Children With Cochlear Implants Recognize Their Mother's Voice

Tara Vongpaisal,[1] Sandra E. Trehub,[1] E. Glenn Schellenberg,[1] Pascal van Lieshout,[2] and Blake C. Papsin[3]

Objectives: The available research indicates that cochlear implant (CI) users have difficulty in differentiating talkers, especially those of the same gender. The goal of this study was to determine whether child CI users could differentiate talkers under favorable stimulus and task conditions. We predicted that the use of a highly familiar voice, full sentences, and a game-like task with feedback would lead to higher performance levels than those achieved in previous studies of talker identification in CI users.

Design: In experiment 1, 21 CI users aged 4.8 to 14.3 yrs and 16 normal-hearing (NH) 5-yr-old children were required to differentiate their mother's scripted utterances from those of an unfamiliar man, woman, and girl in a four-alternative forced-choice task with feedback. In one condition, the utterances incorporated natural prosodic variations. In another condition, nonmaternal talkers imitated the prosody of each maternal utterance. In experiment 2, 19 of the child CI users and 11 of the NH children from experiment 1 returned on a subsequent occasion to participate in a task that required them to differentiate their mother's utterances from those of unfamiliar women in a two-alternative forced-choice task with feedback. Again, one condition had natural prosodic variations and another had maternal imitations.

Results: Child CI users in experiment 1 succeeded in differentiating their mother's utterances from those of a man, woman, and girl. Their performance was poorer than the performance of younger NH children, which was at ceiling. Child CI users' performance was better in the context of natural prosodic variations than in the context of imitations of maternal prosody. Child CI users in experiment 2 differentiated their mother's utterances from that of other women, and they also performed better on naturally varying samples than on imitations.

Conclusions: We attribute child CI users' success on talker differentiation, even on same-gender differentiation, to their use of two types of temporal cues: variations in consonant and vowel articulation and variations in speaking rate. Moreover, we contend that child CI users' differentiation of speakers was facilitated by long-term familiarity with their mother's voice.

(Ear & Hearing 2010;31;555–566)

INTRODUCTION

This study examined whether deaf children with cochlear implants (CIs) can recognize their mother's voice, presumably the voice they had heard most frequently. The research was motivated both by the prevailing evidence of poor voice recognition (Cleary & Pisoni 2002; Fu et al. 2004, 2005; Cleary et al. 2005) and by the absence of research on familiar voices among CI users.

Traditionally, speech recognition and talker recognition are viewed as entirely separate domains, with attention focused largely on the former. Mutually exclusive sets of cues in the acoustic spectrum, linguistic and paralinguistic, are thought to make independent contributions to the processing of speech and vocal identity (Klatt 1987). Forensic concerns (e.g., ear-witness testimony) have sparked renewed interest in the critical cues for talker identification (Read & Craik 1995), but these cues remain elusive. Nevertheless, interactive aspects of speech and voice processing have become apparent. For example, the adverse consequences of talker variation (Ryalls & Pisoni 1997) and the favorable consequences of talker familiarity (Nygaard et al. 1994) on speech processing imply that linguistic and indexical (i.e., talker specific) cues are perceived and stored together (Yonan & Sommers 2000). The emerging consensus is that talker-specific information is available at the segmental level, specifically, in the timing of consonant and vowel production, otherwise known as dynamic articulatory cues inherent to speech gestures (Remez et al. 1997; Sheffert et al. 2002).

Even when reduced spectral cues eliminate natural voice quality, similar to that in the case of sine-wave analogues of speech, listeners can still identify specific talkers, albeit at modest levels (Remez et al. 1997; Sheffert et al. 2002). The stimulus synthesis scheme in these studies is similar to signal-processing schemes used with conventional CIs. Cochlear prostheses restore auditory sensations to individuals with profound sensorineural hearing loss by direct stimulation of the auditory nerve. The signal-processing schemes preserve spectrotemporal features that are critical for speech recognition, but these are presented in an alternative carrier signal that lacks the temporal fine structure found in natural voices. Although many congenitally deaf children acquire good speech perception and production skills on the basis of input from their implant (Svirsky et al. 2000; Geers 2006), they have difficulty perceiving melody (Vongpaisal et al. 2006, 2009), talker identity (Cleary & Pisoni 2002), and speech in noise (Schafer & Thibodeau 2006).

The phonetic code that is rooted in the dynamics of speech production seems to provide cues that serve talker as well as speech recognition. For example, talker identification is optimal when talkers and listeners share the same language (Perrachione & Wong 2007), perhaps because of listeners' sensitivity to subtle variations in phoneme articulation. At times, talker identification is possible with an unfamiliar language (Lander et al. 2007), which may implicate the role of global cues such as intonation, rhythm, and speaking rate. Interestingly, practice in identifying talkers from spectrally degraded speech has positive transfer to subsequent decoding of spectrally degraded speech (Loebach et al. 2008), which attests to the tight coupling of speech and voice processing. In short, familiarity with a language facilitates talker identifica-

Departments of [1]Psychology and [2]Speech-Language Pathology, University of Toronto; and [3]Department of Otolaryngology, The Hospital for Sick Children, Toronto, Ontario, Canada.
Additional material related to this article can be found at http://links.lww.com/EANDH.

0196/0202/10/3104-0555/0 • Ear & Hearing • Copyright © 2010 by Lippincott Williams & Wilkins • Printed in the U.S.A.

tion, and familiarity with a talker facilitates speech recognition (Nygaard & Pisoni 1998; Perrachione & Wong 2007).

Familiar talkers, mothers in particular, are of critical importance for young children. For example, the maternal voice enhances speech processing in infancy (Barker & Newman 2004). Although hearing newborns differentiate their mother's voice from other voices (DeCasper & Fifer 1980), it is unclear whether congenitally or prelingually deaf children who are successful CI users can do so after implantation. The signal processing strategies used with CIs, which result in degraded spectral information, pose particular difficulty for talker differentiation, especially when differences in fundamental and formant frequencies are modest among talkers (Cleary & Pisoni 2002; Fu et al. 2004; Cleary et al. 2005). Nevertheless, temporal pitch cues may support gender discrimination, which typically involves large F0 differences (an octave).

Within-gender contrasts pose the greatest difficulty for CI users (Cleary & Pisoni 2002; Fu et al. 2004; Cleary et al. 2005). Fu et al. (2004) examined spectral and temporal contributions to gender identification and speech (phoneme) recognition in adult CI users and hearing adults exposed to CI simulations. CI simulations, or vocoders, operate in a similar manner to commercially available implants. A series of band-pass filters separate the acoustic input into several frequency bands, and the temporal envelope from each band is extracted. The temporal envelope information is then used to modulate noise-band or sine-wave carriers of similar frequencies. In vocoder simulations, spectral content is typically manipulated by varying the number of spectral bands—corresponding to the stimulating electrodes of CIs—while temporal information is manipulated by increasing the envelope cutoff frequencies. Fu et al. found that phoneme perception improved with increasing numbers of spectral channels, but no further improvement resulted from increasing temporal information. In contrast, talker identification improved with increases in temporal as well as spectral information. These findings highlight the different informational demands of speech perception and talker identification. Fu et al. also found that CI users' performance on talker identification was comparable with that of hearing controls listening to stimuli with sparse spectral information (4 to 8 channels) but good temporal information (160 to 320 Hz envelope cutoff frequencies).

Vongphoe and Zeng (2005) extended these findings by examining adult CI users' ability to identify specific male and female talkers and the vowels they produced after an hour of training on initially unfamiliar talkers. CI users achieved only 20% correct on talker identification, which was comparable with the performance of hearing adults on simulations involving a single channel of information. Despite poor talker identification, CI users succeeded in identifying vowels from the same stimuli. It is likely that the brief stimuli and 10-alternative forced-choice task contributed to CI users' difficulty with talker identification. In the context of such brief stimuli, CI users are more accurate at identifying vowels produced by men than by women, boys, and girls (Loizou et al. 1998). Greater success on talker identification may be possible when redundant acoustic cues in the long-term speech spectrum are made available.

