Age-related changes in talker recognition with reduced spectral cues

Tara Vongpaisal, Sandra E. Trehub,[a] E. Glenn Schellenberg, and Pascal van Lieshout
Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

(Received 16 March 2010; revised 16 February 2011; accepted 18 November 2011)

Temporal information provided by cochlear implants enables successful speech perception in quiet, but limited spectral information precludes comparable success in voice perception. Talker identification and speech decoding by young hearing children (5–7 yr), older hearing children (10–12 yr), and hearing adults were examined by means of vocoder simulations of cochlear implant processing. In Experiment 1, listeners heard vocoder simulations of sentences from a man, woman, and girl and were required to identify the talker from a closed set. Younger children identified talkers more poorly than older listeners, but all age groups showed similar benefit from increased spectral information. In Experiment 2, children and adults provided verbatim repetition of vocoded sentences from the same talkers. The youngest children had more difficulty than older listeners, but all age groups showed comparable benefit from increasing spectral resolution. At comparable levels of spectral degradation, performance on the open-set task of speech decoding was considerably more accurate than on the closed-set task of talker identification. Hearing children's ability to identify talkers and decode speech from spectrally degraded material sheds light on the difficulty of these domains for child implant users. © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3669978]

PACS number(s): 43.71.Bp, 43.71.Ky, 43.71.Ft, 43.66.Ts [JES] Pages: 501–508

I. INTRODUCTION

Talker recognition remains a major challenge for cochlear implant (CI) users, even for those who are capable of good speech perception and production (Cleary and Pisoni, 2002; Loizou *et al.,* 1998; Vongphoe and Zeng, 2005). Such difficulty is attributed primarily to the limited temporal fine structure or spectral detail transmitted by CIs; this precludes natural voice quality (Fu *et al.,* 2004). Unlike hearing adults, CI users have considerable difficulty differentiating same-gender voices in the context of single-syllable stimuli, even though they readily identify the syllables (Fu *et al.,* 2004; Vongphoe and Zeng, 2005). The presumption is that unique, or person-specific, spectro-temporal features enable hearing adults to identify talkers from intact or spectrally degraded syllables (O'Toole *et al.,* 2002; Remez *et al.,* 1997; Sheffert *et al.,* 2002). CI users have difficulty perceiving these brief spectro-temporal cues to identity.

Although single syllables have been the stimuli of choice in many adult studies of voice recognition, they have not been used in studies of talker recognition with children, perhaps because of their poor ecological validity. With sentence-length utterances, child CI users are usually able to differentiate male from female voices (Osberger *et al.,* 1991; Staller *et al.,* 1991; but see Kovačić and Balaban, 2009), which is in line with the adult CI findings with single syllables. Unlike hearing children, however, child CI users have difficulty differentiating unfamiliar female talkers from one another (Cleary and Pisoni, 2002). The situation is very different for highly familiar talkers for which CI children presumably have robust, long-term representations. Child CI users can differentiate their mother's voice from that of an unfamiliar man, woman, or child even when unfamiliar actors mimic the mother's prosody (Vongpaisal *et al.,* 2010). Although implanted children are not as accurate as their hearing counterparts in this regard, they perform well above chance levels. The implication is that they rely primarily on dynamic variations in articulation, secondarily on variations in speaking rate (Vongpaisal *et al.,* 2010), and little or not at all on the pitch and timbre differences traditionally associated with voice differentiation (Van Lancker and Kreiman, 1987).

Investigators have attempted to simulate the input available to CI users by means of *vocoders,* which extract the amplitude contour or temporal envelope of the acoustic waveform and replicate this profile on an alternative carrier such as noise bands (Shannon *et al.,* 1995) or sine waves (Dorman *et al.,* 1997). The vocoded stimuli provide a coarse representation of the signal that enables hearing listeners to decode speech under ideal listening conditions but not in noise (Dorman *et al.,* 1998). Adults have some difficulty differentiating talkers from vocoded syllables, but their difficulty is less than that experienced by implant users tested with non-vocoded stimuli (Fu *et al.,* 2004; Fu *et al.,* 2005; Vongphoe and Zeng, 2005). The use of vocoded sentences rather than syllables could shed light on children's identification of talkers. Accordingly, the present study examined children's and adults' ability to identify male, female, and child talkers from spectrally degraded sentences.

The available research indicates that talker identification with vocoded stimuli requires greater spectral and temporal detail than that required for word recognition with the same stimuli (Fu *et al.,* 2004; Vongphoe and Zeng, 2005). Although four spectral bands are adequate for decoding speech in quiet (Shannon *et al.,* 1995), they are inadequate for voice

[a]Author to whom correspondence should be addressed. Electronic mail: sandra.trehub@utotoronto.ca

recognition. The number of spectral channels required for voice recognition seems to depend on the available temporal cues (Fu *et al.*, 2004; Fu *et al.*, 2005). When gender differences are distinctive (e.g., F0 separation of approximately an octave, or 12 semitones), hearing individuals can achieve near-perfect gender discrimination from simulations with 16 channels, which permit the resolution of formant structures. When the spectral differences between talkers are more modest, talker differentiation becomes increasingly difficult (Fu *et al.*, 2005). Fu *et al.* (2005) suggest that CI users' discrimination of male from female voices is equivalent to the performance of hearing listeners with vocoded stimuli constructed with four to eight spectral bands.

