RESEARCH ON
SPOKEN LANGUAGE PROCESSING

Technical Report No. 11

January 30, 2003

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

Supported by:

Department of Health and Human Services
U.S. Public Health Service
National Institutes of Health
Research Grant DC00111
and
Training Grant DC00012

PERCEPTION OF TALKER DIFFERENCES IN NORMAL-HEARING CHILDREN AND HEARING-IMPAIRED CHILDREN WITH COCHLEAR IMPLANTS

Miranda Cleary

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

In the Department of Psychology and Program in Cognitive Science

Indiana University

February 2003

Acknowledgements

Many people gave generously of their time and energy to help me complete this project. Firstly, I would like to thank the children and parents who participated in this research. I enjoyed working with them and they taught me a great deal. This project could not have been completed without the help of my research advisory committee, David B. Pisoni, Karen Iler Kirk, Diane Kewley-Port, and Tom A. Busey. Without the benefit of Karen's expert advice regarding children with cochlear implants, I would not have been able to do this research. Karen's help in recruiting hearing-impaired participants, and in making facilities at DeVault available to me, was invaluable. Diane has generously advised me on a variety of technical issues related to this project. I have appreciated being able to consult with her as the project developed. Through the years, Diane's graduate-level courses, together with David's, have been a huge motivating factor in making me believe that a future in the speech sciences would be exciting, challenging, and worthwhile. Tom offered a number of valuable methodological suggestions during the development of this project. The results of my research are much more interesting, I believe, as a result of this input. David has provided boundless energy, enthusiasm, advice, and support for this project, and I will always appreciate the time and effort he has invested in my training. I am particularly grateful for his careful reading of my manuscript drafts, the detailed feedback, and the many opportunities I have been given to get involved in a variety of different projects within the lab. I also deeply appreciate his efforts to make everything I needed for my research available to me and for his initial invitation, some years ago, to become part of what has consistently been a smart, fun, and vibrant laboratory community. I would like to extend a special thanks to Darla Sallee, Luis Hernandez, Linette Caldwell, Terri Kerr, Ann Geers, Chris Brenner, and Caitlin Dillon, for providing invaluable assistance with specific aspects of this dissertation. Many clinicians and researchers associated with DeVault Otologic Research Lab helped me out at crucial times, among them, Becky, Bev, Randy, Amy, Cara, Helen, Tara, & Liz. Additionally, I would like to thank Linda Smith and the Developmental Area for allowing me to use their birth-record files to find and recruit potential normal-hearing participants. Hideki Kawahara kindly allowed me use his STRAIGHT speech resynthesis algorithms. I also express my appreciation for two wonderful tools that have been generously made available to the research public: the Praat speech analysis software package by Paul Booersma & David Weenick and the CHILDES database managed by Brian MacWhinney. I would also like to acknowledge a number of fellow lab members who took time to help me in my research or taught me a lot through just being around them: Rose, Lorin, Cynthia, Richard, Gina, Janna, Brenda, Sonya, Jeff, Rebecca, Damon, Steve, Mario, Allyson, Derek, Jimmy, Stefan, Tonya, Amano-san, Mike, Connie, Kristin, Nathan, & Melissa. Lana, Bob, & Ken have also been consistently friendly and helpful faces on campus. Winston has helped me innumerable times over the years and I have valued his advice and encouragement. Caitlin and Tamara have each gone out of their way to make sure I survived graduate school, and I thank Heather for looking in on me to see how I was doing at all the right times. Finally, I extend my thanks and appreciation to my wonderful husband Jack, and to my parents and little sisters, Nicole and Muriel. This work was supported by NIH-NIDCD Training Grant T32DC00012 and NIH-NIDCD Research Grants R01DC00111 and R01DC00064 to Indiana University, and also by an Indiana University Grant-in-Aid of Research to the author.

Abstract

The perception of talker differences was examined in normal-hearing five-year-old children and in hearing-impaired pediatric cochlear implant users, five to twelve years of age. For all hearing-impaired children, the onset of deafness occurred prior to age three years, and each child had at least two years of implant experience at time of testing. In two preliminary studies, children who use cochlear implants were found to have difficulty discriminating among talkers that were easily discriminated by normal-hearing children. The main experiment examined the influence of voice similarity on talker discrimination. Recorded sentences were manipulated using a high-quality speech resynthesis procedure to form a stimulus continuum of similar-sounding voices. An adaptive procedure for eliciting judgments of "same talker" vs. "different talker" was developed to determine how acoustically different, in terms of average fundamental and formant frequencies, two sentences needed to be for a listener to categorize the sentences as spoken by two different talkers. We found that the spectral envelopes of two utterances, including their fundamental frequencies, needed to differ by at least 11-16%, or 2-2.5 semitones, for normal-hearing children to perceive the voices as belonging to different talkers. Normal-hearing children maintained the perception of voice individuality over a larger frequency shift than previously studied normal-hearing adults. Although several individual children with cochlear implants exhibited response patterns that resembled those of normal-hearing children, most of the hearing-impaired children experienced difficulty perceiving even large differences between talkers. Within the implant group, better talker discrimination performance was associated with higher scores on spoken word recognition tests. Our results suggest that poor discrimination of mean fundamental and formant frequencies, the primary acoustic correlates of voice pitch and voice timbre, is a factor contributing to implanted children's difficulty with talker discrimination tasks. Additionally, the present findings offer specific predictions regarding the degree to which natural voices must differ, with respect to average fundamental and formant frequencies, for normal-hearing children to perceive a change in talker. We discuss possible implications for understanding the difficulties in speech perception experienced by normal-hearing and hearing-impaired listeners when frequent talker shifts or multiple voice sources are present in the speech signal.

PERCEPTION OF TALKER DIFFERENCES IN NORMAL-HEARING CHILDREN AND HEARING-IMPAIRED CHILDREN WITH COCHLEAR IMPLANTS

TABLE OF CONTENTS

CHAPTER I - GENERAL INTRODUCTION AND OVERVIEW

For normal-hearing adult listeners, the human speech signal represents the encoding of many different types of information transmitted in parallel. Guided by their implicit knowledge of how energy patterns in the acoustic speech signal systematically reflect stable as well as more transitory characteristics of the human vocal tract, normal-hearing speech perceivers can make reliable inferences regarding the source and intent of new utterances they encounter. These inferences related to speech source include perceptual judgments about whether or not a change of talker has occurred while listening to speech. Although there are many cues that can help a listener to detect talker changes, among the most basic are certain stabilities and distributional regularities found in an individual talker's long-term average speech spectrum. More specifically, as examined in the present report, factors such as the mean fundamental and formant frequencies of a speech sample can serve to individualize a voice, and can aid a listener in discriminating between talkers.

A basic tenet of cognitive function is that in order for an intelligent organism to cope well with variety, it must first be exposed to variety (Gauthier, Williams, Tarr, & Tanaka, 1998; O'Toole, Peterson, & Deffenbacher, 1996; Pisoni & Lively, 1995). Therefore, in order to help understand the criteria by which listeners effortlessly process the numerous shifts of talker encountered in any given day, it may prove fruitful to study this perceptual decision-making process more directly in listeners with different histories of listening experience. If the assumption is correct that prior listening experience contributes to how a listener decides about talker identity, then normal-hearing children, for example, might be expected to perform differently from adults on a range of talker discrimination tasks.

Still more radical to consider is the case of the prelingually/congenitally deaf child who has acquired all of his/her auditory speech perception skills while using a cochlear implant. The hearing-impaired children studied in the present report all have audiological histories typical of children fitted with cochlear implants during the mid- to late 1990's: severe-to-profound bilateral sensorineural hearing loss greater than 90 dB HL, minimal-to-no-benefit demonstrated from traditional hearing-aid use, and evidence of some neuronal survival within the auditory nerve. This type of hearing-impaired listener clearly cannot draw on the same past experience of talker-related attributes as a normal-hearing child. Although the cochlear implant is an amazing technological achievement, the speech signal being received by today's generation of pediatric implant users is still quite spectrally degraded, and may not contain all the detailed acoustic information needed for perceiving talker differences.

Chapter II reports two preliminary studies of talker discrimination in normal-hearing children and hearing-impaired children with cochlear implants. Chapter IIA is included in the present report primarily because unlike Chapter III, it contains results obtained using only natural speech. For the goals of this thesis it is important to demonstrate that the difficulty displayed by the hearing-impaired participants in perceiving talker-specific acoustic attributes is not related to the use of artificially resynthesized natural speech. The results of Chapter IIB help to explain the choice of various parameters characterizing the stimulus continuum of similar-sounding voices used in Chapter III.

Chapter III constitutes the heart of the thesis project. Here we investigate the perception of talker voice similarity in normal-hearing children and hearing-impaired children with cochlear implants by manipulating certain basic acoustic properties of a voice and examining the effect of this manipulation on the perception of voice individuality. More specifically, we assess how acoustically different in terms of average fundamental and formant frequencies, two sentences must be for children to perceive the utterances as spoken by two different talkers. The introduction to this chapter includes a review of the developmental literature on children's perception of voices, together with background information about how cochlear implants process the speech signal. This chapter also describes the development of a new set of sentences

recorded specifically for use in the study, as well as a novel variant on traditional adaptive testing procedures that was designed to quickly and efficiently collect the children's responses.

In Chapter IV, our concluding section, practical and theoretical implications of the findings from this project are reviewed. We discuss several new issues raised by certain aspects of our results and outline future directions for this research. In connection with these future directions, we summarize some recent findings in normal-hearing adults extending the methods developed in Chapter III to study the relative contributions of mean fundamental frequency versus mean formant frequencies to the perception of talker-voice individuality.

The present report is intended as a dissertation in the Cognitive Sciences, and represents an attempt to integrate concerns and methodologies from the various speech science-related disciplines. Much of what is known about the spectral characteristics of speech and how these emerge from the shape, composition, and activity of the human vocal tract, is drawn from the field of acoustic phonetics. The literature on the language development of children with cochlear implants comes largely from the work of medical and clinical professionals associated with otolaryngology departments and institutions devoted to the education of the hearing-impaired. A cochlear implant is a highly sophisticated electronic device, and to consider the possible ramifications of its design, I relied on the reports of individuals whose background is primarily in electrical engineering and signal processing techniques. The research of sensory psychologists in the area of psychophysical testing has influenced the design of the discrimination task used in Chapter III. Finally, the fields of cognitive and developmental psychology have much to say regarding how humans create and maintain perceptual categories; many ideas from these fields have influenced my conception of how different listener populations approach the task of discriminating between talkers.

References

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38,* 2401-2428.

O'Toole, A.J., Peterson, J., & Deffenbacher, K.A. (1996). An 'other-race effect' for categorizing faces by sex. *Perception, 25,* 669-676.

Pisoni, D.B., & Lively, S.E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (ed.), *Speech Perception and Linguistic Experience (pp. 433-459).* Timonium, MD: York Press.

CHAPTER II - PRELIMINARY STUDIES

IIA. TALKER DISCRIMINATION IN PRELINGUALLY DEAF CHILDREN WITH COCHLEAR IMPLANTS

Abstract. Forty-four eight- and nine-year-old children who had used a multi-channel cochlear implant (CI) for at least four years were tested to assess their ability to discriminate differences between recorded pairs of female talkers uttering sentences. Children were asked to decide on each trial whether the two sentences were both spoken by the same talker or by two different talkers. Two conditions were examined. In the "fixed sentence" condition, the linguistic content of the sentences was always held constant and only the talkers differed from trial to trial. In the "varied sentence" condition, the linguistic content of the utterances also varied so that in order to correctly respond "same" on a trial, 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 were used to establish that these tasks were well within the cognitive and sensory capabilities of typically-developing children without hearing impairment (mean proportion correct on the varied sentence condition = 89%). For the CI children tested, the mean proportion correct in the "fixed sentence" condition was 68% 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%, indicating that the children were essentially unable to identify two voices as the same or different, when the linguistic content of the two sentences in a pair varied. Correlations with other speech and language measures are also reported. We discuss these findings on talker discrimination relative to data previously reported on voice gender discrimination in this clinical population and in light of interactions known to exist between the perception of linguistic and indexical properties of speech.

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 talker. These properties of speech are sometimes referred to as "indexical" properties of the signal (Pisoni, 1997) and convey information regarding the talker's identity, gender, age, regional background, and emotional state of mind (Bricker & Pruzansky, 1976; Kramer, 1963; Kreiman, 1997; Ptacek & Sanders, 1966). Indexical information is usually conceptualized as contrasting with symbolically-encoded "linguistic" information about the intended pattern of phonemes/phonemic contrasts (see Ladefoged & Broadbent, 1957; Pisoni, 1997). Since 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 (Mullennix, 1997; Sommers, Kirk, & Pisoni, 1997).

The ability to use indexical information in the speech signal to perceptually discriminate between the voices of different talkers is typically taken for granted in normal 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 detect a change of speaker when it occurs. For both normal-hearing and hearing-impaired individuals, this task becomes more difficult when associated visual cues are unavailable. Auditory-only situations that involve listening to multiple talkers, such as might result from group

teleconferencing, listening to a radio, or simply facing away from the faces of participants in a conversation, pose a particular challenge.

In the present study, we asked children with cochlear implants and a comparison group of younger normal-hearing children, to make perceptual judgments about whether or not pairs of recorded sentences were spoken by the same talker. Two conditions were explored. 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 to the children, we reasoned that in order to perform the varied sentence condition it would be necessary for the listeners to form and maintain 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 comparative 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 our talker discrimination task through several methodological simplifications. In addition to acoustic similarities between voices, 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 a small set of three talkers.

Our choice of three female speakers as our test talkers was motivated by several considerations. Previous research has suggested that prelingually deafened children who use cochlear implants reliably discriminate male voices from female voices at above-chance levels after about two years of implant use. Early studies by Osberger, Miyamoto, Zimmerman-Phillips, et al. (1991), and Staller, Dowell, Beiter, and Brimacombe (1991), for example, reported mean scores two years post-implantation, of 70 to 80% correct for discrimination of talker gender amongst groups comprised primarily of prelingually-deafened children. Several studies have also demonstrated that normal-hearing children can very accurately discriminate the large acoustic differences that distinguish the speech of male versus female talkers. Bennett and Montero-Diaz (1982), for example, found that normal-hearing children ages 6 to 8 years of age, could categorize speech samples as spoken by either a male or female talker with an accuracy rate of over 97% correct.

In designing the present study we therefore decided to focus on assessing the children's ability to make more subtle perceptual discriminations between talkers, such as those involved in discriminating between talkers of the same gender. Moreover, our initial reading of the prior clinical studies led us to believe that ceiling effects might be observed in the group of pediatric cochlear implant users, as well as in the comparison group of normal-hearing children, if the task was simply one of gender discrimination.

Our manipulation of the linguistic content of the utterances used in the talker discrimination task was motivated in part by recent findings on the interaction in perception between the linguistic and indexical properties of speech (Pisoni, 1997). Research has shown that the presence of indexical variability influences the speed and accuracy with which linguistic information is perceived, 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 indexical properties of speech.

A number of related studies involving both normal-hearing and hearing-impaired children have been conducted. In a series of papers, Jerger and her colleagues investigated whether children demonstrate the same degree of interaction between linguistic and indexical processing as shown by adults. The primary methodology used in these studies, the Garner speeded classification task, requires subjects to rapidly categorize stimuli along a particular dimension, while attempting to ignore stimulus variability along other task-irrelevant dimensions (Garner, 1974). 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 gender identification 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 milliseconds 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 milliseconds.

In a subsequent study, Jerger, Martin, Pearson, and 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 the age of two years. 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 indicated 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 employed in the present study differs considerably from the speeded classification task used by Jerger et al., the theoretical issues involved are fundamentally the same. 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 cognitive processing strategies, we are also interested more generally in the perception of similarity between talkers, and in determining which acoustic properties of the speech signal cochlear implant users are able to use to differentiate between talkers. The present study begins to address some of these larger issues.

The eight- and nine-year-old hearing-impaired children who participated in our study had all used a multi-channel cochlear implant for at least four years. Furthermore, at time of data collection all children were using a recent "state-of-the-art" coding strategy in their implant, capable of conveying a fairly detailed spectral representation of the speech signal.[1] Based on the design of these devices, the children's history of use, and previously published reports on talker gender discrimination performance in this population, we fully expected that many of the pediatric cochlear implant users would be able to make the simple discriminations presented under the "fixed sentence" condition. 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. We anticipated that the "varied sentence" condition, in contrast, might prove more difficult, since a generalizable representation of each talker's voice is presumably necessary to accomplish this task. If, however, the children were able to ignore the linguistic variability as directed, performance in the varied

[1] A summary of how cochlear implant speech processors filter and transform the speech waveform in order to convert this signal into electrical pulses delivered to the cochlear nerve can be found in Chapter III of the present report (Cleary, 2003).

sentence condition should not differ significantly from that observed in the fixed sentence condition. For the varied sentence condition, data collected from a group of younger normal-hearing children would help us to assess the hearing-impaired children's performance.

Method

Participants

Normal-Hearing Children. Twenty-one normal-hearing (NH) preschoolers were tested as part of a larger project on speech perception 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). The data reported here were gathered from 22 consecutively recruited children with the data from one child eliminated from the final analysis due to experimenter error.

Pediatric Cochlear Implant Users. Forty-four hearing-impaired children with cochlear implants participated as part of a larger study that was being conducted at Central Institute for the Deaf (CID) (Geers, Nicholas, Tye-Murray, et al., 1999). As shown in Table 1, the hearing-impaired children ranged in age from 7.92 years to 9.91 years at time of testing (mean age = 8.76 years). All pediatric CI 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 children who use total communication (TC), i.e., who rely on manual signs to supplement spoken language. Forty-one of the hearing-impaired children used a Nucleus-22 cochlear implant. Of the three remaining children, two used a Clarion implant and one child was a former Nucleus-22 user who had switched to the newer Nucleus-24 cochlear implant four months prior to testing. All of the children at time of data collection were using a recent "state-of-the-art" coding strategy (either SPEAK or CIS).

TABLE 1

SUMMARY OF DEMOGRAPHIC CHARACTERISTICS FOR THE PEDIATRIC COCHLEAR IMPLANT GROUP

N=44	MEAN	MINIMUM	MAXIMUM	SD
AGE AT TESTING IN YEARS	8.76	7.92	9.91	0.53
AGE AT ONSET OF DEAFNESS IN MONTHS	2.52	0	36	7
DURATION OF DEAFNESS IN YEARS	2.94	0.58	5.17	1.11
DURATION OF CI USE IN YEARS	5.60	4.09	6.87	0.66
NUMBER OF ACTIVE CI ELECTRODES	18.20	8	22	2.82

The normal-hearing children were not recruited to serve as a direct comparison group to the CI group, and were, in fact, tested six months prior to beginning our research with the CI group. Nevertheless, we feel that reporting the normal-hearing 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 who were three to four years younger than the children with CIs in this study.

Stimulus Materials

The stimuli were selected from the Indiana Multi-Talker Sentence Database (Karl & Pisoni, 1994; Bradlow, Torretta, & Pisoni, 1996), a compact disc 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 and Pisoni (1994). Eight sentences were used for the practice trials and another twenty-four sentences were selected for use during the test trials. A list of these sentences can be found in the Appendix to this chapter.

The "medium" speaking rate from the sentence database 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").

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 (gravelly), and #21 (deeper). The three female talkers used for the test stimuli were Talker #06 (smooth, deeper, somewhat older), Talker #07 (gravelly, young, unmelodious), and Talker #23 (higher-pitched, young, more melodious). The three female talkers were judged by the experimenter to differ, at least impressionistically, along the dimensions of age, and roughness of voice. The type of indexical variation represented by these three female talkers is, therefore, 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 been shown to have generally higher speech intelligibility scores than the male talkers (see Bradlow et al., 1996). We reasoned that use of less intelligible sentences 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 et al., 1996).

In addition, the three female 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 female talkers in the database. As reported by Bradlow et al. (1996), the mean fundamental frequencies of the three females 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.

A same-different discrimination task with an equal number of "same voice" versus "different voice" pairings was employed. 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 done 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 received only one condition of the voice discrimination task, namely, the "varied sentence" condition. In this condition, no sentence was repeated within a trial, or across trials. To correctly respond "Same" the child therefore needed to recognize the talker in two linguistically different utterances. The discrimination trials were administered using a PC computer running a control program written in C.

After the child was instructed about the basic nature of the task, four practice trials were administered. All children received the same ordering of practice trials (same, different, same, different) with 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 then 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, the child 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. The child received no explicit feedback during the test trials regarding the accuracy of his/her performance.

Pediatric Cochlear Implant Users. The pediatric cochlear implant users were tested in a manner that was very similar to the normal-hearing 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, as spoken by the three different talkers, across all twelve trials. 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 two children.

All children with CIs first received the same four fixed sentence practice trials using a single sentence and two 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 received the same four varied sentence practice trials using eight different sentences and two different male voices. Finally, twelve randomized test trials using the three female talkers and 24 different sentences were administered. 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. The children received no explicit feedback during any of the test trials regarding the accuracy of their responses. The children were given as much time as they needed to make each response (within the limits of the 15 minute testing session) and were told to guess if they were unsure as to the correct response.

The pediatric cochlear implant users were tested using a Macintosh G3 portable laptop computer running a Psyscope script written to mimic the C program used with the normal-hearing children. Stimuli were presented via a loudspeaker (Advent AV280) at approximately 70 dB SPL. 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 these procedures were slightly 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.

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

Results

Normal-Hearing Children. 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 NH children is shown on the bottom 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. This pairing involved the two talkers with the most similar average fundamental frequency. 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 top and center panels of Figure 1. The mean accuracy for the group in the fixed sentence condition was 68% correct which is significantly above 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 chance ($t(43) = 3.10$, $p = .003$), indicates that the children with cochlear implants encountered considerable difficulty with this task. A paired-samples t-test between scores in the two conditions showed a significant decrease in scores for the varied sentence task as compared to the fixed sentence task ($t(43) = 3.66$, $p = .001$). Scores in the two conditions showed a weak but significant positive correlation ($r = +.30$, $p = .049$).

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. Although we did not test the normal-hearing 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 illustrates the distribution of the two possible error types in each condition for the CI users. One pattern that emerges from 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 hearing-impaired children in the CI group may have found it difficult to ignore the linguistic variability present in the varied sentence condition.

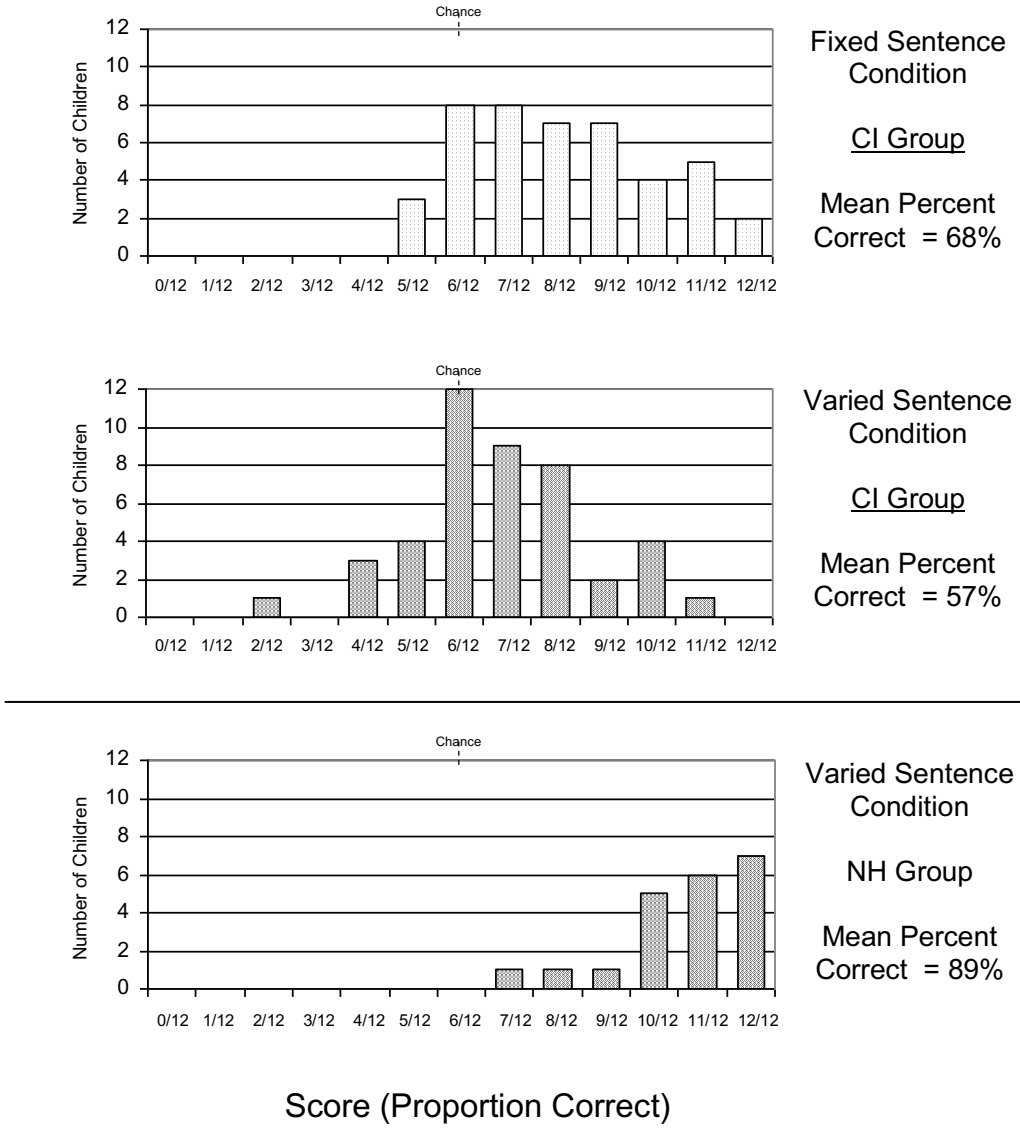


Figure 1. Distribution of talker discrimination scores in the "fixed" and "varied" sentence conditions for the pediatric cochlear implant users (N = 44), top and center panels. Distribution of scores in the "varied" sentence condition for a group of normal-hearing five-year olds (N = 21), bottom panel.

TABLE 2

TALKER DISCRIMINATION TASK, MISS RATES AND FALSE ALARM RATES FOR CI PARTICIPANTS

ERROR TYPES	FIXED SENTENCE CONDITION	VARIED SENTENCE CONDITION
RESPONDED "SAME" WHEN DIFFERENT MISS	.30	.34
RESPONDED "DIFFERENT" WHEN SAME FALSE ALARM	.35	.52

Although the observed talker discrimination scores were rather discontinuously distributed due to the small number of trials, the variability present in the obtained scores was sufficient to calculate correlations between talker discrimination scores and several other measures available for the children in the CI group. The measures included background demographic information as well as several independently collected measures of speech and language performance. Before proceeding, it should be noted however, that because the pediatric CI users were nearly at chance in the varied sentence talker discrimination condition, meaningful correlations obtained with scores in this condition 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 testing, age at onset of deafness, duration of deafness, or duration of implant use. Recall, however, that the children were pre-selected before testing to demonstrate relatively little variability along these dimensions. Weak evidence was found for a greater number of active electrodes and more exposure to oral-only communication being positively associated with better talker discrimination scores. 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. 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-TC" group's mean of 52% correct ($t(42) = 2.41$, $p = .021$) .

As shown in Table 4, talker discrimination scores were positively correlated with three measures of spoken word recognition gathered by CID clinicians for another project (see Geers et al., 1999) 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 cues to talker identity in the speech signal.

TABLE 3

SIMPLE BIVARIATE CORRELATIONS BETWEEN TALKER DISCRIMINATION PERFORMANCE AND DEMOGRAPHIC VARIABLES

	PROPORTION CORRECT	PROPORTION CORRECT

	FIXED SENTENCE CONDITION	VARIED SENTENCE CONDITION
AGE AT TESTING IN YEARS	-.13	.06
AGE AT ONSET OF DEAFNESS IN MONTHS	.19	.16
DURATION OF DEAFNESS IN YEARS	-.12	.06
DURATION OF IMPLANT USE IN YEARS	-.06	-.19
NUMBER OF ACTIVE ELECTRODES	.26	.19
DEGREE OF EXPOSURE TO AN ORAL-ONLY COMMUNICATION ENVIRONMENT	.27	.32*

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

TABLE 4

SIMPLE BIVARIATE CORRELATIONS BETWEEN TALKER DISCRIMINATION PERFORMANCE AND WORD RECOGNITION MEASURES

	FIXED SENTENCE CONDITION	VARIED SENTENCE CONDITION
WORD INTELLIGIBILITY BY PICTURE IDENTIFICATION, CLOSED-SET WORD IDENTIFICATION (WIPI) (ROSS & LERMAN, 1970)	.60 **	.36 *
BAMFORD-KOWEL-BENCH OPEN-SET SENTENCE TEST, KEY WORD IDENTIFICATION (BKB) (BENCH, KOWAL, & BAMFORD, 1979)	.44 **	.16
LEXICAL NEIGHBORHOOD TEST, EASY WORD LISTS OPEN-SET WORD IDENTIFICATION (LNTE) (KIRK, PISONI, & OSBERGER, 1995)	.48 **	.32 *

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Discussion

The talker discrimination results reported in this paper demonstrate that hearing-impaired prelingually-deafened children who have acquired language via a multi-channel cochlear implant have difficulty discriminating between similar-sounding talkers, particularly under conditions where the linguistic content of the message is varied. Depending on how the talker discrimination task in the varied sentence condition is conceptualized, the hearing-impaired children who participated in this study appear clearly less able than normal-hearing children to either, “generalize” their representations of a talker’s voice

to a new utterance, or "ignore" the presence of linguistic variability that is largely irrelevant to the task at hand.

Previously published studies regarding the discrimination of talker voice gender in children with cochlear implants had led us to expect somewhat better performance from the CI group than we, in fact, observed. Upon further reflection however, this expectation of ours may have been unwarranted. Although the existing literature typically describes the ability of children with CIs to discriminate between talkers of different genders, as "good" relative to their open-set speech perception skills (Osberger, Miyamoto, Zimmerman-Phillips, et al., 1991; Staller, Dowell, Beiter, & Brimacombe, 1991), the 70 to 80% accuracy levels for gender discrimination reported in these previous studies can be viewed as actually rather poor given the large acoustic differences in fundamental and formant frequencies that typically distinguish male from female speech. Because our stimulus set required a more difficult discrimination than has been typically used in previous studies that have asked children with CIs to make judgments about the indexical properties of speech, we did expect that the children would find our talker discrimination tasks more difficult. Nevertheless, we were surprised by just how difficult many of the CI children found the varied sentence condition, particularly given that this particular discrimination was well within the sensory and cognitive abilities of normally developing children.

The data presented here are inconsistent with the view suggested by some of the findings of Jerger and colleagues, that hearing-impaired children may find it relatively easy to ignore linguistic variability and instead attend to the more coarse-grained and temporally extended spectral information conveying voice quality in speech (Jerger, Martin, Pearson, & Dinh, 1995). We have found no evidence that the severe hearing-impairments and spoken language delays of the pediatric CI users in the present study cause these children to treat the speech signal as primarily a source of non-linguistic information about a talker. Discriminating between talkers of the same gender appears to be quite challenging for hearing-impaired children who use cochlear implants and the presence of linguistic variability simply makes the task harder for them.

We suggest that, although the current generation of cochlear implants code sufficient spectral detail (Loizou, 1998) to, in principle, permit talker discrimination, hearing-impaired children with CIs have difficulty interpreting the acoustic information in the speech signal that normal-hearing listeners routinely use to recognize subtle indexical differences between talkers. Given that the ability to explicitly discriminate between talkers develops over time even in normal-hearing children, with the ability to recognize briefly studied unfamiliar voices continuing to improve throughout the school-age years and reaching adult levels by adolescence (Mann, Diamond, & Carey, 1979), 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 and routine listening activities using the implant, or whether explicit training and feedback are necessary. As prelingually-deaf children with cochlear implants begin to enter main-stream classrooms in larger numbers, it will become increasingly important to understand how these children encode and process the wide indexical variability present in spoken language.

References

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for hearing-impaired children. *British Journal of Audiology, 13,* 108-112.

Bennett, S., & Montero-Diaz, L. (1982). Children's perception of speaker sex. *Journal of Phonetics, 10,* 113-121.

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.

Bricker, P.D. & Pruzansky, S. (1976). Speaker recognition. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics (pp.295-326).* NY: Academic Press.

Cleary, M. (2003/Chapter III). Influence of voice similarity on talker discrimination in normal-hearing children and hearing-impaired children with cochlear implants. This volume.

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope, 58,* 955-991.

Garner, W. (1974). *The processing of information and structure.* Potomac, MD: Lawrence Erlbaum.

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.

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 & Psychophysics, 54,* 310-320.

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.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16,* 470-481.

Kramer, E. (1963). Judgment 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.

Ladefoged, P. & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98-104.

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.

Miller, J.L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics, 24*, 175-180.

Mullennix, J.W. (1997). On the nature of the perceptual adjustments to voice. In K. Johnson & J.W. Mullennix (Eds.) *Talker Variability in Speech Processing (pp. 67-84).* San Diego: Academic Press.

Mullennix, J., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & 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 & 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., & Sumby, 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.

Ross, M., & Lerman, J. (1970). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13,* 44-53.

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 & Hearing, 18,* 89-99.

Staller, S.J., Dowell, R.C., Beiter, A.L., & Brimacombe, J.A. (1991). Perceptual abilities of children with the Nucleus 22-channel cochlear implant. *Ear & Hearing, 12 (Supplement),* 34S-47S.

Appendix

The stimuli were selected from the Indiana Multi-Talker Database, a compact disc containing recordings of 100 of the Harvard Sentences as spoken by 21 talkers (see Karl & Pisoni, 1994).

Practice stimuli: Speaking rate #02, Male talkers #01 and #21

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

Test Stimuli: Speaking rate #02, Female talkers #06, #07, and #23

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

IIB. TALKER DISCRIMINATION IN IMPLANTED CHILDREN ATTENDING AN ORAL SCHOOL FOR THE HEARING-IMPAIRED: A SHORT REPORT

Purpose

Cleary (2003, Chapter IIA) examined the degree to which children with cochlear implants are able to discriminate between talkers based on sentence-length utterances. The results of this previous study suggested that even after over four years of implant use, prelingually-deafened children who use cochlear implants experience difficulty in tasks that require them to explicitly discriminate between talkers whose voices are very easily discriminated by young normal-hearing children. From that earlier report, it was concluded that children with cochlear implants appear to do especially poorly at a same-different talker decision task when the linguistic content differs between the two speech samples to be compared.

One purpose of the present study was to verify that the difficulty in talker discrimination first reported in Cleary (2003, Chapter IIA) for children who use cochlear implants was, in fact a real and replicable result. Assuming that these perceptual difficulties were again observed, the other primary objective was to obtain an estimate of the degree to which two voices needed to differ in their average spectral content for implanted children to begin to be able perceive these differences. Because we had several future studies in mind that involved tests of talker discrimination using a stimulus continuum based on mean formant and fundamental frequency differences, some basic information regarding implanted children's ability to discriminate differences along these dimensions was needed in order to help us calibrate the difficulty of our future task.

Method

Participants

Fourteen children with cochlear implants participated in the study. The range of ages represented in the group was large (5.8 years to 12.67 years), with an average chronological age of 9.3 years. This mean age is therefore comparable to that of the 8- and 9-year-old children reported on in the previous study. All but one of the children were reported as congenitally deaf since birth. The remaining child became profoundly deaf at age 25 months. Age at implantation ranged from 2;0 to 7;8, with a mean of 4;6. Every child had used his or her cochlear implant for at least two years at the time of testing, with the average length of experience being almost five years (Mean = 4.93 years, range = 2.25 to 7.75 years). This sample of children therefore represents a fairly experienced group of pediatric cochlear implant users.

Mean unaided pure tone average threshold was 108 dB HL (range = 102-118 dB HL). For eight of the children, the etiology of hearing loss was unknown. In the remaining children, the etiology was listed as cytomegalovirus in three cases, of genetic origin in two cases, and as meningitis in the remaining case. The group was evenly split among users of the Cochlear Corp. Nucleus device using the SPEAK processing strategy (6 Nucleus 22 users and 1 Nucleus 24 user), and the Advanced Bionics Clarion device programmed with the CIS strategy (n = 7). However, in this sample, chronological age was strongly correlated with device type, such that the Clarion CIS group tended to be younger than the SPEAK group, thereby preventing a direct comparison between the two devices.

All children in the present study used primarily Oral communication methods. Three of the children were male; the eleven remaining children were female. All data collection was carried out as part of a larger project being conducted by the Indiana University School of Medicine in collaboration with St. Joseph Institute for the Deaf, an oral school for deaf children in St. Louis, Missouri.

Stimulus Materials

Six separate sets of stimuli were assembled for six different testing conditions. For the "fixed sentence" and "varied sentence" conditions, the stimuli were exactly the same as those described in Chapter IIA, namely, recorded tokens from three female talkers. All of the test sentences were selected from the Harvard Sentence List (Egan, 1948; IEEE, 1969), but had been screened for similarity of syntactic form and familiarity of lexical content. Recordings of 8 practice and 24 test sentences were selected for use from the Indiana Multi-Talker Sentence Database (IMTSD), a compact disc of recorded utterances described in Karl and Pisoni (1994) and Bradlow, Torretta, and Pisoni (1996). The selected sentences from IMTSD Talkers #06, #07, and #23 were all 8 to 11 syllables in length and were 1.61 to 2.16 seconds in duration. Practice trials, as before, used the voices of two male talkers (Talker #01 and Talker #21).