In one such attempt, Cleary and Pisoni (2002) examined implanted children's ability to distinguish two unfamiliar talkers on the basis of sentence-length utterances. Implanted children were significantly less accurate than hearing children. They were even less accurate when the sentences from the two speakers differed in linguistic content. In a follow-up study with systematic alterations of sentences produced by a single talker, Cleary et al. (2005) attempted to ascertain the minimum fundamental and formant frequency differences that children required to judge two sentences as produced by different talkers. For 5-yr-old hearing children, the minimum difference was 2 to 2.5 semitones. Overall, older child CI users performed at chance levels, but the best performers were within the range of the younger hearing children.

Acoustic transformations such as those used by Cleary et al. (2005) make it possible to examine the role of spectral features in talker differentiation, but they eliminate the natural variability that provides other cues to identity. Talkers are distinguished not only by spectral features but also by other characteristics such as breathiness and nasality (Eskenazi et al. 1990) or by rhythm and speaking rate (Van Lancker et al. 1985). It is likely, then, that listeners in general and child CI users in particular would differentiate talkers more readily in the context of natural acoustic variations. However, the available evidence indicates that CI children have difficulty in differentiating unfamiliar female talkers from one another, even in the context of naturally spoken utterances (Cleary & Pisoni 2002).

For young hearing children, the identification of familiar talkers is accomplished with ease. For example, preschoolers identify their classmates and teachers from natural vocal samples (Bartholomeus 1973), and they also identify the voices of familiar cartoon characters (Spence et al. 2002). These studies indicate the advantages of highly familiar voices as experimental stimuli. As noted, however, there is no research on implanted children's identification of familiar talkers and only one study of implanted adults involved short-term familiarization (Vongphoe & Zeng 2005). Moreover, the prevailing view is that current device limitations impede the perception of cues critical to the differentiation of talkers, especially same-gender talkers. It is clear that adult CI users can differentiate male and female talkers from syllable-length utterances (Fu et al. 2004; Vongphoe & Zeng 2005) and that their performance improves with sentence-length utterances (Meister et al. 2009). It is also clear that child CI users can differentiate male and female talkers from sentence-length utterances (Osberger et al. 1991; Staller et al. 1991). On the basis of the available evidence from implanted adults and children, one would expect child CI users to have little difficulty with cross-gender contrasts but considerable difficulty with same-gender contrasts.

The principal goal of this study was to determine whether child implant users could distinguish their mother's voice from unfamiliar voices including those of a man, a girl, and other women. Their ability to do so would indicate that algorithms designed for optimizing speech perception enable talker identification under favorable circumstances. Moreover, success on the part of CI users would provide support for the interdependent nature of speech perception and talker identification (Sheffert et al. 2002). On the basis of previous research with implanted adults (Fu et al. 2004; Vongphoe & Zeng 2005) and children (Cleary et al. 2005), child CI users were expected to differentiate their mother's voice from the voices of other talkers differing in gender and age. Our use of a highly familiar talker and sentence-length utterances was designed to provide a variety of cues that were unavailable to CI users in previous studies. Because of implant users' difficulty in resolving fine spectral differences, we predicted that same-gender differentiation would be more difficult than cross-gender or cross-age differentiation. Because the man's voice was the most acoustically distinct from the other voices (all female), higher levels of identification were predicted for the male talker than for the others.

A secondary goal was to determine whether global features of speech such as prosody (intonation, rhythm, and speaking rate) would facilitate implanted children's differentiation of their mother's voice from the voices of unfamiliar talkers. Prosodic features involving intensity and duration are transmitted reasonably well by CIs, but those involving voice pitch are not (Green et al. 2004). As a result, child implant users can differentiate utterances on the basis of number of syllables (Carter et al. 2002; Most & Peled 2007), but they have considerably more difficulty differentiating utterances with contrastive intonation contours or affective prosody (Most & Peled 2007; Peng et al. 2008; Hopyan-Misakyan et al. 2009). Although they can distinguish statements from yes/no questions at modest but above chance levels (Most & Peled 2007; Peng et al. 2008), they are unable to distinguish among happy, sad, angry, and fearful versions of the same statements (Hopyan-Misakyan et al. 2009) unless provided with extensive (long-term) training (Klieve & Jeanes 2001). In this study, higher levels of talker identification were expected when natural prosodic variations across talkers were available than when such variations were reduced.

EXPERIMENT 1

In this experiment, we examined talker identification in child CI users and a control sample of younger children with normal hearing (NH). The control group was included primarily to confirm the ease of the task for young hearing listeners. Children were asked to identify the talker of sentence-length utterances from a set of four alternatives: mother, other woman, girl, and man. We predicted that CI children would differentiate their mother's voice from those of other talkers but that their performance would be poorer for talkers matched on gender and age. We assessed the contribution of global speech style to talker identification by comparing children's performance on naturally varying speech samples and on those that imitated the mother's intonation and speaking rate. Voice samples that reduced the natural prosodic variations among talkers were expected to yield poorer performance than those that incorporated these variations.

Materials and Methods

Participants • The participants included 21 CI children who had Nucleus 24 implants with Freedom, ESprit 3G, or SPrint processors programmed with the Advanced Combination Encoder strategy (see Table 1). These children, who were 4.8 to 14.3 yrs of age (M = 8.9 yrs, SD = 3.0), were congenitally or prelingually deaf, except for one child (CI 5) who had a progressive hearing loss since birth and had some residual, low-frequency hearing in the unimplanted ear. Two CI children (CI 14 and 15) were siblings, which meant that there were 20 mothers of CI children. No CI child used a hearing aid in the unimplanted ear, in line with local clinical practices. Five

TABLE 1. Demographics of CI participants in experiment 1

Participant	Age (yr)	Length of Device Use (yr)
1*	10.7	7.7
2†	11.4	8.4
3*	10.7	5.7
4*	6.2	3.2
5‡	12.3	3.6
6†	5.4	3.0
7*	11.5	9.6
8*	4.9	3.8
9*	10.3	7.4
10*	4.7	3.6
11*	11.8	8.8
12*	5.2	2.9
13*	9.8	7.2
14*	4.8	3.8
15*	7.8	4.8
16*	14.3	8.3
17‡§	10.4	6.9; 0.1
18*§	6.2	4.3; 1.1
19*§	11.5	7.5; 1.5
20*§	6.8	6.1; 2.8
21*§	9.7	7.6; 2.6

** CI processor: Freedom.*
† CI processor: Sprint.
‡ CI processor: ESprit 3G.
§ Bilateral CI participants.
CI, cochlear implant.

previous unilateral users (CI 17, 18, 19, 20, and 21) had received a second implant and were regular bilateral users at the time of the study. Bilateral users had their first implant for an average of 6.5 yrs (SD = 1.4) and their second device for an average of 1.6 yrs (SD = 1.1). Unilateral CI users had their implant for an average of 5.7 yrs (SD = 2.4). All CI children communicated exclusively by oral means and were enrolled in age-appropriate grade levels with their hearing peers. The comparison sample of 16 NH children had a mean age of 5.6 yrs (SD = 0.3), which approximated the age of the youngest children in the sample of CI users. Children in the NH sample had no personal or family history of hearing problems and were free of colds at the time of testing.

Apparatus • Voice recordings were made in a 3 × 2.5 m double-walled, sound-attenuating chamber (Industrial Acoustics Corporation, Bronx, NY). Voice samples were recorded with a microphone (Sony F-V30T, San Diego, CA) connected directly into a computer workstation operating Windows XP. High-quality digital sound files (44.1 kHz, 16-bit, Mono) were created, and the average amplitude of all voice samples was equated with a digital audio editor (Sound Forge 6.0). Colored digital photographs (headshots) of the mothers and actors were taken against a white background for subsequent use in the speaker-identification task.

Testing was conducted in the sound-attenuating chamber. A computer workstation and amplifier (Harmon-Kardon 3380, Stamford, CT) outside of the booth interfaced with a 17" touch-screen monitor (Elo LCD TouchSystems, Berwyn, PA) and two wall-mounted loudspeakers (Electro-Medical Instrument Co., Mississauga, Ontario, Canada) inside the booth. The loudspeakers were mounted at the corners of the sound booth, each located at 45 degrees azimuth to the participant, and the

TABLE 2. Scripted common phrases recorded by mother and voice actors

Questions
1. How was school today?
2. Would you like to go to the park?
3. What would you like for breakfast?
4. Did you brush your teeth?
5. Would you like a snack?