Vocoder simulations indicate that men's vowels are easier to decode than those of women and boys, which are easier to decode than those of girls (Loizou *et al.*, 1998). The greater ease of decoding men's speech is attributable to better preservation of formant frequencies in the narrower bandwidths of the lower frequency channels of commercial devices. Analyses of channel output patterns indicate that difficulty with female voices stems from poorer resolution of formant frequencies in the wider bandwidths of the higher frequency channels (Loizou *et al.*, 1998).

Voice and speech decoding have been examined in tasks featuring voice recognition training. Even after considerable training on vowel tokens from 10 different talkers (three men, two boys, three women, two girls), adult CI users only managed to achieve 20% correct performance in a 10-alternative forced-choice task (Vongphoe and Zeng, 2005). Although their performance was above chance levels, it contrasted markedly with that of hearing adults, who achieved over 80% correct on unprocessed stimuli and over 40% correct on vocoded stimuli with eight spectral bands. In such contexts, hearing adults' performance improved with training, but CI users' performance did not. In fact, CI users' talker identification was comparable to the performance of hearing adults on simulated voices involving a single spectral band. For CI users, within-gender confusions were considerably greater than cross-gender confusions. In contrast to poor performance on talker identification, their performance on vowel identification for the same single-syllable stimuli was comparable to that of hearing adults on processed stimuli with eight spectral bands. In short, talker recognition is more compromised than speech recognition under comparable degrees of spectral degradation. It is likely that the short-duration samples (e.g., single syllables) in previous research exacerbated the voice-processing difficulties of CI users, who exhibit better performance with sentence-length samples (Vongpaisal *et al.*, 2010).

Little is known about the consequences of spectral degradation on talker identification by hearing children. What is clear is that 5- to 7-yr-old children require more spectral detail than 10- to 12-yr-old children and adults to decode vocoded sentences, words, and phonemes; this is attributed, in part, to younger listeners' poor use of contextual cues (Eisenberg *et al.*, 2000). There are also suggestions that hearing children require F0 and formant-frequency differences of two semitones or more to ascertain whether pairs of spoken sentences emanate from the same talker or from different talkers (Cleary *et al.*, 2005). Children 5-8 yr of age also require greater frequency separation (11 semitones or more) than 9- to 11-yr-olds and adults to segregate tone sequences into separate streams on the basis of frequency similarity (Sussman *et al.*, 2007). We know, however, that 5-yr-old children can identify the direction of pitch shifts in tone sequences on the basis of differences as small as 0.3 semitones and that 8-yr-old children can do so with differences of 0.1 semitones (Stalinski *et al.*, 2008). Although absolute sensitivity does not become adult-like until the school years (Trehub *et al.*, 1988), selective attention difficulties and inflexible use of cues are greater impediments to optimal processing of complex auditory information in early childhood (Hazan and Barrett, 2000; Nittrouer, 2005; Nittrouer *et al.*, 2000; Stollman *et al.*, 2004; Wightman and Kistler, 2005). Accordingly, we predicted that 5- to 7-yr-old hearing children would have more difficulty identifying talkers than older children and adults.

As noted, evidence from adult CI recipients and from vocoder simulations indicates that talker identification is severely compromised by reduced spectral input. These adverse effects may be magnified in young children. Vongpaisal *et al.* (2010) demonstrated, however, that child CI users succeeded in differentiating their mother's voice from those of other women whose mean F0 differed by as little as 1 semitone. These findings are at odds with the voice discrimination difficulties of adult CI users and hearing adults with simulations involving F0 differences smaller than 12 semitones (Fu *et al.*, 2005; Vongphoe and Zeng, 2005). The unexpectedly good performance of child CI users implies that they capitalized on temporal cues and on gross spectral cues in the long-term speech spectrum. These cues may be equally informative for hearing children in the context of reduced spectral cues.

In the present investigation, we sought to ascertain the effects of age and spectral degradation on talker identification and speech decoding in the context of sentence-length samples. In line with previous findings from child CI users (Vongpaisal *et al.*, 2010), hearing children as well as adults were expected to differentiate cross-gender, cross-age, and same-gender talkers from spectrally degraded sentences. On the basis of adults' difficulty identifying voices from spectrally degraded syllables (Fu *et al.* 2004, 2005; Vongphoe and Zeng, 2005), children were expected to have comparable, if not greater, difficulty in the context of such stimuli. As noted, however, they were expected to benefit from the use of sentence-length stimuli. As in previous research with adults (Vongphoe and Zeng, 2005), spectral degradation was expected to have more adverse consequences for talker recognition than for speech decoding. Moreover, the adverse effects were expected to be greater for younger than for older children (Eisenberg *et al.*, 2000).

II. EXPERIMENT 1

Children and adults were required to identify the talkers (man, woman, and girl) of sentences presented at various degrees of spectral degradation. Better performance was expected for older than for younger children, for male than

for female talkers, and when more spectral information was available.

A. Method

1. Participants

The participants were children 5–7 yr of age ($M = 6.5$, SD = 1.0, $n = 19$), children 10–12 yr of age ($M = 11.3$, SD = 0.82, $n = 22$), and college students ($M = 21.2$, SD = 2.0, $n = 16$). No hearing tests were administered. However, criteria for inclusion in the sample included no colds on the day of testing (or very recent colds), no family history of hearing impairment, and no personal history of ear infections or hearing difficulties as indicated by parental report in the case of children and self-report for adults.

