



Published in final edited form as:

Ann Otol Rhinol Laryngol Suppl. 2002 May ; 189: 113–118.

TALKER DISCRIMINATION BY PRELINGUALLY DEAF CHILDREN WITH COCHLEAR IMPLANTS: PRELIMINARY RESULTS

Miranda Cleary and David B. Pisoni, PhD

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

Abstract

Forty-four school-age children who had used a multichannel cochlear implant (CI) for at least 4 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” on each trial. Two conditions were examined. In one condition, the linguistic content of the sentence was always held constant and only the talker’s voice varied from trial to trial. In another condition, the linguistic content of the utterance also varied so that to correctly respond “same voice,” the child needed to recognize that two different sentences were spoken by the same talker. Data from normal-hearing children were used to establish that these tasks were well within the capabilities of children without hearing impairment. For the children with CIs, in the “fixed sentence condition” the mean proportion correct was 68%, which, although significantly different from the 50% score expected by chance, suggests that the children with CIs found this discrimination task rather difficult. In the “varied sentence condition,” however, the mean proportion correct was only 57%, indicating that the children were essentially unable to recognize an unfamiliar talker’s voice when the linguistic content of the paired sentences differed. Correlations with other speech and language outcome measures are also reported.

Keywords

child; cochlear implant; indexical property; speech perception; talker discrimination; voice discrimination

INTRODUCTION

A large body of research has shown that normal-hearing (NH) listeners are sensitive to properties in the speech signal that provide information about the talker who produced the utterance. These properties are sometimes referred to as “indexical” properties of speech and can convey, although imperfectly, information about the talker’s gender, age, regional background, emotional state, etc.^{1–3} Indexical information is usually conceptualized as contrasting with “linguistic” information about the intended pattern of phonemes/phonemic contrasts.⁴ Because linguistic and indexical information are both simultaneously encoded in the acoustic waveform, 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.⁵

The ability to use indexical information to perceptually discriminate between the utterances of different talkers is often taken for granted in communicative situations. In order to

interpret what is being said in the larger context of a spoken conversation, a listener must be able to keep track of the current speaker and register a change of speaker when it occurs. For both NH and hearing-impaired persons, this task becomes more difficult when associated visual cues are unavailable, such as in communicating via telephone, in listening to the radio, or if the listener has momentarily turned away from the speaker's face.

In the present study, we asked pediatric users of cochlear implants (CIs) to make perceptual judgments about whether pairs of recorded sentences were spoken by the same or different talkers. Two conditions were explored. In one condition, the linguistic contents of the paired utterances were identical (referred to henceforth as the "fixed sentence condition"). In the other condition, the linguistic contents of the two sentences always differed (the "varied sentence condition"). In this 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. Because the talkers used in this study were previously unfamiliar to the listeners, 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.

The 8- and 9-year-old children who participated in this study had all used a multichannel CI for at least 4 years. Forty-one of the children used a Nucleus-22 device. Two children used a Clarion implant. One child was a former Nucleus-22 user who had switched to the Nucleus-24 implant 4 months before testing. All of the children at the time of data collection were using a coding strategy (either spectral peak or continuous interleaved sampling) capable of conveying a fairly detailed spectral representation of the speech signal. Given the design of the device and the children's history of use, we expected that at least some of the pediatric CI users would be able to make the simple voice discriminations presented under the "fixed sentence" condition. Because the linguistic content of the sentence was held constant across all comparisons, intertalker differences should constitute the primary source of any perceived acoustic variation between sentences. We anticipated that the "varied sentence" condition would prove more difficult, because a generalizable representation of each talker's voice is presumably necessary to accomplish this task. However, if for the purpose of the task at hand the children were able to ignore the linguistic variability as directed, the signal provided by the implant should be sufficient to permit some children to form the necessary representations of the different voices. Data from a group of younger NH children helped us assess the overall difficulty of this "varied sentence" condition.

METHODS

PARTICIPANTS

NH Children—Twenty-one NH children were tested as part of a larger project being conducted at the Indiana University Speech Research Laboratory. The children ranged in age from 5 years 3 months to 5 years 8 months (mean, 5 years 6 months; SD, 2 months).

Pediatric CI Users—Forty-four hearing-impaired children with CIs participated as part of a larger study currently being conducted at Central Institute for the Deaf.⁶ As shown in Table 1, the children ranged in age from 7.92 to 9.91 years at the time of testing (mean, 8.76 years). All pediatric CI users in this study had lost their hearing before 3 years of age; the majority were reported as congenitally deaf. The duration of deafness before implantation averaged approximately 3 years, and every child had used his or her implant for at least 4 years before the present testing. The group included both children who used oral communication and children who used total communication.