Statements
6. It's time to go to bed.
7. Look at the cute puppy.
8. Good job on your homework.
9. You can watch one TV show.
10. You did a great job.

touch-screen monitor was placed at the midpoint. Sound files were presented at approximately 70 dB (A).

Stimuli • A few weeks before the test session, mothers of the CI and NH children visited the laboratory to record 10 scripted utterances (5 questions and 5 statements) with familiar vocabulary (see Table 2 and Audio, Supplemental Digital Content 1, 2, 3, and 4, http://links.lww.com/EANDH/A14, http://links.lww.com/EANDH/A15, http://links.lww.com/EANDH/A16, and http://links.lww.com/EANDH/A17 for natural samples of a mother, man, girl, and woman, respectively). Mothers were instructed to use a natural child-directed style while expressing each utterance. Child-directed speech is more animated than typical adult-directed speech, but it is similar to emotional adult-directed speech (Trainor et al. 2000). A man, a woman, and a 10-yr-old girl (the actors) were instructed to produce their own child-directed (i.e., animated) versions of the same utterances (natural condition) as well as other versions of the sentences in which they imitated each mother's renditions (imitation condition). The result was affectively positive speech rather than the affectively neutral utterances of typical adult-directed speech. Mean F0 of these actors was 134.0 Hz (SD = 12.7 Hz) for the man, 241.9 Hz (SD = 32.0) for the woman, and 258.5 Hz (SD = 7.2) for the girl. Mean F0 values of mothers are provided in Table 3.

To generate imitations of maternal utterances, the actors listened to each target voice sample and matched, as closely as possible, the intonation and speaking rate (i.e., global speech style) of each utterance (see Audio, Supplemental Digital Content 5, 6, and 7, http://links.lww.com/EANDH/A18, http://links.lww.com/EANDH/A19, and http://links.lww.com/EANDH/A20 for demonstrations of maternal utterance imitations by the man, girl, and woman, respectively). The adequacy of each imitation was evaluated by the experimenter and confirmed by an independent listener. Figure 1 provides an example of pitch contours extracted from the time waveform of an utterance produced by all four speakers in both conditions (natural and imitation). The same maternal utterances were used in the natural and imitation conditions of the present experiment. As can be seen in Figure 1, overall durations and pitch contours across speakers were more variable in the natural condition than that in the imitation condition. In fact, durations and pitch contours of all nonmaternal utterances in the imitation condition closely resembled those of the mother.

Procedure • Children were tested individually and the experimenter remained beside the child for the entire test session.

TABLE 3. Mean utterance-length F0 (Hz) for each mother

Mother	Mean F0 (Hz)	SD
1	230	21
2*	217	20
3	280	30
4	243	48
5	273	28
6	232	22
7	213	17
8	232	18
9	228	18
10	222	39
11*	208	10
12	238	22
13	212	51
14	268	35
15	207	19
16	225	23
17	223	24
18	251	13
19	210	55
20	220	11

** Participants in experiment 1 only.*

The natural and imitation conditions were blocked and repeated for each participant in ABBA order (A = natural, B = imitation). All 40 samples (10 sentences × 4 speakers) were randomized individually within each condition, resulting in a total of 160 sentences. For the natural condition, children were instructed that they would hear the voice of their mother or that of an unfamiliar man, woman, or girl. Their task was to identify the talker by selecting the appropriate picture on the monitor. During each trial, a four-picture display arranged in a 2 × 2 grid on the monitor showed each of the possible talkers (mother, other woman, man, and girl). To hear the voice sample, the child was instructed to press the "play" button below the picture display. With the exception of a few cases when a child attempted to talk to the experimenter or was distracted during a trial, voice samples were played only once. Immediately after hearing the stimulus, children were required to identify the talker by selecting the corresponding picture on the display. The computer registered a touch response only after a voice sample occurred. Feedback was provided in the form of a schematic happy face on the monitor for a correct response and a blank screen for an incorrect response. Such feedback contributed to the game-like nature of the task and helped maintain children's attention. The child pressed a "continue" button on the monitor to proceed to the next trial. For the imitation condition, children were told that they would hear the same talkers as before. However, this time the speakers would sound more similar to one another, so children were encouraged to listen carefully. In other respects, the procedure was the same as in the natural condition.

Results and Discussion

Visual inspection of the data from unilateral and bilateral implant users revealed a virtually identical range of performance. Accordingly, data from unilateral and bilateral CI users were considered together. Figure 2 shows the performance of NH and CI children in both global speech styles (natural and

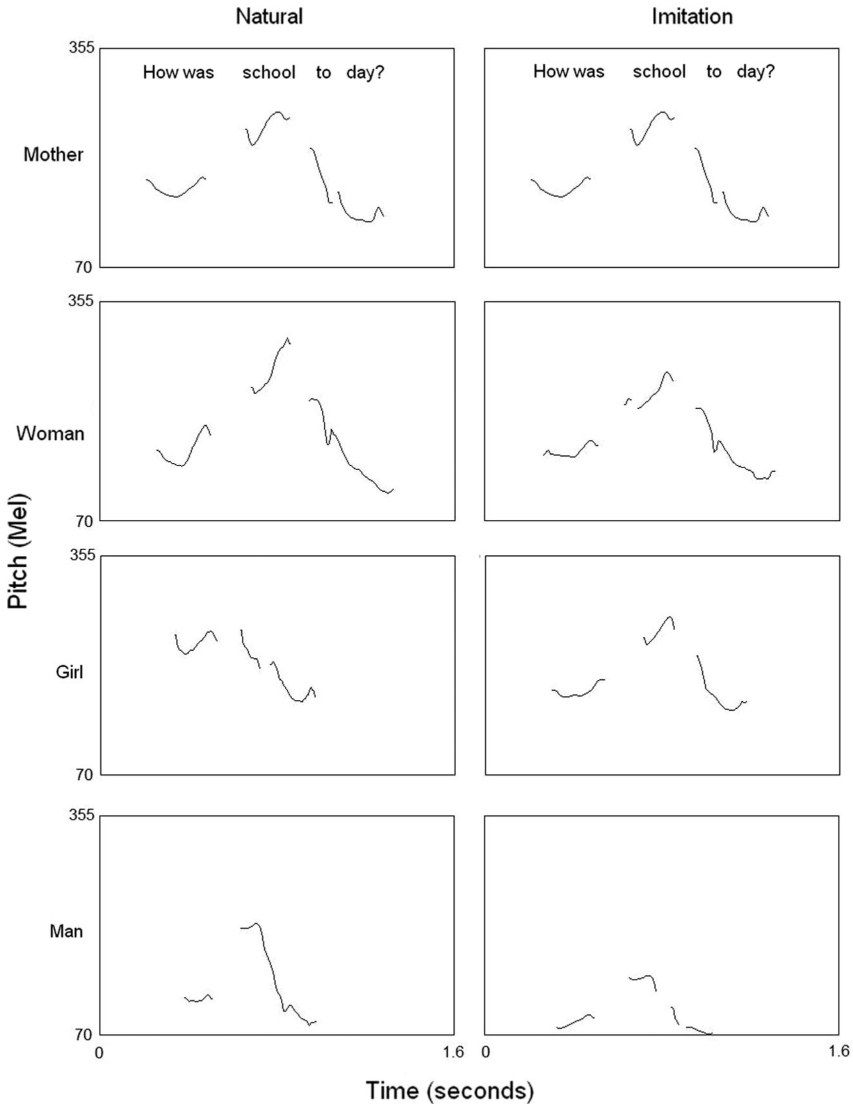

Fig. 1. Fundamental frequency (F0) contours of a typical utterance in the natural and imitation conditions of experiment 1.

imitation) and presentation orders. NH children performed at ceiling in all circumstances. In contrast, CI children's performance varied as a function of global speech style and presentation order, with the difference in global speech style emerging in the second presentation, when CI children performed 9% better in the natural condition than in the imitation condition. These observations were confirmed by means of a two-way analysis of variance (ANOVA) on the scores of the CI group, with presentation order (first and second) and global speech style (natural, imitation) as repeated measures. A main effect of global speech style revealed higher accuracy in the natural condition (M = 0.80, SD = 0.15) than in the imitation condition (M = 0.74, SD = 0.16), $F(1, 20) = 6.66$, $p = 0.018$. A significant two-way interaction confirmed that the difference between conditions varied as a function of presentation order, $F(1, 20) = 4.78$, $p = 0.041$. Follow-up paired *t* tests indicated that performance in the first presentation of natural (M = 0.76, SD = 0.17) and imitation conditions (M = 0.75; SD = 0.15) did not differ. In contrast, performance in the second presentation was better in the natural condition (M = 0.83, SD = 0.14) than in the imitation condition (M = 0.74; SD = 0.17), $t(20) = 3.35$, $p = 0.003$. The findings indicate progressive learning about individual differences in global speech style over the course of the test session, leading to more accurate judgments of speaker identity. However, there was no evidence of improved performance over quartiles or halves of the first