2. Apparatus

Testing was conducted in a double-walled sound-attenuating booth. A PC workstation and amplifier (Harmon-Kardon 3380) located outside of the booth interfaced with a 17-in. touch screen monitor (Elo LCD TouchSystems) and two loudspeakers (Electro-Medical Instrument Co.) inside the booth. The loudspeakers were mounted at the corners of the booth, each at 45° azimuth to the participant with the touch-screen monitor at the midpoint.

3. Stimuli

Sentence-length recordings of a man, woman, and girl from Vongpaisal *et al.* (2010) were used in the present experiment (see Table I). Mean F0 of the talkers was 134.0 Hz (SD = 12.7 Hz), 241.9 Hz (SD = 32.0), and 258.5 Hz (SD = 7.2), for the man, woman, and girl, respectively. Mean sentence duration for the man, woman, and girl was 1.02 (SD = 0.22), 1.15 (SD = 0.21), and 1.20 (SD = 0.26) seconds, respectively. Although the man's average speaking rate was somewhat faster than the other two talkers, duration differences among talkers were not significant, $F(2, 29) = 1.16$, $P > 0.2$.

The present noise-band vocoder (Sheldon *et al.*, 2008) was implemented according to the general principles outlined by Shannon *et al.* (1995) and Eisenberg *et al.* (2000). Sound files were converted to binary form and processed through a pre-emphasis filter implemented in MATLAB. They were then passed through a series of bandpass filters (4, 8, 16, 32, and 64 bands for increasing spectral resolution), with the series spanning a 300- to 6000-Hz frequency range. Figure 1 illustrates the cross-over frequencies and bandwidths of the filters in the five spectral conditions. The time-amplitude envelope of the acoustic signal was detected and extracted using the Hilbert transform. The temporal envelope was used to modulate narrow-band Gaussian white noise. The product was subsequently processed through the original filters used to analyze the input signal. The outputs of these filters were summed to form a noise-band signal that preserved the same temporal envelope and amplitude profile of the original signal, but the fine structure of the acoustic and vocoded signals was uncorrelated. Figure 2 shows the waveforms, envelope extraction, and spectrograms of natural and vocoded samples. All sound files were equated for amplitude and presented through loudspeakers at 65 dB A-level, as measured at the ear level of participants.

TABLE I. Scripted common phrases recorded by the 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.

4. Procedure

Participants were told that the voices of a man, woman, and girl would be presented and that many of the voices had been altered so that they sounded unnatural or strange. Their task was to indicate whether the talker was a man, woman, or girl by selecting the appropriate picture on the touch-screen monitor. Each spectral-band condition (including the unaltered speech samples) was presented in blocks of 15 trials, with five speech samples selected randomly from each actor. The order of these six blocks and the order of stimuli within were randomized separately for each participant.

The experimenter was seated next to child participants during the test session. When the child was ready to listen to a sound sample, he or she pressed a "play" button below a display of faces of a man, woman, and girl. Each sample was presented once except for a few instances when a child was distracted during a trial. After hearing a voice sample, children judged the identity of the talker by selecting one of the three faces on the monitor. Feedback was provided in the form of a schematic happy face for a correct response and a blank screen for an incorrect response. Such feedback was

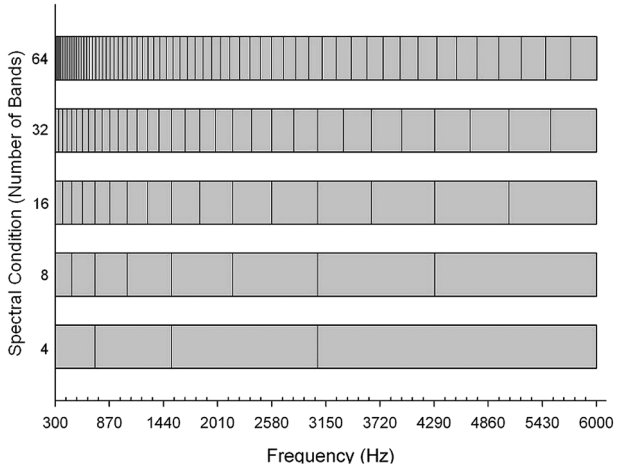

FIG. 1. Cross-over frequencies of channels in each spectral condition.