The NH children were *not* recruited as a direct comparison group for the CI group. Nevertheless, we believe that reporting the NH children's performance here is useful at this time to establish that the experimental procedure used was within the perceptual and cognitive abilities of normally developing children 3 to 4 years younger than the CI users in this study.

STIMULUS MATERIALS

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.^{7,8} We therefore used relatively long sentence-length stimuli and only a very small set of 3 female talkers. The stimuli were selected from the Indiana Multi-Talker Sentence Database (IMTSD),^{9,10} a CD-ROM containing digital recordings of 21 talkers each uttering 100 sentences selected from the Harvard Sentence lists.^{11,12} All sound files were sampled at 20 kHz with 16-bit amplitude quantization and normalized such that the average root-mean-square values for all files were equated.

Eight sentences were used for the practice trials, and another 24 sentences were selected for use during the test trials. The sentences were selected to have roughly similar construction, and all tokens were between 1.50 and 2.25 seconds in duration (8 to 11 syllables; eg, "The lazy cow lay in the cool grass"). An effort was made to select sentences that contain vocabulary the children would be familiar with; however, because of the nature of the available database, there remain some words that are probably unfamiliar to hearing-impaired children (eg, "colt").

Tokens from 2 male talkers were selected for the practice trials, and tokens from 3 female talkers were selected for the test trials. The 3 female talkers used for the test stimuli were talkers 6, 7, and 23 from the IMTSD. This set was deliberately chosen to include some separation between the talkers' mean fundamental frequency (F0) values. As reported by Bradlow et al,¹⁰ the mean F0 values for each talker over the full set of sentences contained in the database were as follows: talker 6, approximately 168 Hz; talker 7, approximately 179 Hz; and talker 23, approximately 237 Hz. The recordings from the 3 talkers were otherwise similar in that all were produced by highly intelligible female adults with similar speaking rates, similar regional accents, and no marked emotional quality.

Because a same-different discrimination task was to be used, 6 trials representing every possible ordered pairing of the 3 voices were used for the "different voice" trials. For the 6 "same voice" trials, each of the 3 voices was paired with itself twice. This arrangement was used in both the fixed sentence and varied sentence conditions. Within each pair, a 1-second silent interval was inserted between the offset of the first sentence and the onset of the second sentence.

PROCEDURE

NH Children—Each of the 21 NH children passed a hearing screening at 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz at a level of 20 dB hearing level on a portable pure tone audiometer (MA27; Maico, Minneapolis, Minnesota) and TDH-39P headphones. A response at 25 dB hearing level was accepted for 250 Hz. Left and right ears were tested separately.

The NH children received only the varied sentence condition of the talker discrimination task. After the child was instructed about the basic nature of the task, 4 practice trials were administered with a personal computer and a tabletop loudspeaker. All children performed the same ordering of practice trials (same, different, same, different) with stimuli from 2

male talkers. On the first 2 practice trials, the experimenter modeled the task by giving the correct answer after the pair of sentences was played. The child was then encouraged to do the last 2 practice trials on his or her own, and feedback was provided. During the practice period, the experimenter explained that if the child was not sure about the correct answer, he or she could ask for the same pair of sentences to be presented again, up to 2 additional times. This option was available for both the practice and test trials. The 12 test trials were presented via headphones (DT100; Beyerdynamic, Farmingdale, New York), and the examiner was unable to hear the current trial as it was played. The child was asked to verbally report whether the 2 talkers were the “same” or “different.” Assignment of the 24 different sentences to the 12 test trial pairs, and the order of presentation of the test trials were pseudorandomized by the computer. The child received no explicit feedback during the test trials regarding the accuracy of his or her performance.

Pediatric CI Users—The pediatric CI users were tested in a manner very similar to that of 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 1 sentence across all 12 trials, as spoken by the 3 different talkers. The assignment of this sentence was balanced such that each of the 24 sentences selected for use in the varied sentence condition was heard in the fixed sentence condition by approximately 2 children.