To create a "four-way talker discrimination condition," tokens from male Talkers #01 and #21 were used together with tokens from female Talkers #06 and #07. In this condition, the task was to discriminate among the voices of all four talkers. Within each pair of talkers from a given gender, the voices were selected to be somewhat confusable, although still easily discriminable by normal-hearing listeners. The voices of two different male and female talkers were used to construct a set of practice trials.

For the three "voice manipulation" conditions, a new set of stimuli was derived from the existing stimuli. In order to try to make the differences along the talker voice continuum as salient as possible for the cochlear implant users, and in light of the difficulty the CI group had discriminating the pairs of natural female talkers in Chapter IIA, we decided to have the stimulus continuum extend from formant and fundamental frequency values typical of a middle-aged female talker to values more typical of an adult male speaker. We reasoned that even if there were some children who could only resolve enough spectral detail to treat the task as a gender discrimination task, and not, more specifically as directed, as a talker discrimination task, use of tokens that straddled the female/male continuum would make the task easier for these children.

To construct a simple set of artificially similar-sounding "talkers," the tokens from IMTSD Talker #06, the female talker with the lowest average f0, were resynthesized using the proprietary "pitch" shifting routine included in CoolEdit 2000, a sound analysis software package from Syntrillium Corporation. Our examination of its function suggests that their routine is a variant on Pitch Synchronous Over-Lap and Add (PSOLA) techniques and not an LPC-based method, although the company would not confirm this. Using this routine, we were able to rescale the spectral characteristics of the recorded speech tokens in order to obtain "new" voices having different fundamental and formant frequencies than present in the original tokens. This rescaling process retained the sentence and word durations as well as amplitude envelope patterns that typically differentiate recordings of the same sentence collected from different talkers.

All sentence stimuli from Talker #06 used in the previous study were resynthesized with a shift of minus 4 semitones, a shift of minus 2 semitones, and a shift of minus 1 semitone (1 semitone equaling about a 7-9 Hz downward shift in the f0 range of this particular talker). A semitone difference is defined as a ratio relation of $2^{1/12}$ (= ~ 1.0595), between a higher frequency to a lower frequency and thus represents differences defined along a logarithmic scale (Dowling & Harwood, 1986). Somewhat analogous stimulus construction methods involving semitone differences constructed using PSOLA can be found in Bird and Darwin (1997) and in Assmann (1999). Acoustic analysis using the speech analysis software package PRAAT (Boersma & Weenink, 2001) was used to confirm that the intended shift in both formant and fundamental frequencies was actually present in the resulting tokens.

The speech stimuli created using the downwards pitch shift of 4 semitones were not recognizable as natural productions from female Talker #06 and were plausible instances of speech from a male talker. The 1-semitone shift was somewhat difficult even for a normal-hearing listener to detect, and did not appear to

destroy the individuality of the original talker's voice for a normal-hearing listener. The 2-semitone shift fell somewhere between these two extremes. A somewhat similar voice continuum involving simultaneous manipulation of both formant and fundamental frequencies is described in Mullennix, Johnson, Topcu-Durgun, and Farnsworth (1995), although in their case, synthesized vowels (rather than resynthesized sentences) were used as the stimulus materials.

Testing Procedure

The same basic procedure as used in Chapter IIA was used here. In all conditions, the child was asked to listen to two sentences played one after another and to decide if it sounded like the same person saying both sentences, or more like two different people. A one-second delay separated the offset of the first sentence and the onset of the second sentence. The children verbally indicated their responses which were then recorded by the experimenter. Although direction and feedback were given during the practice trials, no explicit feedback regarding the accuracy of the children's responses was provided by the experimenter during the test trials.

Due to time constraints, not all children could be run under all six conditions of the original design. The following procedure was used: the fixed-sentence condition using the 4-semitone difference was administered first in all cases. If the child became confused and frustrated during this condition, an easier condition, not completed by the other better-performing children, using a fixed sentence and a four-way discrimination among four "natural" talkers, two male and two female, was next administered. Following this, testing was then ceased. If, however, the child was able to complete the initial 4-semitone fixed sentence condition without obvious problems in making this discrimination, the child was given the fixed and varied sentence conditions described in Chapter IIA, then the 2-semitone fixed sentence condition, and finally the 1-semitone fixed sentence condition. Each condition took only a few minutes to administer, for a total testing time of about 15 minutes in all cases.

The sentences used in the five conditions involving a fixed sentence were a subset of 14 sentences taken from the larger group of 24 utterances used in the varied sentence condition. Each child heard the same sentence in all of the fixed sentence conditions in which he/she was tested.

For the fixed sentence conditions that tested the differences of 4, 2, and 1-semitone, each child completed 12 trials per condition. In these manipulated speech conditions, in six of the trials, the two sentence stimuli heard by the child on each trial differed physically in mean pitch by the amount specified. On the remaining six trials, no pitch difference was present. In the conditions that used natural, non-manipulated speech, 12 test trials were used to test discrimination of the three female talkers in each of the fixed and varied conditions, and 24 test trials were used in the condition which tested discrimination among four natural talkers, two male and two female. The greater number of trials in the four-way natural talker case was necessary to test every pairing of the four talkers and also include an equal number of trials in which a correct response of "same talker" was expected.

Each child was tested individually in a quiet room. Stimuli were presented using a loudspeaker (Advent AV280) located on the side of the child's implanted ear, at approximately 70 dB SPL as determined by a sound level meter located at approximately the level of the child's head. Presentation of all stimuli was audible to the examiner. Children who completed both fixed and varied sentence conditions were warned for these conditions in advance that the sentences they were about to hear would either "change all the time" or " be repeated over and over" as appropriate. The children were also told, however, that this change in procedure was "not important," and should just be ignored. It was reiterated that their "job was to do just like we were doing before, listen to the voices and decide if it sounds like the same person saying both sentences, or more like two different people."

Although the experimenter provided explicit feedback regarding the correct response in each of the four practice trials for each natural-speech condition, during the practice trials for the 4-semitone manipulated speech conditions, the experimenter avoided saying what was "correct" and simply modeled the predicted response of a normal-hearing person on the practice trials by saying, "I think that sounded like the same person," and "I think that sounded more like two different people talking that time." Because the "correct" answer is inherently undefined in the more difficult 2- and 1-semitone conditions, no practice trials were included for these conditions. Since those children who were tested in these conditions had, however, already completed three other conditions using the same response format, it was judged that this would not affect their understanding of the task.

Results and Discussion

Group means for each condition are shown in Figure 1. The overall group mean in the 4-semitone condition used to divide the children into two groups was 71% correct, where "correct" was defined as responding "different talker" for trials in which the 4-semitone pitch difference was present. This is shown in the top panel of Figure 1. There were more than twice as many incorrect responses of "same talker" in cases where the voices in fact differed than there were incorrect responses of "different talker" in cases where the same voice had been presented.

Of the fourteen children who participated, six encountered problems in completing the initial 4-semitone condition. As shown in lower left hand panel of Figure 1, this group scored an average of 54% correct in discriminating this large 4-semitone difference. These six children also completed the four-way talker discrimination task involving the natural voices of 2 women and 2 men, scoring 65% correct on average. When the data were re-scored by adding credit for different voices identified as "same" when the two talkers were of the same gender, the children averaged 78% correct (although it is important to emphasize that discriminating gender was not part of the instructions given to participants). These results are shown in the dotted and clear bars in the lower left-hand panel of Figure 1.

Eight children completed the larger set of five conditions as shown in the lower right-hand panel of Figure 1. These children displayed an average score of 84% in the fixed sentence condition involving the 4-semitone difference. For the 2-semitone and 1-semitone differences, however, their performance was essentially at chance (50%). These results suggest that the manipulation of the acoustic similarity of the voices was effective in the case of the 4-semitone difference as compared to the two smaller differences. There was no apparent difference, however, between these children's ability to discriminate the 2-semitone difference and their ability to discriminate the 1-semitone difference.

This same group of 8 children obtained a mean score of 71% correct on the fixed sentence condition that used natural female voices, very similar to the 68% correct previously reported in Chapter IIA for the same test. In the varied sentence condition, however, the children as a group also scored 71% correct, on average, which is substantially better than the 58% correct obtained by the previously studied group of pediatric CI users in that condition. The cochlear implant users in the present study appear to be better able to recognize a talker's voice across utterances that differ in linguistic content. These last results did not, therefore, replicate the difference between the fixed and varied sentence conditions that was obtained in the previous study.

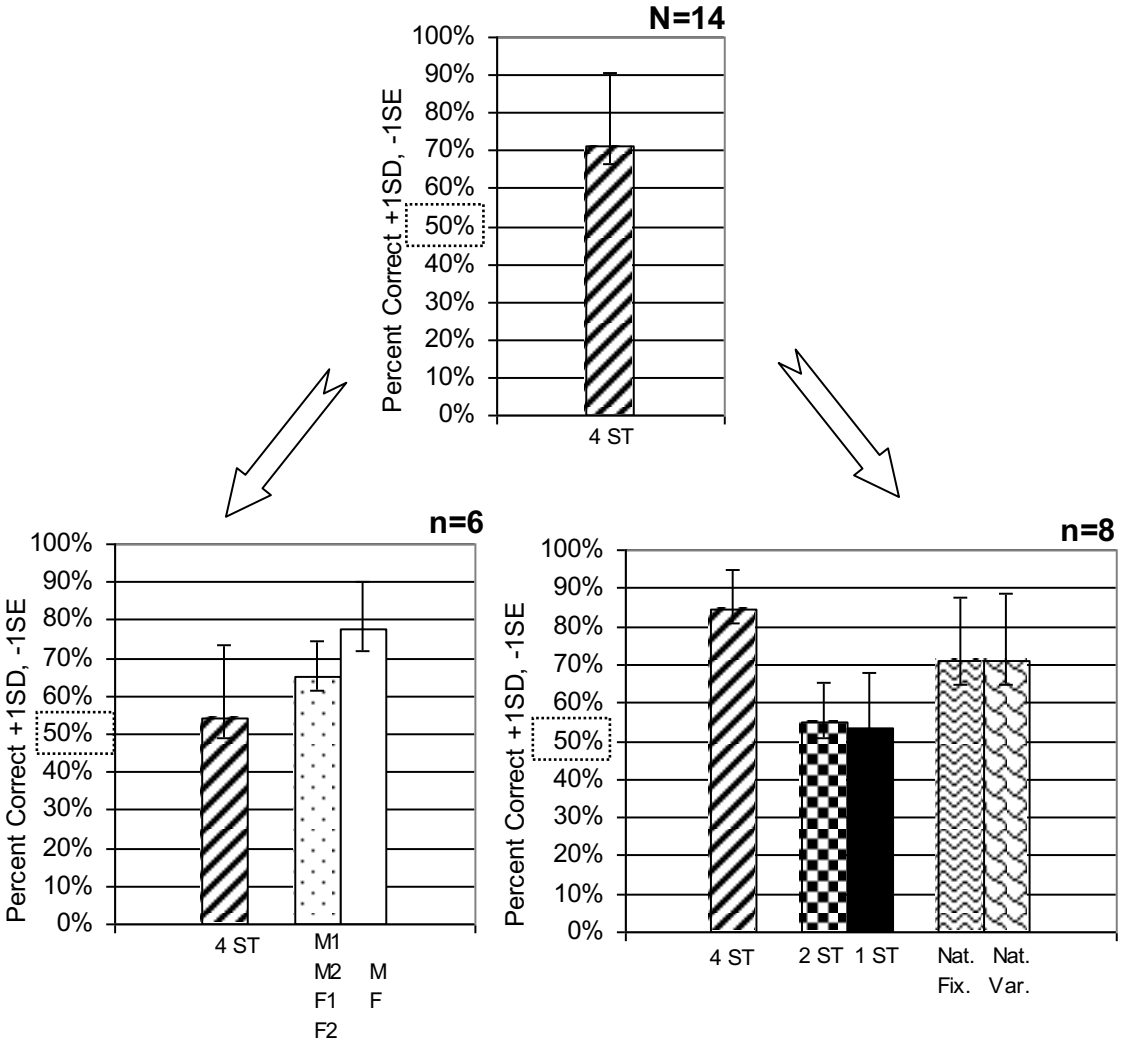


Figure 1. Mean percent correct in each talker discrimination task condition. Results from the 4-semitone difference condition ("4 ST") are shown in the top graph for all 14 children. Results for the subgroup of 6 children who struggled with the 4-semitone condition are shown at bottom left. Results from the remaining 8 children are shown at bottom right. "M1 M2 F1 F2" = four-way discrimination, natural talkers. "M F" = performance on the four-way discrimination task, rescored for discriminating the male from female talkers. "2 ST" = 2 semitone condition. "1 ST" = 1 semitone condition. "Nat. Fix." = Discrimination among 3 natural female talkers, fixed sentence condition. "Nat. Var." = Discrimination among 3 natural female talkers, varied sentence condition.

Although many different post-hoc analyses were tried in attempting to account for the finding of equivalent performance in the fixed and varied sentence conditions in these eight children, no single factor could be identified to fully account for this finding. The fact that the present sample of children obtained higher scores under the varied sentence condition than the children in Cleary (2003, Chapter IIA), may be partly accounted for by the inclusion of only children who use oral communication in the present study. The previous study reported that Oral children averaged 62% correct on the varied sentence condition as

compared to 52% correct in the TC group, and thus, it might be reasonable to expect an all-Oral group of children to do better, in general, on the task.

The finding of equivalent performance in the fixed and varied sentence conditions cannot, however, be easily accounted for in terms of communication mode differences, because regardless of how communication mode was defined for the participants in the prior study, performance in the varied sentence condition was always substantially worse than performance in the fixed sentence condition. That is to say, in that previous study even the children who used exclusively oral communication showed significantly poorer performance in the varied than in the fixed sentence condition.

The present results are likely due to factors such as the small sample size in this preliminary study ($n = 8$, for the fixed vs. varied comparison), the wide range in chronological ages within the group, and the selective pre-screening of participants based on the 4-semitone discrimination condition. Regarding this last possibility, it is conceivable that this selection process yielded a relatively non-representative group of eight children who were able to ignore the linguistic variability well enough to score as highly in the varied sentence condition as they did in the fixed sentence condition.

Given the large differences between the groups and testing conditions used in Cleary (2003, Chapter IIA) and the present study, it was striking, however, to note how nearly equivalent the average scores for the fixed sentence condition were across the two studies. Both studies clearly demonstrate that talker discrimination is not an easy task for children with cochlear implants, even after several years of implant use.

Despite the obvious difficulty many of the children encountered on this task, it is worth noting that one implanted child did extremely well in all five conditions in which he was tested, obtaining the following set of scores: 4-semitone difference = 11/12, 2-semitone difference = 9/12, 1-semitone difference = 8/12, Fixed Sentence (natural) = 11/12, Varied Sentence (natural) = 11/12. When interacting with this child, it was difficult to tell that he had a hearing impairment. His speech was highly intelligible, he was able to converse fluently, and he displayed a noticeable Texas accent.

A variety of correlational analyses were attempted using demographic information and other speech perception measures gathered from these same children within a few days of the present testing as part of the larger project being conducted at the Indiana University School of Medicine. As sample sizes for the two subgroups were too small to conduct meaningful analyses, only the correlations with performance in the 4-semitone discrimination condition warrant any kind of report here. Performance on the 4-semitone pitch difference talker discrimination task was significantly correlated with the chronological age of the child ($r = +.66$, $p < .05$) as well as with the child's duration of implant use ($r = +.60$, $p < .05$). No other correlations of note emerged from examination of participant or implant characteristics.

Somewhat contrary to expectation, we also found that none of the correlations calculated between discrimination of the 4-semitone difference and any of the three auditory-only spoken word recognition measures available for the 14 children reached statistical significance. The word recognition tests examined were all open-set tests of isolated word identification, and included the Phonetically Balanced Kindergarten (PBK) Test (Haskins, 1949), and the Lexical Neighborhood Test Easy and Hard Word Lists (LNT-E, LNT-H, Kirk, Pisoni, & Osberger, 1995). (Word recognition scores from spoken sentence contexts were not available for most of the children.) As was also the case in comparing between condition means, the small sample size and heterogeneity of the group hampered the correlational analyses. One possible reason for the failure to find a relationship between talker discrimination performance and word recognition scores in this study, as compared to the previous study, may be the inclusion in the present sample of five children who were implanted after the age of six years. This subset of children who experienced a comparatively

lengthy period of auditory deprivation between onset of deafness and cochlear implantation, scored very poorly relative to the larger group, on the word recognition tests, but performed at an average level on the voice discrimination task.

In summary, the results of this study serve to replicate the finding in Chapter IIA of particularly poor performance of pediatric cochlear implant users on a very simple talker discrimination task. Even after several years of cochlear implant use, prelingually-deaf children who use CIs appear to experience great difficulty explicitly discriminating between talkers who are very easily discriminated by young normal-hearing children (see Chapter IIA). However, the present study failed to replicate the previously reported finding that children with cochlear implants appear to do especially poorly at talker discrimination tasks when the linguistic content of the speech samples to be compared differs. Due to constraints on the sample size and available participant pool we were unable to find any measurable relationships between performance on the talker discrimination tasks and spoken word recognition scores. The present results do, however, suggest that while some implanted children may be able to detect a 4-semitone difference between voices, an even larger pitch difference may be necessary to simplify the task. The data also indicate that it is unlikely that pitch shifts smaller than 2 semitones will be perceived by any child with a cochlear implant as indicating a change of talker. These findings provide us with some initial information regarding the spectral discrimination abilities of this clinical population that will be useful in the design of new studies examining how talker discrimination in children with cochlear implants is influenced by manipulations of voice similarity.

References

Assmann, P.F. (1999). Fundamental frequency and the intelligibility of competing voices. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco, CA, August.

Bird, J. and Darwin, C.J. (1997). Effects of a difference in fundamental frequency in separating two sentences. In A.R. Palmer, A. Rees, A.Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing (pp.263-269).* London: Whurr.

Boersma, P. & Weenink, D. (2001). Praat Version 3.9.27: A system for doing phonetics by computer. www.praat.org.

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. (2003/Chapter IIA). Talker discrimination in prelingually deaf children with cochlear implants. This volume.

Dowling, W.J., & Harwood, D.L. (1986). *Music Cognition.* San Diego, CA: Academic Press.

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope, 58,* 955-991.

IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Report No. 297*.

Haskins, H. (1949). A phonetically balanced test of speech discrimination for children. Unpublished master's thesis. Evanston, IL: Northwestern University.

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.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing, 16,* 470-481.

Mullennix, J., Johnson, K., Topcu-Durgun, M., & Farnsworth, L. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 98,* 3080-3095.

CHAPTER III - INFLUENCE OF VOICE SIMILARITY ON TALKER DISCRIMINATION IN NORMAL-HEARING CHILDREN AND HEARING-IMPAIRED CHILDREN WITH COCHLEAR IMPLANTS

Abstract. In this study we investigated how manipulations of voice similarity affect talker discrimination judgments in children. More specifically we addressed the question, how different must two sentences be in terms of their average fundamental and formant frequencies for children to categorize these utterances as spoken by different talkers? Same-different judgments were obtained using an adaptive testing procedure in which the similarity of voice pairs was systematically varied. Participants included 5-year-old normal-hearing (NH) children and cochlear implant (CI) users ages 5-12 years, implanted at or before age 6 years with 2+ years of implant experience. Stimuli consisted of natural sentence-length utterances resynthesized to form a stimulus continuum of similar-sounding voices. The probability of "different talker" judgments was examined for each listener group as a function of pitch difference size. We compared this measure of "acceptable within-talker variability" under two different presentation conditions, one in which the linguistic content of the sentence was held constant and one in which the linguistic content was varied. We found that the spectral envelopes of two utterances, together with their average fundamental frequencies, needed to differ by at least 11-16% for normal-hearing children to perceive the voices as belonging to different talkers. Normal-hearing children maintained the perception of voice individuality over a larger frequency shift than previously studied normal-hearing adults. Although several individual children with implants exhibited response patterns that resembled those of normal-hearing children, most of the hearing-impaired children experienced difficulty perceiving even large differences between talkers. Both groups of children displayed essentially the same pattern of responses in both the "fixed sentence" and "varied sentence" conditions, indicating no effect of linguistic variability on the children's criteria for identifying two talkers as the same or different. Greater variability in performance was observed, however, in the varied sentence condition for both listener groups. Finally, within the CI group, higher spoken word recognition scores were found to be associated with better talker discrimination performance. Our results suggest that poor discrimination of pitch and timbre-based differences is a factor contributing to implanted children's difficulty with talker discrimination tasks. Additionally, the present findings offer specific predictions regarding the degree to which natural voices must differ for normal-hearing children to perceive a change in talker. We discuss our findings with reference to how children with implants process cues to pitch and timbre in other complex sounds such as music, and speculate regarding how it is that some implanted children appear be able to perceive phoneme contrasts but still have difficulty discriminating between talkers. The relevance of voice similarity perception to current developmental theories of speech perception and spoken word recognition is also reviewed. Finally, we consider how source monitoring and perception of vocal signature are skills relevant to all species that use acoustic signals to communicate.

Introduction

Although the speech signal is generally thought of as a channel for communicating linguistic information coded by word and phoneme contrasts, it has been known for many years that normally-produced speech also contains a wealth of extra-linguistic information about the talker (Ladefoged & Broadbent, 1957; Peters, 1954; Voiers, 1964). Through these so-called "indexical" properties of speech (Pisoni, 1997), listeners can frequently make inferences concerning a talker's unique identity (in the case of a familiar talker), as well as more general categorizations such as gender, age, regional background, emotional state of mind, etc. (Kramer, 1963; Voiers, 1964; see reviews in Kreiman, 1997; Bricker & Pruzansky, 1976). Even given relatively short samples of speech, lasting only a few

seconds, normal-hearing listeners can often make judgments (though not necessarily accurate judgments) about indexical properties of the individual talker (Bricker & Pruzansky, 1966; Pollack, Pickett, & Sumby, 1954). When complementary visual cues are unavailable, such as during telephone or radio use, the ability to extract basic indexical information from the auditory speech signal can become particularly important for successful communication (e.g., Daly-Jones, Monk, & Watts, 1998; Williams, 1977).

Because linguistic and indexical information are both simultaneously encoded in the acoustic waveform, a fundamental question of great interest to speech researchers is how the parallel extraction of these two types of information takes place, and the degree to which these perceptual processes interact with each other (e.g., Mullennix & Pisoni, 1990; Pisoni, 1997). A large and growing literature has begun to address the relationships between the perception of linguistic and indexical information (Johnson & Mullennix, 1997). Building on a number of studies with normal-hearing adults, new research has examined interactions between the perception of linguistic and indexical information in normal-hearing children (Ryalls & Pisoni, 1997), as well as in special populations such as elderly normal-hearing listeners, and adults and children with hearing impairments (Sommers, 1997; Jerger, Martin, Pearson, & Dinh, 1995; Kirk, Pisoni, & Miyamoto, 1997). Taken together, these studies have shown that speech perception cannot be studied in isolation from how listeners process indexical information in the acoustic signal.

Perhaps the simplest form of indexical processing allows a listener to perceive that a particular talker is the same or has changed from one utterance to another (Bricker & Pruzansky, 1976). This ability to discriminate between different talkers does not require that a listener be able to explicitly recognize or name a particular talker or to even be able describe the perceptual qualities that differ between talkers, but merely to notice that a voice is the same or different, irrespective of the linguistic content of the utterances. This particular ability can be aptly characterized as the capacity to discriminate between within-talker variability and cross-talker variability. That is, the listener's perceptual system must be able to estimate the likelihood that a given amount and type of acoustic variation is due to natural/ordinary fluctuations in a single talker's articulatory productions as opposed to a change in the individual vocal source/organism producing the speech. Thus, in monitoring talker voice, there are actually two complementary tasks involved—discriminating between talkers when a change occurs, and identifying a given talker as the same individual across different utterances. Given the amount of variability that exists in the speech signal across talkers as well as within the speech of an individual talker, these are not trivial problems, yet the normally-developed human perceptual system usually accomplishes these tasks very quickly with relative ease and little conscious awareness.

The ability to identify and discriminate between talkers is influenced by a variety of factors, including characteristics of the stimuli, characteristics of the listening environment, and characteristics of the listener (Bricker & Pruzansky, 1976). The present study examined the influence of two of these factors, the degree of acoustic similarity between voice pairs, and the hearing status of a child listener. Judgments of perceptual similarity were obtained from normal-hearing children and children with cochlear implants to assess how similar two sentences needed to be in terms of average fundamental and formant frequencies for a listener to categorize both utterances as spoken by the same talker. The present study also assessed the ability of these two groups of children to distinguish within-talker variability from between-talker variability by examining same-different talker judgments obtained when the linguistic content of two voice samples remained the same compared to when the linguistic content of the two voice samples differed.

Although a sizable literature now exists on the ability of infants to discriminate between voices using habituation techniques (e.g., DeCasper & Fifer, 1980; Miller, 1983; Mills & Melhuish, 1974; see Locke, 1993 for review), there are only a relatively small number of studies on preschool and school-age normal-hearing children's perception of talker similarity. The literature that has been published on this problem tends to categorize itself as either a study of voice recognition or a study of voice discrimination.

Typical voice recognition tasks require the matching of a name, face, or other label, to a sample of an already familiarized voice, often after extensive opportunity to learn the tested voice and some substantial period of delay intervening before testing. Voice discrimination tasks, on the other hand, tend to de-emphasize the role of memory and learning by using pairs of utterances presented one after the other in quick succession, with the task being to decide if the utterances have been produced by the same talker or by two different talkers (Bricker & Pruzansky, 1976; Kreiman & Papcun, 1991). A same-different talker discrimination task can, however, be viewed as an extreme case of a voice recognition task, offering minimal opportunity to learn the first presented voice, if it is not already familiar, and only a short delay before presentation of the second voice sample. Taking the perspective that studies using both recognition and discrimination methodologies may be relevant to the present study, a review of the existing literature suggests that most typically-developing young children understand the tasks of recognizing and discriminating between voices and can perform these tasks at a level that is generally above chance performance though not at the level expected of adult listeners given the same stimuli.

For example, Bartholomeus (1973) examined whether four- and five-year-old normal-hearing children were able to recognize the familiar voices of their classmates five months into the academic year of a Canadian nursery school. Children were tested on several different tasks. In the first task, participants were shown photos of each classmate's face and asked to say the name of the pictured child. In the second task, the children listened to a recorded voice sample that was two sentences in length, and were asked to point to the correctly matched photo out of a set of pictures showing all twenty children. In the third and final task, the children listened to each two-sentence-long recorded voice sample, and were told to simply say the name of the classmate to whom the voice belonged. This voice-naming task was also carried out under a condition in which each voice sample was played backwards, thus making the lexical information unintelligible. Bartholomeus found that the children demonstrated nearly perfect performance on photo labeling, but scored only 55 to 60 percent correct on average on both the voice labeling and voice-face matching tasks. On the backwards-voice-labeling task, the children scored an average of 40 percent correct. Bartholomeus observed large individual differences among the children in performance on all the voice-related tasks.

In a similar study, Murry and Cort (1971) examined the ability of nine-year-old fourth-graders to recognize the voices of their classmates. Performance was compared under three different presentation conditions, representing differing amounts of voice information defined in terms of quantity and variety: recognition was tested using a five-second-long utterance, a five-second-long "ah" vowel sound, and a reading of the Rainbow Passage lasting approximately 20 seconds. The twenty children studied scored 47 percent correct on average for voice recognition of the vowel sample, but they were better than 95 percent correct when provided with either the 5-second utterance or the 20-second paragraph. Furthermore, Murry and Cort found that providing multiple repetitions of the speech sample before allowing the children to respond did not help performance.

More recently, Spence, Rollins, and Jerger (2002) assessed the voice identification skills of 24 children in each of three age groups, ages 3, 4 and 5 years. Twenty popular television cartoon character voices were selected as stimulus materials. Children at all three ages were able to match pictures of the cartoon characters to a four-second sample of each character's voice at better than chance levels of performance. Even three-year-olds were able to do the twenty-trial, six-alternative picture-pointing task at well above chance levels, i.e. averaging 61 percent correct. The five-year-olds performed the best, scoring 86 percent correct on average. Spence et al.'s results demonstrate that preschool-age children are able to recognize and label familiar voices learned only from listening to recorded speech materials and not in interactive communicative language situations.

In a somewhat more complicated study, Bennett and Montero-Diaz (1982) tested children ages 6 to 8 years of age on their ability to categorize one-second-long sustained vowel utterances from unfamiliar adult speakers as spoken by either a male or female talker. The experimenters used both phonated (voiced) speech samples as well as whispered (unvoiced) speech samples. The whispered unvoiced speech is an interesting case to study because under these conditions, only the talker's supralaryngeal vocal tract resonances are present, without the talker's fundamental frequency characteristics.

The children studied by Bennett and Montero-Diaz were able to quite easily discriminate the male from female voices, scoring better than 97 percent correct on both the male and female phonated samples. For the whispered unvoiced speech samples, gender categorization scores were somewhat lower but still well above chance. The children averaged 87 percent correct identification of the male whispered voices and 71 percent correct for the female whispered vowel samples. Bennett and Montero-Diaz also examined how well the children could accurately categorize the unfamiliar voices of other six- and seven-year-old children as either male or female. As was found also for adult listeners tested in another part of the study, the children had much more difficulty categorizing the gender of these preadolescent talkers, scoring about 65 percent correct on phonated samples and about 55 percent correct on the whispered samples.

Mann, Diamond, and Carey (1979) also examined children's perception of previously unfamiliar voices. However, in this study, an extremely simplified version of an old/new voice recognition memory paradigm was used. On each trial, the children heard either one or two target talkers for study and then after a three-second delay, they were asked to listen to either two test talkers or four test talkers and to select the target talker(s) from the test set. One hundred and forty children ranging in age from six to sixteen years of age were tested using a rather elaborate procedure. Although Mann et al. were particularly interested in the effect of target-voice set size, more interesting from the point of view of the present research was their investigation of the presence or absence of linguistic variability as a between-subjects factor. Half of the children performed the task under conditions in which the target utterances and the utterances in the test set were all linguistically the same sentence (although a different recorded token was used in the test-set utterances of the target talker(s)). The remaining half performed the task under conditions in which the target utterances differed in linguistic content from the utterances in the test set—that is, the sentence varied between study and test.

All age groups tested performed above chance levels, with the exception of the six-year-old group. Performance increased as a function of age between the ages of 6 and 10, with the ten-year-olds performing as well as adult undergraduate students also tested. Although the resulting data were rather noisy, performance was found to generally be worse in the "varied" sentence condition than in the "fixed" sentence condition, with this effect clearly evident for only the older children

The results of the Mann et al. study are worth noting here because the authors attempted to examine the perception of far more subtle indexical differences than have been employed in earlier studies; the stimulus set used by Mann et al. consisted of twenty-two adult female talkers between the ages of 25-45 years, screened for homogeneity of accent, with each talker presented on only one trial during the experiment. However, a weakness of the study is that performance measures were based on only four or fewer test trials per experimental cell/condition.

Intrigued by the results initially reported by Mann et al., Cleary (2003/Chapter IIA) tested a group of 21 normal-hearing five-year-olds using a simple discrimination task that required the children to categorize pairs of sentences as spoken either by the same talker or by two different talkers. A small set of three female talkers with similar speaking rates, similar regional accents and no marked emotional quality were selected for use in the study from an existing stimulus database. Twelve randomized test trials were used. The correct response was "same talker" for half the trials, and "different talkers" for the remaining

half. The stimuli were presented over headphones, with the examiner blind as to the nature of each trial. All testing was conducted under a "varied" sentence condition, so that in order to correctly respond "same talker" the child needed to recognize a given speaker in a new utterance that had not been heard prior. Cleary found that the normal-hearing five-year-olds had very little difficulty with the task, scoring 89 percent correct on average. Most of the errors were incorrect responses of "same talker" for pairs of sentences uttered by the two talkers with the most similar fundamental frequencies.

Building on the these preliminary results and on earlier work suggesting a need to further study the abilities of children with cochlear implants to process indexical information in speech (Osberger, Miyamoto, Zimmerman-Phillips, et al., 1991; Staller, Dowell, Beiter, & Brimacombe, 1991; Tyler, 1993), Cleary (2003/Chapter IIA) used the same stimulus materials to assess the talker discrimination skills of a group of eight- and nine-year-old hearing-impaired children each of whom had used a multi-channel cochlear implant for at least four years. In addition to the "varied" sentence condition already described, Cleary also tested each child using a "fixed" sentence condition in which the linguistic/lexical content of the utterances to be compared was always held constant. Each of the 24 sentences used in the "varied" sentence condition was used as the "fixed" sentence for approximately two participants.

Cleary (2003/Chapter IIA) found that the hearing-impaired children with cochlear implants had a great deal of difficulty in discriminating between the three female voices. In the fixed sentence condition, average performance was only 68 percent correct, and in the varied sentence condition, performance averaged 57 percent correct. The hearing-impaired children with cochlear implants performed far more poorly than the normal-hearing children previously studied, despite both their substantial amount of experience with the implant, and their older chronological age.

Previous studies have reported that after two years of implant use, prelingually deafened children with cochlear implants typically discriminate male voices from female voices at rates of only about 70 to 80% correct (Osberger, Miyamoto, Zimmerman-Phillips, et al., 1991; Staller, Dowell, Beiter, & Brimacombe, 1991), where chance performance is, of course, 50%. When compared to rates of 97% correct reported for young normal-hearing children on similar tasks (Bennett & Montero-Diaz, 1982), these data and the results of Cleary (2003/Chapter II) strongly suggest that talker discrimination skills in pediatric cochlear implant users require further examination.

Cleary (2003/Chapter IIB) therefore conducted a follow-up study assessing the ability of a small group of pediatric cochlear implant users to categorize utterance pairs as spoken by the same or different talkers given differing degrees of acoustic similarity between the voices in the test set. Unlike the previous study, the 14 participants in this follow-up study included only children who used oral communication methods. The primary task on which all 14 children were tested used a combination of recorded sentences from an adult female speaker and resynthesized versions of these same sentences after a four semitone downwards pitch shift of the entire speech spectrum, including fundamental frequency. The resynthesized tokens were heard by normal-hearing listeners as plausible instances of utterances spoken by a male talker. To simplify the task, the linguistic content of the sentences to be compared was always held constant.

Cleary (2003/Chapter IIB) found that the children with cochlear implants scored an average of 71 percent correct under this condition involving a difference of 4 semitones, where chance performance, as before, was 50 percent. A subset of the children who completed the 4-semitone condition more easily than the other participants were also tested on the same basic task using smaller pitch differences. These children performed at chance levels when asked to discriminate differences of one and two semitones. The children who had had difficulty with the 4-semitone discrimination were additionally tested on their ability to discriminate between four natural talkers, two male and two female. Even on this easier task, the children

scored an average of only 65 percent correct. The male talkers could be successfully discriminated from the female talkers 78 percent of the time.

Although a comprehensive literature review regarding the perceptual abilities of adult cochlear implant users is beyond the scope of this paper, one recent study on talker discrimination in adult cochlear implant users provides some important and useful benchmark data for the present investigation. Kirk, Houston, Pisoni, Sprunger, and Kim-Lee (2002) examined the effect of linguistic variability on same talker vs. different talker judgments for pairs of sentences and pairs of isolated words. Pairings among ten voices, five male and five female, were used in the isolated word condition. Pairings among eight voices, four male and four female, were used in the sentence condition. Kirk et al. found that adult implant users performed significantly worse on the categorization task than a comparison group of normal-hearing adults and were significantly less accurate in the isolated word condition than in the sentence condition. Overall, the cochlear implant users had particular difficulty in correctly responding "different" on trials in which the voices did, in fact, differ. The implant users also had more difficulty with the talker discrimination task when linguistic variability was present--that is, when the linguistic content of the paired utterances was different. This effect was substantially larger in the condition in which isolated words were used than when sentences were used. The normal-hearing controls, in contrast, showed no effect of linguistic variability in either the word or sentence condition. Taken together with the results from children reported earlier by Cleary (2003/Chapter IIA, Chapter IIB), these new findings with adults demonstrate that cochlear implant users do not very accurately perceive talker voice differences.

Most of the studies reviewed above used natural stimuli drawn from a fixed set of talkers contained in a pre-recorded database. One drawback of using utterances selected in this manner is that the experimenter has relatively little control over the perceptual similarity relations between the talkers in the sample. At best, a few of the studies have controlled similarity by recruiting talkers meeting narrowly defined criteria, or by hand-selecting test voices from a larger set of recorded talkers. This is largely also the case in recent studies on the effects of indexical variability on linguistic processing in normal-hearing adults, although some of these have obtained perceptual similarity data on the selected talkers using multidimensional scaling methods (e.g., Goldinger, 1996; Nygaard & Pisoni, 1998).

Adding to the problem is the fact that, as in the bulk of the published literature on talker identification and discrimination, the results of the earlier studies are reported primarily in terms of percent correct responses (see Kreiman, 1997, for discussion). However, unless the particular sample of talkers is somehow argued to be representative of an "ideal" sample, having the same degree of variability as one might be likely to encounter in everyday life, a percent correct measure is only minimally informative in the absence of a detailed description of the acoustic characteristics of the talkers in the corpus. Task performance will necessarily depend heavily on the similarities and differences among the talkers within the sample, and comparisons of results across different studies become more difficult when only percent correct is reported.