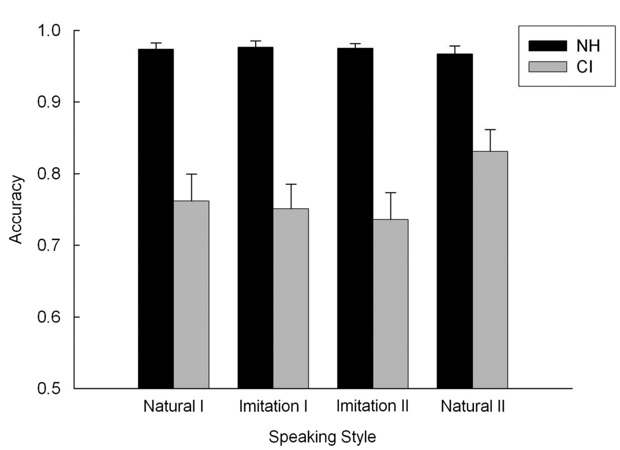

Fig. 2. Performance (proportion correct) of NH and CI children in natural and imitation conditions of experiment 1. CI, cochlear implant; NH, normal hearing.

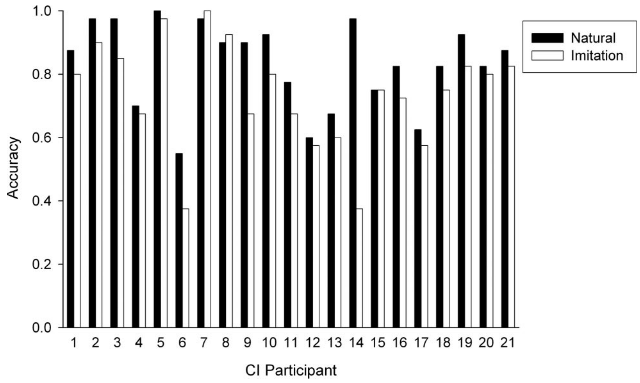

Fig. 3. Individual accuracy scores of CI participants on each condition (second presentation). CI, cochlear implant.

presentation, which indicates that proficiency in talker identification was not a simple consequence of reinforcement. Instead, considerably greater exposure to the speech samples was necessary before CI children showed differential performance on naturally varying samples and imitations.

Figure 3 shows the scores of individual CI children in the second presentation of both conditions. It is evident that performance was consistently better in the natural condition than in the imitation condition. In fact, 18 of the 21 children showed this pattern, which is highly significant ($p < 0.001$) by a sign test. The modest sample size and large variability in performance limited the utility of statistical analyses aimed at identifying links between demographic variables and performance. No significant correlations were evident between performance in each condition (natural or imitation) and age or duration of implant use. Incidentally, the imitation condition posed the greatest difficulty for participants CI 6 and CI 14, who were among the youngest children and had the least implant experience. Nevertheless, their performance on this task was well above chance levels.

Figure 4 depicts performance on each talker in the second presentation of the natural and imitation conditions. In the natural condition, performance was best for the man (M = 0.99, SD = 0.02), followed by the girl (M = 0.90, SD = 0.25), the mother (M = 0.77, SD = 0.26), and the other woman (M = 0.65, SD = 0.28). In the imitation condition, performance was best for the man (M = 0.95, SD = 0.17), followed by the mother (M = 0.73, SD = 0.27), the girl (M = 0.71, SD = 0.30), and the other woman (M = 0.56, SD = 0.28). In both conditions, there was near-perfect accuracy in identifying the man and lowest accuracy on the unfamiliar woman. In this context, highest accuracy for the male talker is attributed to its unique status among the all-female alternatives in the response set.

A two-way repeated-measures ANOVA was used to compare performance on the mother, other woman, and girl talkers in both global speech styles (second presentation only). The man was excluded from this analysis because of consistently high levels of recognition. There was a robust main effect of global speech style, $F(1, 20) = 16.08$, $p = 0.001$. Specifically, performance was better in the natural than in the imitation condition. There was also a main effect of talker, $F(2, 40) = 4.52$, $p = 0.017$, which was qualified by an interaction between talker and global speech style, $F(2, 40) = 4.00$, $p = 0.026$. One-way repeated-measures ANOVAs were performed separately across female talkers for both conditions. Differences were significant in the natural condition, $F(2, 40) = 5.89$, $p = 0.006$, but just beyond the cusp of statistical significance in the imitation condition, $F(2, 40) = 3.16$, $p = 0.053$. Differences in the natural condition were attributable to higher accuracy in identifying the girl (M = 0.90, SD = 0.25) than the other woman (M = 0.65, SD = 0.28), $p = 0.008$, neither of which differed from performance on the mother (M = 0.77, SD = 0.26), with p values >0.20. In short, cues to the identity of female talkers, particularly the girl, were more salient in the natural condition for the CI children.

To examine the source of errors in the second presentation of natural and imitation conditions, we formed stimulus–response matrices by pooling responses from all 21 CI children (see Table 4). Diagonal entries (bolded) indicate proportions of correct responses, whereas off-diagonal entries indicate proportions of errors for the different speakers. Marginal totals indicating proportions of selected responses for each speaker are listed at the row ends. Examination of hits showed that CI children successfully identified each speaker at levels that were well above chance (i.e., 0.25), with p values <0.001. Inspection of off-diagonal entries between matrices reveals that overall error rates were more pronounced in the imitation condition than in the natural condition. The stimulus–response matrices were submitted to a Sequential Information Transfer Analysis (Wang & Bilger 1973) to investigate the distribution

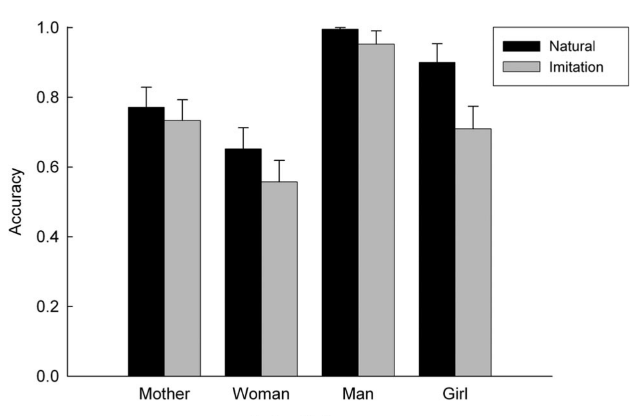

Fig. 4. Performance on each talker in natural and imitation conditions of experiment 1 (second presentation).

TABLE 4. Stimulus–response matrices of CI children for the (a) natural and (b) imitation conditions of experiment 1

	Mom	Woman	Man	Girl	
a					
Mom	**0.77**	0.22	0	0.05	**0.26**
Woman	0.18	**0.65**	0	0.04	**0.22**
Man	0.01	0.01	**1.00**	0	**0.27**
Girl	0.04	0.11	0	**0.90**	**0.26**
b					
Mom	**0.73**	0.29	0.02	0.14	**0.29**
Woman	0.17	**0.55**	0.01	0.15	**0.22**
Man	0.04	0.01	**0.95**	0	**0.25**
Girl	0.06	0.15	0.02	**0.71**	**0.23**

Columns represent stimulus speaker categories and rows represent the responses. Marginal proportions (bolded) are listed at the row ends.
CI, cochlear implant.