All children with CIs first underwent the same 4 fixed sentence practice trials with a single sentence and 2 different male voices. Twelve randomized test trials were then administered in the fixed sentence condition. The child next received the revised instructions for the varied sentence condition. These instructions alerted the child that the sentence content of the trials would vary, but emphasized that the primary task of “listening to the voice” had not changed. All children then underwent the same 4 varied sentence practice trials with 8 different sentences and 2 different male voices. Finally, 12 randomized test trials with the 3 female talkers and 24 different sentences were administered. Four different pseudorandom assignments of the 24 different sentences to the 12 available test pairs were generated before testing, and nearly equal numbers of children were tested with each randomization. The children received no explicit feedback during any of the test trials regarding the accuracy of their responses.

The pediatric CI users were tested with a Macintosh portable laptop computer. The stimuli were presented via a loudspeaker (AV280; Advent, Benica, California) at approximately 70 dB sound pressure level. In some cases, the level was adjusted upward at the request of the child. The 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 method 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.

The procedures followed with the CI 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 NH children.

RESULTS

NH Children

The NH 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 of 12 or 11 of 12 correct. The distribution of scores obtained from the NH children is shown on

the bottom panel of the Figure. The scores of the children as a group differed significantly from a chance performance of 50% ($t(20) = 15.28$; $p < .001$). The few errors that were observed primarily involved children's incorrectly responding "same" for different voice pairs involving comparisons between talkers 6 and 7. Very few other errors were obtained.

Pediatric CI Users

The score distributions obtained from the CI users for both the fixed and varied conditions are shown in the top and center panels of the Figure. The mean accuracy for the group in the fixed sentence condition was 68%, which is significantly above a chance performance of 50% ($t(43) = 7.13$; $p < .001$). The mean accuracy for the group in the varied sentence condition was 57% correct, which, although significantly above a chance performance of 50% ($t(43) = 3.10$; $p = .003$), indicates that the pediatric CI users encountered considerable difficulty with this task. A paired-samples t -test between the scores in the 2 conditions showed a significant decrease in scores for the varied sentence task as compared to the fixed sentence task ($t(43) = 3.66$; $p = .001$). The scores in the 2 conditions showed a weak but significant positive correlation ($r = .30$, $p = .049$).

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

Table 2 shows the miss and false alarm error rates obtained 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. This pattern of results suggests that the children in the CI group may have found it difficult to ignore the linguistic variability present in the varied sentence condition as instructed.

Despite the fact that the talker discrimination scores obtained were not very continuously distributed because of 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 3 and 4,^{13–15} must be interpreted cautiously.

The results shown in Table 3 indicate that within this sample of CI users, talker discrimination performance was not significantly correlated in either condition with age at testing, age at onset of deafness, duration of deafness, or duration of implant use. Recall, however, that the children were preselected to demonstrate relatively little variability along these dimensions. Weak evidence was found that a greater number of active electrodes and more exposure to oral-only communication were positively associated with better talker discrimination scores. Exposure to oral-only communication was quantified by the communication mode scoring procedure described by Geers et al,⁶ which takes into account the type of communication environment experienced by the child in the year just before implantation, in each year over the first 3 years of CI use, and in the year just before the current testing. An independent-samples t -test using a median split along the variable of communication mode score indicated that the "primarily oral" group's mean of 62% correct on the varied sentence condition was, in fact, significantly higher than the "primarily total communication" group's mean of 52% correct ($t(42) = 2.41$; $p = .021$).

As shown in Table 4, the talker discrimination scores were positively correlated with three outcome measures of spoken word recognition gathered by clinicians at Central Institute for the Deaf for another project within a few days of the talker discrimination data. These correlations were moderately large and statistically significant in the case of the fixed sentence condition, suggesting that children who are better able to identify spoken words are also better equipped to perceive acoustic information useful for discriminating between talkers.

DISCUSSION

The talker discrimination results reported in this paper are preliminary, and we are currently in the process of expanding the scope of this research project in a number of directions. Our results do, however, confirm the expectation that prelingually deafened children who have acquired language via a multichannel CI have more difficulty discriminating between similar-sounding talkers than do NH children, particularly under conditions in which the linguistic content of the message is varied.