One of the main goals of the present study was therefore to examine the perception of voice similarity when acoustic dimensions known to be important to inter-talker differences are systematically manipulated. Defining the dimensions that underlie perceived differences among talkers is, however, a complicated issue that remains a matter of considerable debate and a topic of current research (Fellowes, Remez, & Rubin, 1997; Kreiman, 1997; Nolan, 1997). Despite much research in this area, the perceptual factors that children and adults use to discriminate between talkers are far from completely understood. Speech researchers have managed to identify a number of measurable acoustic dimensions along which talkers tend to differ, but the degree to which listeners "perceptually weigh" the importance of each dimension has not yet been resolved (see Kreiman, 1997 for discussion).

One potentially useful distinction can be made between voice characteristics determined largely by physiology/anatomy (for example, the lowest resonance of a neutrally positioned open vocal tract), versus voice characteristics which are largely learned, such as regional accent (Ladefoged & Broadbent, 1957). The present paper will focus on the former, that is, characteristics determined primarily by anatomy. More specifically, this study involves manipulation of what are among the more stable of these acoustic dimensions: a talker's average fundamental frequency (the rate at which the vocal folds open and close during voiced speech), and a talker's average formant frequencies (the frequencies at which harmonic (or nonharmonic) components of the source signal are maximally amplified).[2]

These acoustic values are determined largely although not exclusively by the physical dimensions and composition of the vocal folds and larynx in the case of fundamental frequency, and of the supra-laryngeal vocal tract in the case of formant frequencies (Fant, 1960; Stevens, 1998). Although these characteristics can be altered considerably within a single speaker, either involuntarily (e.g., from an illness such as a sore throat or nasal congestion), or more deliberately when some kind of vocal imitation or disguise is intended, these dimensions do constrain the possible range of sounds produced. Fundamental frequency and formant frequencies are only two of the many variables that can differ across talkers (see Carrell, 1984; Klatt & Klatt, 1990; Matsumoto, Hiki, Sone, & Nimura, 1973; Murry & Singh, 1980). These two dimensions however, have been well studied and a variety of sophisticated methods have been used over the years to investigate how the perception of voice quality is influenced by variation in these parameters (Carrell, 1984; Coleman, 1971; Fellowes, Remez, & Rubin, 1997; Kuwabara & Takagi, 1991; Lavner, Gath, & Rosenhouse, 2000; Matsumoto, Hiki, Sone, & Nimura, 1973).

Fundamental frequency varies considerably within any one talker's speech productions, mostly due to changes in the tension of the muscular vocal folds. Nevertheless, individual talkers can be characterized fairly reliably using measures of f0 central tendency and variability. Typical values for a 35-year-old adult male are 120-125 Hz, while for an adult female, of the same age, a typical f0 might be 200-210 Hz (Kent, 1997). While the distributions of average f0 for males versus females show very little overlap, within gender there is considerable variability from talker to talker. For example, the ten female talkers in the Indiana Multitalker Sentence Database exhibit average f0s ranging from 163 Hz to 237 Hz, while the ten male talkers exhibit average f0s ranging from 100 Hz to 142 Hz (Bradlow, Torretta, & Pisoni, 1996). For any individual speaker, the standard deviation around his/her average f0 has been argued to be about 3 semitones (Orlikoff & Kahane, 1996; although Fleetwood (1990) reports standard deviations of ~3.9 STs), corresponding to a 1 SD range of about 105-149 Hz for an average male talker with an f0 of 125 Hz, or about a 1 SD range of 172-243 Hz for an average female with an f0 of 205 Hz. An individual's minimum and maximum f0 can obviously fall well outside this 1 SD range, and male maximum f0 values can easily reach typically female minimum and average f0 values. Likewise, female minima typically dip into the upper half of the male range. Familiarity with a given talker's fundamental frequency characteristics can be assumed to include knowledge regarding the talker's average f0 and usual range of variation.

Formant frequencies can also aid listeners in discriminating between talkers and genders. Due to the larger dimensions of the typical male vocal tract than the typical female vocal tract, female formant frequencies tend to be about 15-20% higher on average than male formant frequencies (Fant, 1973). Thus, while an average male might have the F1 = 625 Hz, F2 = 1200 Hz, F3 = 2550 Hz for the central vowel [^], typical female formant values for the same vowel might be F1 = 750 Hz, F2 = 1425 Hz, F3 = 2930 Hz. The degree of difference observed tends to vary across the nature of the articulation for which the formants are

[2] Although the typical definition of a speech formant assumes that a formant is the result of a subset of enhanced harmonics of a fundamental frequency arising from vocal fold vibration, in the present discussion we expand the definition to include well-defined areas of spectral intensity in the upper frequencies that result from articulatory noise sources. That is, when we discuss "formant frequencies" we are more accurately referring to the speech spectrum and spectral shape over time, not only the simple resonances of the vocal tract.

measured (Fant, 1973). Again, within gender there is considerable variability between talkers. Although it will not be addressed further in this paper, anatomical variability can also influence the relative intensity of the individual formants and the extent of the formant bandwidths (Fant, 1960; Stevens, 1998).

Formant and fundamental frequency values can each provide information regarding the presence of continuity or change in the identity of the talker. Physiological constraints on these values tend to encode acoustic cues to talker identity in a manner that tends to be quite widely distributed in time across spoken utterances and which relies primarily on the average absolute values of individual spectral landmarks rather than on the relative locations of different spectral prominences considered together. Formant and fundamental frequency values have, however, also been opportunistically appropriated by nature as a means of conveying a "speech code" of linguistic contrasts (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). This code tends to involve more densely packed, rapidly changing information, and relies more on relative positioning or second-order relations both in time and frequency between resonance characteristics (Gerstman, 1968; Ladefoged & Broadbent, 1957).

These apparent differences between the extraction of talker-specific versus linguistic information naturally give rise to questions regarding which type of task is "simpler" for the listener. Over fifty years of speech science research suggests that the linguistic task is, in theory, a more difficult problem for the listener to solve because of the contextually-dependent nature of the densely packed and rapidly conveyed symbolic code. However, what is less clear is the degree to which the perception of indexical properties is actually a prerequisite skill for linguistic decoding, or whether linguistic decoding can take place irrespective of whether the listener is able to resolve the type of spectral detail that differentiates between talkers. There may be a need to more carefully evaluate the assumption that a hearing-impaired listener can be assumed to be able to perceive and recognize indexical characteristics such as talker gender and talker identity, at least as well or better than he/she perceives linguistic information in the speech signal.

Children who use cochlear implants are an interesting population to study with respect to the perception of indexical information in speech because cochlear implants have been developed and evaluated primarily in terms of how successfully they convey the acoustic phonetic information necessary to identify spoken words. Interestingly, one of the factors behind the success of the cochlear implant is the relative robustness of spoken word recognition to spectral shifts such as those that may occur when the various channels/bandwidths of acoustic information are mapped to locations somewhat different from what they would normally map to in a healthy cochlea. That is, much of the information used in coding phonemic contrasts lies in the relations between different spectral peaks/formants over time, not their absolute frequency values--in part, this is what is believed to allow the normal human speech processing system to maintain perceptual constancy of the linguistic message among utterances that differ considerably in their absolute formant frequency values (Peterson & Barney, 1952; Pisoni, 1997; Shankweiler, Strange, & Verbrugge, 1977).

While speech perception and comprehension are understandably the primary objectives in cochlear implant design, as this goal is met, it may become useful to consider how well current speech processing strategies are able to code other types of acoustic information in the speech signal, such as can be used to differentiate between talkers. Although cochlear implants are able to support perception of linguistic contrasts partly because phonemic/lexical information is relatively robust to spectral shifts and lack of fine spectral/harmonic detail (see Assmann & Summerfield, in press), it is not clear if perception of talker differences should necessarily be equally robust under these same types of degraded listening conditions.

The hearing-impaired children who participated in this study had all used a multi-channel cochlear implant with a state-of-the-art speech processing strategy for at least two years. Nucleus-22, Nucleus-24, and Clarion users were represented in the sample. The Nucleus devices manufactured by Cochlear

Corporation used either the SPEAK or ACE processing strategies. The Clarion devices made by Advanced Bionics used the Clarion-CIS processing strategy.

All three of these cochlear implant speech processing strategies filter the incoming acoustic signal into frequency bands, also referred to as "channels," that together span the frequency range of most speech information. Depending on the speech processing strategy, all or a subset of the information contained in these channels is transmitted to the intra-cochlear electrodes. The Clarion CIS strategy continuously transmits the information in all 8 of its frequency bandwidths, whereas Nucleus's SPEAK locates the most active of its 20 available channels from moment to moment, and only transmits the signals contained in this subset of channels. Although as many as 10 channels can be included in this active subset, SPEAK stimulates about 6 channels on average (Skinner, Arndt, & Staller, 2002; Wilson, 2000). The Advanced Combination Encoder (ACE) strategy combines aspects of both SPEAK and CIS. Like SPEAK, at each timeframe ACE selects the most active of its 22 available channels for transmission, but more like CIS, for a particular user, the number of channels selected on each cycle is fixed at a number between 6 and 12 (Skinner, Arndt, & Staller, 2002; Skinner, Holden, Whitford, Plant, Psarros, & Holden, 2002).

In all three processing strategies, the amplitude envelope of the waveform signal within each individual channel is obtained using a low-pass-filter. This signal is then used to modulate the amplitude of brief biphasic pulses delivered to the electrode associated with that channel. The processing strategies vary in the rate at which these pulses are delivered. In SPEAK, the pulse rate tends to range between 180-300 pulses per second per channel, depending on the number of spectral maxima identified and parameters of the individual's mapping. The pulse rate tends to be higher in CIS, generally exceeding 800 pulses per second per channel. ACE uses a pulse rate per channel that is largely determined by the number of maxima selected for a particular patient's map, and a device limit of 14,400 pulses per second per cycle summed across all electrodes. This rate can range between 250-2400 pulses, with rates exceeding 700 pulses per second per channel being typical.

Unlike early cochlear implant speech processing strategies which tried to provide an independent temporal cue to fundamental frequency via stimulation rate, none of the current strategies use stimulation rate to code f0 but instead leave fundamental frequency to be decoded by the listener from the envelope variations conveyed by the fluctuating pulse amplitudes (Jones, McDermott, Seligman, & Millar, 1995; Seligman & McDermott, 1995; Vandali, Whitford, Plant, & Clark, 2000). The higher the pulse rate per channel, the more fine-grained the temporal amplitude structure preserved in this envelope.

Relevant to the representation of fundamental frequency are the lower frequency bound of the bandpass filter corresponding to the most apical electrode, and the cutoff frequency of the amplitude envelope detectors in the individual channels. The lower frequency bound of the most apical electrode is usually described as either 116 Hz or 150 Hz for SPEAK, 250 Hz for the Clarion-CIS strategy, and 187 Hz for ACE. These values are low enough to convey the low F1's expected for vowels such as [i] and [u] spoken by male talkers and the low frequency nasal resonances.

In a cochlear implant, the placement of each electrode along the length of the cochlea codes for frequency while the amplitude envelope of the filtered signal within a channel governs the magnitude of the electrical pulse delivered to the auditory nerve. If the fundamental frequency is higher than the lower frequency bound of the most apical electrode, this f0 will be roughly coded by placement of the associated electrode. Another cue to f0 however can potentially be preserved in the amplitude fluctuations of the pulses in individual channels. This will be the case if the f0 remains below the highest frequency passed by the low-pass filter of the envelope extractors and if the pulse rate is at least double this f0 value.

Typical values for the low-pass-filter upper frequency bound are 200 Hz, for both SPEAK and Clarion-CIS. For ACE, the cutoff frequency is described as 180 Hz, but with a filter roll-off that results in the inclusion of frequencies up to 250 Hz (Vandali, Whitford, Plant, & Clarke, 2000). Thus, it makes sense that SPEAK with its relatively low pulse rates tends also to have a lower overall frequency bound for its most apical filter. Taken together these details imply that almost all f0 information is coded by Clarion-CIS primarily by amplitude envelope, whereas in SPEAK, high f0's are conveyed primarily by electrode placement and low f0's by a combination of both electrode placement and amplitude envelope. In ACE, low f0s below 187 Hz must be coded via amplitude variation, and f0s above this value through a combination of electrode placement and amplitude variation. The typical filter cutoff value for ACE suggests, however, that the strength of f0s above 180 Hz will be attenuated by the low-pass filter.

Finally, it should be noted that although all of these strategies possess sufficient spectral resolution at the higher frequencies to convey formant information, it is however doubtful, how well individual harmonics are preserved. Thus, although the original pitch of a speaker's voice can be perceived by normal-hearing persons even if energy at the fundamental frequency is removed by filtering, presumably from an analysis of the frequency distance between the surviving harmonics (Moore, 1997), such cues to f0 based on distances between harmonics are probably not perceptible to users of cochlear implants.

The description provided above is a simplification that includes only the most basic stages of the signal processing involved in a cochlear implant. Moreover, the parameters of each device can be tailored considerably to the individual patient. Nevertheless, it is hoped that from this brief description, a general sense can be conveyed of the degree to which formant and fundamental frequencies are degraded and/or preserved by the current cochlear implant signal processing strategies

Despite the interesting implications that these speech processing strategies have for cochlear implant users' perception of acoustic differences between talkers, this area has not been very rigorously studied. One reason for this may be the difficulty of controlling the acoustic properties manifested by actual talkers. One technique to try to circumvent this difficulty is to use speech synthesis and/or resynthesis methods to create artificial voices for which the physical differences among the voices are determined by the experimenter. The ability of cochlear implant users to perform a talker discrimination task in which the similarity of the to-be-compared voices was systematically varied in this way has not, to our knowledge, ever been studied before in any previous research.

In the present study, systematic manipulation of voice similarity was carried out by using a recently developed speech resynthesis technique called STRAIGHT (Kawahara, Matsuda-Katsue, & de Cheveigne, 1999). This technique is known for its high quality resynthesis routines that permit independent manipulation of fundamental and formant frequencies (Kawahara et al., 1999; Liu & Kewley-Port, 2001). As described in further detail in the Method section below, a stimulus continuum of similar-sounding "talkers" was constructed using a natural female voice. This voice was resynthesized using a range of incrementally different scaling factors, such that at the smallest scaling factors, the voice became quite male-like in quality.

For the present study, we decided to manipulate both fundamental and formant frequencies in tandem in order to create the continuum of differentially similar-sounding voices. This decision was based on informal listening experiments showing that large changes in only fundamental frequency, or only formant frequencies, yielded relatively unnatural-sounding voices that were, especially in the fundamental frequency-alone case, not terribly convincing as "different talker" exemplars (see Reference Note 1 for some recent unpublished data relevant to this issue, also Lavner, Gath, & Rosenhouse, 2000). This is not entirely unexpected given that f0 and formant frequencies tend naturally to co-vary amongst talkers, due primarily to their shared association with talker gender (Fant, 1960; Bachorowski & Owren, 1999). To try

to simplify the task still further for the hearing-impaired listeners, given the poor performance in discriminating within-gender differences reported by Cleary (2003/Chapter II), it was decided to have the stimulus continuum cross the female-male gender boundary. More details regarding the scaling factors used are provided in the Method section below.

Because of time and attention constraints in testing young children, an efficient method of data collection was needed to measure their perceptual discrimination of similar-sounding talkers. A novel variant on traditional adaptive staircase procedures was therefore designed, so that the size of the pitch difference between paired stimuli would be automatically adjusted according to the response of the listener.[3] Sentence-length stimuli were selected over word-length stimuli both for reasons of greater real-life validity and to increase the ease of the discrimination task. Both a "fixed" sentence condition and a "varied" sentence condition were included in the experimental design, in order to evaluate the ability of each listener group to ignore linguistic variability that was irrelevant to the immediate task at hand.

Although normal-hearing listeners will readily identify two utterances that differ slightly in their pitch and timbre as both the productions of an individual talker, several findings suggest that increasing the fundamental frequency and/or spectral envelope differences beyond a particular size leads to the perception of hearing two different talkers (Kuwabara & Takagi, 1991; Lavner, Gath, & Rosenhouse, 2000). The purpose of the present study was to assess where these boundaries lie for normal-hearing children and for hearing-impaired children with cochlear implants. We expected that the normal-hearing children as a group would begin to favor "different" talker responses at a smaller pitch difference size than the hearing-impaired group, because of the reduced spectral resolution of the cochlear implant users.

Based on some informal pilot testing with normal-hearing children, the pitch difference required for "different talker responses" in the group of normal-hearing children was predicted to be on the order of 2-3 semitones. Pilot testing also suggested that normal-hearing listeners almost never responded "different talker" when the two utterances differed only by a shift of a half semitone (see also Kuwabara & Takagi, 1991; Lavner, Gath, & Rosenhouse, 2000). Additionally, the earlier results of Cleary (2003/Chapter IIB) indicated that hearing-impaired children with cochlear implants were able to discriminate a 4-semitone difference with some modest degree of success. Hence, a stimulus continuum was chosen that progressed from a minimum difference of zero semitones to a maximum difference of 6 semitones, in half-semitone steps.

We planned to quantify performance on the adaptive procedure used in the talker discrimination task by examining the proportion of "different talker" responses obtained for each pitch difference size along the stimulus continuum as well as by using an endpoint-based measure of the pitch difference size reached at the end of the procedure. Because large individual differences were expected among the children who use cochlear implants (Pisoni, Cleary, Geers, & Tobey, 2000), both grouped data and individual subject data were of interest to us. Although the procedure used to collect these data was not a simple discrimination task, considerable experimental precedent exists for using psychometric methods to collect judgments about specific qualities of complex stimuli (Kewley-Port, 2001; Kewley-Port & Watson, 1994; Stevens, 1975).

An auditory-only keyword identification task requiring a picture-pointing response was also developed using the same sentence-length stimulus materials that were used in the discrimination task. The design of this word identification task permitted us to examine for each group of children whether the stimuli that had undergone the most severe degree of resynthesis manipulation were less intelligible than

[3] In this chapter we frequently use the term "pitch difference" to refer to the acoustic differences defined by the stimulus continuum. We choose to do this even though in addition to manipulating fundamental frequency (typically thought of as the physical correlate of pitch), we have also manipulated the spectral envelope of each utterance.

the utterances that had undergone minimal manipulation. More generally, this methodology permitted collection of an independent measure of word identification that could then be compared against performance on the talker discrimination task.

In summary, the purpose of the present study was to investigate how normal-hearing children and prelingually-deafened children who have begun to acquire spoken language using a multi-channel cochlear implant perceive talker voice differences. More specifically, we sought to determine how acoustically different, in terms of their average spectral characteristics, two sentences needed to be for our child listeners to perceive these utterances as spoken by two different talkers. In addition, we planned to examine how performance on the talker voice discrimination task might be related to other speech perception and language measures collected from the children with cochlear implants.

Method

Participants

Thirty five-year-old normal-hearing children were recruited from the local Bloomington, Indiana community to participate in the present study. A database of published birth notices maintained by the psychology department was used to locate children of eligible age. Data from two children were eliminated due to problems with the testing room/equipment, and one due to experimenter error. Data from three additional children were eliminated, two for a child not meeting the hearing screening requirements and one for a child refusing to fully cooperate during testing. Thus, data from 24 children were retained for the final analysis.

The mean age of the normal-hearing group of 24 children was 65.5 months (*SD* = 2.2 months). All children in this group passed a simple pure-tone hearing screening administered at 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz. Left and right ears were individually screened using a Maico Instruments portable audiometer (model MA270A, TDH39 receivers). For frequencies of 1000 Hz and above, a response at 20 db HL was required. For frequencies below 1000 Hz, either a response at 20 or 25 db HL was accepted due to the characteristics of the ambient noise in the testing room. Sixteen male and eight female children were included in the group. The children received ten dollars and a Speech Research Laboratory t-shirt for their participation. All testing of normal-hearing children took place in a quiet testing room at the Speech Research Laboratory in Bloomington, Indiana.

Twenty-one hearing-impaired children with cochlear implants were also recruited for the present study. Recruitment took place through the DeVault Otologic Research Lab in the Department of Otolaryngology-Head and Neck Surgery at the Indiana University School of Medicine in Indianapolis. The twenty-one families who agreed to participate were drawn from a larger subset of forty-one families whose children were already enrolled in a large-scale longitudinal study of speech and language development in prelingually-deaf children with cochlear implants (Miyamoto, Kirk, Svirsky, & Sehgal, 1999; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000), and who met the eligibility requirements of the present study.

To be eligible for the present study, a hearing-impaired child with a cochlear implant had to have experienced his/her onset of deafness prior to age three years, have been implanted at or before age six years, be between five and twelve years of age at time of testing, and have used a cochlear implant for at least two years. Additionally, the child's implant was required to use one of the newer speech processing strategies (i.e., either SPEAK, CIS, or ACE). The child was also required to be a user of spoken and/or signed English as a first language. Demographic and device characteristics for each child are shown in Table 1.

TABLE 1

PARTICIPANT AND DEVICE CHARACTERISTICS, COCHLEAR IMPLANT GROUP

ID#	Age at Test (Y;M)	Sex	Age at Onset of Deafness (Y;M)	Age at CI Fit (Y;M)	CI Use (Y;M)	CI Device (N="Nucleus")	CI Processing Strategy	Commu-nication Mode	Etiology
CI01	8;0	M	0	2;9	5;3	Nucleus22	SPEAK	TC	Unknown
*CI02	4;6	F	0	1;8	2;10	Nucleus24	ACE	Oral	Unknown
CI03	5;10	M	0	1;9	4;1	Nucleus24	SPEAK	Oral	Unknown
CI04	12;7	M	0;10	5;1	7;6	Nucleus22	SPEAK	Oral	Meningitis
CI05	9;7	F	0	2;6	7;1	Nucleus22	SPEAK	TC	Unknown
CI06	8;3	F	0	2;9	5;6	Nucleus22	SPEAK	Oral	Unknown
CI07	10;1	M	0	5;3	4;10	Nucleus24	ACE	Oral**	Unknown
*CI08	5;7	M	0	3;2	2;5	Nucleus24	ACE	Oral	Unknown
CI09	8;5	M	0	1;5	7;0	Nucleus22	SPEAK	Oral	Unknown
CI10	11;6	F	3;0	6;1	5;5	N24#2/N22	ACE	Oral	Unknown
CI11	9;11	F	0	3;4	6;7	Nucleus22	SPEAK	Oral***	Genetic
CI12	8;8	F	2;1	2;10	5;10	Nucleus22	SPEAK	Oral	Meningitis
CI13	7;0	M	0	2;3	4;9	Clarion	CIS	Oral	Unknown
CI14	6;1	M	0	2;1	4;0	Clarion	CIS	Oral	Unknown
CI15	6;10	M	0	2;0	4;10	Clarion	CIS	Oral	Unknown
CI16	6;0	M	0	3;3	2;9	Clarion	CIS	Oral	Unknown
CI17	5;9	M	0	2;6	3;3	Clarion	CIS	Oral	Genetic
CI18	7;1	F	0	2;4	4;9	Clarion	CIS	Oral	CMV
CI19	5;6	M	1;6	2;1	3;5	N24#2/N24	ACE	Oral	Meningitis
CI20	7;10	F	0	3;2	4;8	Clarion	CIS	Oral	Unknown
*CI21	5;7	M	0	3;6	2;1	Nucleus24	ACE	Oral	Unknown
MEAN	8;1			3;0	5;1				
MIN	5;6			1;5	2;9				
MAX	12;7			6;1	7;6				

*Data are not included in final data analysis or group means; child was unable to cooperate during testing.
**Child is exposed to some use of sign at school.
***Child is exposed to some use of sign at home.
A "#2" indicates that this is the second device used by the child. The device used by the child prior to re-implantation is listed after the "/".
"CMV" = Cytomegalovirus

Data from three hearing-impaired children were not used in the final data analysis. One child who had not reached her 5th birthday and struggled with all tasks was not included. Two other children, both males, 67 months of age, could not be convinced to complete any of the tasks. The remaining 18 children ranged in age from 66 to 151 months, with a mean age of 96.6 months (= 8;1) (*SD* = 24.6 months). Fourteen of the children were identified as congenitally deaf. The remaining four children lost their hearing at or before age 3 years. The mean age at cochlear implantation was 36 months (median = 32 months) and the average duration of implant use at time of testing was 5 years, 1 month (median, 4 years, 10 months). The final group of hearing-impaired pediatric cochlear implant users included eleven male and seven female children.

Although we initially planned to attempt to recruit roughly equal numbers of children who used oral/aural language alone and children who used Total Communication (TC), it proved difficult to recruit strongly TC children to participate in this study. Of the 18 children whose data are reported here, only two children used some amount of sign together with speech, as determined through medical records and parental report. A speech-language pathologist trained in total communication assisted in the administration of the tasks to any child designated in his/her medical charts as using total communication. Neither of the two "TC" children tested was observed to use any sign in their spontaneous speech productions, however. One additional child, though designated as an oral communicator, did have some experience using manual

language at home with family members. Although these numbers are not sufficient to make any formal comparisons, the issue of communication mode and the role of early experience will be discussed further in presentation of the results.

The hearing-impaired children received twenty dollars and a laboratory t-shirt for their participation. Although this is a larger amount than that received by the normal-hearing children, this was the recommended payment as established by previous studies done in collaboration with researchers at the DeVault Otologic Laboratory. Parents of children who made a special trip to Riley Hospital, specifically to participate in the present study were also reimbursed for mileage traveled between their home and the Indianapolis laboratory. Testing took place in a quiet testing room either at DeVault Otologic Research Lab at Riley Hospital for Children in Indianapolis, Indiana, or at the Audiology Department of the St. Joseph Institute for the Deaf, an oral school for hearing-impaired children located in St. Louis, Missouri. Additionally, for the convenience of the family, one child was tested at her speech-language pathologist's office in Indianapolis.

Stimulus Selection and Recording

Because no existing corpus of recorded stimulus materials was fully suitable for the present study, a new database of sentence-length speech stimuli was created specifically for this purpose. A single sentence frame was chosen: "The _____ and the _____ are by the _____." This particular sentence frame was selected in order to minimize morphological and grammatical complexity, because severely hearing-impaired children may have difficulty perceiving and learning to process such markers (Robbins, Svirsky, & Miyamoto, 2000; Svirsky, Stallings, Lento, Ying, & Leonard, 2002). Given our choice of keywords, described in more detail below, and each word's random assignment to one of the three locations within each sentence, the sentence frame used resulted in very low predictability for all key words. The use of a single sentence frame with only the keywords varying between sentences also simplified the scoring of each sentence for correct keyword identification during the accompanying auditory comprehension task. The particular sentence structure chosen additionally lends itself well to having all three keyword items interpreted as having equal emphasis/importance in the utterance.

Since the ability to successfully compare the voice characteristics of sample utterances depends considerably on the amount of information provided to the listener (Pollack, Pickett, & Sumby, 1954), it was important that all the sentences be comparable in number of syllables. Thus, all keywords were chosen to be monosyllabic, thereby making all sentences nine syllables in length. Monosyllabic keywords were selected for several reasons. Firstly, a large pool of words was needed, all of which would be highly familiar to preschool age children: the longer the word, the fewer instances are available from the lexicons of young children (Charles-Luce & Luce, 1990). Monosyllabic words are also of special interest to researchers investigating spoken word recognition because they have been shown to have more similar-sounding phonological "lexical neighbors" than do longer words (Luce & Pisoni, 1998). As a consequence, identification of monosyllabic words entails that each phoneme be identified for these items to be discriminated from other similar-sounding words (Luce, 1986). The choice of monosyllables also allows for greater uniformity across the intonation patterns of the sentence set, in that all monosyllabic words follow a single stress pattern. Although the lexical neighborhood characteristics of monosyllabic words present more of a challenge to hearing-impaired children than would equally-familiar longer words (Kirk, Pisoni, & Osberger, 1995), in other ways the task was simplified as much as possible for the hearing-impaired children. All keywords were mono-morphemic and highly picture-able concrete nouns.

In order to have a minimum of 24 trials in the varied sentence condition, 48 different sentences each containing three unique keywords were necessary. Thus, at least 144 keywords were required, plus additional words for use in practice trials. Several hundred candidate monosyllabic nouns were culled from

children's dictionaries of "first words" and from paging through CHILDES transcripts of spontaneous speech produced by normal-hearing preschool-age children (MacWhinney, 2000).

Only words which appeared in the CHILDES transcriptions were retained in the final set of 48 test sentences. Frequency counts for the selected keywords are provided in Appendix A. Details regarding the subset of CHILDES transcripts used and the commands used to tally the counts are also provided in Appendix A. The final set of words ranged quite widely in frequency. As there were more inanimate than animate nouns in the word set, for the sake of uniformity, the final slot in each sentence was filled with a randomly selected inanimate noun. The animate nouns were randomly paired with one another, as were the remaining inanimate nouns, and then assigned randomly to fill the first two slots of each sentence. No two consecutive keywords were permitted to begin with the same phoneme. Keywords that were closely semantically related (e.g., king vs. queen) were also not allowed to appear in the same utterance. The final sentence set is provided in Appendix B.

Greyscale illustrations of each keyword were obtained via the large commercial ArtToday online image database service. The suitability of each image was pilot-tested on normal-hearing adults. The data contained in the present report will serve to establish their appropriateness for normal-hearing preschool-age children. Each image was sized to fit a window 360 pixels wide x 216 pixels high, converted to portable network graphic (.png) image file format, and saved for future use on compact disc. The auditory test stimuli used in this study were spoken by one of twenty female talkers recruited to record the sentence set. Each participant was paid ten dollars for her participation in the one-hour recording session. Talkers were asked to listen to two model utterances previously recorded by the author, demonstrating the desired speaking rate and intonation pattern.

Recording took place inside a sound-attenuated, single walled anechoic recording chamber (Industrial Acoustics Company Audiometric Testing Room, Model 402) using a head-mounted close-talking microphone (Shure, Model SM98). The recordings were digitally sampled online using a 22.05 kHz sampling rate with 16-bit amplitude resolution using a Tucker Davis Technologies (TDT) System II with an A-to-D converter (Stereo-Analog Interface, DD1) and a low-pass filter of 10.4 kHz (anti-aliasing filter FT5) controlled by a Matlab version of the "Speech Acquisition Program (SAP)" (Hernandez, 1995) which is an updated version of the Speech Acquisition Program described in Dedina (1987).

Three recorded repetitions of each sentence were obtained from each talker, collected over the course of three runs through the entire set of sentences. The resulting digital sound files were then individually edited using the CoolEdit2000 waveform editor (Syntrillium Software Corp.) in order to remove silence before and after each utterance. For each talker, the "best" two of the three elicitations were retained in the database for each sentence. "Best" was subjectively defined as having the uttered sentence contain the same words as the intended utterance, produced without dysfluencies such as pauses or repeated syllables. Retained utterances also needed to follow the intonation pattern used throughout the recording session. All retained utterances were saved onto compact disc for further use.

One talker was selected as the primary talker to be used in the present study. This talker is referred to in the compact disc documentation as T-13. T-13 was a 22-year-old right-handed White Caucasian female native to the Bloomington, Indiana area. T-13 reported no history of hearing or speech disorders at the time of testing and was fluent in no other language besides English. This talker spoke clearly and followed our directions very accurately regarding the use of a regular intonation pattern and speech rate. As will be further explained, T-13 also had an average fundamental frequency that lent itself to the planned resynthesis methodology.

For this study, only one token of the two available tokens of each sentence was used. Before any further stimulus preparation took place, the selected sound files were played to six normal-hearing adult listeners via loudspeaker at the same level as would be used during testing (about 70 dB SPL). Listeners were given a numbered response sheet consisting of 60 instances of "The _____ and the _____ are by the _____." Listeners were asked to write in the words they heard for each sentence. No word list was provided from which to choose. Misidentification of the keywords was rare: "bowl" was misheard as "bull" by two listeners, "bath" was heard as "back" by one listener, "pin" for pen by one listener, and "key" for tea by one listener.

Scaling of the Stimulus Continuum

The goal of this procedure was to create a naturalistic-sounding voice continuum that would span acoustic differences large enough to readily elicit "different talker" judgments but also be sufficiently graduated in step size so that normal-hearing listeners could not easily discriminate immediate adjacent stimuli on the continuum. Several alternative manipulation schemes were considered. One decision involved whether to manipulate f0 only, formant frequencies only, or both dimensions simultaneously. Because large changes in one dimension in the absence of change in the other dimension tended to yield very unnatural-sounding tokens (see also Lavner, Gath, & Rosenhouse, 2000), we decided to manipulate both fundamental frequency and formant frequencies along the single stimulus continuum. Independent variation of the two dimensions was explored further in a separate study with normal-hearing adults using these same procedures (Reference Note 1).

As already mentioned, the previous work suggested that some of the hearing-impaired children we were planning to test might find even a 4-semitone difference difficult to discriminate. Thus, the largest difference to be tested was set at 6 semitones. Pilot work had also established that for these complex signals, normal-hearing (untrained) adults could not always distinguish a half-semitone difference even when the task was simple average pitch discrimination, without mention of talker differences. Therefore, we constructed a thirteen-step stimulus continuum that varied in half-semitone increments with the naturally produced pitch characteristics of T-13 used as the "high" pitch end of the continuum. Since T-13's average f0 averaged across all 48 test sentences and 8 practice sentences was approximately 175 Hz (*SD* = 7 Hz), tokens at the "low" pitch of the continuum averaged 123.7 Hz in fundamental frequency—a typical f0 value for a low-pitched male voice. The ratio of 123.7 to 175 corresponds to a scaling factor of .71. Although T-13's average fundamental frequency was relatively low compared to the average female voice, her vowel formant frequencies (as based on vowel midpoint measurements of a subset of the keywords) were found to be fairly typical of the other female talkers from this region of the Midwest who participated in the initial recording session. In light of this fact and after experimenting with various scaling factors, we decided to use .71 also as the scaling factor for the formant frequencies. Although this is a somewhat more extreme conversion than would be needed to transform T-13's formant frequencies into average male formant frequencies (i.e., .85), a scaling factor of .71 yields formant frequencies that are still within the possible range for a male talker.

Use of a common scaling factor meant all frequencies present in the speech stimulus would be shifted by an equal percentage--in this case, along a stimulus continuum defined in half-semitone steps. Semitones are a commonly used unit in studies of intonation and f0 variation and are defined by a ratio relation of $2^{1/12}$, or approximately 1.0595, between a higher frequency to a lower frequency (Dowling & Harwood, 1986). A half-semitone difference therefore corresponds to a ratio relation of $2^{1/24}$, approximately 1.0293. For talker T-13, a shift of 3 semitones along both dimensions yielded stimuli that impressionistically sounded like a young male speaker. A shift of 6 semitones yielded stimuli that impressionistically sounded like a deep voiced adult male. The mappings between the original frequency values and those resulting from the most extreme scaling factor of .71 are shown in Figure 1.

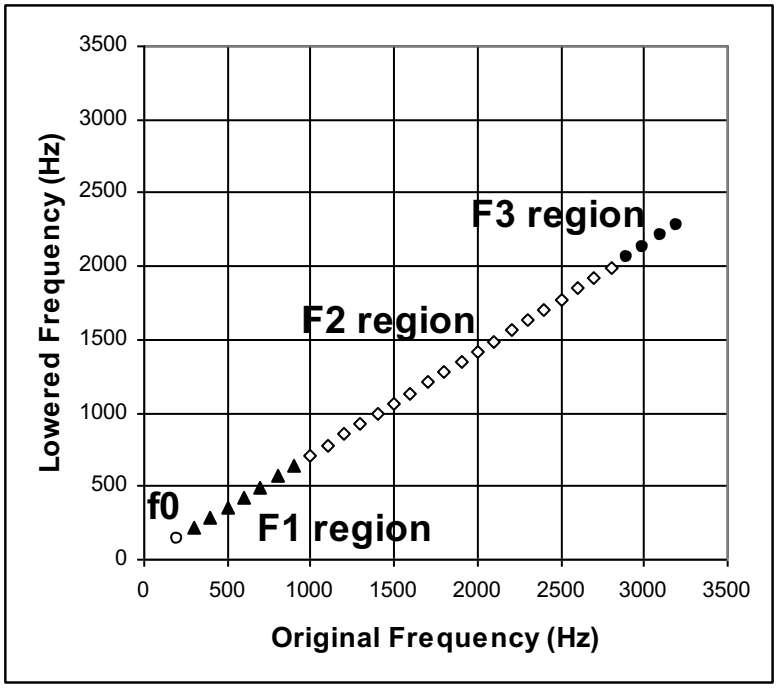


Figure 1. Mapping between the original frequency values and those resulting from the most extreme scaling factor of .71. These values correspond to the two endpoints of the stimulus continuum. The general regions of the first three formant frequencies and the mean fundamental frequencies are indicated by individual marker type.

T-13's average f0 did, as in any natural speaker, vary somewhat across the 48 test sentences. One potential problem, despite T-13's remarkable regularity of intonation across tokens, was that equivalently-sized pitch shifts in different sentences might not result in new utterances that sounded at all as if they were spoken by the same newly synthesized "talker". Auditory examination of the resulting stimulus sets suggested however, that this would not be a problem, although the real test of this was left be seen in the results of the fixed sentence versus varied sentence experimental conditions.