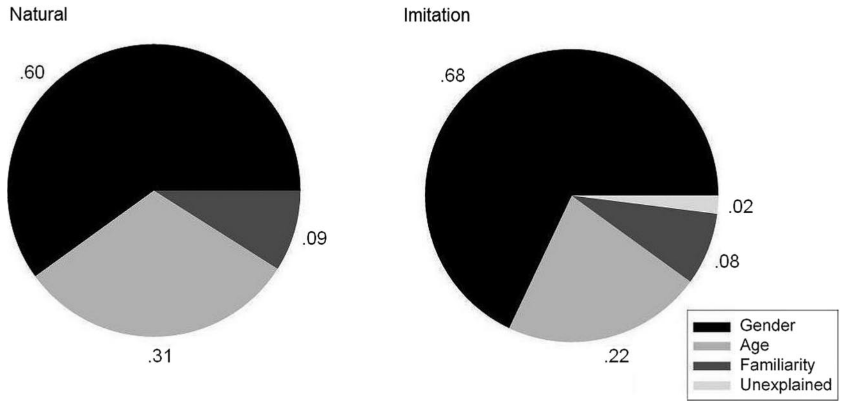

Fig. 5. Contribution of speaker characteristics to response patterns in natural and imitation conditions of experiment 1 (second presentation).

of responses attributable to age (girl versus the other three speakers), gender (man versus the other three speakers), and familiarity (mother versus the other three speakers). The proportion of explained variance for a particular feature is an indication of the relative contribution (or transfer of information) of that feature to the observed distribution of scores. For the purpose of this analysis, familiarity referred to characteristics of the mother's speech beyond those contributed by age and gender. Figure 5 shows the proportion of variance explained by age, gender, and familiarity in the natural and imitation conditions, respectively. These three features accounted for nearly all of the observed variance in performance in both conditions, with gender making the largest contribution, followed by age and familiarity. In line with results of the ANOVA, the reduced contribution of age in the imitation condition probably stemmed from lower identification accuracy of the girl, whose imitations of maternal style seemed to mask her age. The reduced contribution of familiarity in the imitation condition is also consistent with greater difficulty in the context of increasing similarity of global speech style across talkers.

Spectral cues in the long-term spectrum are thought to have little utility for within-gender voice differentiation by CI users (Green et al. 2004), but temporal cues could be relevant. Accordingly, we examined temporal variability for each talker-mother pair by measuring differences in speaking rate, as indexed by sentence duration, between global speech styles. For each sentence, the absolute difference in duration between a talker's and mother's version was calculated. Figure 6 shows mean differences in sentence duration between each talker (woman, girl, and man) and mother in the natural and imitation conditions. It is clear that duration differences between mother-speaker pairs were systematically greater in the natural condition than in the imitation condition. In the overwhelming majority of instances, actors substantially altered their natural speaking rate in the imitation condition to approximate the maternal rate. In fact, the duration of sentence imitations differed from maternal sentence duration, on average, by less than one tenth of a second.

In short, child implant users readily distinguished their mother's voice from those of an unfamiliar man, woman, or girl. These children seemed to make use of gross spectral differences (e.g., fundamental frequency) that arise from gender differences in the size and shape of the vocal tract. They also seemed to derive some benefit from differences in global speech style, as evidenced by the performance decrement in the second imitation condition. Nevertheless, child CI users' success in distinguishing among talkers with similar speaking rates and intonation contour implies that other cues were available. One possibility is that cues at the segmental level, notably the timing of vocal tract articulator movements in vowel and consonant production (i.e., dynamic articulatory cues), played a critical role in the identification of same-gender talkers. Presumably, the use of sentence-length utterances would make it easier to extract systematic differences in articulation than would be the case for single-syllable utterances.

The relatively high levels of performance by implant users in the present experiment are at odds with reports of poor talker recognition in previous research (Cleary & Pisoni 2002; Cleary et al. 2005; Vongphoe & Zeng 2005) and with the acknowledged limitations of current implant processors (Carlyon et al. 2002; Shannon et al. 2004). The maximum pitch conveyed by the speech envelope is capped at approximately 300 Hz in current devices (Burns & Viemeister 1976), which is inadequate for conveying the F0 range of typical female voices. Furthermore, the average F0 difference between the unfamiliar woman and girl in the present experiment was only 1.1 semitones, considerably less than the 3-semitone difference considered necessary for reliable voice differentiation in this population (Cleary & Pisoni 2002).

When global speech style (i.e., intonation contour and speaking rate) was similar across talkers, the decline in accuracy, although significant, was relatively modest (i.e., 9%), and talker identification remained well above chance levels. This finding implies that CI children did not rely primarily on such global differences. It is likely that other systematic differences in natural speech are perceptible to implant users and are the principal basis for talker identification. We explored this issue in experiment 2 by requiring hearing and implanted children to distinguish their mother's voice from other women's voices. We expected that greater spectral similarity between voices would help focus children's attention on person-specific variations in phoneme articulation and global speech style.

EXPERIMENT 2

The findings of experiment 1 revealed that child CI users differentiated their mother's voice from those of other talkers varying in gender and age. Error rates were increased for utterances that mimicked the global style of maternal speech, but performance still remained well above chance levels. In

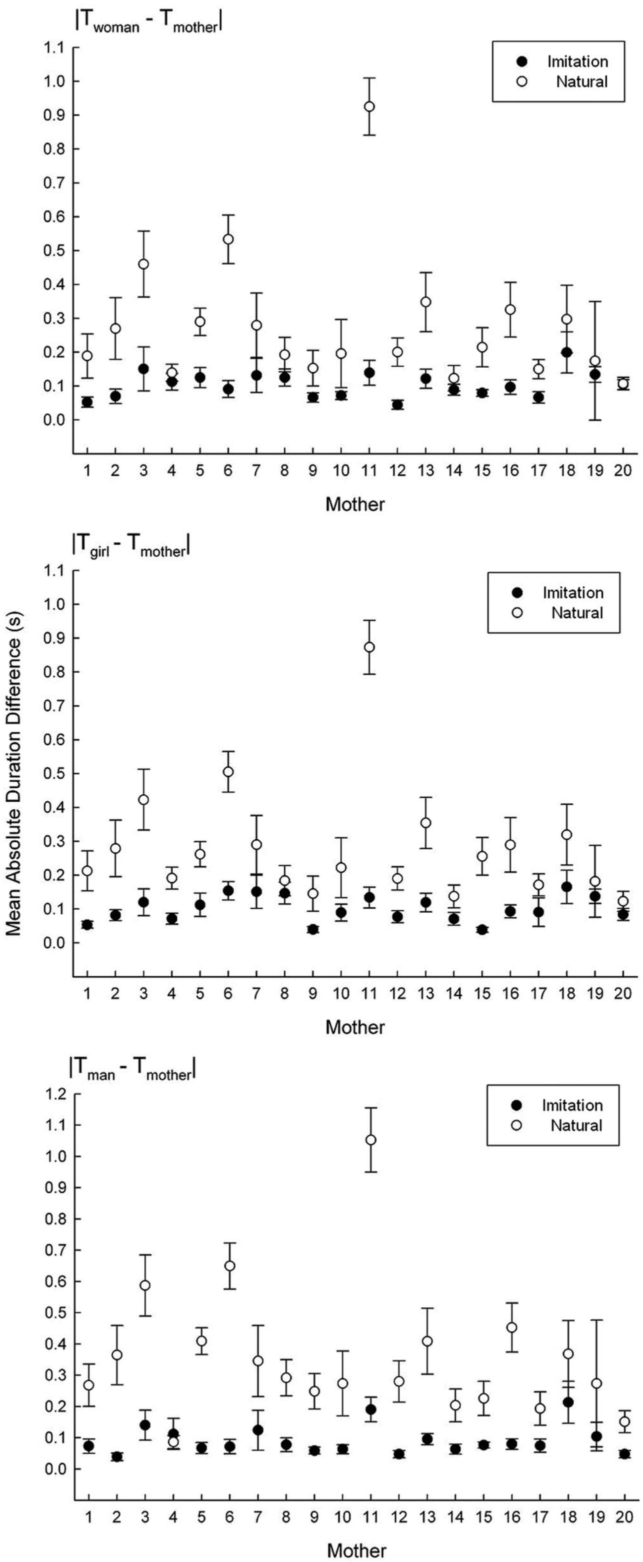

Fig. 6. Absolute duration differences between mother–actor pairs in natural and imitation conditions of experiment 1. Error bars indicate standard errors.

light of CI users' success on same-gender discrimination in experiment 1 and their failure in all published research to date (Cleary & Pisoni 2002; Fu et al. 2004, 2005; Vongphoe & Zeng 2005), it was important to confirm our findings. Accordingly, we focused exclusively on child CI users' differentiation of female voices. To this end, we presented scripted utterances from the mother and three other women in a two-alternative forced-choice task. As in experiment 1, the speech samples included natural variations across speakers as well as imitations of maternal utterances. Because of greater spectral similarity among voices in the present experiment, we expected global speech style to play a more prominent role in voice identification than it did in experiment 1.