In general, the existing literature indicates that for children with CIs, the large acoustic differences in F0 that distinguish declarative versus WH-question intonation and male versus female speech are fairly easily discriminated, even before phonemic distinctions are readily made.¹⁶ Because our stimulus set required a more difficult discrimination than was typically used in the few studies that have asked children with CIs to make judgments about the indexical properties of speech, we did expect that the hearing-impaired children would find our talker discrimination tasks more difficult. Even so, we were surprised at just how difficult many of the children with CIs found the varied sentence condition, particularly given that this particular discrimination was well within the capacities of normally developing children. We suggest that although the device does convey the relevant spectral detail, children with CIs have difficulty interpreting the information that NH listeners routinely use to recognize subtle indexical differences between talkers. Given that the ability to explicitly discriminate between talkers develops over time, even in NH children, with the ability to recognize briefly studied unfamiliar voices continuing to improve throughout the school-age years and reaching adult levels by adolescence,¹⁷ we think it is probable that children with CIs have the ability to improve their skills in this area. It is not known, however, whether this improvement can be brought about simply through greater everyday experience with the implant, or whether explicit training is necessary. As prelingually deafened children with CIs begin to enter mainstream classrooms in larger numbers, it will become increasingly important to understand how these children encode and process the enormous indexical variability present in spoken language.

Acknowledgments

Work supported by Research Grant R01 DCO0111 and T32 Training Grant DC00012 from the National Institutes of Health/National Institute on Deafness and Other Communication Disorders to Indiana University, Bloomington, Indiana, by Research Grant R01 DC00064 from the National Institutes of Health to the Indiana University School of Medicine, Indianapolis, Indiana, and by Research Grant R01 DC03100 from the National Institutes of Health to Central Institute for the Deaf, St Louis, Missouri.

We gratefully acknowledge the kind cooperation of Dr Ann Geers, C. Brenner, and their colleagues at Central Institute for the Deaf in making this project possible.

References

1. Kramer E. Judgement of personal characteristics and emotions from nonverbal properties of speech. *Psychol Bull.* 1963; 60:408–20. [PubMed: 14035446]

2. Kreiman, J. Listening to voices: theory and practice in voice perception research. In: Johnson, K.; Mullennix, JW., editors. Talker variability in speech processing. San Diego, Calif: Academic Press; 1997. p. 85-108.
3. Ptacek P, Sanders E. Age recognition from voice. *J Speech Hear Res.* 1966; 9:273–7. [PubMed: 5925532]
4. Pisoni, DB. Some thoughts on “normalization” in speech perception. In: Johnson, K.; Mullennix, JW., editors. Talker variability in speech processing. San Diego, Calif: Academic Press; 1997. p. 9-32.
5. Sommers MS, Kirk KI, Pisoni DB. 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 Hear.* 1997; 18:89–99. [PubMed: 9099558]
6. Geers, AE.; Nicholas, J.; Tye-Murray, N., et al. Central Institute for the Deaf Research Periodic Progress Report No 35. St Louis, Mo: Central Institute for the Deaf; 1999. Center for Childhood Deafness and Adult Aural Rehabilitation. Current research projects: Cochlear implants and education of the deaf child, second-year results; p. 5-20.
7. Murray T, Cort S. Aural identification of children’s voices. *J Aud Res.* 1971; 11:260–2.
8. Pollack I, Pickett J, Sumby W. On the identification of speakers by voice. *J Acoust Soc Am.* 1954; 26:403–6.
9. Karl, JR.; Pisoni, DB. Research on Spoken Language Processing Progress Report No 19. Bloomington, Ind: Indiana University; 1994. Effects of stimulus variability on recall of spoken sentences: a first report; p. 145-93.
10. Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech. I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun.* 1996; 20:255–72. [PubMed: 21461127]
11. Egan JP. Articulation testing methods. *Laryngoscope.* 1948; 58:955–91. [PubMed: 18887435]
12. IEEE Report No 297. New York, NY: Institute of Electrical and Electronic Engineers; 1969. IEEE recommended practice for speech quality measurements.
13. Ross M, Lerman J. A picture identification test for hearing-impaired children. *J Speech Hear Res.* 1970; 13:44–53. [PubMed: 4192711]
14. Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol.* 1979; 13:108–12. [PubMed: 486816]
15. Kirk KI, Pisoni DB, Osberger MJ. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear Hear.* 1995; 16:470–81. [PubMed: 8654902]
16. Osberger MJ, Miyamoto RT, Zimmerman-Phillips S, et al. Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear Hear.* 1991; 12(suppl):66S–80S. [PubMed: 1955092]
17. Mann VA, Diamond R, Carey S. Development of voice recognition: parallels with face recognition. *J Exp Child Psychol.* 1979; 27:153–65. [PubMed: 458368]