Choice of Resynthesis Technique

Although the choice of scaling factors was an important design decision, the success of the experimental manipulation was largely dependent on the quality of the speech resynthesis method. A number of different methods for modifying speech recorded from natural talkers have been developed over the last two or three decades (e.g., Kawahara et al., 1999; Moulines & Charpentier, 1990; see reviews in Lemmetty, 1999; Schroeder, 1993). The success of these methods is usually judged in terms of how well aspects of the speech signal such as fundamental frequency, formant frequencies, and speech rate can be varied while the linguistic phonemic content of the message remains constant and intelligible.

Speech resynthesis techniques often run into problems when large parameter changes are involved. Essentially, researchers have found that large changes made along one speech dimension usually cause the speech quality to deteriorate along another dimension. The STRAIGHT resynthesis method described by Kawahara et al. (1999) attempts to avoid the worst of these problems by representing the speech signal in a nontraditional fashion. The primary problem, according to Kawahara, is the difficulty of creating a

representation of the vocal tract that is free of the influence of the glottal source. The success of the STRAIGHT method depends on having two relatively "pure" representations--one of the glottal source and one of the supra-laryngeal vocal tract, each free of "trace elements" of the other. Familiar representations of spectral information such as those visualized in wide-band or narrow-band spectrograms inherently lack this purity because in order to create these representations, averaging over the time or frequency domains takes place. Fundamental frequency is therefore typically recoverable either from the amplitude fluctuations corresponding to the individual glottal pulses, or from the relative spacing of individual harmonics. Spectral representations such as LPC (Linear Predictive Coding) according to Kawahara et al. (1999) also carry a remnant of f0 in their patterns of variability and error.

The Time x Frequency x Amplitude representation utilized by STRAIGHT therefore differs in several crucial ways from the other speech representations mentioned above. In order to represent these dimensions free of the influences of fundamental frequency, a good representation of fundamental frequency is required, one which does not simply use f0 measurements at the resolution of the f0 frequency, but rather uses a smooth, interpolated continuous trajectory representing f0 change over time. Armed with this representation of f0, FFT spectra are then calculated using not one preset set of parameters for all time points as in standard spectrographic representations, but rather individualized FFT parameters based in part on the "instantaneous" values for f0 associated with the particular time points/intervals. Thus, an "ideal" "middle-band" "spectrogram" can be built which has "equivalent relative resolution in both the time and frequency domains" (1999, p.190) instead of one which trades off one type of information for accuracy in the other dimension, as is usually the case. Then, using a sophisticated smoothing procedure, the influence of the glottal pulsations can be effectively cancelled out of the spectral representation to arrive at a "fundamental-frequency-free" time x frequency x amplitude representation. Kawahara describes the process of arriving at this final smoothed spectral representation as mathematically a "surface reconstruction problem." This final representation can now be recombined with f0 trajectories that differ radically from the original with quite good results. Alternatively, the spectral information may be stretched in the frequency or time domains and then recombined back with the original f0 trajectory information, depending on the desired outcome.

A version of STRAIGHT last modified in November of 1999 (STRAIGHTV30kr16) was used for this project. The Matlab routines that implement this method were kindly provided by Dr. Kawahara upon my request. The synthesis path followed in the present procedure was the following: (1) Analyze 1 CHX, (2) Bypass, (3) Synthesize. "Analyze 1CHX" uses a binary voiced/unvoiced representation to decide if particular time windows should be used in determining the continuous f0 trajectory. Use of the "Bypass" option results in using the "smoothed and optimally recovered" result as is, skipping the option of applying a not yet fully documented method to further remove some additional spectral interference artifacts.

The parameters used to resynthesize the test sentences using STRAIGHT were as follows (see Kawahara, 1998; 1999, for further explication): The f0 extraction algorithm was provided an f0 lower bound and upper bound of 100 Hz and 375 Hz respectively, selected specifically for talker T-13. After some experimentation, the f0 extraction settings were left at their default values for a sample rate of 22.05 kHz : FFT length = 1024 points, "w stretch in time" eta the time-stretching factor for the complex Gabor function based analyzing wavelet that permits slightly finer resolution in frequency than in time to calculate "fundamentalness" = 1.4, power constant alpha = 0.6, magnification factor = 0.2, and a frame rate of 1 ms. The manipulation and resynthesis parameters were experimented with, but largely left at their default values (aside, of course, from the scaling factors.) Absolute tg dispersion was selected over relative dispersion, and left at its default setting of 2 ms. The corner frequency value was left at 3000 Hz, and the tg smoothness parameter was left at 70 Hz.

Each of the test and practice sentences was resynthesized along the previously described 13-step continuum. Although the resynthesis process was largely automated, every resulting sound file was individually checked for usability. The average RMS amplitude levels of the resulting 728 sounds files were then measured (both with and without taking values approaching zero into account, given that these were sentences), and the full set of resynthesized stimuli re-leveled together at 67 dB RMS.

Comparable examples of sentence-length speech resynthesis with pitch manipulation can be found in Bird and Darwin (1997) and Assmann (1999). In order to confirm the results of the resynthesis manipulation, fundamental and formant frequencies of the final stimulus set were checked using the PRAAT speech analysis package (Boersma & Weenink, 2001). The formant frequency shifts for the sentences at the two ends and mid-point of the stimulus continuum were checked visually using wide-band spectrograms. The mean fundamental frequency was calculated for each token from these three locations along the stimulus continuum and the average of these measurements was then compared across the three locations to verify the shift in f0.

General Testing Procedure

For the normal-hearing group of children, the order of the tasks was as follows: a hearing screening, the talker discrimination task, the keyword identification task, a competing speech task not described in the present paper (see Reference Note 2), and finally, the Peabody Picture Vocabulary Test Third Edition Form A (Dunn & Dunn, 1997) as a general language screening measure. The testing session lasted approximately 45 minutes.

For the hearing-impaired children with cochlear implants, the following task order was used: talker discrimination task, keyword identification task, and optionally, the competing speech task described in Reference Note 2. For three of the youngest hearing-impaired children tested, the order of the talker discrimination and keyword identification tasks had to be reversed because the keyword identification task was more engaging and conceptually simpler, and therefore served as a better opening task for children who were shy or unsure about participating. Depending on the individual child, the testing session lasted between 30 minutes to an hour.

The Talker Discrimination Procedure. The same task instructions were given to both the normal-hearing and hearing-impaired groups. Prior to starting, each child's understanding of the concepts of "same" and "different" was tested using two cards printed with the same symbols as would be later used to label the "same talker" and "different talker" response buttons. The child was told that on one of the cards, there would be two things that were the SAME, and on one of the cards there would be two things that were DIFFERENT. The child was allowed to see the cards, the cards were shuffled, and then the child was asked to indicate, which card had two things that were the same, and then, which card had two things that were different. All the children whose data are included in this report were able to perform this task perfectly with no errors. An illustration of the button labels can be seen in Figure 2.

For the talker discrimination task, the child was told that pairs of sentences would be played through the loudspeaker, and that his/her job was to decide if it sounded like the same person saying both sentences, or if it sounded like two different people. The child was specifically told to listen to the voices of the people talking. As an introduction to the task, the experimenter and the child discussed the fact that the child was able to tell the experimenter's voice from the voice of the child's mom/dad. The precise wording of the directions can be found in Appendix C.

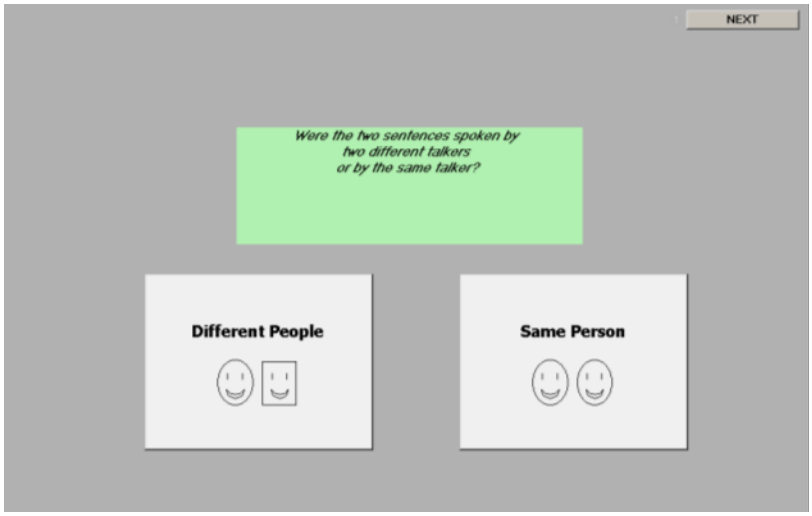


Figure 2. Talker Discrimination Task Same-Different Response Touch Screen Image

The testing began with four practice trials. During the first two practice trials, the experimenter told the child the correct response and demonstrated how two large labeled squares shown on a touch-screen-equipped monitor should be used as response buttons (see Figure 2). The child was encouraged to respond on his/her own to the two remaining practice trials. The correct answers on the practice trials alternated between "same" and "different." The pitch difference size was either zero semitones or 6 semitones. Children who responded incorrectly on the final two practice trials were run through the same identical set of practice trials up to two additional times. For all children the 24 test trials in the condition were then administered.

The experimenter provided general verbal encouragement after each response on the first four test trials, but apart from this, no feedback regarding performance was provided. In the case of the younger children, it was necessary to give some additional general non-contingent verbal encouragement to complete the task. The children made their responses by pressing on one of the two large rectangular "push buttons" displayed on the touch-screen on each trial. The buttons did not appear on the screen until the second utterance in each pair began to play, and disappeared as soon as the child made his/her response. A 500 ms silent interval separated the offset of the first sentence and the onset of the second comparison sentence on all trials. Individual sentences ranged in duration from 1.88 to 2.38 seconds. The children were given as much time as they needed to make each response (within the limits of the testing session) and were asked to "make their best guess" if they were unsure as to what they heard.

After the 24 test trials were completed in a given condition (fixed sentence or varied sentence), the child was told that for the next task, he/she would be doing the same thing, but was alerted to the fact that he/she might notice that the sentences this time would either "change all the time" or "be the same over and over again." However, it was emphasized that this change was not important, and could be ignored, and that the child's task was exactly the same as before. The basic instructions were then reviewed in somewhat shortened form. Four practice trials and 24 test trials were then administered in the same manner as before.

The testing procedure developed specifically for the present talker discrimination task is a novel variant on traditional adaptive staircase procedures and therefore needs to be described in some detail. At a

general level of description, the method used followed a classic one-up-one-down adaptive staircase procedure (Levitt, 1971). Responses of "same talker" resulted in the next trial having a pairing that was further apart on the stimulus continuum. Response of "different talker" resulted in the next pairing of utterances being closer together on the stimulus continuum. An adaptive procedure was employed in the present study in order to collect each data set quickly and efficiently before a child lost interest in the task or became distracted.

For a given condition, fixed sentence or varied sentence, each subject completed 24 trials, as previously mentioned. Each set of 24 trials was comprised of two interleaved "tracks" of 12 trials apiece (see Levitt, 1971, for a discussion of the advantages of interleaving tracks). One set of trials (a track) started with two utterances both taken from one end of the stimulus pitch continuum, and will be referred to henceforth as the "start-same" track. The other set of trials started with two utterances taken from opposite ends of the stimulus continuum and will be referred to as the "start-different" track. The type of track administered on a given trial ("start-same" or "start-different") was determined pseudorandomly using a random number generator.

The purpose of interleaving these two types of trials was to design a testing procedure in which the "ease" of the task was maximized at the beginning of the procedure and a roughly equal mixture of "same talker" and "different talker" responses would be expected throughout the course of the procedure from a normal-hearing person engaged in the task. That is, if the tracks were not interleaved--for example, if only start-same track was tested--the subject's behavior could be expected to be a consecutive series of "same" responses until some criterial value was eventually reached. Interleaving the track types forces the subject to listen carefully to each pair with no expectation that consecutive trials are conditionally related. In the case of testing young children, an inclination to perseverate on a particular response can be avoided by interleaving the two types of trials, particularly at the beginning of the procedure, in which opportunities for both of the two available responses are provided.

Since testing all pair-wise comparisons along the stimulus continuum was not a feasible option due to time and attention constraints, the stimuli on either end of the stimulus continuum were chosen to serve as standards. In order that the results not be specific to just one particular frequency or frequency range, two standards rather than a single standard were used.

We decided that for each type of track ("start-same" and "start-different"), one of the two standards would be used for all trials. The tested pitch difference would therefore always be relative to this standard. Various methods of selecting the standards were tried, and finally it was decided that for the start-same trials, the standard would be the voice from the low end of the stimulus continuum, and for the start-different trials, the standard would be the voice from the high end of the continuum. (The reverse assignments could also have been used, but were not implemented in the present study.) Selection of two different standards provided two distinct reference points against which the pitch difference could be measured. We would then be able to determine whether our results were specific to the particular pitch standards used, or instead were generalizable across at least two different points along the pitch continuum. This particular choice also seemed appropriate because through its use, the two ends of the continuum were always heard throughout the set of 24 trials, thus providing a consistent set of "anchoring-points" heard throughout the procedure.

Additionally, the location of the standard was randomized, such that sometimes it was heard first in each pair, and sometimes second. This randomization made the first stimulus presentation less predictable, and forced/engaged the listener to listen carefully to every stimulus on every trial without any bias of expectation. The design choices described above rendered the underlying adaptive nature of the task relatively opaque to the participant.

The adaptive rule used was as follows. For the start-different trials, the size of the pitch difference was adjusted in 1-semitone increments until a reversal pattern of same-different-same was obtained, at which point the pitch increment size was reduced to a half-semitone. For the start-same trials, the size of the pitch difference was adjusted in half-semitone increments until three consecutive same-same-same responses were obtained, at which point, 1 semitone increments were used until a reversal pattern was obtained, at which point, the increment size was reduced back to a half-semitone. If a subject responded "different" when the pitch difference was zero, or "same" when the pitch difference was 6 semitones, the same pitch difference size was repeated again on the next test trial.

Individual sample data from a normal-hearing adult, a normal-hearing five-year-old child, and a pediatric cochlear implant user with above-average performance are shown in Figure 3. The trial number is shown on the abscissa, the pitch difference size tested on each trial is plotted on the ordinate. In the adult sample data shown in the top graph, for purposes of illustration, filled markers indicate a subject response of "same," unfilled markers indicate a subject response of "different." (This distinction between marker types is used only in this first illustration because the listener's response can be determined in this type of graph by observing whether the pitch difference was made smaller or larger on the subsequent trial.) For all three data sets, performance on the start-same trials is plotted using solid lines, while performance on the start-different trials is plotted in the dashed lines. Overlap towards the end of the trial-set between the lines representing the two tracks indicates a convergence of the two tracks on a similar-sized pitch difference.

As previously described, both a "fixed sentence" and a "varied sentence" condition were implemented in order to study the effect of linguistic variability on talker discrimination performance. In the fixed sentence condition, the linguistic content was always held constant across the two utterances to be compared, and in the present study, also across all trials. It should also be noted that although multiple tokens of every sentence were recorded and available, for the present experiment we decided to use only one token of each sentence. Although this choice leads to a less-convincing demonstration of discrimination because no abstract representation of a talker need necessarily be used to correctly respond

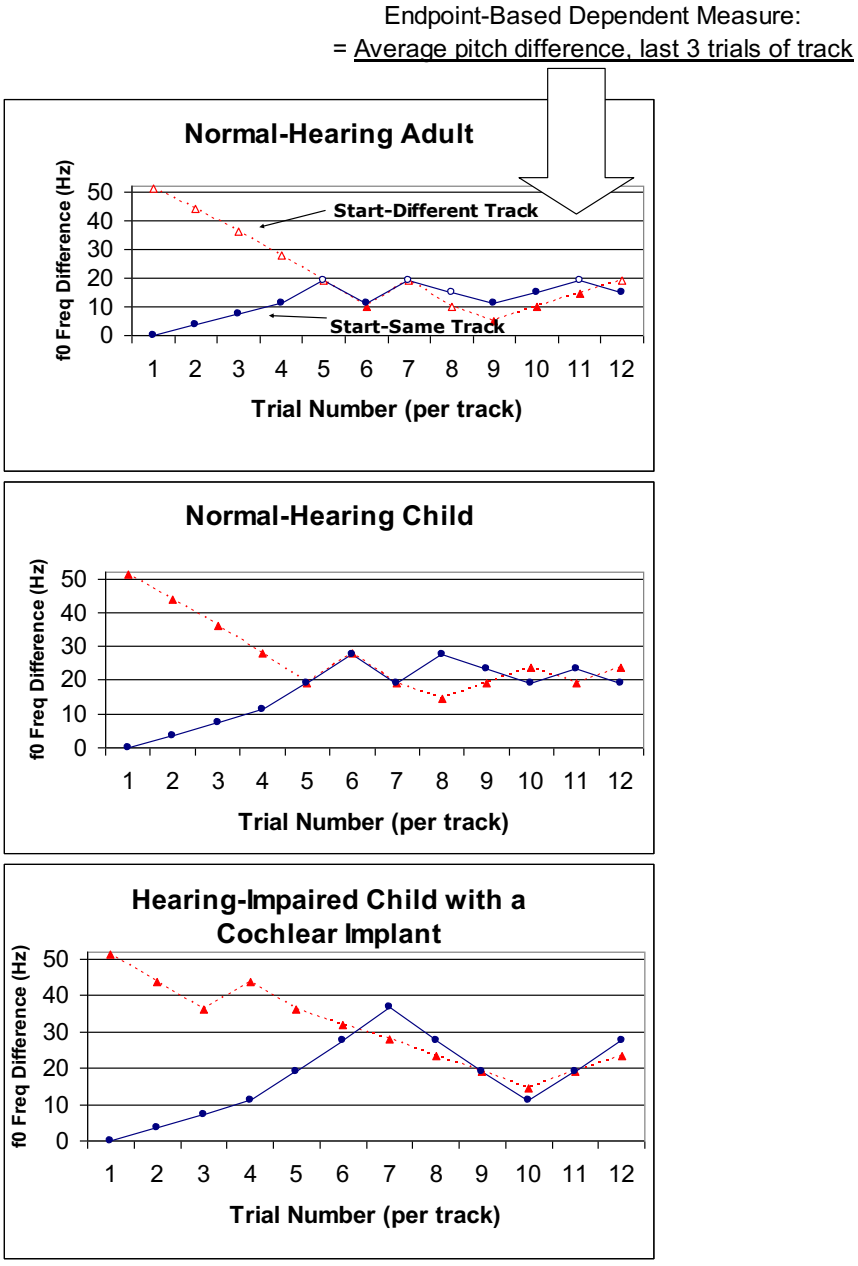


Figure 3. Sample Data. (a) Normal-hearing adult listener. (b) Normal-hearing five-year-old child. (c) Hearing-impaired child with a cochlear implant, displaying above-average performance. In (a) only, responses of "different" are shown by unfilled circles, responses of "same" by filled circles.

"same talker," we wanted to have at least one condition in which the pitch differences between tokens were exactly the size intended. When different tokens are used for the comparison, as in the varied sentence condition, because the voice of the natural talker varies slightly in pitch across her productions, the pitch differences between pairs of tokens lie close to, but are not exactly at the intended intervals along the stimulus pitch continuum. In the varied sentence condition, no sentence was ever repeated within the set of 24 test trials, and the listener was therefore required to make voice comparisons across linguistically different utterances.

The administration of the fixed versus varied sentence conditions was counterbalanced across subjects. Approximately half of the participants in each group completed the fixed condition followed by the varied sentence condition, and approximately half completed the conditions in the reverse order. Ideally, in order to control for stimulus token-specific effects, each sentence used in the varied sentence condition should also be used in the fixed sentence condition. However, due to the modest sample size, this was not possible. Unlike the previous studies using natural talkers however, since the size of the differences between the "talkers" along the stimulus continuum were mathematically equal across different sentences, unless keyword content was unexpectedly a factor, full balancing of the sentence set did not seem crucial.

The implant group therefore received a subset of the sentences heard by the normal-hearing group, in order to equate the two procedures of the two groups as nearly as possible. Approximately half of the sentences used in the varied sentence condition were used in the fixed sentence condition. As added insurance, a separate study, not reported here, was run using a group of 48 normal-hearing adult listeners, in which each of the 48 sentences used in the varied sentence condition was heard in the fixed sentence condition by one subject. No particular effect of individual stimulus tokens was observed in this other study (see Reference Note 1).

Keyword Identification Procedure. Following the talker discrimination task, each child completed a keyword identification task measuring auditory-only comprehension of the materials used in the talker discrimination task. The children were told that although they would still be listening to sentences through the loudspeaker, this time the task was different, and their job was now to listen to the words the people said. The children were shown how a 3 x 3 grid of pictures would appear on the screen immediately followed by the playing of a sentence of the (by now familiar) form, "The ____ and the ____ are by the ____." The children were told to look carefully at all the pictures on each trial and to find and press the three pictures that matched the words they heard in the sentence. The nine pictures included the three target pictures and six distracters illustrating other items in the keyword set. Each picture functioned as a touch-screen button and "bounced" as if depressed when touched. A red border was used to mark the pictures selected by the child. A screen shot of a sample response "plate" is shown in Figure 4.

Four practice trials were used to familiarize the child with the keyword identification task. On the first two practice trials, the experimenter demonstrated the task. The experimenter used the second practice trial to point out to the child that the first keyword would appear somewhere along the left side of the screen, the second keyword somewhere down the center of the screen, and that the final keyword would appear somewhere on the right-hand side of the screen. The child was directed to attempt the third and fourth practice trials on his/her own and was given feedback regarding performance. The test trials were then begun. The child was required to select three pictures, one in each "column" before moving onto the next trial. The location (top, middle, or center of each column) of the target pictures was pseudo-randomized. If the child could not identify a word, he/she was encouraged to guess from among the three picture choices. Although most normal-hearing children tended to select the target pictures from left to right on the screen, in the order in which the words were heard, there was no constraint placed on the order in which the target items could be selected.

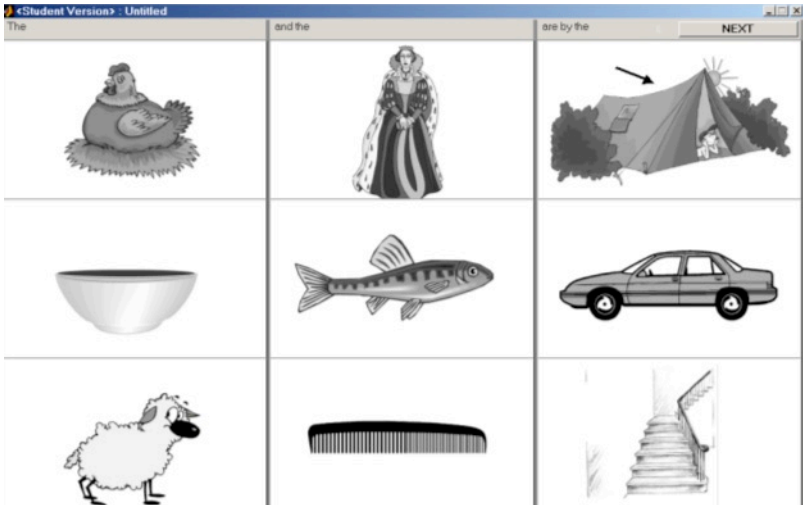


Figure 4. Keyword Identification Task Touch Screen Image

We initially intended to have participants complete the keyword identification task using all 48 sentences. Some pilot testing revealed however that 48 trials were simply too much for the younger children to complete in a single test session. Therefore, two sets of 24 sentences were created. Each child heard a set of 24 sentences that included the sentence he/she heard during the fixed sentence condition of the discrimination task. This was done to provide an informal check of whether the hearing-impaired children could identify the linguistic content of the sentence used during that particular discrimination condition. Within each set of 24 identification trials, participants heard one-third of the sentences under conditions of minimal manipulation in which no pitch shift was implemented but the stimuli had been passed through the resynthesis process with a scaling factor of 1. Another third were stimuli taken from the middle of the stimulus pitch continuum, and the remaining third represented stimuli from the far end of the pitch continuum, i.e., these were utterances that had undergone the most extreme manipulation of their f0 and formant frequencies. The three types of utterances were randomly intermingled during testing.

Including the three different levels of manipulation allowed us to assess whether the degree of manipulation was related to the intelligibility of the utterances. This design allowed us to ascertain whether the stimuli that had undergone the most extreme manipulation were less intelligible than the least manipulated stimuli. If this were found to be the case, it might be problematic for the use of these same stimuli in the talker discrimination task that involved comparisons across stimuli from different points along the stimulus continuum. Having pairs of stimuli differ greatly in intelligibility in addition to differing in their frequency composition might adversely affect listeners' performance on the talker discrimination task.

The keyword identification task was included in the procedure not only to assess the intelligibility of the test stimuli, but also to obtain an independent measure of spoken word recognition from each child. Because all the hearing-impaired children were already taking part in a separate longitudinal study that utilized most of the standard clinical word recognition tests used with this population, an independent measure of word recognition that the children were unfamiliar with was necessary to avoid interfering with the longitudinal study. A closed-set response format was chosen for the present keyword identification task in order to minimize the role of articulation in assessing the child's response.

Hardware and Programming

The computer equipment used to test the children consisted of a Dell Inspiron 4000 PC laptop computer with an Intel Mobile Pentium III processor and 128 MB of RAM, running the Windows ME operating system. The computer was equipped with an ESS Technology Maestro 3i audio controller/PCI audio soundcard with 16 bit D-to-A conversion.

All images were presented on a 15-inch color LCD monitor (ViewSonic VG151 ViewPanel) with 1024x768 XGA resolution. The monitor was located directly in front of the child at eye level (see Figure 5). Participant responses were collected using a Magic Touch touch-screen by Mass Multimedia/KeyTec Inc. (Model KTMT-1315 Pro-E touch screen add-on model for 14-15" monitors, Windows Version). All auditory stimuli for all listener groups were presented using a tabletop loudspeaker (HK185, THD < 1%) located either directly to the right of the computer monitor in the case of the normal-hearing children, or on the side of the implanted ear, in the case of the children with cochlear implants. Test stimuli were presented at approximately 70 dB SPL as determined by a portable sound level meter (RadioShack or Triplett) (A weighting) held at the approximate location of the child's head which was usually at a distance of 10-15 inches from the loudspeaker.

All control programs were written using the Matlab programming language (PC Version 6.0.0.42a Release 12 Student Version).

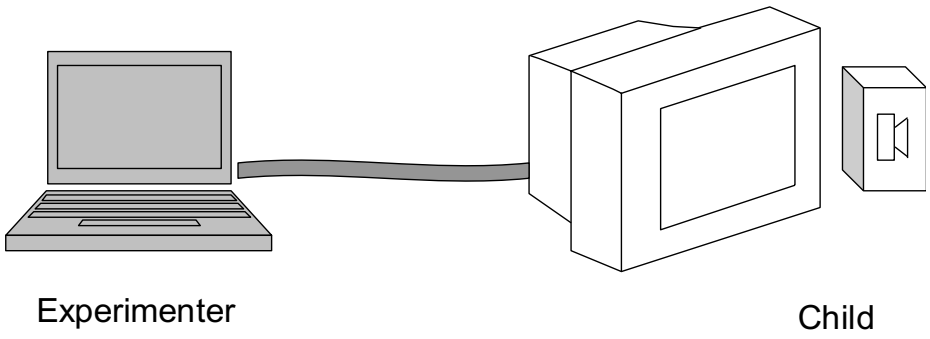


Figure 5. Schematic of testing equipment showing layout of hardware, experimenter, and child.

Receptive Vocabulary Screening Measure

All normal-hearing children were administered Form A of the Peabody Picture Vocabulary Test, Third Edition (PPVT-III Dunn & Dunn, 1997). This widely used receptive language test requires that on each trial the child listen to an auditorily presented word and then choose the correctly matched picture from a set of four illustrations. Raw scores, age equivalents, and standardized scores were tallied using the manual and published norms for the PPVT-III.

The families of the normal-hearing children who volunteered to participate from the Bloomington community tended to have children who were academically somewhat above average, given the current method of subject recruitment. The mean PPVT-III raw score for the five-year-old normal-hearing group was 89, *SD* = 13, *SE* = 3, range = 59-115. The mean PPVT-III standardized score was 113, *SD* = 13, *SE* = 3, range 74-134. An average raw score of 89 corresponds to an age equivalent score of about 6;8.

Form A or Form B of the PPVT-III was also administered to twelve of the children in the cochlear implant group during the same week as the current testing was carried out as part of the longitudinal study

being conducted at DeVault Otologic Research Laboratory. For this testing, the PPVT-III was administered by a clinician using either speech alone or using simultaneous speech and sign, depending on the child's preferred communication mode. The mean PPVT-III raw score for this subgroup of twelve hearing-impaired children was 68, *SD* = 20, range = 40-100, where this group included children who ranged in age from 5;6 to 11;6. The mean age equivalent PPVT-III score for these children with cochlear implants was, on average, 2.35 years below the children's chronological age. Examination of PPVT-III scores that were not obtained within the same week, but within the preceding 24 months from most of the remaining children indicated that a 2-year delay in receptive vocabulary was typical of the group as a whole. The size of the observed language delay conforms very closely to the average 2.60 years of auditory deprivation experienced by the group of hearing-impaired children prior to implantation.

Results and Discussion

Keyword Identification Task

Although the talker discrimination task was described before the keyword identification task in the preceding Method section, we reverse this order in presenting the results from these two tasks. The results from the keyword identification task are presented first because the children's performance on this more linguistically-based measure provides a comparison point against which to evaluate the talker discrimination results. Ordering the results in this manner permits us to establish that the vocabulary content of the new sentence materials was suitable for both groups of children being studied and that the resynthesis method did not greatly degrade the speech intelligibility of the test stimuli. For the hearing-impaired children with cochlear implants, this ordering of the results also allows us to demonstrate the degree to which the children who volunteered for this study are representative of pediatric cochlear implant users more generally, in terms of their spoken word recognition skills.

Normal-Hearing Children. The mean keyword identification score for the normal-hearing group of 5-year-olds was 92% words correct, *SD* = 6%, range = 80.6% to 98.6% words correct. The distribution of the group's scores is shown in the top panel of Figure 6. Because participants were required to choose one picture from each of three columns on each trial, chance performance for percent of keywords correctly identified is 33.3%. On the same task using the same speech materials, a group of 24 normal-hearing adults scored a mean of 99.8% correct, *SD* = 0.6% (Reference Note 1).

For the normal-hearing children, the mean percent correct for stimuli resynthesized by a scaling factor of 1.0 was 88.5% (range = 75% to 100%), by a factor of .84 (-3 semitones) was 92.3% (range = 70.8% to 100%, and by a factor of .71 (-6 semitones) was 93.6% (range = 64.6% to 100%). These means are displayed in Figure 7. A one-way repeated-measures ANOVA showed a significant main effect of synthesis degree $F(2,46) = 3.95$, $p = .026$. Paired sample t-tests indicated that the higher-pitched stimuli yielded significantly more errors than either the mid-range stimuli ($p = .032$) or the low-pitched stimuli ($p = .019$). These differences were small and are probably not of much theoretical importance, given that the differences may simply be an artifact of the assignment of particular sentences to the three synthesis levels: two different pseudo-randomized assignments of sentences to synthesis levels were used for all subjects in order to try to equate the potential difficulty and facilitate comparison of keyword identification scores across subgroups of subjects. The results do, however, contradict our concern that the stimuli with more extreme manipulation might be less intelligible than the stimuli that were only minimally manipulated.

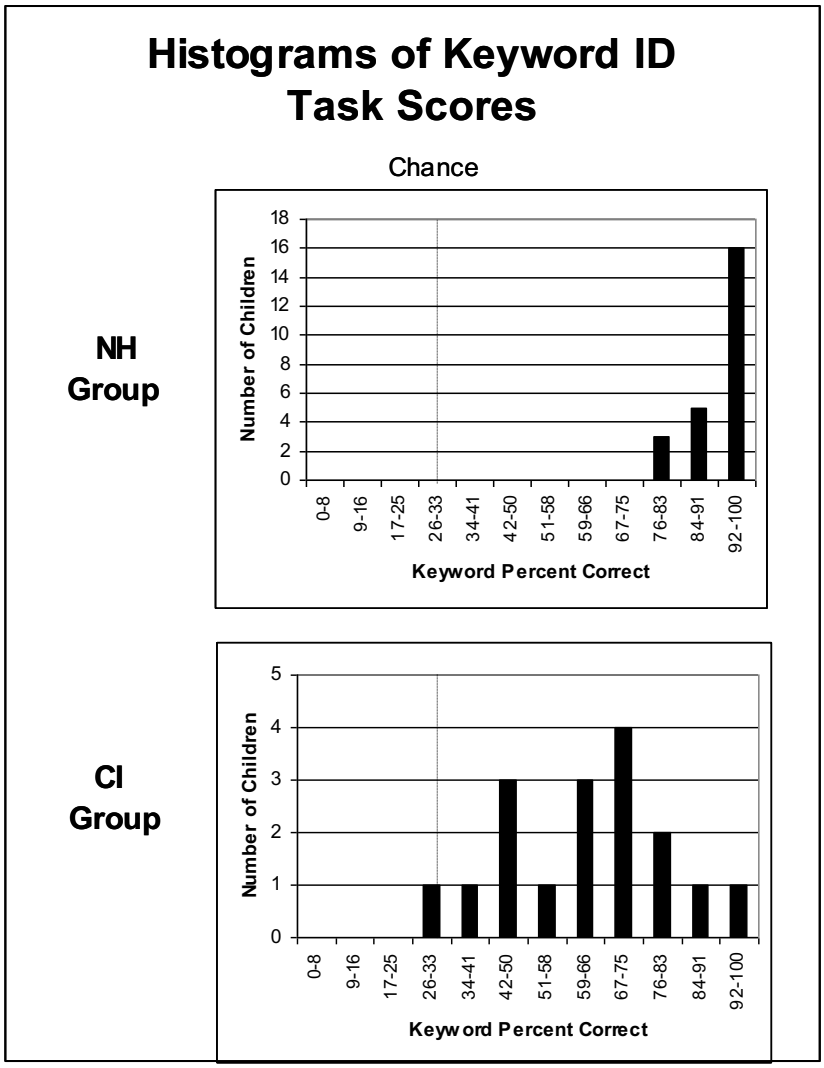


Figure 6. Frequency histograms illustrating the distribution of keyword identification performance within each group. Data from the normal-hearing children are shown in the top graph. Data from the hearing-impaired children with cochlear implants are shown in the bottom graph.

We also thought we might observe differences in keyword identification as a function of position within the test sentence. For example, memory for the third and final keyword might decay as the child made his/her manual responses for the first and second keywords. This pattern was not observed: the average number of words correct in first position was 91.3% correct, in second position was 92.5% correct, and in third position was 90.7% correct ($F < 1$). Thus, perhaps because the picture grid was presented just prior to the onset of the sentence, and remained visible until the child finished responding, memory demands did not appear for normal-hearing children to be a factor in this task.

Hearing-Impaired Children with Cochlear Implants. One child (CI-15, age 6) could not even complete the practice trials. The test trials were attempted, but no usable data was obtained from this child. Three of the youngest children, two five-year-olds and one six-year-old, completed a set of 12 trials (36 keyword responses) rather than the full set of 24 trials (72 keyword responses) because of short attention

spans. One seven-year-old who displayed unusually poor attention skills and who had a prior history of attention problems was also only tested on a half-set of 12 sentences. Given that scores on the first 12 trials correlated well with scores on the full set of 24 trials for the remaining children (r = +0.93), these four half-set scores were retained in most of the analyses reported below.

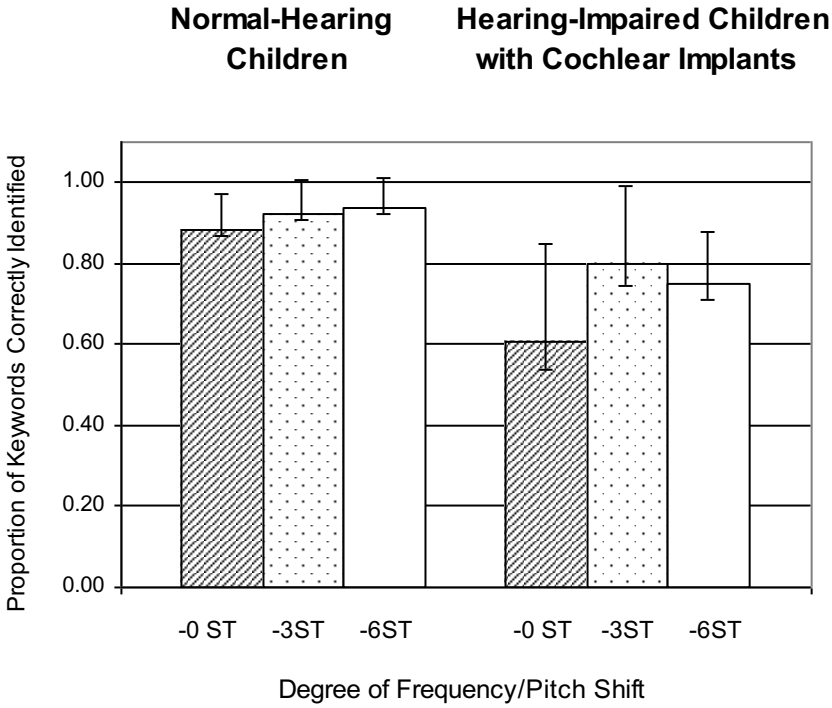


Figure 7. Mean keyword identification performance as a function of degree of manipulation. Speech stimuli were resynthesized with all frequency components shifted either zero semitones, minus three semitones or minus six semitones. Error bars in the positive direction indicate one standard deviation. Error bars in the negative direction indicate one standard error.