TABLE 5. Demographics of CI participants in experiment 2

Participant	Type of User	Age (yr)	Length of Device Use (yr)
1	Bilateral	10.4	6.9; 0.1
2*	Bilateral	5.3	3.3; 0.5
3*	Bilateral	8.3	5.3; 0.3
4	Bilateral	11.5	8.5; 1.5
5	Bilateral	6.2	4.3; 1.1
6*	Bilateral	6.0	3.8; 0.4
7*	Bilateral	5.5	4.4; 1.4
8	Bilateral	6.8	6.1; 2.8
9	Bilateral	9.7	7.7; 2.7
10	Unilateral	10.2	7.2
11	Unilateral	10.2	7.2
12	Unilateral	12.7	9.7
13	Unilateral	10.3	7.4
14	Unilateral	6.43	1.0
15	Unilateral	13.4	5.4
16	Unilateral	11.7	6.8
17	Unilateral	5.93	4.8
18	Unilateral	14.3	8.3
19	Unilateral	12.6	10.7

** Unilateral to bilateral.*
CI, cochlear implant.

Materials and Methods

Participants • Nineteen CI children from experiment 1 returned to participate in the present experiment. Table 5 lists their demographic details (10 unilateral and 9 bilateral). In the time that elapsed between experiments, four of the unilateral CI children had received a second implant (CI 2, 3, 6, and 7). At the time of testing, they had their first implant for 5.7 yrs on average (SD = 1.9, range = 3.3 to 8.5) and their second implant for 1.6 yrs (SD = 1.0, range = 0.13 to 2.7 yrs). Of the 16 NH children from experiment 1, 11 were available to return for the present experiment. Because two of the child implant users (CI 2 and 3) were siblings, there were 18 mothers in the sample.

Materials • We recruited three women whose age and regional dialect matched those of maternal participants and who demonstrated a facility for vocal imitation. The mean utterance-length F0 for the women designated woman 1, woman 2, and woman 3 was 223.6 (SD = 18.2), 234.9 (SD = 22.0), and 239.6 (SD = 30.6) Hz, respectively, which was within the F0 range of mothers (see Table 3). The average F0 difference was 0.85 semitones between woman 1 and woman 2, 0.34 semitones between woman 2 and woman 3, and 1.2 semitones between woman 1 and woman 3. From the set of recordings in experiment 1, nine utterances were selected randomly from each mother, and each actor was assigned to imitate three of

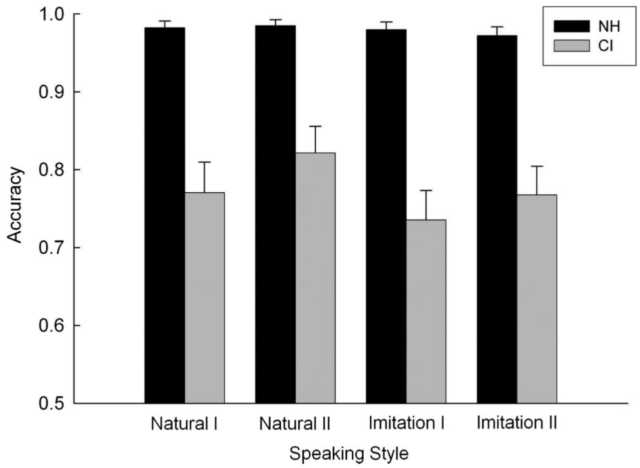

Fig. 7. Performance of NH and CI children in natural and imitation conditions of experiment 2. CI, cochlear implant; NH, normal hearing.

these, selected randomly. For the natural condition, each actor's personal rendition of these sentences was recorded. For the imitation condition, each actor listened to a maternal utterance and imitated its intonation and rate (tempo) as closely as possible. The imitations not only captured the intonation and speaking rate of the mother being imitated but they also reduced the differences in average voice pitch (all <2 semitones) between mothers and actors.

Procedure • The procedure and instructions were similar to those of experiment 1 except for aspects relating to the use of a two-alternative forced-choice task in the present experiment. Children selected one of the two images that appeared on the touch-screen monitor: a picture of their own mother or a featureless outline of a female face, the latter being used for all nonmaternal voices. As in experiment 1, immediate feedback was provided: a schematic happy face for correct responses and a blank screen for incorrect responses. Voice samples were repeated in each condition and presented in a random order for a total of 36 trials (nine samples per mother and three samples for each unfamiliar woman, with each sample presented twice). Children were tested twice in both conditions, with the natural and imitation conditions in ABBA order for half of the children and BAAB order for the other half. Each child heard a total of 144 utterances.

Results and Discussion

A three-way mixed ANOVA, with group (unilateral and bilateral) and presentation order (first and second) as between-subjects factors and global speech style (natural and imitation) as a within-subject factor, revealed no main effects or interactions involving group (Fs <1). Accordingly, the data of unilateral and bilateral CI users were considered together, as in experiment 1. Figure 7 illustrates the performance of NH and CI children in both conditions. CI children's overall accuracy across conditions (M = 0.77, SD = 0.16) was considerably lower than that of NH children (M = 0.98, SD = 0.03), who performed at ceiling, as they did in experiment 1. Consequently, subsequent analyses focused exclusively on CI children. A two-way ANOVA with order (first and second) and global speech style (natural and imitation) as repeated measures revealed a main effect of order, $F(1, 18) = 17.19$, $p = 0.001$. Performance was better in the third and fourth conditions (M = 0.79, SD = 0.15) than in the first and second conditions (M = 0.75, SD = 0.17), and accuracy in all conditions was well above chance levels, with p values <0.001. A main effect of global speech style, $F(1, 18) = 8.15$, $p = 0.011$, confirmed that performance was better in the natural (M = 0.80, SD = 0.16) than in the imitation (M = 0.75, SD = 0.16) conditions. There was no two-way interaction, $F <1$. Because CI children displayed no systematic confusion patterns between the actors and their own mother, there was no indication that performance was related to the magnitude of differences in fundamental and formant frequencies between voices.

Figure 8 shows the scores of individual CI children in the second presentation of each condition. A sign test confirmed that performance was better in the natural condition, $p = 0.022$. Although 11 of the 19 children performed better in the natural condition than in the imitation condition, only two performed better in the imitation condition. Six children performed equivalently in both conditions. Examination of confusion patterns in the second presentation of the natural and imitation conditions, as in experiment 1, corroborated the greater incidence of errors on unfamiliar women's voices in the imitation condition than in the natural condition.

As in experiment 1, CI children's performance was superior for natural variations than for vocal imitations, which implies that CI children capitalized to some extent on global cues to talker identity. An unexpected finding was the high level of

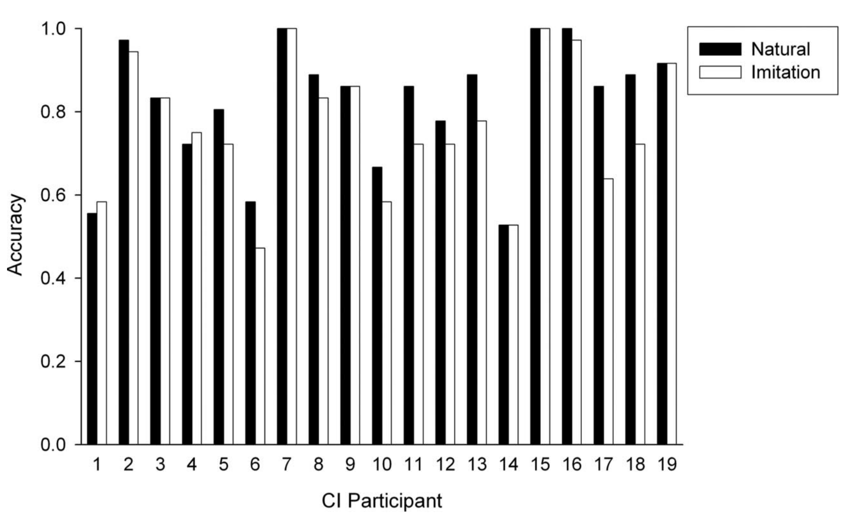

Fig. 8. Individual accuracy scores of CI participants for each condition. CI, cochlear implant.