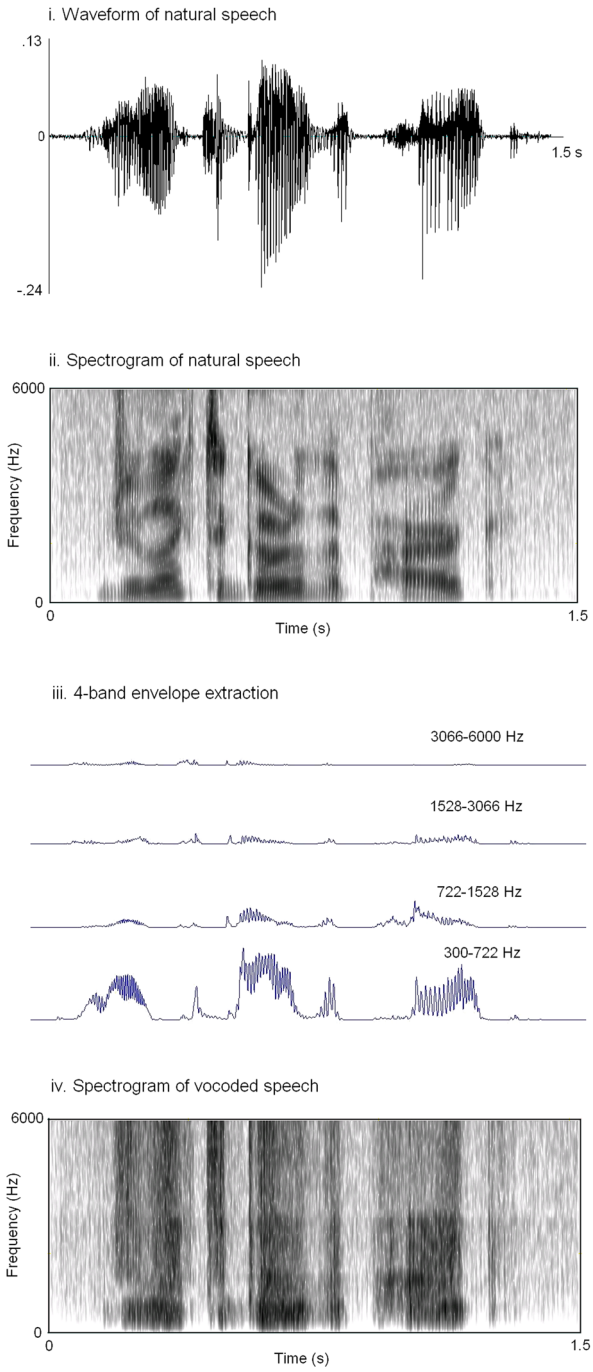

FIG. 2. (Color online) Natural to vocoded voices: (i) Waveform of unprocessed voice, (ii) spectrogram of unprocessed voice, (iii) 4-band envelope extraction, (iv) spectrogram of vocoded voice.

provided to motivate children and to provide guidance about what to listen for. To proceed to the next trial, the child pressed a "continue" button on the monitor. Adults received the same instructions as the children and were tested in the same manner except that the experimenter remained outside the sound-attenuating booth.

B. Results and discussion

Each participant had 18 scores: one for each of the three talkers for each of the six blocks, with each score ranging from zero (no correct responses) to five (perfect identification). For

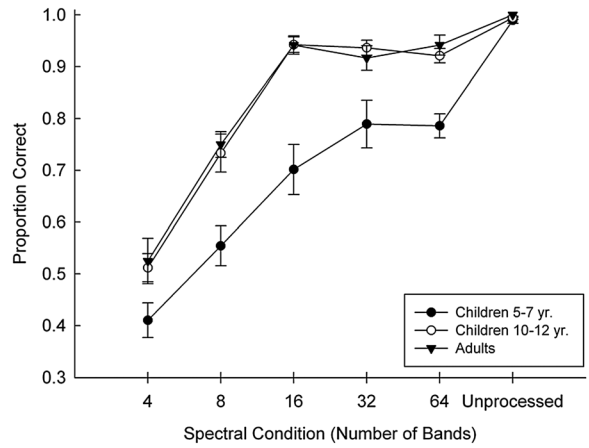

FIG. 3. Voice identification accuracy of young children, older children, and adults across spectral conditions. Error bars indicate standard errors.

ease of interpretation, each score was converted to a proportion by dividing the raw score by five. Figures 3 and 4 illustrate the performance of younger children, older children, and adults as a function of spectral-band condition and talker. All groups performed at ceiling on the unprocessed voices and above chance levels in all conditions of spectral degradation (all P levels < 0.05). Except for three of the youngest children and one older child, who made one error each, participants were 100% correct on the unprocessed voices. Accordingly, we removed this condition from subsequent analyses.

A three-way mixed-design ANOVA, with age group (younger children, older children, and adults) as a between-subjects factor and spectral information (4, 8, 16, 32, 64 bands) and talker (man, woman, girl) as within-subjects factors, revealed significant main effects of age, $F(2, 54) = 24.63$, $P < 0.0001$, spectral information, $F(4, 216) = 127.76$, $P < 0.0001$, and talker, $F(2, 108) = 39.00$, $P < 0.0001$. Follow-up tests (Tukey HSD) revealed that younger children performed significantly more poorly than older children and adults (P levels < 0.001), who did not differ from one another ($P > 0.9$). Performance improved from 4 to 8 to 16 spectral bands (Bonferroni-corrected P levels < 0.001), where it reached a plateau. Moreover, talker identification accuracy was greater for the man than for the woman and girl (P levels < 0.001).