Mean keyword identification performance for the remaining group of 17 children with cochlear implants was 63.8% correct, *SD* = 19.3%, *SEM* = 4.7%. Scores ranged between 27.8% to 98.6%, where chance is 33.3%. A frequency histogram of the hearing-impaired children's scores is shown in the bottom panel of Figure 6. The group mean was significantly higher than chance performance ($t(16) = 6.5$, $p < .001$), but significantly lower than the average performance of the normal-hearing group ($t(39) = 6.6$, $p < .001$). We were pleased to find that despite our concern that the recruiting process might self-select for only exceptionally successful implant users, a wide range of performance was represented in the sample. A normally-distributed set of keyword identification scores was obtained with a mean centered almost directly between chance and ceiling performance.

For the thirteen children who completed a full set of 24 trials (group mean score = 70% correct), performance across the three synthesis conditions was compared. As in the case of the normal-hearing children, it was found that counter to expectation, average performance was worst on the least manipulated stimuli (M = 60.7%). Performance on the mid-range stimuli averaged 79.8% correct, and on the low-pitched stimuli, 74.8% correct (by a repeated measures one-way ANOVA, $F(2,24) = 5.53$, $p = .01$). The means for the hearing-impaired children as a function of degree of manipulation are also shown in Figure 7. In this group, as in the normal-hearing group, performance across the three keyword positions

within the sentences was almost exactly the same, first position = 68.6%, second position= 68.4%, third position = 72.8% ($F < 1$).

Assessment of Stimulus Materials. For the keyword identification task to serve as a word recognition task rather than as an auditory receptive vocabulary test, it was necessary that the children know the meaning of the words used in the sentences. For the normal-hearing group of five-year-olds, the high scores obtained on the keyword identification task indicate that word-knowledge did not play much of a role in determining within-group performance differences. Despite the relatively truncated range of scores in the normal-hearing group, we did, however, attempt to account for what little variability was observed.

A multiple linear regression analysis conducted for descriptive purposes showed that subject gender and age in months together accounted for 35% of the observed variance in the normal-hearing group's keyword identification scores, or almost equivalently, 34% of the variance, when arcsine transformed scores were used. Female gender and older chronological age were associated with higher keyword identification scores. Raw receptive vocabulary scores accounted for an additional 3% of the variance, or, taken alone, about 21% of the observed variance in the normal-hearing children's scores, both when transformed and non-transformed scores were examined. Higher vocabulary scores were associated with better keyword identification performance. Informal observation suggested that factors such as individual differences in attention, and or dexterity with the manual response format might also account for some of the remaining variance in keyword identification scores.

An item error analysis conducted on the responses from the normal-hearing group of five-year-olds showed that none of the 144 keywords were misidentified by more than one-third of participants. Only 3 of the 144 words were misidentified by one-third of participants ("cook" (a few children looked for "Coke", "scarf" (unclear why this word was often misperceived), and "skate" (the image was of an ice-skate)). Four additional words were misidentified by a quarter of participants ("ape", "chair", "hill", and "grape"). Twelve words were misidentified by one-fifth of the participants. The remaining 125 words were correctly identified by 81-100% of the normal-hearing participants. These results may be used in the future to remove/replace items from the current stimulus set to further reduce the role of word familiarity for normal-hearing children of this age.

We also examined the degree to which word-knowledge might have contributed to performance differences observed among the children with cochlear implants. In a multiple linear regression analysis, chronological age accounted for 33% of the observed variance in keyword identification scores but subject gender failed to account for any variance in the implant group. As expected, older children obtained higher keyword identification scores. A measure reflecting extent of implant experience, expressed as either length of implant use or age at time of implantation, accounted for an additional ~20% of variance. Finally, within the group of 14 children for whom the PPVT-III scores were current to within 10 months, raw receptive vocabulary scores accounted for an additional ~36% of the variance. Higher vocabulary scores were again associated with better keyword identification performance in this task.

Taken together, the results from both the normal-hearing and hearing-impaired children provide some measure of the degree to which the receptive language delay observed in the cochlear implant group might have influenced performance on the keyword identification task. Our results suggest that for hearing-impaired children of age 7 to 8 or older, the procedure and vocabulary used in the keyword identification task are probably quite appropriate. However, these data and our experiences in administering the test indicate that, despite the high levels of performance observed in five-year-old normal-hearing children, as a measure of spoken word identification, the picture selection procedure and vocabulary set may be slightly too difficult for hearing-impaired five- and six-year-olds who have delayed language and receptive vocabulary skills.

Nevertheless, as shown in Table 2, the scores of the cochlear implant group on the keyword identification task were strongly correlated with the children's scores on several standard clinically-used word recognition tests administered within 24 months of the present testing. Strong positive correlations were observed between the hearing-impaired children's keyword identification scores and their scores on the Phonetically Balanced Kindergarten (PBK) test (Haskins, 1949), and Lexical Neighborhood Test (LNT) Easy and Hard Word Lists (Kirk, Pisoni, & Osberger, 1995). The PBK and LNT are both widely used open-set tests of word recognition that require the child to repeat back each test item presented in isolation. These results suggest that the keyword identification task introduced in the present report provides a valid measure of speech perception and word recognition skills in this clinical population.

TABLE 2

CORRELATIONS BETWEEN KEYWORD IDENTIFICATION SCORES AND COMMONLY USED CLINICAL SPEECH PERCEPTION MEASURES

CORRELATIONS WITH KEYWORD IDENTIFICATION TASK (AUDITORY-ONLY, PERCENT WORDS CORRECT) N = "AVAILABLE N"	
CHRONOLOGICAL AGE IN MONTHS PARTIALLED OUT OF ALL CORRELATIONS	
PBK WORDS CORRECT, AUDITORY-ONLY, LIVE-VOICE M = 61%, SD = 19%	r = +.72** N = 13
LNT-EASY WORDS CORRECT, RECORDED, MULTI-TALKER M = 62%, SD = 21%	r = +.71** N = 13
LNT-HARD WORDS CORRECT, RECORDED, MULTI-TALKER M = 47%, SD = 22%	r = +.85*** N = 13
HINT-C KEY WORDS CORRECT IN QUIET, RECORDED, SINGLE-TALKER M = 60%, SD = 26%	r = +.94*** N = 9
HINT-C KEY WORDS CORRECT IN NOISE, RECORDED, SINGLE-TALKER M = 49%, SD = 34%	r = +.75* N = 9
* $p < .05$, ** $p < .01$, *** $p < .001$	

Strong correlations were also obtained between the children's keyword identification scores and two versions of the Hearing-in-Noise Test for Children (HINT-C) (Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Gelnett, Sumida, Nilsson, & Soli, 1995). The HINT-C is a standard clinical test of word identification in sentence context that the present keyword identification test was designed in part to emulate. Like the present keyword identification task, the HINT-C test is also intended to be an easy task for "normally-hearing children as young as 5 and 6 years of age" (Eisenberg et al., 2000). The new keyword identification task designed for this investigation was found to correlate most strongly with the HINT-C test administered under quiet listening conditions. For the subgroup of nine children for whom scores on both the HINT-C in quiet and the present keyword identification task were available, the mean difference between the two scores was only two percentage points (i.e. 60% correct on the HINT-C in quiet

vs. 62% correct on the current keyword identification task). The close correspondence between the scores obtained in the two tests indicates that the keyword identification task introduced in the present study measures essentially the same skills as the HINT-C administered under quiet listening conditions.

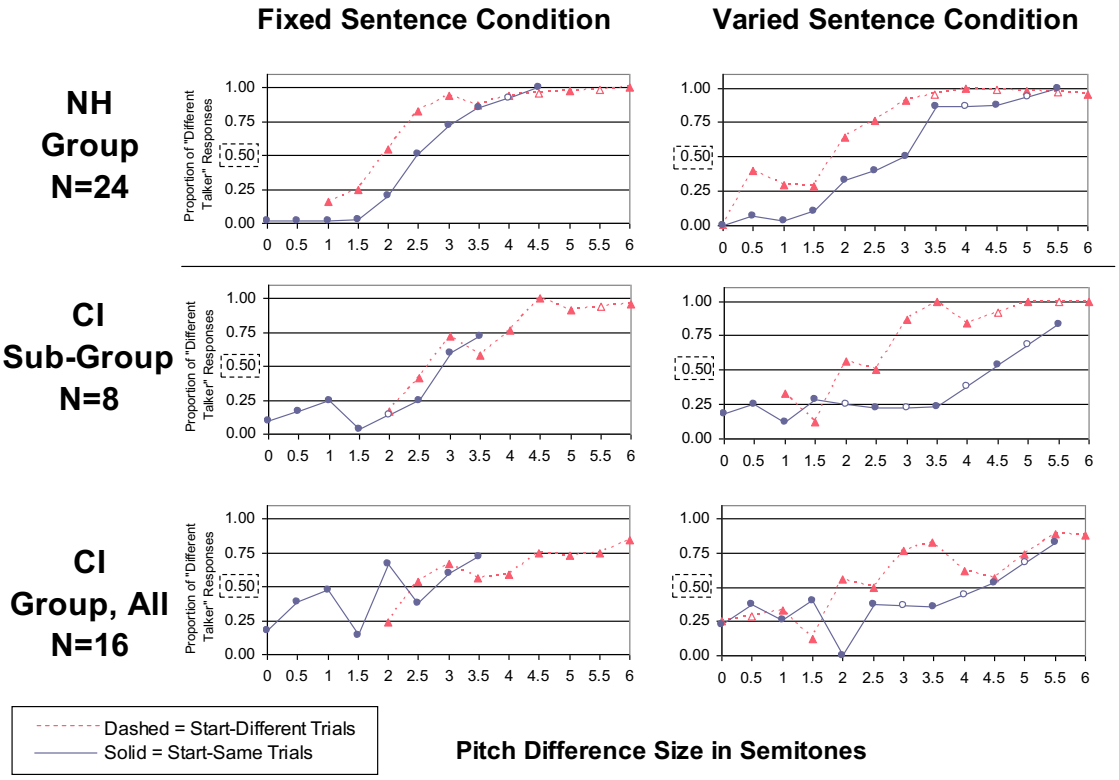
Talker Discrimination Task

Data Limitations. In the normal-hearing group, all 24 children were tested in both the fixed sentence and varied sentence conditions. In the cochlear implant group, four of the 18 children in the sample only completed either the fixed sentence condition or the varied sentence condition, but not both tasks. These were the same four children who completed a half-set of trials on the keyword identification task. Two children were tested in the fixed sentence condition only, and two were tested in the varied sentence condition. Thus, for the cochlear implant group there were 16 sets of data available per condition, 14 of which were within-subject sets, and 2 of which were between-subjects sets.

Simulated Comparison Data Representing Chance Responding. In addition to the actual empirical data collected from the two groups of children, we also generated simulated "data" representing the chance level of performance expected from random responding. To obtain estimates of the scores predicted by chance performance, the control program used to test the children was run multiple times using same/different responses determined by a random number generator as its input. In order to determine the variability in performance that might be expected within a data set of the size represented by the sample of normal-hearing group of children, we tabulated the simulated data from 24 tracks of each type using randomly determined same/different responses. These values are shown for comparison purposes in several of the figures presented below.

In order to obtain a more accurate estimate of the average pitch difference size expected by chance on the last three trials of each track according to random responding, we ran an additional 1000 simulated tracks of each type. This "end-point"-based value was of special interest to us because in analyzing actual subject data we planned to use this value as one of our primary dependent measures. For each track type (start-different and start-same), we calculated the average pitch difference between the standard and the comparison stimulus across the last three trials, and averaged these values over the 1000 simulated data sets. The resulting values will also be shown for comparison purposes in several of the figures below.

Group Psychometric Functions. We first examined the performance of each group of children averaged across all twelve trials in each track type and condition. Figure 8 shows the proportion of "different talker" responses obtained as a function of pitch difference size in semitones for both groups of children. Data from the normal-hearing children are shown at the top of the figure. Data from the group of hearing-impaired children with cochlear implants are shown in the middle and bottom panels of the figure: the middle panel displays the data from a better-performing subset of the hearing-impaired children who met certain "minimum performance" criteria that will be described later in further detail for the talker discrimination task, while the bottom panel presents the results obtained from the entire group of hearing-impaired children. Results from the fixed sentence condition are shown on the left; results from the varied sentence condition are shown on the right. Responses for start-different trials using the higher-pitched standard are shown by the triangles and dashed lines; responses for start-same trials using the lower-pitched standard are shown by the circles and solid lines. Chance performance would entail a 0.50 probability of a "different talker" response at all pitch differences sizes, and therefore would appear as a horizontal straight line if it were plotted in each graph.



*Unfilled markers are interpolated from immediately adjacent values. Only difference values with 3+ data points available are plotted using filled markers. Chance performance = a horizontal line at the .50 level.

Figure 8. Group psychometric functions. Proportion of "Different Talker" responses are shown as a function of pitch difference size in semitones. Results from "Start-Different" tracks are shown in dashed lines, results from "Start-Same" tracks, in solid lines.

One pattern clearly evident in these data are systematic differences between the results obtained using the two different standards. For each graph shown in Figure 8, the data points connected by the dashed lines lie consistently above the data points connected by the solid lines, corresponding to a smaller 50% intercept for the dashed line function showing the start-different trial results. More specifically, when expressed in terms of semitones, the frequency difference needed to elicit any fixed proportion of "different talker" responses was consistently larger when the start-same track with the lower-pitched standard was used as compared to when the start-different track with the higher-pitched standard was used. The solid-line functions therefore appear shifted to the right of the dashed line functions in each graph. The shift is about one-half semitone in size for the normal-hearing group of children, regardless of presentation condition (fixed sentence vs. varied sentence). The differences between the start-different and start-same functions are quite well defined in the normal-hearing group of children, and are also present although somewhat less clearly, for the children with cochlear implants.

This same relationship between the functions generated using the two different standards was also observed in analogous data obtained from 48 normal-hearing adults (see Reference Note 1). Once this pattern was observed in both sets of normal-hearing listeners, it became clear that this was not a random discrepancy, but rather that both normal-hearing adults and normal-hearing children were using a linear Hertz scale in their responses rather than a logarithmic scale. We therefore converted all frequency difference measures into linear Hertz units, expressed relative to the mean fundamental frequency of each standard. The pitch differences obtained using the two different types of tracks then fell into fairly close agreement for all our normal-hearing participants and there was no longer a significant difference between

the measures of performance gathered using the two standards. The group data shown in Figure 8 are reproduced in Figure 9 using Hertz units to illustrate this point.

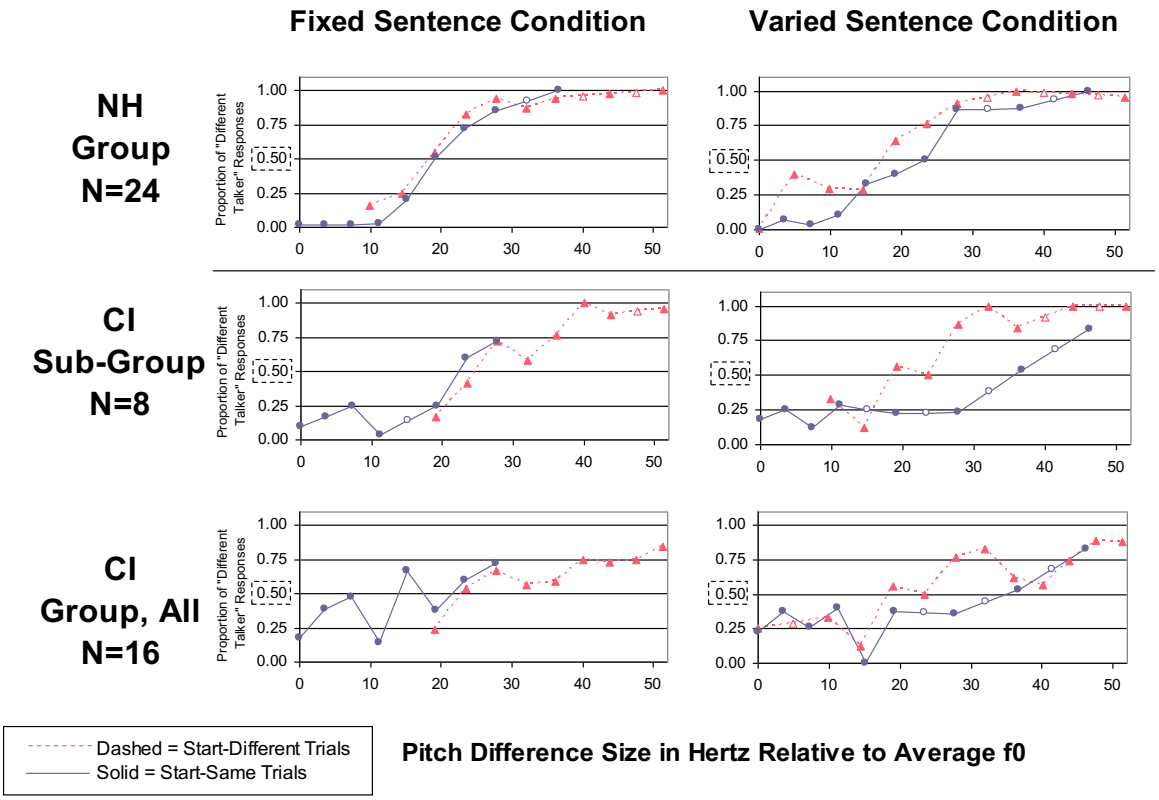


Figure 9. Group psychometric functions. Proportion of "Different Talker" responses are shown as a function of pitch difference size in Hertz relative to the mean fundamental frequency of the standard. Results from "Start-Different" tracks are shown in dashed lines, results from "Start-Same" tracks, in solid lines.

As shown in Figure 9, the solid and dashed-line functions now display considerably greater overlap for the normal-hearing group, particularly in the fixed sentence condition. For the hearing-impaired children with cochlear implants, the conversion between semitone and Hertz units had a somewhat less dramatic effect. Better agreement between the functions was, however, clearly evident in the fixed sentence condition for the implant group, as shown in the lefthand panels of Figure 9. In light of these results, although we initially planned to report our results in terms of the semitone (ST) units used to define the stimulus continuum, in the remainder of our analyses, we adopted linear Hertz units to express our dependent measures.

Once we addressed these largely unanticipated effects of standard type, we could then examine the effects of primary interest to us, specifically, those of listener group (normal-hearing versus hearing impaired) and presentation condition (fixed sentence versus varied sentence). Inspection of Figures 8 and 9 reveals some clear differences between the two groups of children. The functions from the normal-hearing group are relatively smooth and increase steadily as a function of pitch difference size in a near monotonic fashion, particularly in the fixed sentence condition. The group functions from the hearing-impaired children with cochlear implants, in contrast, are considerably more variable and do not increase in smooth

monotonic manner. Although the performance of the implant group is certainly erratic and rather poor overall, their group psychometric functions do, however, differ visibly from the horizontal straight line function that would characterize chance performance at the .50 level for all pitch difference sizes. This suggests that at least some of the hearing-impaired children in the group completed the talker discrimination task at better than chance levels.

Examination of the 50% crossover point in each function indicates that the normal-hearing five-year-old children perceived the paired voices as more often belonging to two different talkers than to the same talker when the pitch difference between the voices was at least 2 to 2.5 semitones, or approximately 19 Hz, relative to the mean fundamental frequency of the standards. Although the functions for the implant group are difficult to interpret due to their irregular shape, the hearing-impaired children considered as a group appear to require a somewhat larger difference than the normal-hearing children to perceive two voices as belonging to two different talkers. In the fixed sentence condition, if a smoothed function were fit to the group data, the 50% crossover point would suggest that a difference of approximately 2.5 to 3 semitones or about 24 Hz is required for children with cochlear implants to perceive two voices as belonging to different talkers. For the varied sentence condition, the irregular shape of the function makes estimation of this value more difficult, but it would appear to be approximately the same as in the fixed case for the implant group, that is, about 2.5 to 3 semitones or about 24 Hz.

Another notable difference between the psychometric functions of the normal-hearing children compared to those from the children with cochlear implants is the hearing-impaired group's higher rate of "different talker" responses when there was minimal to no actual physical difference between the stimuli comprising the test pair. In the hearing-impaired group of children it was also the case that even for the largest pitch differences tested, responses of "same talker" were sometimes obtained, even though this was virtually never found to occur in the normal-hearing group of children. This finding reflects the difficulty experienced by the hearing-impaired children with cochlear implants even when the perceptual judgments in the task were designed to be relatively easy.

The effects of presentation condition were less clear-cut than those of hearing-status. For both the normal-hearing and hearing-impaired groups of children, the 50% cross-over point at which responses of "different talker" and "same talker" were equally likely, occurred at approximately the same pitch difference size regardless of presentation condition. That is, neither group of children appeared to require a larger pitch difference in the varied sentence condition than in the fixed sentence condition. This finding of similar performance in both the fixed and varied sentence conditions is, in the case of the hearing-impaired children, contrary to what previous findings using natural talkers would predict (see Cleary, 2003/Chapter IIA).

Although no differences were evident in the 50% crossover points of the fixed versus varied sentence functions, visual comparisons across the two sets of data revealed that for both groups of children, the functions observed in the fixed sentence condition were smoother and more-regularly shaped than the functions obtained for the varied sentence condition. Not only were the functions more regularly shaped, but the agreement between the "start-different" and "start-same" functions was also much greater in the fixed sentence condition than in the varied sentence condition for both groups of children.

Several different possible explanations can be offered to account for this finding of greater variability in performance within the varied sentence condition. The most interesting proposal is that in the varied sentence condition, both groups of children experienced some difficulty "generalizing" their representations of a talker's voice to a new utterance and "ignoring" the presence of linguistic variability that was irrelevant to the task at hand. This greater difficulty led both groups of children to display more

variable patterns of performance in the varied sentence condition than in the fixed sentence condition which required no such generalization.

Another possible explanation for the greater variability in the varied sentence condition is related to methodology. In the varied sentence condition, the differences in average fundamental and formant frequencies between the sentences in each test pair were more approximate due to natural variation in these values across different utterances. Thus, we might expect performance in this condition to be somewhat more variable across the entire set of trials. Although we cannot rule out this possibility, we would note that because of the manner in which the tokens were recorded and screened before their use, the actual differences in mean pitch between different utterances were in fact, quite small. The extent to which we were able to successfully control this factor is likely to have contributed to high similarity observed between the 50% cross-over points for both the fixed and varied conditions.

The methodological explanation for the finding of greater variability in the varied sentence condition also seems doubtful because in the fixed sentence condition, different sentences were heard by different subjects, whereas in the varied sentence condition, the same ordering of the sentences (in terms of linguistic content) was heard by all subjects. Thus, it might be predicted that cross-subject variability would actually be less in the varied sentence condition than in the fixed sentence condition because all participants received the same ordering of sentences. However, no evidence of this pattern was found in either group of children.

Formal statistical analyses of these effects were next conducted, however, for a variety of reasons, the psychometric functions for individual subjects over all test trials were not a good basis for these new analyses. Because of the small number of trials and the rather poor and erratic performance of the cochlear implant group, the fitting of smoothed functions to individual subject data proved to be extremely problematic. Therefore, instead of a 50% crossover point determined for individual psychometric functions, we employed another measure that was, by definition of the adaptive algorithm, nearly equivalent, namely the average pitch difference size tested over the last three trials of each "start-different" and "start-same" track (see Levitt, 1971).

To further validate using performance on the last three trials of each test track as a primary dependent measure on which to conduct our statistical analyses, we reasoned that it was important to first show that the adaptive tracking methodology did in fact work as designed. More specifically, by reporting performance in each group of children as a function of trial number we would be able to demonstrate that the number of trials included was sufficient to reach a stable level of performance during the adaptive task. Examining both group and individual subject performance as a function of trial number might also help us account for the greater mismatch in the cochlear implant group between the group psychometric functions obtained using the two different standards. The large degree of mismatch that remained between the start-different and start-same functions even after conversion into linear Hertz units for the cochlear implant group led us to be concerned whether the adaptive procedure had been a suitable testing method for some of the children in implant group.

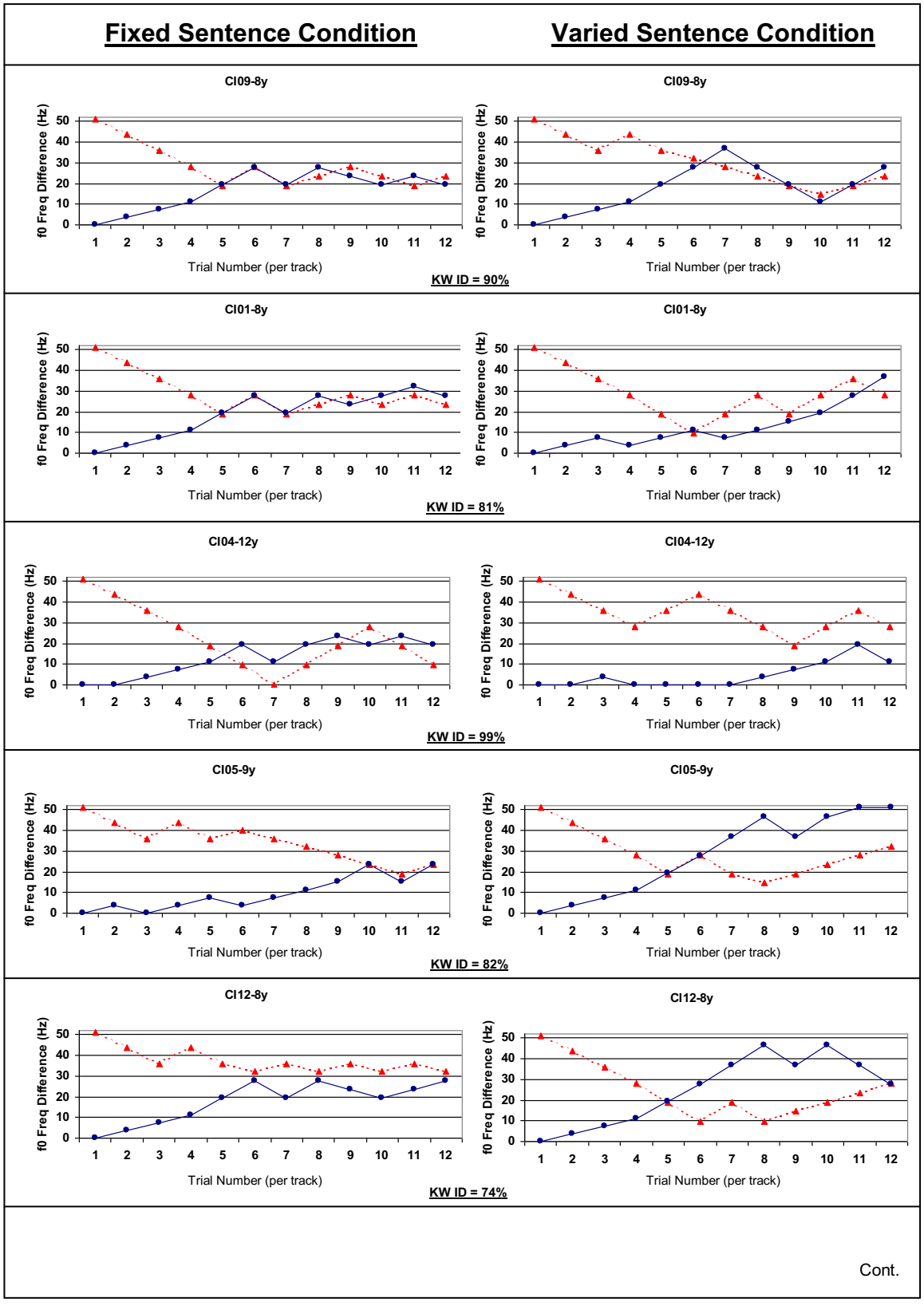
Issue of a Cochlear Implant Subgroup. Individual subject data from each child in the cochlear implant group are displayed as a function of trial number in Figure 10 (pages 61-64). The data sets are presented in rank order of performance. Responses in the fixed sentence condition are shown along the left-hand side of each page. Responses in the varied sentence condition are shown along the right-hand side of each page. Start-different tracks using the higher pitched standard are shown using triangles and dashed lines. Start-same tracks using the lower-pitched standard are shown using circles and solid lines. Each hearing-impaired child's score in the previously described keyword identification task ("KW ID") has been

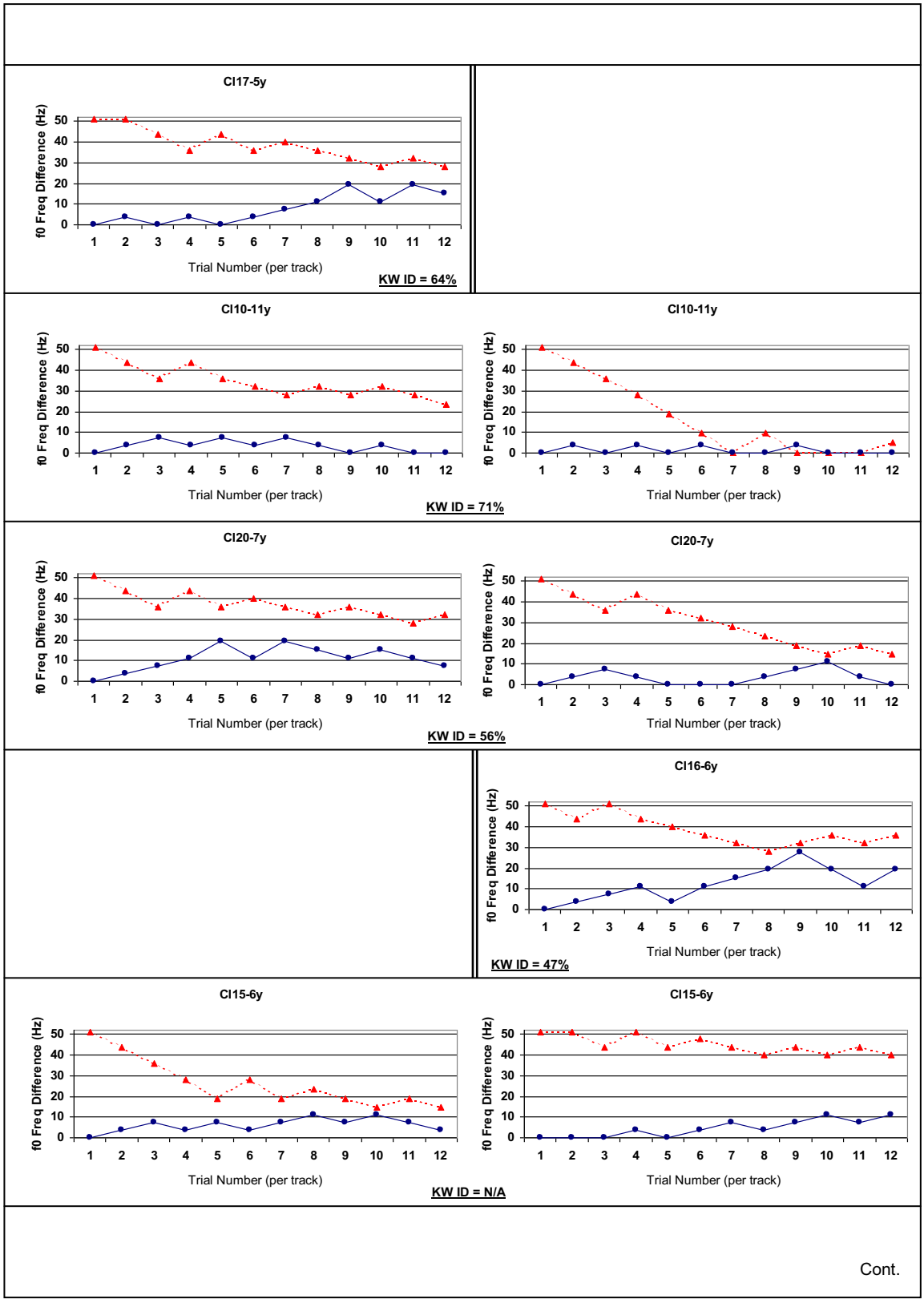
included in Figure 10 in order to facilitate comparison of the children's performance across the different tasks.

Preliminary examination of the data from individual children in the cochlear implant group suggested that the hearing-impaired participants could be subdivided into two groups: one group showed evidence of being able to comply with the talker discrimination task instructions in a manner analogous to that of normal-hearing children, the other group appeared unable to do so. The most obvious difference between the two groups of hearing-impaired children was whether or not the pitch difference values followed by the two different types of tracks converged towards a single value. To quantify these observations, we therefore calculated a "track convergence" score for each child in each condition. The track convergence score was defined as the difference between the average pitch difference tested on the last three trials of the start-different track minus the average pitch difference tested on the last three trials of the start-same track. This was done for both the normal-hearing children and the children with cochlear implants.

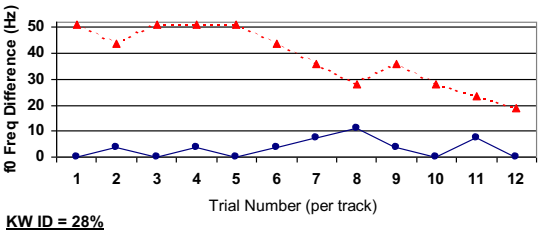
The track convergence score predicted for chance responding is large because random responses lead to poor convergence between the two tracks. More specifically, our simulations showed that with chance responding, the pitch differences tested by the two track types on the last three trials could be expected to differ from each other by about 28 Hz. We found that only eight of the children in the cochlear implant group who were tested in both conditions displayed track convergence scores in both of these conditions that were better than those predicted by chance (i.e., less than 28 Hz). In contrast, all 24 of the normal-hearing children tested displayed track convergence scores in both conditions that were better than chance.[4]

[4] If we use the data from the normal-hearing group as a baseline, and if we assume that similar-sized values should be obtained regardless of the choice of standard, then the track convergence score can also be interpreted as a rough measure of "task compliance." That is, when the track convergence measure is close to zero for a given subject, this is evidence that the listener was following the task directions: Cooperative listeners with good discrimination abilities as well as cooperative listeners with poor discrimination abilities should both exhibit track convergence measures that approach zero. That is to say, a listener with good discrimination capabilities can be expected to reach a pair of small and converging pitch difference values, while a listener with poor discrimination capabilities should exhibit a pair of large but also converging pitch difference values. A listener who does not follow the instructions consistently, on the other hand, should demonstrate a lack of convergence between the two types of tracks, and his/her dependent measures for each track should approach the values predicted by chance. Because of the natural bound on the minimum size of the possible difference to be tested and the selection of a 6 semitone difference as the largest pitch difference tested, random responding predicts pitch difference values at the end of 12 trials that are slightly separated from these floor and ceiling values. More specifically, for Start-Different type tracks, the expected average pitch difference over the last three trials for chance levels of responding was ~37.54 Hz (4.18 ST). For Start-Same type tracks, the expected average pitch difference over the last three trials for chance responding was ~9.57 Hz (1.29 ST). Given the expected values for Start-Different and Start-Different trials, the track convergence score expected by chance is therefore a 28 Hz difference, or ~2.89 ST. Thus, for a particular subject or group of subjects, a track convergence score that approaches 28 Hz (2.89 ST) demonstrates an inability to do the task. A track convergence score that is smaller than 28 Hz, i.e., one that approaches zero as in the case of the normal-hearing children, suggests that subjects are able to comply with the task directions, regardless of their individual discrimination ability. A track convergence score greater than 28 Hz suggests that the subject is doing more poorly than chance; that is, either willfully or unintentionally, the subject is to greater or lesser degree, responding "same" when the correct answer is different, and "different" when the correct response is "same".


Fixed Sentence Condition
Varied Sentence Condition
CI09-8y
CI09-8y
f0 Freq Difference (Hz)
Trial Number (per track)
KW ID = 90%
CI01-8y
CI01-8y
KW ID = 81%
CI04-12y
CI04-12y
KW ID = 99%
CI05-9y
CI05-9y
KW ID = 82%
CI12-8y
CI12-8y
KW ID = 74%
Cont.

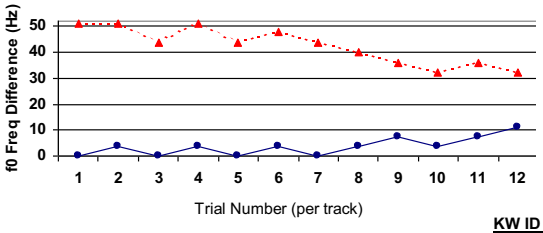

CI17-5y
f0 Freq Difference (Hz)
Trial Number (per track)
KW ID = 64%
CI10-11y
f0 Freq Difference (Hz)
Trial Number (per track)
CI10-11y
f0 Freq Difference (Hz)
Trial Number (per track)
KW ID = 71%
CI20-7y
f0 Freq Difference (Hz)
Trial Number (per track)
CI20-7y
f0 Freq Difference (Hz)
Trial Number (per track)
KW ID = 56%
CI16-6y
f0 Freq Difference (Hz)
Trial Number (per track)
KW ID = 47%
CI15-6y
f0 Freq Difference (Hz)
Trial Number (per track)
CI15-6y
f0 Freq Difference (Hz)
Trial Number (per track)
KW ID = N/A
Cont.