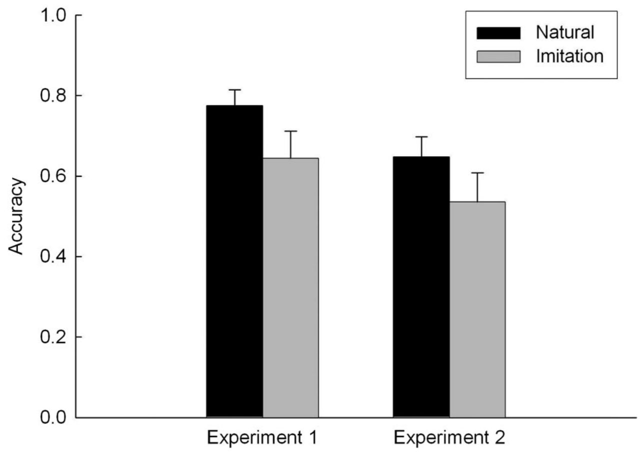

Fig. 9. Performance in natural and imitation conditions of experiments 1 and 2 (second presentation). Error bars indicate standard errors.

performance in both conditions, which was comparable with performance in experiment 1 despite the smaller F0 differences between talkers in the present experiment. The present scores were inflated relative to experiment 1 by differences between chance levels of responding in the present two-alternative task (50%) and the four-alternative task of experiment 1 (25%). To facilitate direct comparisons between experiments, individual scores in the second occurrence of both conditions were transformed to adjust for differences in chance levels by means of the following formula: $Y = (X - p)/(1 - p)$ where Y denotes the transformed score, X is the raw proportion achieved per condition, and p is chance probability for the task. These transformed scores are directly comparable because a score of 1 represents perfect accuracy and 0 represents chance-level responding. Figure 9 presents transformed scores in the natural and imitation conditions for the CI participants who were tested in experiments 1 and 2. A two-way ANOVA with experiment (1 and 2) and global speech style (natural and imitation) as repeated measures revealed a main effect of global speech style, $F(1, 18) = 24.72$, $p < 0.001$, but no effect of experiment and no interaction, with F values <1. In other words, the accuracy of talker identification was similar across experiments after adjusting for differences in chance levels of performance.

Correlational analyses involving the 19 CI children who participated in experiments 1 and 2 revealed positive and statistically significant relations between scores in the natural and imitation conditions across experiments, $r(17) = 0.86$, $p < 0.001$, and $r(17) = 0.50$, $p = 0.031$, respectively. Child CI users' performance was highly consistent for the two tasks with natural speech variations and moderately consistent for the two tasks with reduced cues to talker identity.

To contextualize the perceptual demands of the current experiment, Cleary et al. (2005) found that young hearing children and the highest performing CI children assigned sentence-length utterances to different speakers when F0 and formant frequencies were separated by approximately 2.5 semitones. In this experiment, CI children, on average, differentiated their mother's voice from female foils with even smaller F0 differences. However, we are not suggesting that spectral differences were the basis for CI children's differentiation of talkers. On the contrary, our contention is that they were depending primarily on temporal cues.

In the context of a closed-set task, implanted children differentiated their mother's voice from the voices of unfamiliar women. Undeniably, they were considerably poorer at this task than were their hearing peers. Although global speech style had no effect on talker identification in hearing children, mainly because their performance was at ceiling in the most difficult condition, it influenced the performance of implanted children. Child CI users' performance implies that person-specific cues in natural speech support talker identification, even in the absence of temporal fine structure cues.

GENERAL DISCUSSION

The principal goal of this study was to determine whether child CI users could distinguish their mother's voice from the voices of unfamiliar talkers, including those differing in gender or age. A control sample of young hearing children achieved near-perfect accuracy on talker identification, even when the unfamiliar talkers imitated their mothers' global speech style. Hearing children's performance corroborates previous reports of familiar voice identification (Bartholomeus 1973; Spence et al. 2002), although performance accuracy in this study greatly surpassed that observed previously. Undoubtedly, hearing children's rich representations of their mother's voice contributed to this outcome. Presumably they considered, on each trial, whether the talker was their mother and, if not, whether it was an unfamiliar man, woman, or child in experiment 1 or simply an unfamiliar woman in experiment 2.

Not surprisingly, implanted children's performance was considerably poorer than that of their hearing peers. What was surprising was that CI children went beyond distinguishing their mother's voice from the voices of an unfamiliar man and girl to also distinguishing it from other women's voices. In short, they greatly exceeded expectations based on previous reports of child and adult implant users (Loizou et al. 1998; Cleary & Pisoni 2002; Fu et al. 2004; Cleary et al. 2005; Vongphoe & Zeng 2005). Implanted children's high level of accuracy on the man's voice implies that temporal pitch cues convey cross-gender differences, as suggested in previous research (Fu et al. 2004). Child CI users were less accurate but still relatively successful on same-gender voices, differentiating their mother's voice from the voice of a girl or another woman, even when the speakers mimicked the global style of maternal speech. Despite previous claims that F0 differences of 2 semitones or more are required for speaker differentiation (Cleary et al. 2005), implanted children in this study succeeded in the context of F0 differences as small as 1 semitone. It is likely that child CI users relied primarily on individually distinctive cues to consonant and vowel articulation, especially when differentiating women's voices, and secondarily on individual differences in global speech style when such cues were available.

What accounts for the success of the child CI users in the present study and the difficulties of child and adult CI users in previous studies? In contrast to previous studies that exposed CI children or adults to novel voices that were initially unfamiliar (Cleary & Pisoni 2002; Vongphoe & Zeng 2005), this study used maternal voices, which provided the advantage of long-term exposure in meaningful contexts. Moreover, whereas Cleary et al. (2005) presented systematic alterations of a single voice, which eliminated differences in articulatory timing and speaking rate and forced children to rely on spectral cues, the natural variability of voices in this study provided children with multiple cues to identity. The present use of sentences rather than syllables (Fu et al. 2004; Vongphoe & Zeng 2005) also enabled greater sampling of individual variations in phoneme articulation and global speech style. This is in line with findings on adult CI users, who are also more successful on tasks involving long-duration utterances (i.e., sentences) than short-duration utterances (Meister et al. 2009).

Spectral cues are likely to be of limited use to CI listeners for within-gender differentiation of talkers. Instead, temporal cues—particularly those relating to phoneme articulation and speaking rate—are likely to play a more prominent role. When speaking-rate differences among talkers were minimized, the accuracy of talker identification decreased. Child CI users still performed well above chance levels, however, attesting to the importance of unique aspects of phoneme articulation, which facilitate talker identification for hearing listeners in adverse listening conditions (Remez et al. 1997; Sheffert et al. 2002). Our use of child-directed speech may be relevant to the high levels of performance achieved by implant users. Although hearing adults differentiate talkers more accurately from adult-directed speech than from child-directed speech, they differentiate talkers more accurately from child-directed speech under conditions of CI simulation (Bergeson et al. 2009). Finally, it is likely that the game-like task, which featured a touch-screen, color photographs, and immediate feedback, optimized children's performance. We suggest that the aforementioned stimulus and task factors contributed jointly to child CI users' unexpected high levels of performance.

As noted, we contend that child CI users relied primarily on articulatory timing differences and secondarily on global differences in speech style to differentiate their mother's voice from other women's voices. With respect to the relevant global cues, temporal patterns such as speaking rate are likely to be of greater importance than pitch patterns, which are transmitted poorly by current implants (Green et al. 2004). Person-specific variations in consonant and vowel articulation were available in the imitated as well as natural speech samples, but person-specific variations in global speech style were restricted largely to the natural samples.