The main effects were qualified by significant two-way interactions between age group and talker, $F(4, 108) = 4.08$, $P < 0.005$, and between spectral information and talker, $F(8, 432) = 7.21$, $P < 0.0001$. There was no two-way interaction between spectral information and age group and no three-way interaction. Because both of the significant interactions involved the talker, follow-up analyses examined performance for each talker separately with a two-way (age × spectral band) mixed-design ANOVA. The main effect of age was significant for each talker, Fs$(2, 54) = 6.21$, 12.48, and 13.09, for the man, woman, and child, respectively, P levels < 0.005. For each talker, younger children were less accurate than older children and adults. The interaction between age and talker stemmed from the smaller difference between the youngest group and the other two groups on the man's voice than on the woman's and girl's voices. The main effect

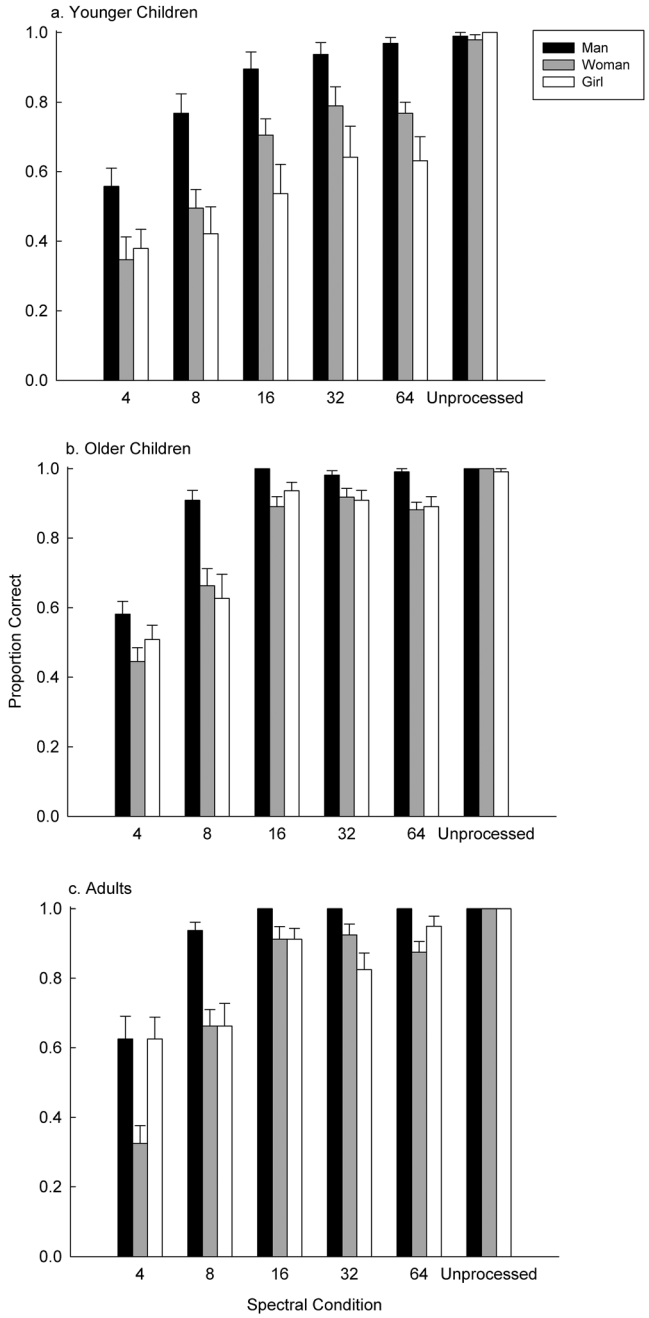

FIG. 4. Voice-identification accuracy of young children, older children, and adults on each talker category across spectral conditions. Error bars indicate standard errors.

of spectral information was also significant for each talker, Fs(4, 216) = 86.04, 83.35, and 34.09, for the man, woman, and girl, respectively, P levels < 0.0001. The interaction between spectral information and talker was a consequence of a larger effect of the spectral manipulation on the man's and woman's voices than on the girl's voice. Specifically, the performance advantage for 8 over 4 spectral bands was evident only for the man and woman talkers.

The possibility of age-related differences in improvement across trials was also explored. Figure 5 shows mean scores of younger and older children in the first block (trials 1–15) and last block (trials 60–75) of the test session. A two-way mixed (age × block) ANOVA revealed that older

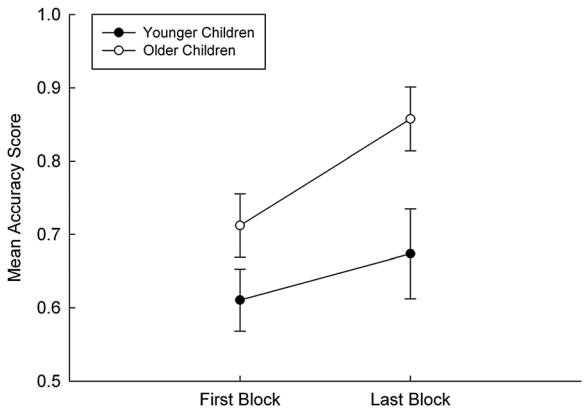

FIG. 5. First and last block mean accuracy scores of younger and older children.

children outperformed younger children, $F(1, 39) = 8.47$, $P < 0.01$, and that accuracy on the last block was higher than on the first block, $F(1, 39) = 4.98$, $P < 0.05$, but the interaction between these factors was not significant, $F < 1$. Taken together, both age groups showed similar improvement across trials, but older children were better able to capitalize on the available cues to outperform younger children.

In short, the results of the present experiment indicated that (1) younger children were relatively poor at identifying talkers when the speech samples were spectrally degraded, (2) the man was easier to identify than the woman and girl, and (3) spectral degradation was equally detrimental to talker identification for all three age groups. The relative ease of identifying the man also meant that the performance decrement for the youngest listeners was attenuated for the man relative to the other talkers.