CI13-7y

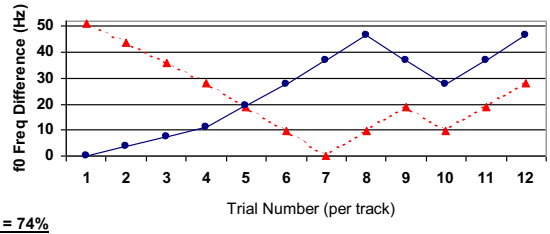


KW ID = 28%

CI06-8y

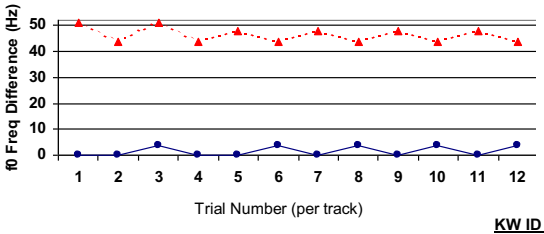


CI06-8y

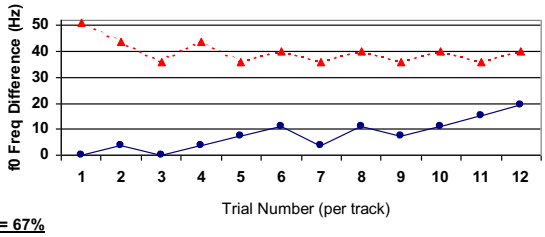


KW ID = 74%

CI14-6y

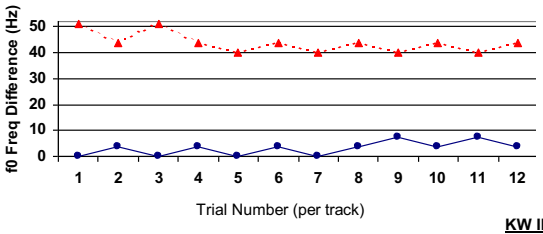


CI14-6y

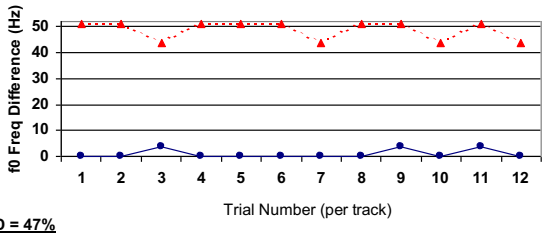


KW ID = 67%

CI03-5y

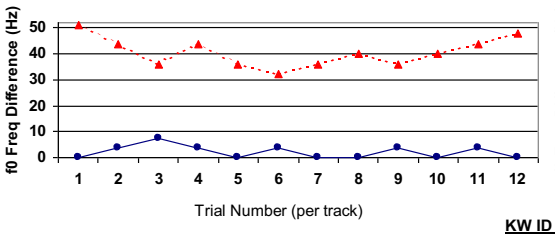


CI03-5y

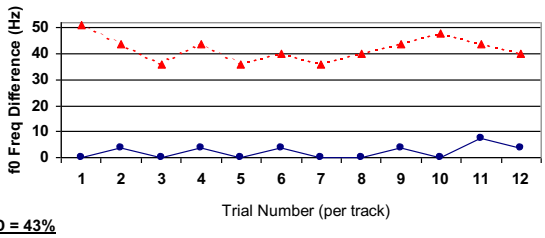


KW ID = 47%

CI07-10y



CI07-10y



KW ID = 43%

Cont.

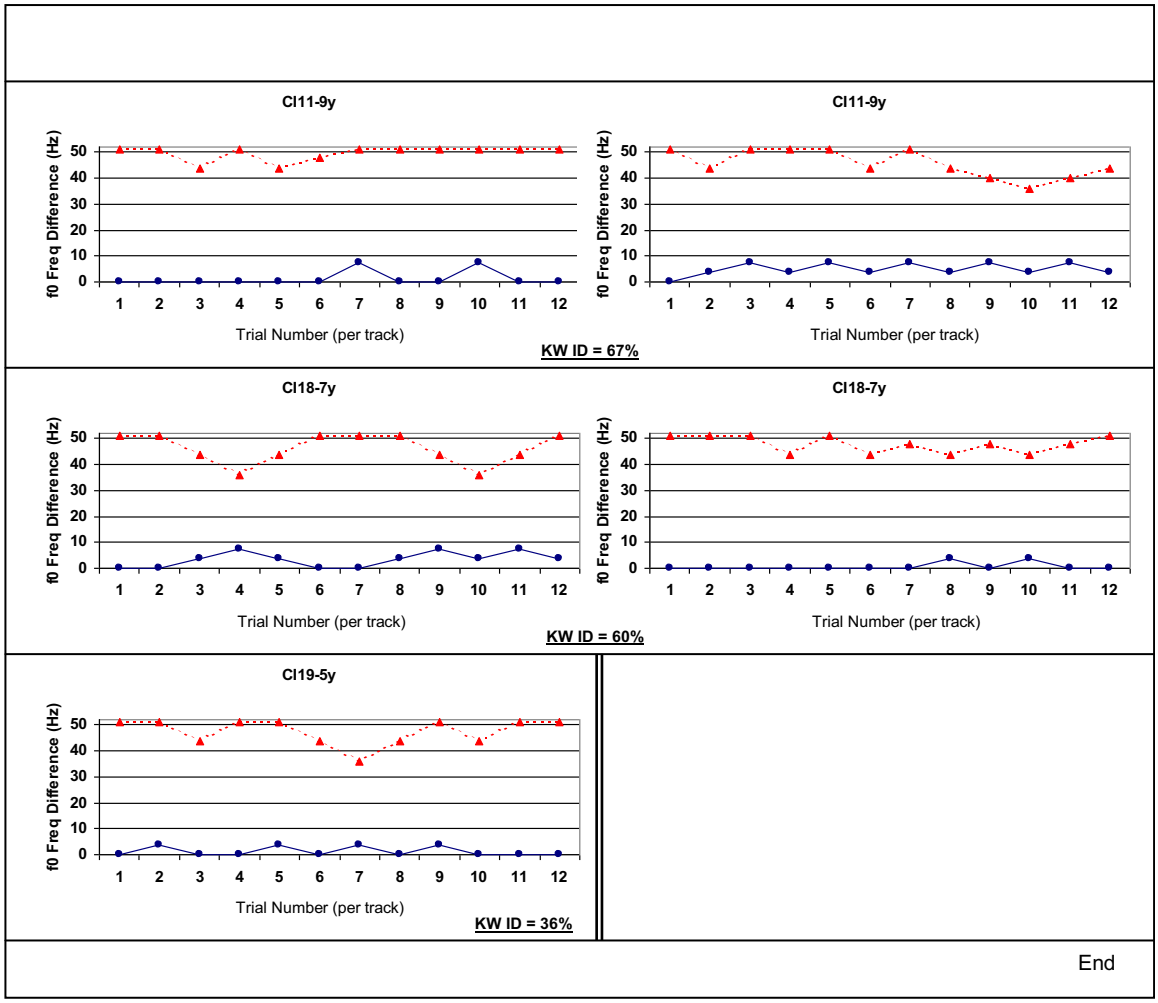


Figure 10. Individual Participant Data, Hearing-Impaired Children with Cochlear Implants. For each child the pitch difference size (in Hz relative to the f0 of the standard) is shown as a function of trial number for the two different track types. Start-Different tracks using the higher pitched standard are shown using triangles and dashed lines. Start-Same tracks using the lower-pitched standard are shown using circles and solid lines. The participants are ordered in terms of their track convergence scores. For each child, the absolute values of their track convergence scores were recorded and then averaged across each condition in which the child was tested. The individual data were then sorted from low to high according to this measure. Better performers therefore appear earlier in the figure, poorer performers, later in the figure. KW ID = "Keyword identification score"

Visual comparisons between the individual subject data from the normal-hearing children included for reference in Appendix D and the data shown in Figure 10 above from the hearing-impaired children suggest that only a few data sets from the implant group resemble the data collected from normal-hearing children. The data from child CI#09, age eight years, for example, is basically indistinguishable from that of a normal-hearing child. To a lesser degree the data from participants CI#01, CI#04, CI#05, and CI#12 also approximate the behavior shown by normal-hearing children. The majority of the remaining data sets in the cochlear implant group do not, however, resemble the individual data sets obtained from normal-hearing children.

The differences we observed among participants in the cochlear implant group suggested that it might be advisable to examine the data both in terms of the performance of the entire implant group, and also in terms of the performance of only those implanted children who performed the talker discrimination task at above chance levels. Therefore, for most of our analyses, data from the subset of eight hearing-impaired children who obtained track convergence scores in both the fixed and varied sentence conditions reflecting better than chance performance are presented separately, in addition to the data from the implant group considered as a whole.

Group Performance as a Function of Trial Number. In order to establish that the adaptive procedure was capable of quickly and efficiently determining the pitch difference for which the children were equally likely to respond "different talker" versus "same talker," we examined each group's performance as a function of trial number. Group mean performance on the talker discrimination task as a function of trial number is displayed in Figure 11 for each group of children. Each line graph represents an averaging of the individual subject data shown for the normal-hearing children in Appendix D and for the hearing-impaired children in Figure 10. Trial number is shown on the abscissa of each line graph; pitch difference size in Hertz is shown on the ordinate. Data from the "start-different" tracks using the higher-pitched standard are shown using dashed lines and triangles, data from the "start-same" tracks using the lower-pitched standard are shown using solid lines and circles. Results for the "Fixed" sentence condition are displayed on the left; results from the "Varied" sentence condition are displayed on the right.

For the normal-hearing children whose data is shown in panel (a) of Figure 11, the adaptive procedure employed in this study yielded a clearly stable pattern of performance in the group data after only approximately six trials. Midway through the trial set, the pitch differences arrived at using the two different types of tracks already show close agreement for the normal-hearing group of children. The normal-hearing children displayed a mean track convergence score of less than 1 Hz (-.72 Hz) in the fixed sentence condition and a mean track convergence score of -4.5 Hz in the varied sentence condition. Similarly, a group of 48 normal-hearing adults recently tested on the same task and whose data are reported in Reference Note 1 were found to have mean track convergence scores of essentially 0 Hz in the fixed sentence condition and -2.4 Hz in the varied sentence condition. These findings indicate that our normal-hearing child listeners were able to perform the talker discrimination task in the manner expected, and that both track-types resulted in essentially identical results when the pitch differences were expressed in linear frequency units rather than semitones.

The group data from the children with cochlear implants, in contrast, displayed relatively poor track convergence as shown in panel (c) of Figure 11. The data from the implant group resembles the pattern of performance predicted by chance responding as displayed in panel (d) of Figure 11. For the hearing-impaired children considered as a group, in the fixed sentence condition, their mean track convergence score of 22.6 Hz did not differ significantly from the value of 28 Hz predicted by chance responding ($p = .27$). In the varied sentence condition, the implant group's mean track convergence score of 15.1 Hz did differ significantly from that expected by chance responding ($p = .03$), but it was still quite poor, relative to the track convergence scores observed for normal-hearing listeners.

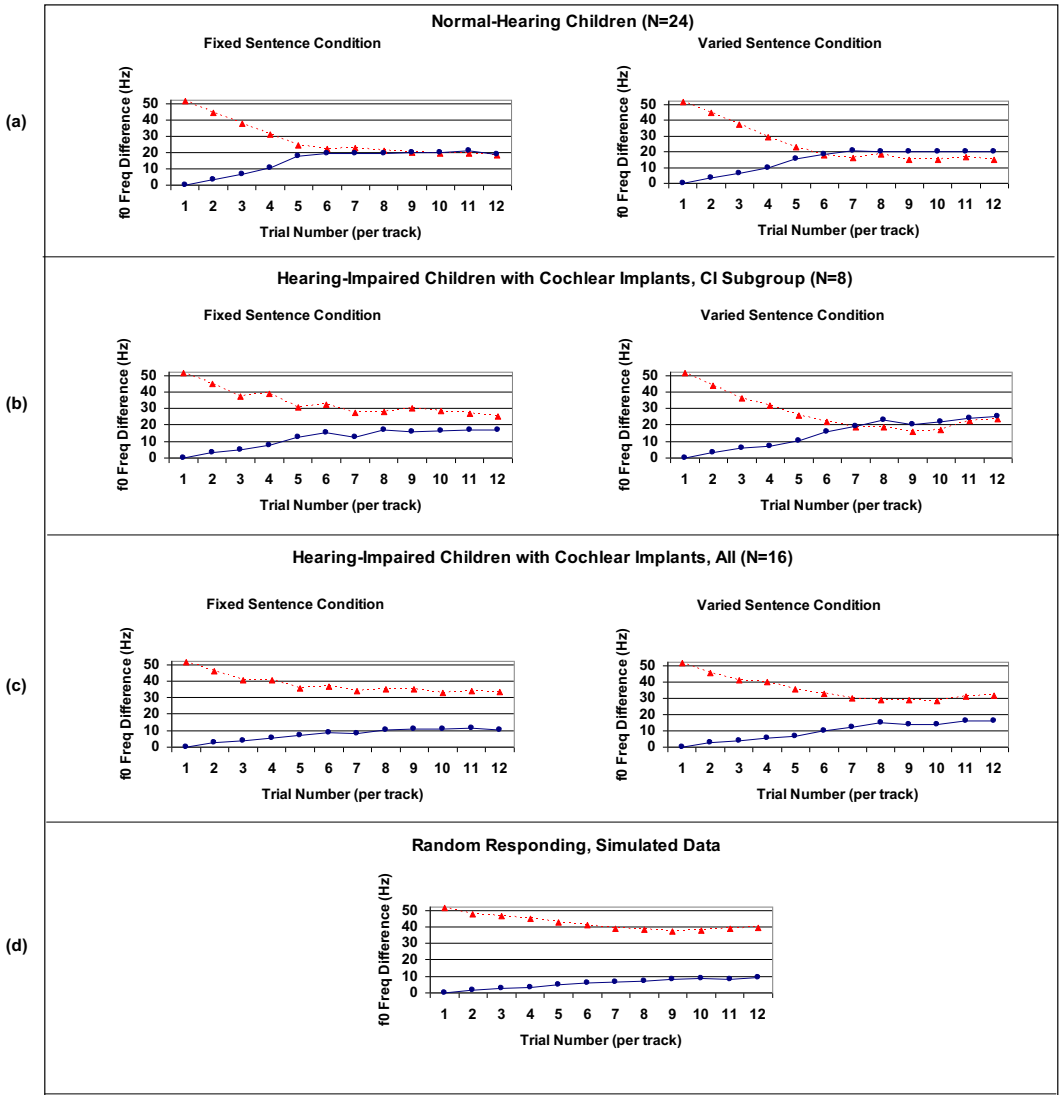


Figure 11. Average performance on the talker discrimination task as a function of trial number. The data from the normal-hearing children (*N*=24) are shown in (a). The data from the better-performing subgroup of hearing-impaired children with cochlear implants (*N*=8), are shown in (b). The performance of the full group of 16 children with cochlear implants is shown in (c). Results based on random responses from 24 simulated "participants" are shown in (d) to illustrate the average performance across trials expected by chance. Trial number is shown on the abscissa and the average pitch difference size tested on each trial number is shown in Hz on the ordinate. Start-Different tracks are shown using dashed lines and triangles. Start-Same tracks are shown using solid lines and circles.

We also examined the data from the subgroup of eight children with cochlear implants who were able to complete both conditions of the talker discrimination task and who showed evidence of being able to comply with task directions. For this group of hearing-impaired children, their mean track convergence score of 9.9 Hz in the fixed sentence condition and -2.5 Hz in the varied sentence condition both differed significantly from the values predicted from chance responding ($p < .01$). This result follows logically from the fact that the track convergence measures were the basis for selecting the individuals comprising this

subgroup. The better convergence between tracks for this subgroup of hearing-impaired children can be clearly seen in panel (b) of Figure 11. Somewhat contrary to expectation however, better track convergence was observed in the varied sentence condition than in the fixed sentence condition for this group of eight children with cochlear implants. In fact, the performance of this group across trials in the varied sentence condition quite closely resembled that of the normal-hearing children when examined in this manner.

In summary, for the normal-hearing children, once the differences were expressed in Hertz, the tracks using the two different standards converged on a single value, suggesting that a mean pitch difference of 20+ Hz has the effect of inducing perception of a different talker regardless of whether the voices involved have the frequency characteristics of male or female talkers. These results also suggest that for the normal-hearing participants it might be valid to combine the results using the two track types into a single composite dependent measure. For the hearing-impaired listeners however, a simple averaging of the two values obtained for the two different tracks would not, however, be appropriate, given the poor convergence between the two measures in this group. The group data shown in Figure 11 and the individual data reported in Figure 10 and Appendix D also suggest that the procedure includes sufficient trials to be able to assess a child's ability to do the task, and that within each track, a relatively stable level of performance (be it poor or high) is reached at or before the 10th trial. Using the average pitch difference size tested over the last three trials of each track as our primary "endpoint-based" dependent measure therefore appeared to be reasonable. Our statistical analyses were conducted using this endpoint-based measure which represents the pitch difference size for which a listener was equally likely to respond "same talker" or "different talker".

Statistical Analyses of Endpoint-Based Measures. Figure 12 displays the average pitch difference tested over the last three trials of each track type for each condition, track type, and listener group. Results for the fixed sentence condition are shown in the top panel. Results for the varied sentence condition are shown in the bottom panel. The bars plotted in this figure represent the pitch difference size for which listeners were equally likely to respond "same talker" or "different talker." Although Figure 12 is somewhat redundant with the data shown in Figure 11, Figure 12 specifically illustrates the endpoint-based dependent measure that was submitted for statistical analysis and also contains important information about within-group variability through the inclusion of standard deviation and standard error bars.

In order to judge whether parametric statistical tests could be applied, we first examined the distributional characteristics of the scores in the various experimental conditions for each group of children. The skewness and kurtosis of the scores from the normal-hearing group were between -1 and +1 for all cells in the experimental design. For the implant group, the skewness values fell between -1.0 and +1.04, and the kurtosis values between -1.32 and +1.0. These values were judged to meet the conventional criteria for the suitability of parametric tests.

Although our unequal and rather modest sample sizes require that these results be interpreted somewhat cautiously, an analysis of variance for a mixed factorial design was conducted on the endpoint-based dependent measure of pitch difference size expressed in terms of Hertz. The hearing-impaired children were treated as a single subject group for this analysis, with no distinction made for the subgroup of better-performing hearing-impaired children. The ANOVA indicated the presence of a significant main effect of the children's hearing status ($F(1,36) = 6.62$, $p = .01$), a significant main effect of track type ($F(1,36) = 14.05$, $p = .001$, and no significant main effect of presentation condition (fixed versus varied) ($F < 1$). Significant two-way interactions between hearing status and track type ($F(1,36) = 25.27$, $p < .001$) and between track type and presentation condition ($F(1,36) = 5.10$, $p = .03$) were also found. The two-way interaction between hearing status and presentation condition did not reach statistical significance ($p = .13$), nor did the three-way interaction between hearing status, track type, and presentation condition ($p = .50$).

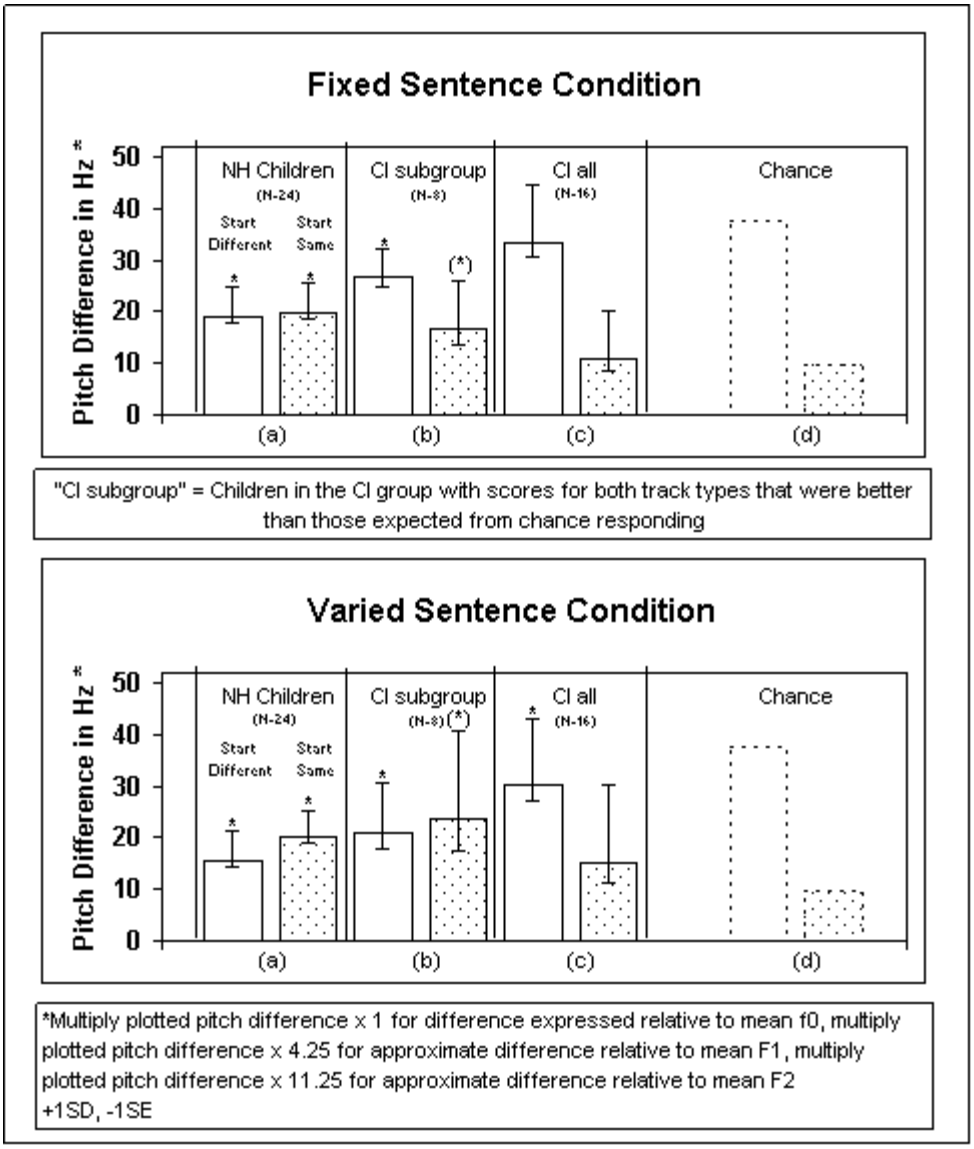


Figure 12. Group means, endpoint-based measure of discrimination performance. Means that differed significantly from chance performance are indicated by "*". Differences that reached marginal significance are indicated by "(*)".

The nature of these effects is best conveyed by describing the performance of each group of children in greater detail. For the normal-hearing five-year-olds, in the fixed sentence condition, the mean endpoint-based pitch difference in Hertz was 19.2 Hz (2.02 ST) when the high-pitched standard was used, and 19.9 Hz (2.57 ST), when the low-pitched standard was used. These means are shown in the upper panel (a) of Figure 12. Both means differed significantly from the values expected from chance responding ($p <$ .001).

The means observed in the fixed sentence condition for this group of normal-hearing five-year-old children can be directly compared to results recently obtained in our lab for a group of 48 normal-hearing

adults who completed the same task using the same stimulus materials and procedures. These adults required a 14.0 Hz difference for the same high-pitched standard, and a 13.9 Hz difference for the same low-pitched standard (see Reference Note 1). These data suggest that two voices that differ in mean pitch by more than 19.5 Hz will tend to be heard by normal-hearing children of this age as different talkers, whereas two voices differing in mean pitch by less than 19.5 Hz will tend to be heard as the same talker. Moreover, a comparison with the results reported for adult listeners suggests that children require a somewhat larger difference to favor the "different talker" response, on the order of about an additional 5.5 Hz.

For the varied sentence condition, the results from the normal-hearing children were slightly more variable: the mean difference in Hertz was 15.5 Hz (1.62 ST) when the higher-pitched standard was used, and 20.0 Hz (2.58 ST), when the lower-pitched standard was used. These results are shown in the lower panel (a) of Figure 12. Both values differed reliably from those expected by chance ($p < .001$). The results can be also be compared to data recently collected at our lab from normal-hearing adults who completed the same talker discrimination task in a varied sentence condition (see Reference Note 1). These adult listeners required mean differences that were somewhat smaller than those obtained under analogous conditions from the normal-hearing children, specifically, a 13.3 Hz difference for the same higher-pitched standard, and a 15.7 Hz difference for the same lower-pitched standard.

As shown in both Figure 11(a) and Figure 12(a), for the normal-hearing five-year-old children we observed better agreement between start-different and start-same tracks in the fixed sentence condition than in the varied sentence condition. From these endpoint-based measures there was no evidence, however, that the varied sentence condition was more difficult for the normal-hearing children than the fixed sentence condition; the children did not require a larger pitch difference to favor different talker responses in the varied sentence condition. Although for the normal-hearing children, the mean pitch difference for the start-different track type in the varied sentence condition was reliably lower than the other three means, none of the other pairwise comparisons between means reached significance. The lower mean obtained for the start-different track in the varied sentence condition was unexpected and may simply be a result of the particular sentence orderings used, rather than a theoretically important finding.

In general, the finding of similar levels of performance in both the fixed and varied sentence conditions replicates a pattern of results that we have obtained also for normal-hearing adults (see Reference Note 1). The results indicate that given the sentence-length stimuli used in the present study, normal-hearing listeners are able to follow the same response rule/strategy both in cases where the linguistic content of the sentence is held constant, and in cases where this content is allowed to vary.

Although the above results for the normal-hearing children have been presented in terms of the pitch difference size relative to the mean fundamental frequency of each standard, given the method by which the stimulus continuum was actually constructed, the obtained differences can also be expressed relative to mean formant frequency values. For example, in the general region of F1 (530-750 Hz), a 2 to 2.5 semitone difference corresponds to a difference of about 84 Hz, and in the general region of F2 (1414-2000 Hz), to a difference of about 223 Hz. As a practical matter, meaningful measures of mean formant frequencies averaged across sentence-length utterances are difficult to quantify, but the approximate values given here can provide a rough estimate of the formant frequency differences involved. The mean pitch differences plotted in Figure 12 can therefore be multiplied by 1 for the difference expressed relative to mean f0, by 4.25 for the approximate difference relative to mean F1, and by 11.25 for the approximate difference relative to mean F2. A better summary of the results in Figure 12 would be that normal-hearing children will favor "different talker" responses when the average spectral envelopes of two sentences, including mean fundamental frequency, differ by more than 11% to 16% (or 2.0-2.5 semitones), more accurately described along a linear scale as a ~19.5 Hz difference in mean f0, ~84 Hz in mean F1, and ~223 Hz in mean F2.

The endpoint-based mean differences for the hearing-impaired children with cochlear implants are also shown in Figure 12. The means for the entire group of 16 hearing-impaired children are shown in panel (c) of each figure. These averaged data closely resemble the data expected by chance responding as shown in the dashed-line bars of panel (d). The values obtained using the two different track types differ quite obviously from each other for this group of hearing-impaired children, as would be also be the case for chance responding. For the full set of 16 implant participants, only the mean of 30.5 Hz (3.35 ST) for the start-different track in the varied sentence condition was significantly different from the value expected by chance ($t(15) = 2.28$, $p = .038$). All three other mean scores failed to differ reliably from chance responding. Enormous within-group variability in performance was observed in this task for the children with cochlear implants, as shown by the large standard deviations (positive error bars) and standard errors (negative error bars).

For the subgroup of eight hearing-impaired children who met the minimal performance criteria already described, the mean pitch difference in the fixed sentence condition for start-different tracks using the higher-pitched standard was 26.7 Hz (2.87 ST). This value, shown in the upper panel (b) of Figure 12, was statistically better, even with this small sample size, from the value expected by chance ($t(7) = 5.58$, $p = .001$). For start-same tracks using the lower-pitched standard, the mean pitch difference was 16.8 Hz (2.17 ST). This difference was only marginally better than that predicted by chance ($t(7) = 2.19$, $p = .065$). The difference between the values obtained for the two different tracks in the fixed sentence condition was quite large (~ 10 Hz), and although there was better agreement observed here than in the larger full group of pediatric cochlear implant users, generalizing across the tracks by averaging the two means still does not appear appropriate.

As shown in the lower panel (b) of Figure 12, in the varied sentence condition, however, there was considerably better agreement for this subgroup of implanted children across the two different types of tracks. The mean pitch difference for start-different tracks was 21.1 Hz (2.25 ST), and for start-same tracks, 23.6 Hz (2.92 ST). The mean for the start-different track differed significantly from the value predicted by chance responding ($t(7) = 4.85$, $p = .002$) while the mean for start-same tracks reached marginal significance ($t(7) = 2.31$, $p = .054$). These values begin to resemble those obtained with the normal-hearing children, although the average pitch differences are still somewhat larger for the implant subgroup, on average, as might be expected. Expressed relative to an approximate mean F1, the 2.25 to 2.9 semitone difference corresponds to a difference of about 91 Hz, and relative to an approximate mean F2, to a difference of about 250 Hz.

Finally, although the children with cochlear implants demonstrated consistently larger standard deviations around the means in the varied sentence condition than in the fixed sentence condition, statistical comparisons of the group means for each track type between the fixed and varied sentence conditions revealed no significant effect of presentation condition for either the full group or smaller subgroup of hearing-impaired children.

General Observations. Because no implanted child under the age of 7;6 was clearly able to comply with task instructions, and because the mean age of the subset of hearing-impaired children with track convergence scores better than chance was 112 months (9;4), it is probably reasonable to conclude that pediatric cochlear implant users who are younger than about seven years of age typically do not yet have the cognitive skills needed to successfully complete a talker discrimination task of the type used here. Although the procedure used appears quite appropriate for five-year-old normal-hearing children, a different methodology is probably required to measure the talker discrimination skills of very young hearing-impaired children with cochlear implants.

Also, it should be kept in mind that given the present methodology it is logically impossible to rule out that some hearing-impaired children with cochlear implants might have better-than-normal ability to discriminate voices that have mean f0s falling between 125-135 Hz, but be unable to perceive even very large talker differences when these voices have mean f0s that fall between 135-175 Hz. Although the present methodology cannot exclude this possibility, this would seem to be highly unlikely. Additional data collection, reversing the use of the two standards in the two different track types could, however, be used to rule out this possibility.

Intercorrelations Among Tasks: Normal-Hearing & Cochlear Implant Groups

Scoring Adjustment. Because of the poor agreement between scores in the two different track types and the overall difficulty of the talker discrimination task for many of the children in the cochlear implant group, correlational analyses using the endpoint-based pitch differences scores in their existing form would have been inappropriate and uninformative. Instead, in order to capture the performance differences among children in the implant group that were clearly obvious in the individual data, a new score was tabulated that, in effect, "rewarded" cochlear implant users who showed good track convergence in their scores—i.e., evidence of being able follow task directions, even if their voice discrimination skills were poor. For each child, the endpoint-based scores for the two different track types were averaged, and then a "penalty" consisting of the subject's track convergence score was added to this average to form an "adjusted" score. Thus, random responders would receive the largest (poorest) adjusted scores, children requiring large pitch differences but who showed good track convergence would score moderately well, and finally, children with good discrimination abilities who also showed good track convergence would receive the lowest (best) adjusted scores.

Correlations between these adjusted talker discrimination scores and several spoken word and sentence recognition measures are shown Table 3. Correlations with the keyword identification scores collected in the present study are presented first in the table, followed by correlations with several standard clinical speech perception measures independently collected from the implanted group by researchers at DeVault Otologic Research Laboratory within 24 months of the present testing. Unless otherwise stated, chronological age in months was statistically partialled out of all correlations reported here.

In both listener groups, children with worse (larger) adjusted talker discrimination scores tended to do more poorly on the auditory keyword identification task. For the normal-hearing children, a significant negative correlation was only observed in the varied sentence condition. This was not altogether surprising because the normal-hearing children demonstrated a restricted range of scores on the keyword identification task, limiting the possibility of observing meaningful correlations because of a ceiling effect.

In the cochlear implant group, correlations between adjusted talker discrimination scores and keyword identification performance were observed in the expected direction for both the fixed and varied sentence conditions of the talker discrimination task, although only the correlation in the fixed sentence condition reached statistical significance. Children with better keyword identification scores tended to favor "different-talker" responses at smaller pitch differences. An examination of correlations between adjusted talker discrimination scores and performance on standard clinical speech perception measures showed a similar pattern: correlations with scores in the fixed sentence condition tended to be larger than those with scores in the varied sentence condition. However, none of these correlations reached statistical significance. The amount of intervening time between the administration of the talker discrimination task and the standard clinical measures for about half of the children may have reduced the size of the observed correlations.

TABLE 3

CORRELATIONS BETWEEN ADJUSTED TALKER DISCRIMINATION SCORES AND WORD AND SENTENCE IDENTIFICATION MEASURES

CORRELATIONS WITH ADJUSTED TALKER DISCRIMINATION SCORES (N = "AVAILABLE N")				
CHRONOLOGICAL AGE PARTIALLED OUT OF ALL CORRELATIONS	NH GROUP		CI GROUP	
	FIXED	VARIED	FIXED	VARIED
KEYWORD IDENTIFICATION SCORE	r = -.06 N = 24	r = -.41* N = 24	r = -.79*** N = 15	r = -.29 N = 15
PBK WORDS CORRECT, AUDITORY-ONLY, LIVE-VOICE	n/a	n/a	r = -.38 N = 12	r = -.19 N = 13
LNT-EASY WORDS CORRECT, RECORDED MULTI-TALKER	n/a	n/a	r = -.48 N = 12	r = -.03 N= 12
LNT-HARD WORDS CORRECT, RECORDED MULTI-TALKER	n/a	n/a	r = -.52 N = 12	r = -.06 N = 12
HINT-C KEY WORDS CORRECT IN QUIET, RECORDED	n/a	n/a	r = -.02 N = 8	r = -.10 N = 8
HINT-C KEY WORDS CORRECT IN NOISE, RECORDED	n/a	n/a	r = -.22 N = 7	r = -.20 N = 8
* p < .05, ** p < .01, *** p < .001				

Nevertheless, the general pattern of results for this group of pediatric cochlear implant users is similar to the talker discrimination findings reported earlier in Cleary (2003/Chapter IIA) that also showed stronger correlations with auditory word recognition when a fixed linguistic context was used in a talker discrimination task than when a varied sentence context was used. Therefore, although there is some evidence within the cochlear implant group that similar skills are necessary to extract linguistic as compared to indexical information from the speech signal, the size of the correlations suggests that the two different abilities do not entirely coincide. To the extent that the adjusted scores are truly capturing meaningful differences in performance among the hearing-impaired children, the correlational analyses suggest that performance on the varied sentence condition of the talker discrimination task is mediated by skills that are not typically tapped by the standard clinical word identification measures.

Correlations between adjusted talker discrimination scores and two potentially important demographic characteristics are shown for the implant group in Table 4.

TABLE 4

CORRELATIONS BETWEEN ADJUSTED SCORES IN THE TALKER DISCRIMINATION TASK AND PARTICIPANT CHARACTERISTICS FOR THE CHILDREN WITH COCHLEAR IMPLANTS.

CORRELATIONS WITH ADJUSTED TALKER DISCRIMINATION SCORES (PENALIZED SCORE), N = "AVAILABLE N"		
	CI GROUP	
CHRONOLOGICAL AGE PARTIALLED OUT OF ALL CORRELATIONS	FIXED	VARIED
AGE AT IMPLANTATION	r = +.35 N = 16	r = +.17 N = 16
DURATION OF CI USE	r = -.40 N = 16	r = -.16 N = 16

As shown in Table 4, the correlations calculated between performance in the talker discrimination task and age at implantation indicated a tendency for later-implanted children to do more poorly on the talker discrimination task. Greater implant experience as measured by duration of CI use was also correlated in the direction suggested by past research; longer durations of CI use were associated with better (smaller) scores on the talker discrimination task. None of these correlations were large enough to reach statistical significance given the small sample size, but they are consistent in direction with previous findings on the effects of age at implantation and duration of implant use.

Unfortunately, the variable of cochlear implant speech processing strategy was too highly correlated with chronological age at time of testing and amount of CI use to draw any conclusions regarding effects of individual processing strategies. The SPEAK users were on average 8.5 years old and had 5.8 years of CI experience, whereas the CIS users were on average 6.6 years old and had 4.1 years of implant experience. The children using the SPEAK speech processing strategy therefore tended, on average, to do better than the children using CIS, on both the keyword identification and talker discrimination tasks reported on in this study.

It was also not possible to study the effects of communication mode using this sample of children, because only two hearing-impaired children who were designated as total communicators volunteered for the study. An examination of the data from the two children who used total communication indicated that these particular children were quite successful implant users, even judged relative to the children who were primarily oral communicators. Both of the children who used total communication scored well above the group means for both the keyword identification and talker discrimination tasks. It might be noted, however, as mentioned previously, that neither of these two children was observed to produce any manual signs during his/her (albeit limited) spontaneous communication with the signing clinician who was present. Thus, these children may not be representative of the larger population of hearing-impaired children with cochlear implants who use total communication.

