiers past the next hill. (57)
7. Place a rose bush near the porch steps. (59)
8. See the cat glaring at the scared mouse. (82)

Test Stimuli: Speaking rate #02, Female talkers #06 (older), #07 (young, unpleasant), #23 (young, sweet)

1. The juice of lemons makes fine punch. (06)
2. The box was thrown beside the parked truck. (07)
3. The boy was there when the sun rose. (11)
4. The soft cushion broke the man's fall. (18)
5. The salt breeze came across from the sea. (19)
6. The small pup gnawed a hole in the sock. (21)
7. The colt reared and threw the tall rider. (28)
8. The meal was cooked before the bell rang. (39)
9. The ship was torn apart on the sharp reef. (42)
10. The wide road shimmered in the hot sun. (44)
11. The lazy cow lay in the cool grass. (45)
12. The frosty air passed through the coat. (51)
13. The crooked maze failed to fool the mouse. (52)
14. The wagon moved on well-oiled wheels. (56)
15. The set of china hit the floor with a crash. (67)
16. The two met while playing on the sand. (72)
17. The ink stain dried on the finished page. (73)
18. The horn of the car woke the sleeping cop. (77)
19. The pearl was worn in a thin silver ring. (79)
20. The fruit peel was cut in thick slices. (80)
21. The hat brim was wide and too droopy. (84)
22. The slush lay deep along the street. (91)
23. A wisp of cloud hung in the blue air. (92)
24. A pound of sugar costs more than eggs. (93)