III. EXPERIMENT 2

As noted, there is relatively limited information about age-related changes in decoding spectrally degraded sentences (Eisenberg *et al.*, 2000). It was of particular interest to examine such decoding skills among hearing listeners with sentences of comparable vocabulary, structure, and levels of degradation as those in Experiment 1. In line with the findings of Eisenberg *et al.* (2000), the results of Experiment 1, and other research on complex auditory processing (e.g., Stalinski *et al.*, 2008), older children and adults were expected to decode spectrally degraded speech more accurately than younger children.

A. Method

1. Participants

As in Experiment 1, participants had no current or recent colds and no individual or family history of hearing problems. The participants were young children 5–7 yr of age ($M = 6.5$, SD = 0.6, $n = 20$), older children 10–12 yr of age ($M = 11.3$, SD = 0.8, $n = 20$), and college students ($M = 21$, SD = 3.6, $n = 9$).

2. Apparatus

The apparatus was identical to that of Experiment 1.

TABLE II. Sample sentences from the Common Phrases Test.

Sentence
1. When is your birthday?
2. I like ice cream.
3. Wait for me!
4. Open the door.
5. What is your favorite TV show?
6. I'm fine.
7. What did you eat for breakfast?
8. Clap your hands.
9. Clean your room
10. I'll call you.

3. Stimuli

Recordings of sentences from the Common Phrases Test (DeVault Otologic Research Laboratory) were made by the talkers in Experiment 1. The Common Phrases Test is a corpus of 60 simple statements and questions developed for assessing decoding accuracy in young children with profound hearing loss. Sentences from this corpus, which had two to six words ($M = 3.7$, SD = 1.3), were selected because of their similarity in structure and vocabulary level to the sentences used in Experiment 1. Sample sentences from the corpus are shown in Table II. Fifteen sentences (five each for the man, woman, and girl actors) were selected randomly for each of the vocoded conditions (4, 8, 16 spectral bands), and 15 were left intact for the unprocessed condition. We used the vocoding method described in Experiment 1 to generate speech samples with 4, 8, and 16 spectral bands. Stimuli were presented at 65 dB A-level.

4. Procedure

To familiarize participants with the sound of vocoded voices, six samples were generated for a brief practice session. The specific sentences and the levels of spectral degradation differed from those used at test. As in the test session that followed, participants were instructed to repeat as much of the sentences as they could. In the test session, participants repeated the 60 test stimuli, which were presented in random order. Their responses were recorded with a microphone (Sony F-V30T) connected directly to a PC workstation. High-quality digital sound files (44.1 kHz, 16-bit) were created with sound-editing software (Sound Forge 6.0) and saved for off-line scoring.

B. Results and discussion

Accuracy (i.e., average proportion correct) was calculated separately for each participant and stimulus. For each stimulus, the number of correctly repeated whole words was divided by the total number of words. Figure 6 illustrates the performance of younger children, older children, and adults as a function of the number of spectral bands. Figure 7 shows these scores for each talker category. Because performance was at ceiling for unprocessed voices across age groups and talkers, scores from this condition were not included in the analyses.

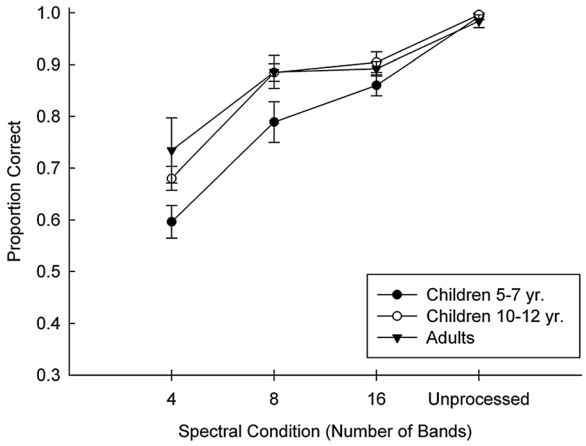

FIG. 6. Speech decoding performance of young children, older children, and adults as a function of spectral condition. Error bars indicate standard errors.

Response patterns were examined with a three-way mixed-design ANOVA, with age group (younger children, older children, and adults) as a between-subjects factor, and spectral information (4, 8, and 16 bands) and talker (man, woman, and girl) as within-subjects factors. Each of the three main effects was significant. Differences due to age group, $F(2, 46) = 5.14$, $P < 0.01$, were the consequence of poorer performance for younger children than for older children and adults (P levels < 0.05), who did not differ from one another ($P > 0.05$). The effect of spectral bands, $F(2, 92) = 62.87$, $P < 0.0001$, was a consequence of poorer performance in the 4-band condition than in the 8- and 16-band conditions (P levels < 0.001), which did not differ ($P > 0.10$). Finally, the talker effect, $F(2, 92) = 39.22$, $P < 0.0001$, revealed that the girl's utterances were less intelligible than those of the man and woman (P levels < 0.001), who were decoded with similar accuracy ($P > 0.05$). The only significant interaction was between spectral information and talker, $F(4, 184) = 3.42$, $P < 0.05$. Separate analyses for each of the three talkers revealed that the spectral-band manipulation was significant for each voice, Fs(2, 92) = 11.37, 58.52, and 14.67, for the man, woman, and girl, respectively, P levels < 0.001. In each instance, performance on the 4-band stimuli was poorer than performance on the 8- and 16-band stimuli, which did not differ from one another. The interaction stemmed solely from the