General Discussion

The present study examined the perception of talker similarity in normal-hearing five-year-old children and in hearing-impaired children with several years of experience using a cochlear implant. Using a novel variant on traditional adaptive staircase procedures, we measured how acoustically different, in terms of their average spectral characteristics, two sentences needed to be for the children to categorize these utterances as spoken by two different talkers. We found that normal-hearing five-year-old children required a frequency shift of at least 2 to 2.5 semitones to categorize a pair of voices as belonging to two different talkers. Differences of one semitone or less were nearly always heard as originating from the same talker, while differences of 3 semitones or more were almost always heard as indicating two different talkers. By comparing the data obtained for two different voice standards, we also found, however, that the logarithmically-defined semitone units used in stimulus set construction did not fit the observed pattern of results nearly as well as measures of frequency difference defined using a linear scale. The normal-hearing children in this study displayed essentially the same response pattern regardless of whether or not the sentences to be compared differed in their linguistic content, and therefore demonstrated no measurable difficulty generalizing their perceptual representations of a previously unfamiliar talker's voice across two linguistically different utterances.

Many of the pediatric cochlear implant users, in contrast, found the talker discrimination task very difficult under both the fixed and varied sentence conditions. Half of the hearing-impaired children, primarily the younger children in the group, responded no differently from chance performance on the talker discrimination task, despite every child having a minimum of two years of implant experience. The remaining children in the implant group displayed performance that approached that shown by the normal-hearing children, although there was considerable within-group variability. For the hearing-impaired children with cochlear implants, better performance on the talker discrimination task in the fixed sentence condition was modestly associated with higher scores on several different measures of spoken word recognition including the newly developed keyword identification task, indicating some degree of overlap in the perceptual skills required for spoken word recognition and talker discrimination.

The results of this investigation suggest that previously reported difficulties in talker discrimination by children who use cochlear implants (e.g. Cleary 2003/Chapter II) may partially be a consequence of an inability to accurately perceive and encode pitch and timbre-related cues to talker identity. Although cochlear implants allow many profoundly deaf children to acquire spoken language skills far beyond what use of conventional hearing aids would permit this population, it should emphasized that the basic auditory capabilities of these children are still quite atypical.

The difficulties encountered by the hearing-impaired children in the present study are perhaps less surprising if one considers the historical development of cochlear implants and how these devices have been designed specifically with the perception of linguistic contrasts in mind (Wilson, 2000). Cochlear implants are able to support perception of linguistically significant contrasts partly because phonemic/lexical information in the speech signal is extremely robust to spectral shifts and lack of fine spectral/harmonic detail (see Assmann & Summerfield, in press, for review). Studies with adult implant users have demonstrated, for example, relatively rapid adaptation to the "basally-shifted re-mapping" of the lower speech frequencies (e.g., Harnsberger, Svirsky, Kaiser, Pisoni, Wright, & Meyers, 2001). Still other research with normal-hearing listeners attending to speech passed through a cochlear implant simulator has shown that little additional benefit in speech intelligibility is gained by subdividing the signal for greater spectral resolution into more than six to eight frequency bandwidths (Dorman, Loizou, & Rainey, 1997; Shannon, Zeng, & Wygonski, 1998). It remains an empirical question, however, whether talker information should be as robust as linguistic information to these types of signal degradation.

Other recent findings provide additional clues suggesting that many cochlear implant users encounter significant problems in perceiving non-linguistic attributes of complex sounds. Unlike the

perception of talker attributes, music perception has attracted a small but sustained amount of attention in the cochlear implant research community. These reports indicate that the perception of musical pitch and timbre is quite atypical even among cochlear implant users who are relatively successful at perceiving linguistic speech information. Although most of this research has focused on hearing-impaired adults with cochlear implants, some unpublished recent developmental work suggests that prelingually-deafened early-implanted children also have markedly abnormal musical pitch, timbre, and melody perception (Gfeller, Mehr, Stordahl, & Tomblin, 2001 talk). This appears to be the case despite findings obtained via parental report, that the majority of children with cochlear implants voluntarily listen to music for enjoyment and some proportion (e.g., 10-20% of children) even participate in musical activities at school (Gfeller, Witt, Spencer, Stordahl, & Tomblin, 1998).

Gfeller and her colleagues have reported that recognition of familiar instrument timbres is markedly impaired in adults who use cochlear implants compared to normal-hearing non-musician listeners (Gfeller, Witt, Woodworth, Mehr, & Knutson, 2002). These authors have also found, however, that large individual differences exist among adult implant users on such timbre perception tasks. They report, for example, that one implanted individual who was determined to return to her job as a music professional and who had worked to reacquaint herself with the sounds of various instruments, did at least as well on an instrument identification test as some normal-hearing listeners. Considered together with the case of the one hearing-impaired child in the present study whose data appeared indistinguishable from those of normal-hearing children, these findings suggest that although cochlear implants may electrically code sufficient acoustic detail to capture spectral differences such as would distinguish the voices of different talkers, in order for this information to be effectively used by implant users, other as yet unidentified requirements need also to be met.

Characteristics of the children's compromised auditory pathways, such as degree of auditory nerve survival, may, for example, be contributing to the difficulties encountered by most of the implant users on our voice discrimination task. The possibility also needs to be considered that the average of five years of implant use exhibited by our sample of hearing-impaired children is still insufficient for adequate development of the relevant auditory processing skills, but that further improvement might be seen over time.

Five years of normal, uninterrupted, auditory experience since birth have evidently been sufficient, however, to provide the young normal-hearing children in the present study with a common set of expectations regarding the degree to which the voice of an individual talker typically varies in its average fundamental and formant frequencies. Our data clearly show that when voices differ from each other in their average fundamental and formant frequencies beyond 2 to 2.5 semitones, normal-hearing children perceive these voices as belonging to different talkers. How children acquire the knowledge that guides this behavior remains to be determined.

A large body of evidence suggests, however, that the perceptual learning responsible for this behavior begins very early in human development, probably during early infancy (DeCasper & Fifer, 1980; Miller, 1983; Mills & Melhuish, 1974). Although the human auditory system is capable of making basic auditory discriminations at birth (Werner & Bernstein, 2000), there is solid evidence for the important role played by experience in the formation of the complex auditory categories encountered in daily life (Dowling, 1999). Numerous studies have shown that infants and children become attuned (and dis-attuned) over time to the relevance of particular physical properties for the classification of stimulus inputs into behaviorally-relevant categories (see Aslin, Jusczyk, & Pisoni, 1998; Jusczyk, 1997). In light of their atypical history of auditory experience, including, quite possibly, less exposure to the voices of many different talkers, it is perhaps not surprising that prelingually-deaf children with cochlear implants find it difficult to perceive talker-voice categories.

In the present study, we have observed for two different listener groups, how acoustic similarity maps onto the perceived category membership of a voice. The perception of voice and talker similarity is indicated as a crucial area of research by a host of recent findings showing that recognition memory for spoken words in normal-hearing infants and adults is systematically influenced by the perceptual similarity between the voices heard during the word familiarization stage and the voices producing those same lexical items during the recognition test phase (Goldinger, 1996; Houston & Jusczyk, 2000). In general, these studies have not addressed in great detail why the particular voices adopted for use in the experiments were perceived by listeners as similar, and therefore, a clear statement of how the size of the effect might vary by actual acoustic distance has not been provided.

Because it is often difficult logistically to obtain precise measures of perceived similarity from young children, much of the previous research on the role of talker variability on word identification and recognition memory for words has relied on hypotheses generated from assessments of talker similarity gathered from adult, normal-hearing listeners (e.g., Houston & Jusczyk, 2000). However, it is risky to assume that children are attuned to the same acoustic properties of voices, to the same extent, as adults. Establishing the degree of sensitivity of normal-hearing children and children with cochlear implants to acoustic similarities between talkers seems like a natural first step before making strong claims about the effects of talker variability on linguistic perception in either of these populations. Basic assumptions about how well each of these information types are preserved in the signal and encoded by listeners still need to be verified.

A better understanding of how child listeners process voice differences may also prove useful in trying to explain the mechanisms that cause words heard as part of a list in which multiple talkers are included to be less accurately identified relative to when the same tokens are presented in single-talker lists (see Creelman, 1957; Mullennix, Pisoni, & Martin, 1989; Ryalls & Pisoni, 1997). Our results regarding the perception of talker similarity may additionally prove to be of some interest to researchers studying interactions between the perception of linguistic and indexical information in normal-hearing children and children with hearing impairments (e.g., Jerger et al., 1995). More specifically, although these studies have demonstrated that listeners' perceptual judgments regarding word identity are typically slowed by the presence of uncorrelated task-irrelevant variability in the voice dimension, how this slowing might be modulated by the degree of indexical variability in the to-be-ignored channel has yet to be examined.

The present research on vocal source attribution also addresses issues which are a mainstream concern in the study of vocal communication in animals. An active area of research in animal communication is whether and how individuals of a particular species establish the identity of con-specifics. Many studies have documented that acoustic dimensions could, in theory, serve as cues to individual identity (e.g., Suthers, 1994), but relatively few have also perceptually tested whether these cues elicit behavior in line with predictions. For some species, however, there is evidence of sensitivity to attributes of vocal productions determined by relatively stable physiological attributes of the communicator such as vocal tract size and shape (e.g., Rendall, Rodman, & Emond, 1996). For example, a recent study by Fitch and Kelley (2000) used playback techniques and a habituation paradigm to examine the whooping crane's perception of naturally recorded whooping crane contact calls, resynthesized with a 0%, -10%, and +10% shift in formant frequencies correlating with typical size differences in crane anatomy. Fitch and Kelley reported that, after habituation to the contact call of an unfamiliar crane, crane listeners reliably dishabituated to the spectrally-shifted resynthesized versions of the contact calls, thus demonstrating sensitivity to the formant frequency changes alone.

Unlike species that may rely primarily on an individual's vocal call or scent, humans generally rely heavily on visual appearance to establish identity in day-to-day activities (technologically sophisticated

methods aside). However, when available, vocal characteristics are often used by humans as another source of confirmatory evidence regarding identity. The reliability of these cues is not always ideal in that although vocal tract characteristics do not lend themselves particularly well to intentional disguise, they are quite susceptible to fairly radical alterations due to their role in respiration and food intake.

Nevertheless, as shown in the present study, normal-hearing children as young as five years of age display clear patterns of agreement regarding whether or not utterances sound as if they were produced by the same individual. Even when irrelevant linguistic variability is present in the signal, young normal-hearing children have little trouble making this type of perceptual judgment. Presumably their already ample history of experience with different talkers and different utterances has equipped these children with certain expectations about the form of the speech signal, more specifically, knowledge about spectral regularities that can potentially help them discriminate between within-talker variability and cross-talker variability. Though situation-based expectations and semantic factors such as message content undoubtedly influence how voices are identified and discriminated under more realistic circumstances, the present experimental methodology largely controlled these potential factors, allowing for a rigorous assessment of what these children were willing to infer strictly from the stimulus alone.

In summary, the goal of the present investigation was to determine how different, in terms of fundamental and formant frequencies, two sentences must be for normal-hearing children and children with cochlear implants to categorize these utterances as spoken by two different talkers. Additionally, this study examined whether these requirements change when the linguistic content of the paired sentences is allowed to vary as compared to when the linguistic content remains constant.

Our results suggest that normal-hearing children will perceive two otherwise identical voices as belonging to two different talkers when the voices differ in their average pitch and timbre characteristics by at least 11-16%, or 2.0-2.5 semitones, more accurately described along a linear scale as a ~19.5 Hz difference in mean f0, ~84 Hz in mean F1, ~223 Hz in mean F2. Compared with previously studied normal-hearing adults, normal-hearing children require a larger acoustic difference to categorize two voices as belonging to different talkers. Among the 18 children with cochlear implants tested, some of these children performed in a manner that was similar to normal-hearing children, and appeared, like the normal-hearing children, to require a difference of at least 2.5 semitones to perceive the sentences as spoken by different talkers. In general, however, the children in the implant group experienced much more difficulty with the voice discrimination task than normal-hearing children. About half of the cochlear implant group responded at chance levels.

Finally, we found that for both listener groups, similar response patterns were observed in both the fixed and varied sentence conditions, although somewhat more variability was observed in the varied sentence condition. The normal-hearing children's equally good performance in both the fixed and varied sentence conditions suggests that these children are able to quickly form representations of a previously unfamiliar talker's voice and can effectively use these representations to determine whether a novel utterance is spoken by that talker.

To better understand the significance of the present results for the perception of talker similarity by normal-hearing and hearing-impaired children, we will need to determine the degree to which each of the altered dimensions (f0, formant frequencies, and even individual formant frequencies) contributes to perceived similarity (see Reference Note 1; also Kuwabara & Takagi, 1991; Lavner, Gath, & Rosenhouse, 2000). Clearly, although spectral characteristics have been the main focus here, a variety of other acoustic cues are usually present in naturalistic circumstances to help children discriminate between talkers. It will be useful to study how well each of these different types of cues can be perceived by implant users and how they can be used in combination. Some of these, such as localization cues for sound sources, are already

being vigorously studied, as current implant designs and the impending move towards bilateral implantation are assessed (Tyler et al., 2002). Other potential cues, such as level differences between simultaneously-heard voices, need to be considered in light of processing manipulations such as automatic gain control often implemented on cochlear implant speech processors. In some current research, we are beginning to assess how talker similarity may interact with factors such as relative level, in determining pediatric cochlear implant users' ability to perceptually segregate the speech of different talkers (Reference Note 2).

Although further work clearly remains to be done to establish the degree of sensitivity of normal-hearing children and children with cochlear implants to acoustic differences between speakers, the present investigation provides some first insights into the talker discrimination skills of these children. In this work we were able to directly relate degree of acoustic similarity with degree of perceived voice similarity by exercising greater experimental control over the stimulus materials than in previous studies which have typically used a convenience sample of natural talkers. The results of the present study are useful in that they have yielded a particular value that can now be tested for quantitative accuracy using new talkers and other voice categorization paradigms.

References

Aslin, R.N., Jusczyk, P.W., & Pisoni, D.B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In W. Damon (Ed.), *Handbook of child psychology, fifth edition: Volume 2, cognition, perception, and language (pp.147-198).* New York: John Wiley & Sons.

Assmann, P.F. (1999). Fundamental frequency and the intelligibility of competing voices. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco, CA, August.

Assmann, P., & Summerfield, Q. (in press). Perception of speech under adverse acoustic conditions. In S. Greenberg, W. Ainsworth, A. Popper, & R. Fay (Eds.), *Springer handbook of auditory research: Speech processing in the auditory system.* (www .pdf file)

Bachorowski, J., & Owren, M.J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106,* 1054-1063.

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 27,* 464-472.

Bennett, S., & Montero-Diaz, L. (1982). Children's perception of speaker sex. *Journal of Phonetics, 10,* 113-121.

Bird, J. and Darwin, C.J. (1997). Effects of a difference in fundamental frequency in separating two sentences. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing, (pp.263-269).* London: Whurr.

Boersma, P. & Weenink, D. (2001). Praat Version 3.9.27: A system for doing phonetics by computer. www.praat.org.

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. Communication, 20, 255-272.

Bricker, P. & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America, 40,* 1441-1450.

Bricker, P.D. & Pruzansky, S. (1976). Speaker recognition. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics (pp. 295-326).* New York: Academic Press.

Carrell, T. (1984). Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. *Research on Speech Perception Technical Report No. 5.* Bloomington, IN, Speech Research Laboratory, Indiana University.

Charles-Luce, J., & Luce, P.A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language, 17,* 205-215.

Cleary, M. (2003/Chapter IIA). Preliminary studies: Talker discrimination in prelingually deaf children with cochlear implants. This volume.

Cleary, M. (2003/Chapter IIB). Preliminary studies: Talker discrimination in implanted children attending an Oral school for the hearing-impaired: A short report. This volume.

Coleman, R.O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research, 14,* 565-577.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29*, 655.

Daly-Jones, O., Monk, A., & Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: Fluency and awareness of attentional focus. *International Journal of Human-Computer Studies, 49,* 21-58.

DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174-1176.

Dedina, M.J. (1987). SAP: A speech acquisition program for the SRL-VAX. In *Research on Speech Perception Progress Report No. 13 (pp.331-337).* Bloomington, IN: Speech Research Laboratory, Indiana University.

Dorman, M.F., Loizou, P.S., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America, 102,* 2403-2411.

Dowling, W.J. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music (2nd ed.) (pp. 603-625).* San Diego, CA: Academic Press.

Dowling, W.J., & Harwood, D.L. (1986). *Music Cognition.* San Diego, CA: Academic Press.

Dunn, L.M., & Dunn, L.M. (1997). *Peabody Picture Vocabulary Test, Third Edition.* Circle Pines, Minnesota: American Guidance Service.

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.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague, Netherlands: Mouton.

Fant, G. (1973). A note on vocal tract size factors and nonuniform f-pattern scalings. In G. Fant, *Speech sounds and features (pp.84-93).* Cambridge, MA: MIT Press.

Fellowes, J.M., Remez, R.E. & Rubin, P.E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics, 59,* 839-849.

Fitch, W.T., & Kelley, J.P. (2000). Perception of vocal tract resonances by whooping cranes Grus Americana. *Ethology, 106,* 559-574.

Fleetwood, S. (1990). Habitual speaking fundamental frequency in percent of total phonational frequency range for normal young adults. Unpublished Master's Thesis, Indiana University, Bloomington, IN.

Gelnett, D., Sumida, A., Nilsson, M., & Soli, S.D. (1995). Development of the Hearing In Noise Test for Children (HINT-C). Paper presented at the American Academy of Audiology, Dallas, TX, April.

Gerstman, L.J. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics, AU-16, 1,* 78-80.

Gfeller, K., Mehr, M., Stordahl, J., & Tomblin, B. (2001). Talk presented at the 8th Symposium on Cochlear Implants in Children. Los Angeles, CA, February 28-March 3, 2001.

Gfeller, K., Witt, S.A., Spencer, L. J., Stordahl, J., & Tomblin, B. (1998). Musical involvement and enjoyment of children who use cochlear implants. *Volta Review, 100,* 213-234.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M.A., & Knutson J. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Annals of Otology, Rhinology & Laryngology, 111,* 349-356.

Goldinger, S.D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166-1183.

Haskins, H. (1949). A phonetically balanced test of speech discrimination for children. Unpublished master's thesis. Evanston, IL: Northwestern University.

Harnsberger, J.D., Svirsky, M.A., Kaiser, A.R., Pisoni, D.B., Wright, R., & Meyers, T.A. (2001). Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *Journal of the Acoustical Society of America, 109,* 2135-2145.

Hernandez, L.R. (1995). Current computer facilities in the Speech Research Laboratory. In *Research on Spoken Language Processing Progress Report No. 20 (pp.389-393).* Bloomington, IN: Speech Research Lab, Indiana University.

Houston, D.M., & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1570-1582.

Jerger, S., Martin, R., Pearson, D.A., & Dinh, T. (1995). Childhood hearing impairment: Auditory and linguistic interactions during multidimensional speech processing. *Journal of Speech & Hearing Research, 38,* 930-948.

Johnson, K., & Mullennix, J.W. (1997). *Talker Variability in Speech Processing.* San Diego, CA: Academic Press.

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 Otology, Rhinology, & Otolaryngology (Supplement), 166,* 363-365.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language.* Cambridge, MA: MIT Press.

Kawahara, H. (1998). *GUI-STRAIGHT: Getting started.* Draft of STRAIGHTV30kr16 documentation.

Kawahara, H., Masuda-Katsue, I., & de Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication, 27,* 187-207.

Kent, R. (1997). *The Speech Sciences.* San Diego, CA: Singular.

Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America, 110,* 2141-2155.

Kewley-Port, D, & Watson, C.S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America, 95,* 485-496.

Kirk, K.I., Houston, D.M., Pisoni, D.B. Sprunger, A. & Kim-Lee, Y. (2002). Talker discrimination and spoken word recognition by adults with cochlear implants. Poster presented at ARO, Florida.

Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech & Hearing Research, 40,* 1395-1405.

Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16,* 470-481.

Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87,* 820-857.

Kramer, E. (1963). Judgment 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.

Kreiman, J., & Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication, 10,* 265-275.

Kuwabara, H., & Takagi, T. (1991). Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method. *Speech Communication, 10,* 491-495.

Ladefoged, P. & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98-104.

Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication, 30,* 9-26.

Lemmetty, S. (1999). Review of speech synthesis technology. Unpublished Masters Thesis, Helsinki, Finland: Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467-477.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Liu, C. & Kewley-Port, D. (submitted). Vowel formant discrimination in natural speech. *Journal of Acoustical Society of America.*

Locke, J. L. (1993). *The Child's Path to Spoken Language.* Cambridge, MA: Harvard University Press.

Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39,* 155-158.

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1-36.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27,* 153-165.

Matsumoto, H., Hiki, S., Sone, T., & Nimura, T. (1973). Multidimensional representation of personal quality of vowels and its acoustic correlates. *IEEE Transactions on Audio and Electroacoustics, AU-21,* 428-436.

Miller, C.L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development, 6,* 313-330.

Mills, M., & Melhuish, E. (1974). Recognition of mother's voice in early infancy. *Nature, 252,* 123-124.

Miyamoto, R.T., Kirk, K.I., Svirsky, M.A. & Sehgal, S.T. (1999). Communication skills in pediatric cochlear implant recipients. *Acta Oto-Laryngologica, 119,* 219-224.

Moore, B.C.J. (1997). *An Introduction to the Psychology of Hearing, Fourth Edition.* San Diego, CA: Academic Press.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453-467.

Mullennix, J.W., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47,* 379-390.

Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365-378.

Murry, T., & Cort, S. (1971). Aural identification of children's voices. *Journal of Auditory Research, 11,* 260-262.

Murry, T., & Singh, S. (1980). Multidimensional scaling analysis of male and female voices. *Journal of the Acoustical Society of America, 68,* 1294-1300.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60,* 355-376.

Orlikoff, R.F., & Kahane, J.C. (1996). Structure and function of the larynx. In N.J. Lass (Ed.), *Principles of Experimental Phonetics (pp.112-181).* St. Louis, MO: Mosby.

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 & Hearing, 12(Supplement),* 66S-80S.

Peters, R.W. (1954). *Studies in extra messages: Listener identification of speakers' voices under conditions of certain restrictions imposed upon the voice signal (Joint Project Report No. 30, Project No. NM 001-064-01).* Pensacola, FL: Naval School of Aviation Medicine, Naval Air Station.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America, 24,* 175-184.

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, CA: Academic Press.

Pisoni, D.B., Cleary, M., Geers, A., & Tobey, E. (2000). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *Volta Review, 101,* 111-164.

Pollack, I., Pickett, J.M., & Sumby, W.H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26,* 403-406.

Rendall, C.A., Rodman, P.S., & Emond, R.E. (1996). Vocal recognition of individuals and kin in free ranging rhesus monkeys. *Animal Behaviour, 51,* 1007-115.

Robbins, A.M, Svirsky, M.A., & Miyamoto, R.T. (2000). Aspects of linguistic development affected by cochlear implantation. In S.B. Waltzman & N.L. Cohen (Eds.), *Cochlear Implants (pp. 284-287).* New York, NY: Thieme.

Ryalls, B.O., & Pisoni, D.B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33,* 441-452.

Schroeder M. (1993). A brief history of synthetic speech. *Speech Communication, 13,* 231-237.

Seligman, P., & McDermott, H. (1995). Architecture of the Spectra 22 speech processor. *Annals of Otology, Rhinology, and Laryngology, Supplement, 166,* 139-141.

Shankweiler, D., Strange, W., & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing (pp. 315-335).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shannon, R.V., Zeng, F.G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104,* 2467-2476.

Skinner, M.W., Arndt, P.L., & Staller, S.J. (2002). Nucleus 24 advanced encoder conversion study: Performance versus preference. *Ear & Hearing, 23(Supplement),* 2S-17S.

Skinner, M.W., Holden, L.K., Whitford, L.A., Plant, K.L, Psarros, C., & Holden, T.A. (2002). Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear & Hearing, 23,* 207-223.

Sommers, M.S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *Journal of the Acoustical Society of America, 101,* 2278-2288.

Spence, M.J., Rollins, P.R., & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language, and Hearing Research, 45,* 214-222.

Staller, S.J., Dowell, R.C., Beiter, A.L., & Brimacombe, J.A. (1991). Perceptual abilities of children with the Nucleus 22-channel cochlear implant. *Ear & Hearing, 12(Supplement),* 34S-47S.

Stevens, K.N. (1998). *Acoustic Phonetics.* Cambridge, MA: MIT Press.

Stevens, S.S. (1975, reprinted 1986). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* Piscataway, NJ: Transaction Publishers.

Suthers, R.A. (1994). Variable asymmetry and resonance in the avian vocal tract: A structural basis for individually distinct vocalizations. *Journal of Comparative Physiology, 175,* 457-66.

Svirsky, M.A., Robbins, A.M., Kirk, K.I., Pisoni, D.B., & Miyamoto, R.T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science, 11,* 153-158.

Svirsky, M.A., Stallings, L.M., Lento, C.L., Ying, E.Y., & Leonard, L.B. (2002). Grammatical morphological development in pediatric cochlear implant users may be affected by the perceptual prominence of the relevant markers. *Annals of Otology, Rhinology, & Laryngology, Supplement: Proceedings of the 8th Symposium on Cochlear Implants in Children, March 2001, 111 (5-Supplement),* 109-112.

Tyler, R.S. (1993). Speech perception by children. In R.S. Tyler (Ed.), *Cochlear Implants: Audiological Foundations (pp.191-256).* San Diego, CA: Singular.

Tyler, R.S., Gantz, B.J., Rubinstein, J.T., Wilson, B.S., Parkinson, A.J., Wolaver, A., Preece, J.P., Will, S., & Lowder, M.W. (2002). Three-month results with bilateral cochlear implants. *Ear & Hearing, 23 (Supplement),* 80S-89S.

Vandali, A.E., Whitford, L.A., Plant, K.L., & Clarke, G.M. (2000). Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system. *Ear & Hearing, 21,* 608-624.

Voiers, W.D. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America, 36,* 1065-1073.

Werner, L.A., & Bernstein, I.L. (2000). Development of the auditory, gustatory, olfactory, and somatosensory systems. In E.B. Goldstein (Ed.), *Handbook of Perception (pp.669-708).* Oxford, UK: Blackwell.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84,* 963-976.

Wilson, B.S. (2000). Strategies for representing speech information with cochlear implants. In J. Niparko et al. (Eds.), *Cochlear Implants: Principles and Practices (pp.129-170).* Philadelphia, PA: Lippincott Williams & Wilkins.

Reference Notes

1. Cleary, M. (2003). Influence of voice similarity on talker discrimination in normal-hearing adults. Unpublished Manuscript.
2. Cleary, M. (in prep). Perception of speech with a competing talker masker by normal-hearing children, normal-hearing adults, and hearing-impaired children with cochlear implants.

APPENDIX A. FREQUENCY COUNTS FOR KEYWORDS USED IN STIMULUS MATERIALS, AS BASED ON SELECTED CHILDES CORPORA

Frequency counts were compiled using the set of English language corpora listed in the CHILDES documentation as of October 2001. Forty-four sets of corpora, all involving normally-developing children are listed in this section. Eliminated were corpora which contained speech from non-American families, and which involved primarily infant-directed speech. Also eliminated were corpora consisting primarily of a case study of only a single child, or which lacked sufficient documentation. Corpora consisting of data collected prior to 1970 were also eliminated. The eight remaining corpora, listed below, were used for the frequency counts:

1) Garvey, 1986, N=49, ages 2;10-5;7, data collected 1970's
2) Gleason, 1988, N=24, ages 2;1-5;2, data collected mid-1970's
3) Hall, 1984, N=39, ages 4;6-5;0, data collected late 1970's-early 80's
4) Morisset, no date, N=263, ages 2;6-3;6 , data collected mid-1980's.
5) New England (Pan & Snow), no date, N=52, ages 1;2-2;8, data collected late-1980's
6) Van Houten, no date, N=~50, ages 2;0-3;7, data collected early-mid-1980's
7) Van Kleeck, no date, N=37, ages ~3;0-3;11, early 1980's (?)
8) Warren-Leubecker, no date, N=15 (20 minus 5 > 5;7), ages 1;6-5;3, data collected early 1980's

Frequency counts were based only on the children's productive utterances. Counts were based on orthographic matching. Each word had to appear exactly as typed (not within a word or with affixes). Part of speech was not checked, hence words with multiple uses have combined counts for the various uses—i.e., "cook" or "tape" as a verb versus as a noun. Counts for such items may be inflated on this basis. Because plurals were not accepted as instances, some counts are probably lower than would be expected for words that are more often used in the plural—i.e., "egg." One "plural" was retained, "stairs"—otherwise, all words were singular.

Practice Items:

Word	Occurrences
ant	9
bat	28
boat	105
book	308
bow	13
bulb	1
cap	16
coin	0, 2 coins
dance	10
dress	67
fridge	0
game	137
hat	303
hose	6
lamb	11
log	1
man	650
mask	16

net	3
nut	15
peach	3
pear	19
pie	42
pin	9
roof	26
shell	5
smoke	20
square	56
stage	0
stove	61
thief	1
watch	454
worm	27
wrench	36
yard	16
yarn	0

Test Items:

Word	Occurrences
ape	5
bag	109
ball	492
bath	48
beach	38
bear	97
bed	440
belt	48
bench	52
bike	45
bird	122
block	87
bone	72
boot	15
bowl	42
box	277
boy	411
bread	76
bridge	18
broom	25
brush	90
bug	48
bus	86
cage	36
cake	138
car	731
card	46
cat	339
cave	6
chair	368
cheese	97
chick	14
clock	33
cloud	1
clown	56
coat	91
comb	53
cook	155
corn	55
couch	30
cow	76
crow	5
crown	3
deer	5
desk	6
dog	216
doll	73
door	246
drink	252
drum	4
duck	167
egg	72
farm	40
fence	28
fish	174
flag	4
fork	87
fox	16
frog	70
fruit	32
ghost	8
gift	2
girl	224
glass	48
glove	9
glue	26
goat	10
grape	35
grass	27
heart	39
hen	3
hill	10
horse	95
house	600
key	26
king	25
kite	12
knife	125
lamp	16
leaf	4
map	14
milk	307
moon	47
mouse	93
nest	1
nurse	32
paint	52
pan	39
pen	66
phone	61
pig	91

plane	30
plant	17
plate	110
pond	2
pool	31
purse	41
queen	10
rat	17
ring	62
rock	83
rope	21
rose	4
rug	15
salt	35
scarf	6
school	429
shark	42
sheep	13
shelf	14
ship	19
shirt	178
shoe	95
sink	29
skate	14
skirt	9
skunk	2
slide	160

snake	172
snow	75
soap	39
sock	21
soup	68
spoon	132
stairs	22
stamp	8
star	26
stick	58
store	154
straw	8
street	111
sun	39
swan	1
swing	37
sword	1
tape	207
tent	5
tie	67
toy	163
train	128
tree	115
truck	282
wolf	26
zoo	49

APPENDIX B. SENTENCE SET USED IN STIMULUS MATERIALS

Sentences used in Talker Discrimination Task

Practice Items:

1	The	bat	and the	thief	are by the	net
2	The	cap	and the	bulb	are by the	pear
3	The	game	and the	peach	are by the	boat
4	The	pin	and the	bow	are by the	mask
5	The	smoke	and the	nut	are by the	hose
6	The	watch	and the	shell	are by the	log
7	The	worm	and the	lamb	are by the	yard
8	The	pie	and the	hat	are by the	roof**
9x	The	coin*	and the	book	are by the	stove
10x	The	man	and the	ant	are by the	stage*
11x	The	dress	and the	square	are by the	fridge*
12x	The	wrench	and the	yarn*	are by the	dance

* - did not occur in frequency count
** - wide local regional variation in pronunciation
x - recorded but did not use in present testing

Test Items:

	The	NOUN	and the	NOUN	are by the	NOUN
1	The	ape	and the	swan	are by the	paint
2	The	bear	and the	king	are by the	truck
3	The	bird	and the	deer	are by the	beach
4	The	boy	and the	ghost	are by the	swing
5	The	bug	and the	fish	are by the	car
6	The	cat	and the	pig	are by the	store
7	The	chick	and the	rat	are by the	bridge
8	The	clown	and the	frog	are by the	street
9	The	cook	and the	skunk	are by the	grass
10	The	cow	and the	nurse	are by the	bed
11	The	dog	and the	crow	are by the	zoo
12	The	girl	and the	fox	are by the	house
13	The	goat	and the	shark	are by the	broom
14	The	hen	and the	queen	are by the	tent
15	The	horse	and the	doll	are by the	hill
16	The	mouse	and the	snake	are by the	slide
17	The	nest	and the	wolf	are by the	pool
18	The	sheep	and the	duck	are by the	rock
19	The	bag	and the	soap	are by the	lamp
20	The	block	and the	shirt	are by the	corn
21	The	boot	and the	tape	are by the	sink
22	The	bowl	and the	comb	are by the	stairs

23	The	box	and the	plate	are by the	ship
24	The	cage	and the	cheese	are by the	phone
25	The	cake	and the	shoe	are by the	flag
26	The	coat	and the	spoon	are by the	fence
27	The	card	and the	grape	are by the	skate
28	The	crown	and the	bread	are by the	sun
29	The	drink	and the	glue	are by the	clock
30	The	egg	and the	rope	are by the	train
31	The	fork	and the	cloud	are by the	pond
32	The	fruit	and the	map	are by the	bike
33	The	gift	and the	bone	are by the	school
34	The	glove	and the	soup	are by the	stick
35	The	key	and the	drum	are by the	shelf
36	The	milk	and the	belt	are by the	snow
37	The	pan	and the	chair	are by the	bath
38	The	pen	and the	moon	are by the	door
39	The	kite	and the	tie	are by the	sword
40	The	plane	and the	brush	are by the	couch
41	The	plant	and the	star	are by the	bench
42	The	ring	and the	ball	are by the	tree
43	The	rose	and the	purse	are by the	farm
44	The	skirt	and the	glass	are by the	desk
45	The	sock	and the	knife	are by the	bus
46	The	stamp	and the	leaf	are by the	rug
47	The	straw	and the	heart	are by the	scarf
48	The	toy	and the	salt	are by the	cave

APPENDIX C. INSTRUCTIONS TO CHILD PARTICIPANTS

Instructions for Talker Discrimination:

"In a minute, we will use the computer, but before we start, I need to make sure that you understand what I mean when I ask you if two things are the SAME or if they are DIFFERENT. See these two cards? One of these cards has two things that are the SAME. One of these cards has two things are DIFFERENT. Which card has two things that are the SAME?" (Allow child to respond.) "Which card has two things that are DIFFERENT?" (Allow child to respond.) (If child does not properly label as same or different, experimenter may decide to drop the session, or decide to go to Part 2.)

"Good. The next thing we're going to do has to do with listening to people's VOICES. I need you to tell me if the voices you're going to hear are the SAME, or if they are DIFFERENT. Let me explain."

"For example, if you had your eyes closed and I said (NAME)! and then your mom(/dad) said (NAME)! you could figure out when it was your mom(/dad) and when it was me. You could tell, because our voices sound different. I don't sound the same as your mom(/dad)."

"So, during this part, you're going to hear some people saying sentences. First I'll play one sentence and you need to listen and remember that voice. Then I'll play you ANOTHER sentence. If you think that both sentences sound like the SAME person talking both times (show two fingers, aligned), I want you to say SAME. If the two sentences sound like two DIFFERENT PEOPLE talking (two fingers not aligned) then I want you to say DIFFERENT."

"Let's practice so you get the idea."

FOUR PRACTICE TRIALS

First: "Did you hear that? That was the SAME! That was the same lady speaking both sentences!" "So, look, we press this button here that looks like this (press SAME button). See, this is a special screen--this is like a button! Let's do the next one."

Second: "That was DIFFERENT. Did you hear? How it was a high lady's voice and then a deep (imitate) man's voice? So now we press the DIFFERENT button (show)—go ahead, press the button! (The pictures labeling the buttons are just like the cards the child sees at the beginning.) …..Okay, why don't you do the next one by yourself?"

Third:

Fourth:

(If child cannot do these last two correctly, OK to repeat the set of four practice trials up to two more times. Then, go on to test trials.)

"Okay, this is the real thing now. Remember, your job is to listen to the two sentences. Pay attention to the voice speaking each sentence. Decide if both sentences sound like the same person talking. Or if the two sentences sound more like two different people. Then press the button on the screen that matches what you decide. Are you ready?"

Prompts:
"You need to do this by yourself. If you're not sure, make your best guess."
"Go ahead, press the button."

The other condition (counterbalanced in order of administration):

"Now, we're going to do the same thing, except that, you know how when you did it before, the words in the sentences (were always the same/changed all the time)?…Well, this time, the words in the sentences (are going to change/are going to stay the same), but THIS ISN'T IMPORTANT! Your job, just like before is to pay attention to the VOICES of the people talking. You need to decide if both sentences sound like the same person talking. Or if the two sentences sound more like two different people---just like you were doing before. Let's practice!"

First: "Did you hear that? That was the SAME! That was the same lady speaking both sentences!" Let's do the next one."

Second: "That was DIFFERENT. Did you hear? How it was a high lady's voice and then a deep (imitate) man's voice? So now we press DIFFERENT. Okay, why don't you do the next one by yourself."

Third:

Fourth:

(If child cannot do these last two correctly, OK to repeat the set of four practice trials up to two more times. Then, go on to test trials.)