According to traditional accounts of speaker recognition, which consider the acoustic correlates of voice pitch and timbre as the principal basis of voice identification (Van Lancker et al. 1985), within-gender talker identification should not be possible for CI users. Obviously, the findings of this study are at odds with such views. Instead, they suggest that when fundamental frequency and timbre are degraded or unavailable, listeners can identify speakers on the basis of individual differences in consonant and vowel articulation (Remez et al. 1997) and global temporal cues in the long-term spectrum. Remez et al. found that when a limited number of frequency-specific sine waves replace natural voices, listeners can still identify the talkers at greater than chance levels. The ability of adults to link silent videos of unfamiliar talkers to previously heard voices (Kamachi et al. 2003) underlines the importance of temporal cues to identity across vision and audition. The present findings lend credence to the view that articulatory timing patterns contribute to voice processing and speech processing (Remez et al. 1997; Kamachi et al. 2003; Lachs & Pisoni 2004), even for CI users. The findings are also consistent with the view that variations in global speech style facilitate speaker identification (Lander et al. 2007).

One surprising outcome was implanted children's comparable performance accuracy across experiments despite greater acoustic similarities between talkers in the second experiment, which featured women's voices only. It is unlikely that positive transfer across experiments, which occurred approximately 12 months apart, offset the substantial increase in task difficulty. In principle, greater implant experience could have contributed to performance at the later test date even though implant experience and overall performance were uncorrelated in both experiments. Our prediction of greater difficulty in the second experiment was based on the reduced spectral cues that were available. However, if children were relying primarily on differences in articulatory timing, then the reduction of spectral cues would be irrelevant to their performance.

In experiment 1, child CI users readily identified the male voice, differentiating it from the woman's and girl's voice. It remains to be determined whether children with CIs can distinguish familiar male voices (e.g., their father) or children's voices (e.g., their siblings or playmates) from other voices of the same gender and age level. It is also of interest to ascertain the amount of exposure required for accurate talker recognition. Just as talker-identification training in spectrally degraded contexts improves hearing adults' decoding of speech in such contexts (Loebach et al. 2008), talker-identification training may have comparable benefits for child CI users. Loebach et al. (2008) suggest that these benefits arise, in part, from the depth of processing required to focus attention on a challenging task. Finally, the maternal voice may confer special benefits by optimizing arousal and mood, thereby facilitating learning or performance (Barker & Newman 2004).

In summary, the present findings suggest that recognition of familiar talkers is possible for CI users on the basis of temporal cues in the speech envelope. We contend that these cues include individually distinctive variations in the timing of phoneme articulation and in overall speaking rate. Our findings, which are the first to document relatively high levels of talker identification by CI users, are attributed to the use of a well-known talker, natural sentence-length stimuli, and a highly engaging task.

ACKNOWLEDGEMENTS

Supported by the Natural Sciences and Engineering Research Council of Canada.

Address for correspondence: Sandra E. Trehub, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, ON, Canada L5L 1C6. E-mail: sandra.trehub@utoronto.ca.

Received October 16, 2009; accepted February 13, 2010.

REFERENCES

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, B45–B53.

Bartholomeus, B. (1973). Voice identification by nursery school children. *Can J Psychol*, *27*, 464–472.

Bergeson, T. R., Till, S., Bent, T. (2009, April). *Perception of indexical information in infant-directed speech using cochlear implant simulations.* Paper presented at the biennial meetings of the Society for Research in Child Development, Denver, Colorado.

Burns, E. W., & Viemeister, N. F. (1976). Nonspectral pitch. *J Acoust Soc Am*, *70*, 1655–1660.

Carlyon, R. P., van Wieringen, A., Long, C. J., et al. (2002). Temporal pitch mechanisms in acoustic and electric hearing. *J Acoust Soc Am*, *112*, 621–633.

Carter, A. K., Dillon, C. M., Pisoni, D. B. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: Suprasegmental analyses. *Clin Linguist Phon*, *16*, 619–639.

Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Ann Otol Rhinol Laryngol*, *189*, 113–118.

Cleary, M., Pisoni, D. B., Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *J Speech Lang Hear Res*, *48*, 204–223.

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

Eskenazi, L., Childers, D. G., Hicks, D. M. (1990). Acoustic correlates of vocal quality. *J Speech Hear Res*, *33*, 298–306.

Fu, Q. J., Chinchilla, S., Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol*, *5*, 253–260.

Fu, Q. J., Chinchilla, S., Nogaki, G., et al. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *J Acoust Soc Am*, *118*, 1711–1718.

Geers, A. E. (2006). Spoken Language in Children with Cochlear Implants. In P. E. Spencer, & M. Marschark (Eds). *Advances in the Spoken Language Development of Deaf and Hard-of-Hearing Children. Perspectives on Deafness* (pp, 244–270). New York, NY: Oxford University Press.

Green, T., Faulkner, A., Rosen, S. (2004). Temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am*, *116*, 2298–2310.

Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., et al. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychol*, *15*, 136–146.

Kamachi, M., Hill, H., Lander, K., et al. (2003). "Putting the face to the voice": Matching identity across modality. *Curr Biol*, *13*, 1709–1714.

Klatt, D. H. (1987). Review of text-to-speech conversion for English. *J Acoust Soc Am*, *82*, 737–793.

Klieve, S., & Jeanes, R. C. (2001). Perception of prosodic features by children with cochlear implants: Is it sufficient for understanding meaning differences in language? *Deaf Educ Int*, *3*, 15–37.

Lachs, L., & Pisoni, D. B. (2004). Cross-modal source information and spoken word recognition. *J Exp Psychol Hum Percept Perform*, *30*, 378–396.

Lander, K., Hill, H., Kamachi, M., et al. (2007). It's not what you say but the way you say it: Matching faces and voices. *J Exp Psychol Hum Percept Perform*, *33*, 905–914.

Loebach, J. L., Bent, T., Pisoni, D. B. (2008). Multiple routes to the perceptual learning of speech. *J Acoust Soc Am*, *124*, 552–561.

Loizou, P. C., Dorman, M. F., Powell, V. (1998). The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor. *J Acoust Soc Am*, *103*, 1141–1149.

Meister, H., Landwehr, M., Pyschny, V., et al. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol*, *48*, 38–48.

Most, T., & Peled, M. (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing aids. *J Deaf Stud Deaf Educ*, *12*, 350–361.

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

Nygaard, L. C., Sommers, M. S., Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychol Sci*, *5*, 42–46.

Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., et al. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus-22 channel cochlear implant. *Ear Hear, 12*(suppl 4), 66–80.

Peng, S.-C., Tomblin, J. B., Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hear*, *29*, 336–351.

Perrachione, T., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*, 1899–1910.

Read, D., & Craik, F. I. M. (1995). Earwitness identification: Some influences on voice recognition. *J Exp Psychol Appl*, *1*, 6–18.

Remez, R. E., Fellowes, J. M., Rubin, P. E. (1997). Talker identification based on phonetic information. *J Exp Psychol Hum Percept Perform*, *23*, 651–666.

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

Schafer, E. C., & Thibodeau, L. M. (2006). Speech recognition in noise in children with cochlear implants while listening in bilateral, bimodal, and FM-system arrangements. *Am J Audiol*, *15*, 114–126.

Shannon, R. V., Fu, Q. J., Galvin, J. J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol Suppl*, *552*, 1–5.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., et al. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *J Exp Psychol Hum Percept Perform*, *28*, 1447–1469.

Spence, M. J., Rollins, P. J., Jerger, S. (2002). Children's recognition of cartoon voices. *J Speech Lang Hear Res*, *45*, 214–222.

Staller, S. J., Dowell, R. C., Beiter, A., et al. (1991). Perceptual abilities of children with Nucleus 22-channel cochlear implant. *Ear Hear, 12*(suppl. 4), 34–47.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., et al. (2000). Language development in profoundly deaf children with cochlear implants. *Psychol Sci*, *11*, 153–158.

Trainor, L. J., Austin, C. M., Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol Sci*, *11*, 188–195.

Van Lancker, D. R., Kreiman, J., Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters: Part I—Recognition of backward voices. *J Phon*, *13*, 19–38.

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G. (2006). Song recognition by children and adolescents with cochlear implants. *J Speech Lang Hear Res*, *49*, 1091–1103.

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G. (2009). Identification of TV tunes by children with cochlear implants. *Music Percept*, *27*, 17–24.

Vongphoe, M., & Zeng. F. G. (2005). Speaker recognition with temporal cues in acoustic and electric hearing. *J Acoust Soc Am*, *118*, 1055–1061.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *J Acoust Soc Am*, *54*, 1248–1266.

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychol Aging*, *15*, 88–99.