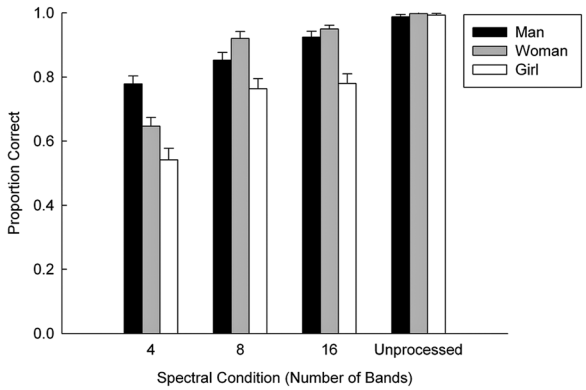

FIG. 7. Speech decoding performance on each talker as a function of spectral condition. Error bars indicate standard errors.

greater decrement in decoding the woman's speech at four spectral bands than the man's or girl's speech. In other words, the woman's speech showed the greatest benefit from increases in spectral information. Although one would expect similar consequences of spectral degradation on the woman and girl's speech, it is possible that more precise articulation or expressive prosody enhanced decoding accuracy of the woman's speech.

Comparisons of performance in Experiments 1 and 2 revealed that the adverse consequences of spectral degradation were much greater for talker identification than for speech decoding. Although the speech decoding scores were approximately 20% higher than the talker identification scores in the four-band condition, such comparisons grossly underestimate the differences. Chance-level responding (i.e., guessing) in the three-alternative, forced-choice task of Experiment 1 was 33.3% correct in contrast to a chance level near zero in the open-set, speech-decoding task. Considered together, the findings from Experiments 1 and 2 indicate that (1) compared to older children and adults, younger children have more difficulty decoding words and identifying talkers from spectrally degraded speech, (2) the three age groups benefit similarly from increased spectral information, (3) increased spectral information is more advantageous for decoding men's and women's speech than girls' speech, and (4) spectral degradation is much more detrimental to talker identification than to speech decoding.

IV. GENERAL DISCUSSION

The principal goal of the present study was to examine effects of age and spectral degradation on talker identification. In Experiment 1, 5-to 7-yr-old children, 10-to 12-yr-old children, and adults were required to indicate whether the talker of each of several sentences presented at different levels of spectral degradation was a man, woman, or girl. In Experiment 2, the same age groups were required to decode sentences by the same talkers and at comparable levels of spectral degradation.

With respect to talker identification, younger children performed more poorly than older children and adults, but listeners of all ages derived similar benefit from increases in spectral information. Younger children's near-perfect accuracy on natural or unprocessed stimuli indicates that their poor performance on the processed stimuli stemmed from reduced spectral cues in voices rather than from talker-identification difficulties in general. Nevertheless, younger as well as older listeners performed above chance levels, even on stimuli with only four spectral bands. Accuracy for all age groups improved up to 16 bands, after which no further improvement was evident. Examination of talker recognition over the course of the test session indicated that both groups showed similar improvement. Older children were better able, however, to capitalize on the available cues to outperform their younger counterparts across conditions and test trials.

Reduced spectral cues had differential consequences across talkers and age groups. Specifically, identification of the man and woman improved with a twofold increase in the number of available spectral channels (from 4 to 8), but improvement in identifying the girl required a fourfold increase in spectral information. Age differences in identifying talkers from spectrally impoverished voices were smallest for the man and more substantial for the woman and girl.

Younger children's difficulty with talker identification in the context of spectrally degraded speech parallels their difficulty decoding such speech (Eisenberg *et al.,* 2000). Moreover, the relative ease with which listeners of all ages identified the man is in line with the findings of Loizou *et al.* (1998) with adult listeners. The woman's and girl's voices were more similar to each other in fundamental frequency than they were to the man's voice; this contributed to the difficulty of differentiating those voices. Moreover, spectral information from female voices is thought to be less well preserved than that of male voices in the input provided by conventional CIs or vocoders (Loizou *et al.,* 1998).

Difficulty differentiating women from girl talkers is consistent with previously reported problems with same-gender differentiation by child CI users (Cleary and Pisoni, 2002; Vongpaisal *et al.,* 2010) and by hearing adults with vocoder simulations (Loizou *et al.,* 1998). It is also consistent with greater difficulty decoding women's utterances than men's utterances, presumably because of poorer resolution of formant frequencies in high-frequency channels (Green *et al.,* 2007).

With respect to speech decoding, young children fared poorly compared to older children and adults, but all groups showed similar benefit from increasing spectral information. Young children's ability to achieve near-perfect accuracy on the unprocessed sentences indicates that their difficulty decoding the sentences did not stem from working-memory limitations but rather from their difficulty with reduced spectral cues. This limitation on the part of young children is consistent with the findings of Eisenberg *et al.* (2000). It also echoes the age-related differences in voice identification in Experiment 1 of the present study. An important difference between the findings of Experiments 1 and 2 was that participants required 8 channels to reach a performance plateau on speech decoding instead of the 16 required for a plateau in talker identification.

As was the case for voice identification, spectral degradation had more adverse consequences on decoding the girl's utterances than the man's or woman's utterances. Moreover, performance on the girl's utterances benefited least from increases in spectral information. Because the woman and girl had a similar speaking rate and F0, it is unlikely that these factors contributed to the observed differences in decoding. Instead, speech clarity or more precise articulation may account for greater intelligibility of the woman's speech under conditions of degradation. Specifically, the woman may have been more sensitive than the girl to instructions to speak in a child-directed style. Attempts to maximize clarity, as when speaking in noisy conditions, result in speech that is more intelligible than otherwise (Goy *et al.,* 2007; Van Summers *et al.,* 1988).