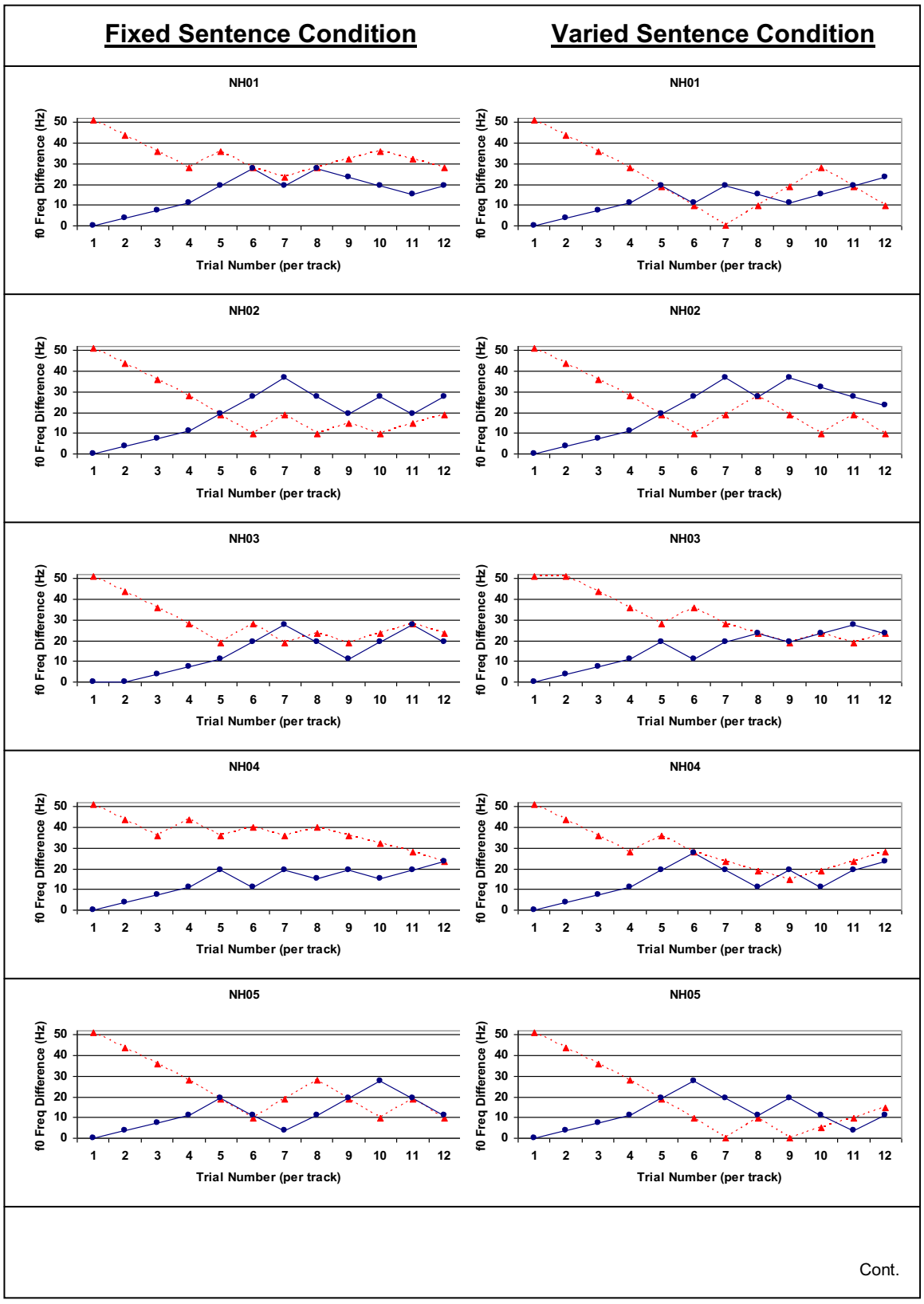
Instructions for Keyword Identification:

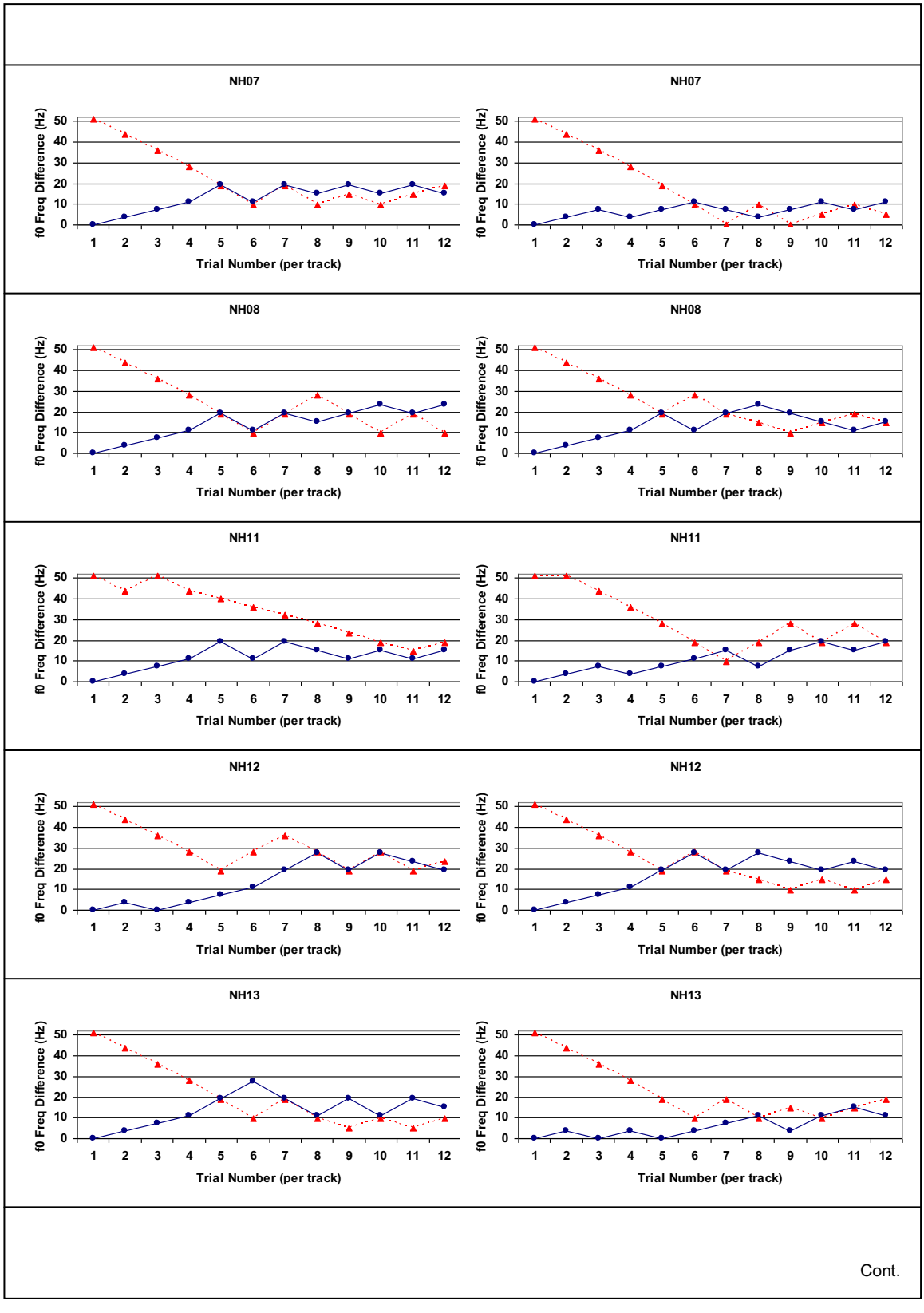
"Next you will hear some more sentences and see some pictures. This time I DO want you to listen to the WORDS in each sentence because after each sentence is played, I want you to point and touch the three pictures that match the words you hear in the sentence. The first word will be one of the pictures in the first column over on this side of the screen (show). The second word will be one of the pictures in the second column here in the middle of the screen (show). The last word will be somewhere in the last column here on this side of the screen. A red line will appear around each picture you choose. We will practice this so that you get used to how it works. (After you press your three choices I will play the next sentence.) If you choose something by mistake, it is okay to go ahead and press a different choice. The computer will remember your new choice. It is important that you press your choices just with one finger and not with your whole hand. Any questions? Ready?"

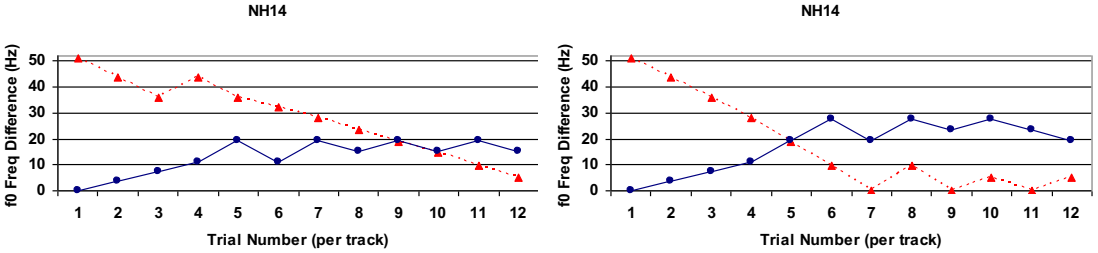
"Okay, good. Now let's do the real thing. Remember, your job is to listen carefully and find the pictures that match the sentence."

Other: "Press the buttons carefully—use just one finger!" "Okay, let go of the button!"

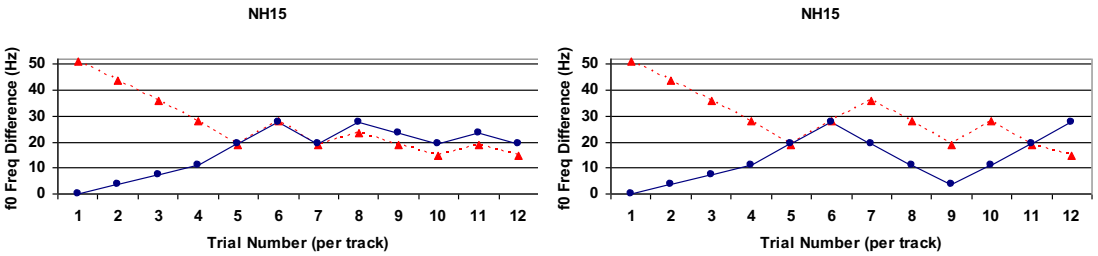
APPENDIX D. INDIVIDUAL SUBJECT DATA, NORMAL-HEARING CHILDREN



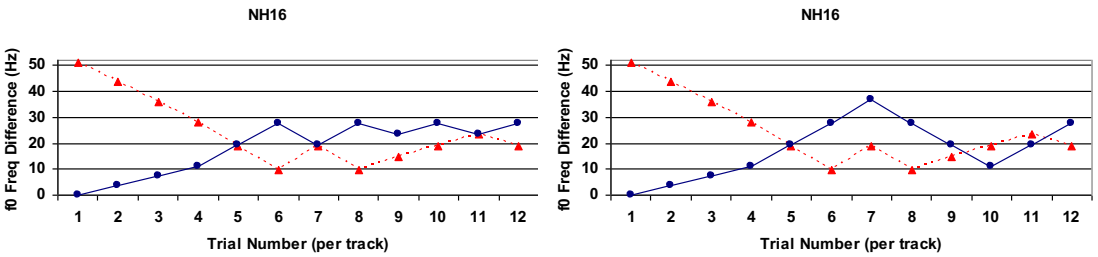

NH07
NH07
NH08
NH08
NH11
NH11
NH12
NH12
NH13
NH13
f0 Freq Difference (Hz)
Trial Number (per track)
0
10
20
30
40
50
1
2
3
4
5
6
7
8
9
10
11
12
Cont.


NH14
f0 Freq Difference (Hz)
Trial Number (per track)

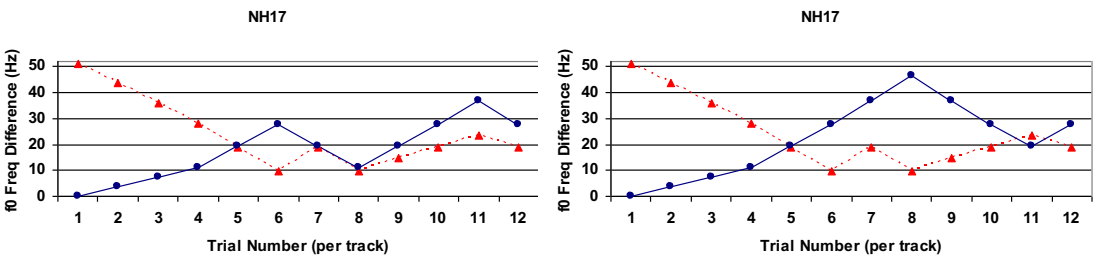
NH14
f0 Freq Difference (Hz)
Trial Number (per track)


NH15
f0 Freq Difference (Hz)
Trial Number (per track)

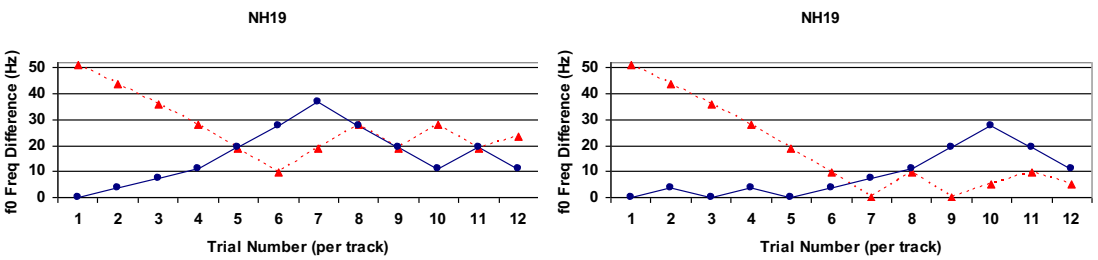
NH15
f0 Freq Difference (Hz)
Trial Number (per track)


NH16
f0 Freq Difference (Hz)
Trial Number (per track)

NH16
f0 Freq Difference (Hz)
Trial Number (per track)

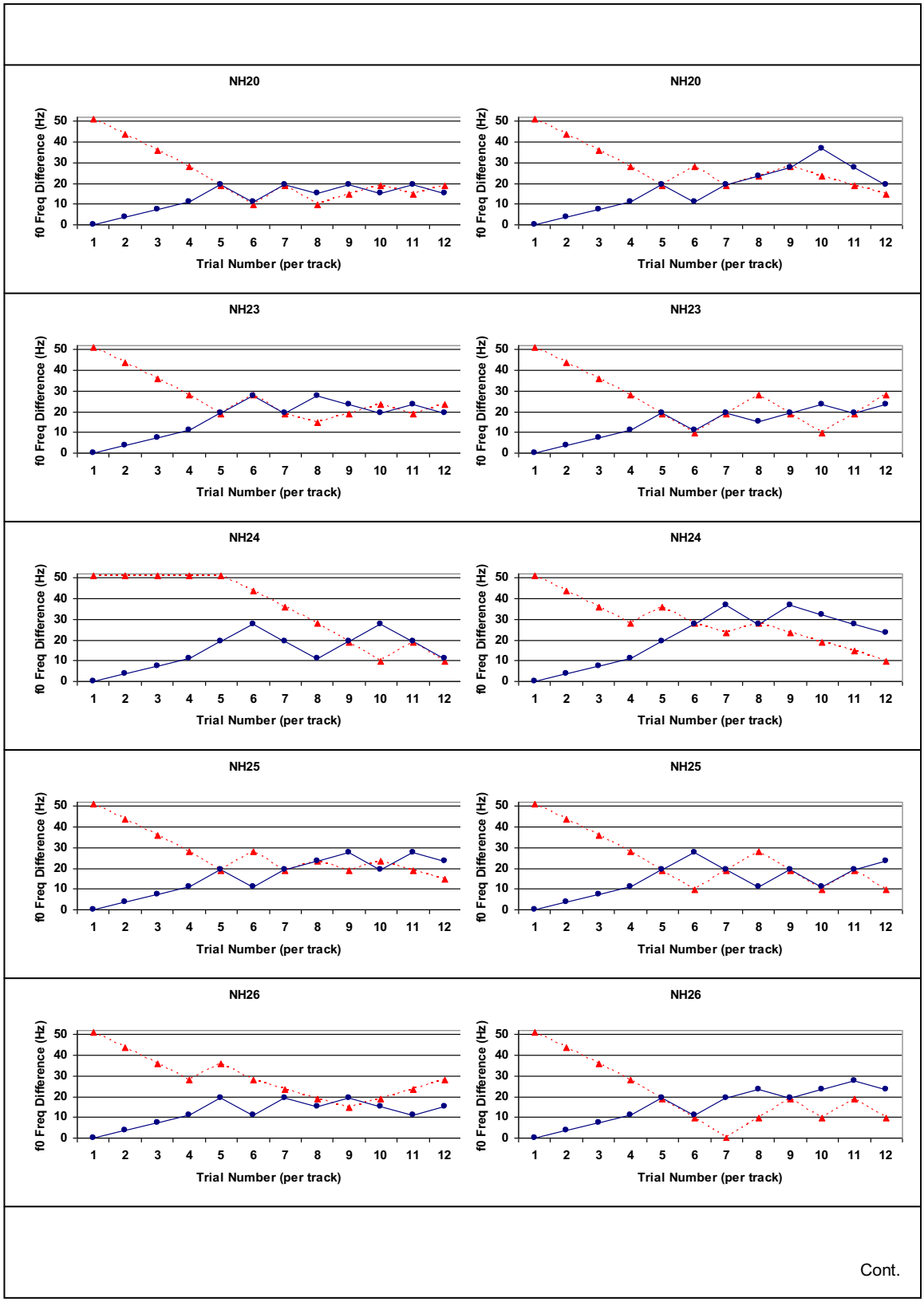

NH17
f0 Freq Difference (Hz)
Trial Number (per track)

NH17
f0 Freq Difference (Hz)
Trial Number (per track)

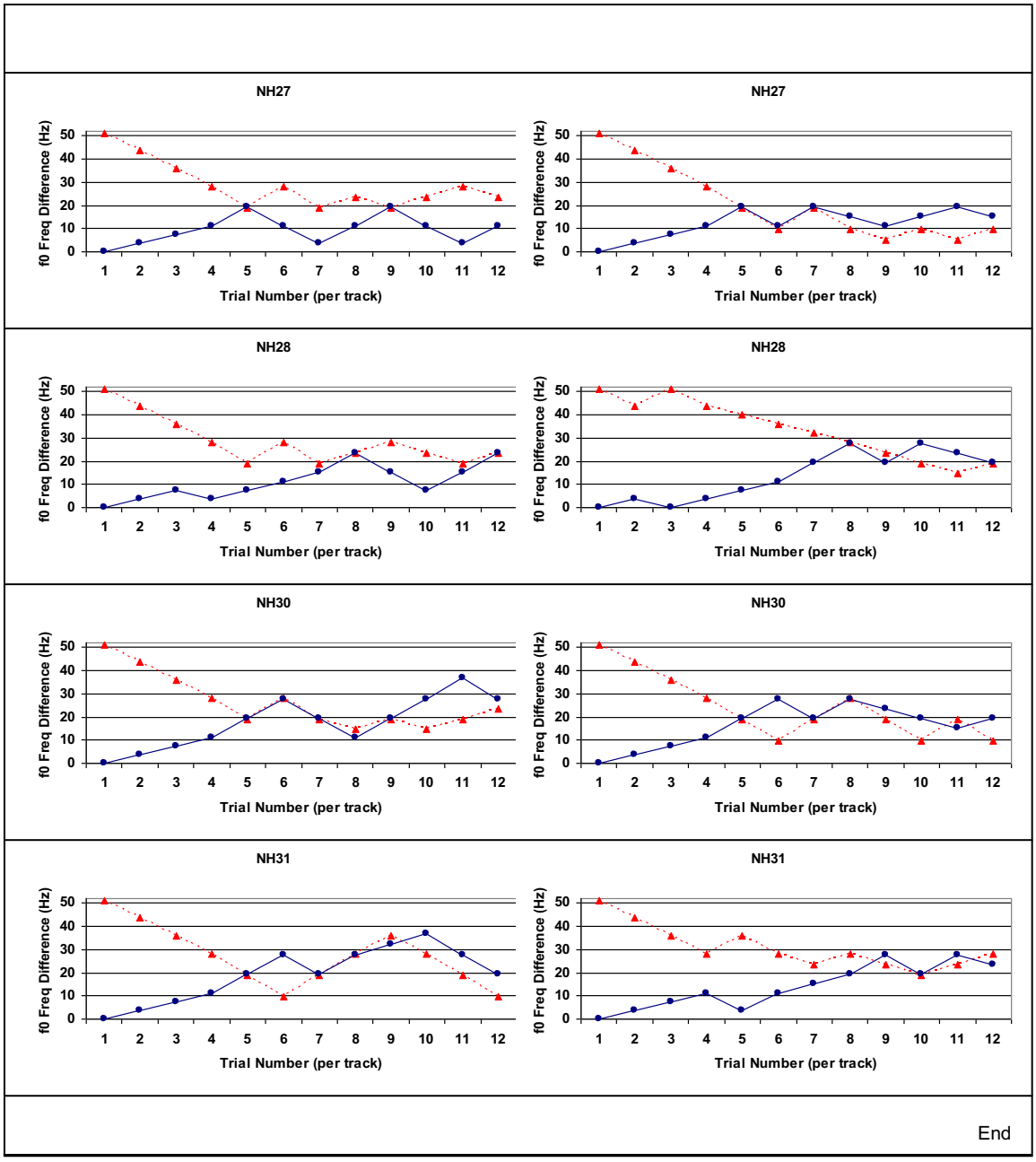

NH19
f0 Freq Difference (Hz)
Trial Number (per track)

NH19
f0 Freq Difference (Hz)
Trial Number (per track)

Cont.


NH20
NH20
NH23
NH23
NH24
NH24
NH25
NH25
NH26
NH26
f0 Freq Difference (Hz)
Trial Number (per track)
0
10
20
30
40
50
1
2
3
4
5
6
7
8
9
10
11
12

Cont.


NH27
f0 Freq Difference (Hz)
Trial Number (per track)
NH27
f0 Freq Difference (Hz)
Trial Number (per track)
NH28
f0 Freq Difference (Hz)
Trial Number (per track)
NH28
f0 Freq Difference (Hz)
Trial Number (per track)
NH30
f0 Freq Difference (Hz)
Trial Number (per track)
NH30
f0 Freq Difference (Hz)
Trial Number (per track)
NH31
f0 Freq Difference (Hz)
Trial Number (per track)
NH31
f0 Freq Difference (Hz)
Trial Number (per track)
End

CHAPTER IV - SUMMARY AND CONCLUDING REMARKS

The present project examined the ability of normal-hearing children and children with cochlear implants to discriminate acoustic differences associated with cross-talker variability. In our preliminary studies, we found that although normal-hearing children were quite adept at discriminating natural talkers, hearing-impaired children with cochlear implants experienced much more difficulty with the same task. Our main experiment was therefore designed to examine one possible source of the difficulties encountered by children with cochlear implants, namely, the ability to discriminate differences in the average spectral content of different voices. The methodology used in this experiment involved systematically manipulating average fundamental and formant frequencies to create a stimulus continuum of similar-sounding voices and then measuring precisely how different two voices needed to be along these co-varied acoustic dimensions, for the children to perceive otherwise identical voices as belonging to different talkers.

In describing the present research, selecting the most appropriate terminology to refer to the perceptual phenomenon of interest has sometimes been difficult because the talker discrimination task employed here can be conceptualized in several different ways. When recordings from actual talkers are used as the stimuli, "correct" and "incorrect" responses are clearly defined and the task can unequivocally be described as measuring talker discrimination. From a somewhat different perspective, however, the task of asking listeners to decide if two utterances were spoken by same actual talker can also be viewed as a means of obtaining pairwise similarity judgments using a two-point ratings scale. Under this view, incorrect responses of "same talker" for talker pairs that actually differ can be taken as a metric of how similar-sounding the talkers are perceived to be. Incorrect responses of "different" for two different tokens recorded from the same talker may also be taken to reflect the ability of listeners to generalize their representations of a talker's voice across utterances, thereby ignoring within-talker variability.

When resynthesized speech is used instead of natural speech of real talkers, the relationships among test stimuli are much simpler, but the talker discrimination task itself is more complex conceptually. In using resynthesized tokens created from the recorded tokens of a natural talker, we have been careful in the present study to include large frequency manipulations that naïve normal-hearing listeners reliably perceive as indicating two different talkers. We have also included small frequency shifts that listeners do not perceive as sufficient to indicate a change in talker. The data from the talker discrimination task using the resynthesized speech tokens have then been analyzed in terms of the proportion of "different talker" responses obtained as a function of the degree of acoustic dissimilarity. Performance for this form of the talker discrimination task cannot be analyzed in terms of percent correct because such a measure would be meaningless. We have, however, analyzed the performance of a clinical population of listeners in terms of how closely their data resemble data collected from normal-hearing listeners, and in terms of how their response patterns differ from those predicted by chance.

Utilizing a resynthesized speech version of a talker discrimination task has enabled us to examine the criteria by which listeners assess whether or not two voice samples were produced by the same talker. We have studied how listeners perceive the natural boundaries on talker-voice categories imposed by a talker's vocal tract characteristics by manipulating the acoustic correlates of vocal tract size and assessing how this affects the perception of talker voice individuality. Controlled manipulation of mean fundamental and formant frequencies has allowed us to study how much a voice sample can be altered before it no longer sounds as if it was produced by the original talker.

The published literature on normal-hearing children's perception of voices is relatively small. Moreover, previous developmental studies on this issue have not focused on the exact nature of the acoustic similarities among talkers used in particular experiments. Therefore, we really know very little about how normal-hearing children perceive talker similarity. The present set of findings offer new and specific

predictions regarding the degree to which natural voices must differ for normal-hearing children to perceive a change in talker. The value of 2 to 2.5 semitones as the frequency shift difference that must be exceeded for children to reliably discriminate between talkers based on fundamental and formant frequencies is a measurement whose generalizability can be directly checked using other talkers and speech materials.

Our findings obtained from pediatric cochlear implant users show that it is difficult for children with implants to discriminate between and recognize the voices of previously unfamiliar natural talkers. Additionally, the results suggest that poor discrimination of mean fundamental and formant frequencies (the primary acoustic correlates of voice pitch and voice timbre), may be an important factor contributing to implanted children's difficulty with talker discrimination tasks using natural speakers. In normal listening environments, it is likely that other factors such as differences in average speaking rate, and overall amplitude/loudness patterns, also influence how accurately children with cochlear implants are able to perceive differences among talkers. However, in the present study, we chose to focus on mean fundamental and formant frequencies because variation in these dimensions is sufficient alone to powerfully convey, to a normal-hearing listener, the perception of hearing different talkers.

In may be of interest to note that nearly all of the parents whose implanted children participated in the study described in Chapter III reported, in a written questionnaire, that their child was able to distinguish between the voices of different talkers. Virtually every parent stated that their child was almost always able to identify who was speaking when the voice was of a familiar family member, and was usually able to tell apart the voices of unfamiliar talkers. What was obvious in conducting all of the testing sessions, however, was that nearly every child with a cochlear implant found the same-different talker discrimination task to be difficult. Observing the responses of these children, especially on the trials involving the largest and smallest frequency shifts, was a very different experience than watching a younger normal-hearing child perform the same task. To the degree to which the studies can be directly compared, our experiences with the implanted children studied in Chapter III matched the level of performance we observed in testing the children with cochlear implants who participated in our preliminary studies using natural talkers.

The striking mismatch between the parental reports and observed behavior may result from the hearing-impaired children making effective use in day-to-day situations of visual and spatial cues for who is speaking, thereby making difficulties in perceiving acoustic cues to talker identity, non-apparent to parents. Alternatively, perhaps once a hearing-impaired child is extensively familiarized with a particular set of talkers, his/her voice discrimination skills appear more normal. Further studies would need to be undertaken to ascertain if this is the case, and more basically, whether the parents' reports are a reliable assessment of their children's perceptual abilities.

In short, we have found that normal-hearing children have well-defined perceptual criteria regarding the amount of within-talker variability in mean fundamental and mean formant frequencies that is typical of an individual talker. Most children with cochlear implants, on the other hand, seem to lack these clear talker category boundaries and experience difficulty making these same discriminations. A few implanted children, however, do show patterns of performance resembling those of normal-hearing children, suggesting that the design of the cochlear implant itself may not be responsible for these findings. Factors such as auditory nerve survival, and the nature of their language environment after implantation, may instead be responsible for the difficulties most pediatric cochlear implant users encounter in discriminating between talkers.

Our results on the effects of linguistic variability on talker discrimination are somewhat mixed. In our initial experiments using natural speech, we observed that some children with implants had difficulty recognizing an individual talker in the varied sentence condition across two linguistically different utterances. However, in subsequent experiments using both natural and resynthesized speech, we no longer

observed this particular difficulty. Instead, we found rather unequivocally that normal-hearing children and hearing-impaired children with cochlear implants each displayed the same pattern of performance regardless of the presence or absence of linguistic variability.

Although for the implant group the failure to find any differences might be seen as somewhat surprising given their poor performance reported in Chapter IIA on a varied sentence condition, a direct comparison is complicated by differences in the participant populations and methodologies used in the experiments. For the normal-hearing children, their equivalent performance in the fixed and varied sentence conditions of the talker discrimination task using resynthesized speech is less surprising in light of their high level of accuracy in discriminating between natural talkers under a varied sentence condition. Normal-hearing children may display good and consistent performance even when the sentence content is varied because these listening conditions bear a close resemblance to the natural circumstances under which they normally process the voices of talkers they hear in their environment.

The present line of research could be expanded in several new directions. The talker discrimination task developed in Chapter III should, for example, be repeated using recordings from other "base" talkers. To facilitate such a replication, recordings of these same sentences have been collected from nineteen additional female talkers in the course of preparing for the present study. We have experimented informally with a handful of other talkers and expect that discrimination functions will be very similar regardless of the base talker. Also available are multiple tokens of each sentence that may be useful in testing predictions regarding the degree to which listeners store veridical episodic memory traces of all detail in the acoustic waveform when listening to speech.

Also informative might be a series of follow-up experiments varying the time delay between sentences to be compared so that the comparisons between voices would entail more reliance by the listener on a somewhat more abstract memory representation than on what may be an initial, relatively unprocessed sensory representation of the speech tokens. It also might be fruitful to study other populations of hearing-impaired listeners, such as adults with cochlear implants and hearing-impaired children who use hearing-aids. Comparing the performance of postlingually-deafened adults who use cochlear implants with the performance of prelingually-deafened children like those included in the present study might help establish how early auditory experience influences talker discrimination. Children who use hearing-aids to correct less severe forms of hearing-impairment, involving specific speech frequency ranges, might also be a population that could provide additional insight into the nature of the perceptual difficulties encountered by children with cochlear implants.

As mentioned briefly in Chapter III, we have already conducted two experiments in normal-hearing adult listeners that are closely related to the present project (see Reference Note 1). In the first experiment, we tested 48 normal-hearing adults using the same stimuli and procedures that were used with the children studied in Chapter III. We found that adult listeners perceived two different talkers when the voices had their spectral characteristics including fundamental frequency shifted by at least 8-12%, corresponding to a difference of ~14 Hz in mean f0, ~60 Hz in the mean F1 range, and ~160 Hz in the mean F2 range. Differences that did not exceed these values were perceived as acceptable "within-talker" variability. The adult listeners in this new study were therefore found to require smaller differences to perceive two voices as originating from different talkers than the normal-hearing children reported on in Chapter III of the present paper. Like the normal-hearing children, normal-hearing adults displayed the same pattern of responses in both the "fixed sentence" and "varied sentence" conditions, indicating no effect of linguistic variability on listeners' criteria for identifying two talkers as the same or different.

Our second experiment with normal-hearing adult listeners was designed to address the issue of whether the same-different talker judgments elicited from listeners in the first experiment were influenced

primarily by differences in fundamental frequency, differences in formant frequencies, or by differences in both dimensions equally. In this second study, talker discrimination judgments were collected under three different conditions using three different stimulus continua. In one condition, both fundamental and formant frequencies were manipulated, as before. In a second condition, only formant frequencies were manipulated, and in a third, only fundamental frequency was manipulated. Examining the data of 12 participants who completed all three conditions, we found that although formant frequencies were found to have a stronger influence on talker discrimination judgments than fundamental frequency, changes in formant frequencies alone did not influence listeners' judgments as strongly as changes made along both dimensions together. It would be useful to see if these findings can be replicated in normal-hearing children.

Although our talker discrimination task served its intended purpose well for the normal-hearing five-year-old children included in the present study, the results reported in Chapter III for the children with cochlear implants suggest that in future work we may want to revise our current procedure or explore other types of explicit voice discrimination / monitoring procedures in order to make these perceptual judgments easier for hearing-impaired children. Although the parameters of the resynthesized speech stimulus continuum used in Chapter III were selected on the basis of prior empirical findings, incorporating even larger average spectral differences among the test stimuli would, in the future, help to separate out issues of poor discrimination skills versus an inability to comply with task instructions. It might also be useful to see if the performance of the children with cochlear implants can be improved by multiple presentations of a standard stimulus item before listening to the comparison utterance, or by multiple presentations of the test pair of utterances, before a response is allowed. Providing the hearing-impaired children with additional repeated exposure to the test voices without changing the actual amount of talker information contained in the signal might help to rule out whether lapses in attention are contributing to their impaired performance relative to that of normal-hearing children. An alternative task format might involve asking children with cochlear implants to simply monitor a running stream of speech, presented as a narrative, and to hit a button or raise their hand whenever they notice a change in talker.

Another possibility might be to develop a testing procedure that does not explicitly require the hearing-impaired child to attend to acoustic properties of the voice. For example, a child with a cochlear implant could be asked to perform another information processing task such as word identification, while the characteristics of the voices producing these tokens were adaptively varied based on a measure of performance on the primary task. We have already begun some work in this vein. In a recent study (Reference Note 2), we created test trials in which two recorded sentences were presented simultaneously to children who had been cued to attend to one of these sentences and to identify certain keywords contained in that sentence. If successful keyword identification was observed on a given trial, the voices uttering the competing sentences were then made more similar to each other for the next trial. Previously published findings by a number of different researchers suggest that intelligibility of the keywords should be negatively affected under conditions of increased voice similarity (e.g., Assmann, 1999; Brokx & Nooteboom, 1982; Brungart, 2001). In our preliminary study, we found that for children with cochlear implants, poorer talker discrimination performance was modestly associated with poorer performance in keyword identification under conditions of competing talkers. This same relationship was not observed, however, in data collected from normal-hearing children.

As indicated by the direction of these new projects, we believe the perception of talker similarity may be important to understanding the difficulties in speech perception experienced by normal-hearing and hearing-impaired listeners when multiple voice sources are simultaneously present in the speech signal. Investigating the perception of talker similarity may also help to explain the small but measurable decrements in word identification accuracy observed when frequent but non-overlapping talker shifts occur while listening to speech (e.g., Creelman, 1957; Mullennix, 1997; Ryalls & Pisoni, 1997). The nature of the so-called "perceptual adjustment" mechanism responsible for the reduced perceptual identification

accuracy of words presented in the context of multiple-talker word lists might be better understood if, for example, the size of this "multiple talker effect" is modulated by the degree of similarity among talkers.

The issue of talker and voice similarity has also come to play an important role in current theoretical accounts of how prior experience with spoken word exemplars from multiple talkers may serve to build the mental lexicons of young children (Jusczyk, 1997). A better grasp of how listeners process variability in the speech signal related to inter-talker vocal tract differences may also help to solve the longstanding question of how adult listeners are able to maintain stable speech sound categories which transcend the huge acoustic differences that often exist between different talkers' productions of the same phonemic contrast (Pisoni & Lively, 1995).

In conclusion, the present set of findings shows that although young normal-hearing children are quite good at discriminating between talkers based on sentence-length utterances, most prelingually-deafened children who are acquiring spoken language using a cochlear implant find this same task very difficult. Furthermore, by using a talker discrimination task in which voice similarity was systematically varied, we have been able to demonstrate that normal-hearing preschool-age children maintain the perception of talker-voice individuality over a range of acoustic variability in mean fundamental and formant frequencies, and that acoustic differences that exceed this well-defined range reliably induce the perception of different talkers. Our results on talker perception provide a principled basis for making future predictions regarding the degree to which natural speakers must differ in terms of fundamental and formant frequencies for these sources of variation to affect talker discrimination in normal-hearing children. Finally, our results suggest that poor discrimination of pitch and timbre-based voice differences is a factor contributing to the perceptual difficulties in talker discrimination observed in hearing-impaired children with cochlear implants.

References

Assmann, P.F. (1999). Fundamental frequency and the intelligibility of competing voices. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco, CA, August.

Brokx, J.P.L., & Nooteboom, S.G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23-36.

Brungart, D.S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109,* 1101-1109.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29,* 655.

Jusczyk, P.W. (1997). *The Discovery of Spoken Language.* Cambridge, MA: MIT Press.

Mullennix, J.W. (1997). On the nature of the perceptual adjustments to voice. In K. Johnson & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing (pp. 67-84).* San Diego, CA: Academic Press.

Pisoni, D.B. & Lively, S.E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech Perception and Linguistic Experience (pp. 433-459).* Timonium, MD: York Press.

Ryalls, B.O., & Pisoni, D.B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33,* 441-452.

Reference Notes

1. Cleary, M. (2003). Influence of voice similarity on talker discrimination in normal-hearing adults. Unpublished Manuscript.
2. Cleary, M. (in prep). Perception of speech with a competing talker masker by normal-hearing children, normal-hearing adults, and hearing-impaired children with cochlear implants.