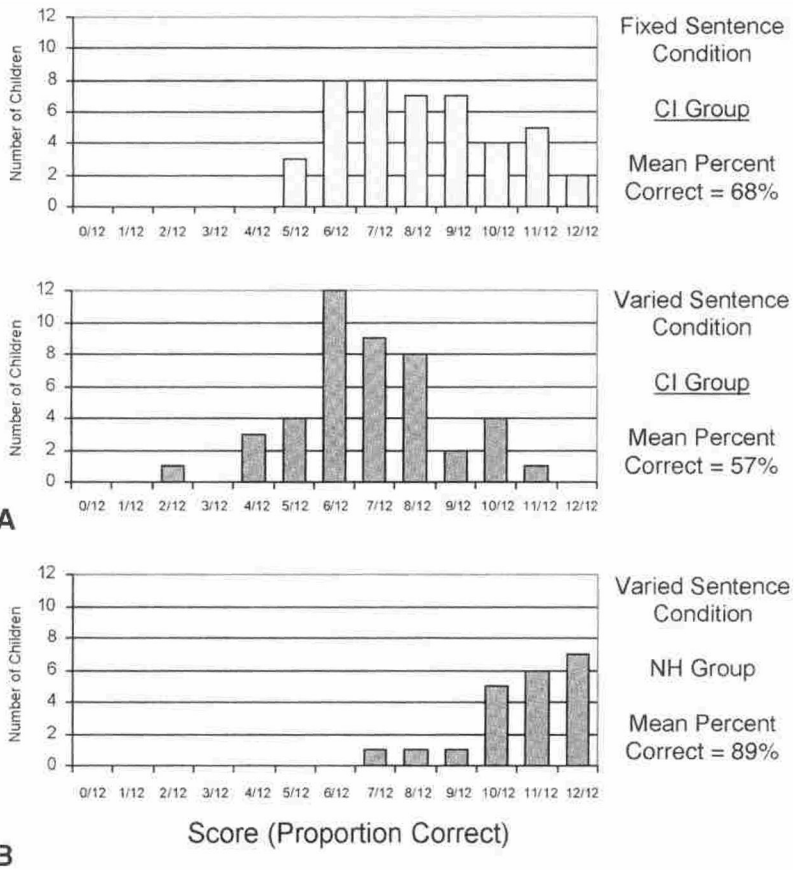


Figure. Distribution of talker discrimination scores **A**) in fixed and varied sentence conditions for 44 pediatric cochlear implant (CI) users and **B**) in varied sentence condition for group of 21 normal-hearing (NH) 5-year-olds.

TABLE 1
DEMOGRAPHIC CHARACTERISTICS OF PEDIATRIC CI GROUP

	Mean	Minimum	Maximum	SD
Age at testing (y)	8.76	7.92	9.91	0.53
Age at onset of deafness (mo)	2.52	0	36	7
Duration of deafness (y)	2.94	0.58	5.17	1.11
Duration of CI use (y)	5.60	4.09	6.87	0.66
No. of active electrodes	18.20	8	22	2.82

CI — cochlear implant.

TABLE 2**MISS RATES AND FALSE ALARM RATES FOR CI GROUP**

Error Type	Fixed Sentence Condition	Varied Sentence Condition
Responded "same" when different (miss rate)	.30	.34
Responded "different" when same (false alarm rate)	.35	.52

TABLE 3**SIMPLE BIVARIATE CORRELATIONS BETWEEN TALKER DISCRIMINATION PERFORMANCE AND DEMOGRAPHIC VARIABLES**

	Proportion Correct in Fixed Sentence Condition	Proportion Correct in Varied Sentence Condition
Age at testing (y)	-.13	.06
Age at onset of deafness (mo)	.19	.16
Duration of deafness (y)	-.12	.06
Duration of CI use (y)	-.06	-.19
No. of active electrodes	.26	.19
Degree of exposure to oral-only communication environment	.27	.32 *

Data are r values.

* Correlation is significant at .05 level (2-tailed).

TABLE 4

SIMPLE BIVARIATE CORRELATIONS BETWEEN TALKER DISCRIMINATION PERFORMANCE AND WORD RECOGNITION MEASURES

	Proportion Correct in Fixed Sentence Condition	Proportion Correct in Varied Sentence Condition
Word Intelligibility by Picture Identification (WIPI) Closed-Set Word Identification ¹³	.60 *	.36 †
Bamford-Kowal-Bench Open-Set Sentence Test, Key Word Identification (BKB) ¹⁴	.44 *	.16
Lexical Neighborhood Test Easy Word Lists Open-Set Word Identification (LNTe) ¹⁵	.48 *	.32 †

Data are r values.

* Correlation is significant at .01 level (2-tailed).

† Correlation is significant at .05 level (2-tailed).