Spectral degradation was much less detrimental to speech decoding than it was to talker identification for all age groups. Despite the considerably greater difficulty of open-set responding relative to closed-set responding, absolute levels of accuracy were substantially greater for speech decoding than for talker identification for all age groups. The

sizable discrepancy between talker identification and speech decoding in the context of spectral degradation is consistent with the relative difficulty of these processes for adult implant users (Fu *et al.*, 2004; Vongphoe and Zeng, 2005).

To some extent, results from the simulations in Experiment 1 provide insight into the voice processing deficits of child CI users. Undoubtedly, sentence-length stimuli, as well as the simple vocabulary and simple sentence structures of materials in the present study resulted in more favorable test conditions than those in previous studies of adult CI users' identification of isolated syllables.

In short, the present findings indicate much graver consequences of spectral degradation for talker identification than for speech decoding. They also draw attention to the possibility that implant simulations—vocoded speech, specifically—provide an overly conservative estimate of child CI users' processing of speech and of talker identity in particular. CI users' greater experience with spectrally degraded input may enable them to make good use of timing cues in the signal to exceed the performance levels predicted by vocoder simulations. Indeed, comparisons of the present data on talker identification with data from child implant users on the same stimuli and task reveal unexpected strengths on the part of implant users (Vongpaisal, 2009). An important challenge for future research is the development of more precise models of talker identification in implant users.

ACKNOWLEDGMENTS

Funding was provided by the Natural Sciences and Engineering Research Council of Canada. Ewan Macdonald provided assistance in the preparation of vocoded speech samples.

Cleary, M., and 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., and 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.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Dorman, M. F., Loizou, P. C., and 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," J. Acoust. Soc. Am. **102**, 2403–2411.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **107**, 2704–2710.

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

Goy, H., Pichora-Fuller, K., Van Lieshout, P., Singh, G., and Schneider, B. (**2007**). "Effects of within- and between-talker variability on word identification in noise by younger and older adults," in *Eighth Annual Conference of the International Speech Communication Association (Interspeech 2007)*, edited by Van hamme, H. and van Son, R. (Causal Productions, Adelaide), pp. 418–421.

Green, T., Katiri, S., Faulkner, A., and Rosen, S. (**2007**). "Talker intelligibility differences in cochlear implant listeners," J. Acoust. Soc. Am. **121**, EL223–EL229.

Hazan, V., and Barrett, S. (**2000**). "The development of phonemic categorization in children aged 6-12," J. Phon. **28**, 377–396.

Kovačić, D., and Balaban, E. (**2009**). "Voice gender perception by cochlear implantees," J. Acoust. Soc. Am. **126**, 762–775.

Loizou, P. C., Dorman, M. F., and 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.

Nittrouer, S. (**2005**). "Age-related differences in weighting and masking of two cues to word final stop voicing in noise," J. Acoust. Soc. Am. **118**, 1072–1088.

Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). "The effect of segmental order on fricative labeling by children and adults," Percept. Psychophys. **62**, 266–284.

Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., Kemink, J. L., Stroer, B. S., Firszt, J. B., and Novak, M. A. (**1991**). "Independent evaluation of the speech perception abilities of children with the Nucleus-22 channel cochlear implant," Ear Hear. **12**(Suppl. 4), 66–80.

O'Toole, A. J., Roark, D. A., and Abdi, H. (**2002**). "Recognizing moving faces: A psychological and neural synthesis," Trends Cogn. Sci. **6**, 261–266.

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

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

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

Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (**2008**). "Priming and sentence context support listening to noise-vocoded speech by younger and older adults," J. Acoust. Soc. Am. **123**, 489–499.

Stalinski, S. M., Schellenberg, E. G., and Trehub, S. E. (**2008**). "Developmental changes in the perception of pitch contour: Distinguishing up from down," J. Acoust. Soc. Am. **124**, 1759–1763.

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

Stollman, M. H. P., van Velzen, E. C. W., Simkens, H. M. F., Snik, A. F. M., and van den Broek, P. (**2004**). "Development of auditory processing in 6-12-year old children: A longitudinal study," Int. J. Audiol. **43**, 33–44.

Sussman, E., Wong, R., Horváth, J., Winkler, I., and Wang, W. (**2007**). "The development of the perceptual organization of sound by frequency separation in 5-11 year-old children," Hear. Res. **225**, 117–127.

Trehub, S. E., Schneider, B. A., Morrongiello, B. A., and Thorpe, L. A. (**1988**). "Auditory sensitivity in school-age children," J. Exp. Child Psychol. **46**, 273–285.

Van Lancker, D. R., and Kreiman, J. (**1987**). "Voice discrimination and recognition are separate abilities," Neuropsychologia **25**, 829–834.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.

Vongpaisal, T. (**2009**). "Children's perception of speaker identity from spectrally degraded input," Ph.D. thesis, University of Toronto, Canada.

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G., van Lieshout, P. H. H. M., and Papsin, B. (**2010**). "Children with cochlear implants recognize their mother's voice," Ear Hear. **31**, 555–566.

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

Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.