Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects[a)]

Helen E. Cullington[b)] and Fan-Gang Zeng
Hearing and Speech Laboratory, University of California, Irvine, 364 Med Surge II, Room 315, Irvine, California 92697

(Received 30 June 2006; revised 12 October 2007; accepted 12 October 2007)

Cochlear-implant users perform far below normal-hearing subjects in background noise. Speech recognition with varying numbers of competing female, male, and child talkers was evaluated in normal-hearing subjects, cochlear-implant users, and normal-hearing subjects utilizing an eight-channel sine-carrier cochlear-implant simulation. Target sentences were spoken by a male. Normal-hearing subjects obtained considerably better speech reception thresholds than cochlear-implant subjects; the largest discrepancy was 24 dB with a female masker. Evaluation of one implant subject with normal hearing in the contralateral ear suggested that this difference is not caused by age-related disparities between the subject groups. Normal-hearing subjects showed a significant advantage with fewer competing talkers, obtaining release from masking with up to three talker maskers. Cochlear-implant and simulation subjects showed little such effect, although there was a substantial difference between the implant and simulation results with talker maskers. All three groups benefited from a voice pitch difference between target and masker, with the female talker providing significantly less masking than the male. Child talkers produced more masking than expected, given their fundamental frequency, syllabic rate, and temporal modulation characteristics. Neither a simulation nor testing in steady-state noise predicts the difficulties cochlear-implant users experience in real-life noisy situations.
 [DOI: 10.1121/1.2805617]

PACS number(s): 43.71.Ky, 43.71.Gv, 43.66.Dc, 43.66.Sr [KWG] Pages: 450–461

I. INTRODUCTION

Speech recognition in background noise depends on the properties of the interfering sounds. It is usually characterized in terms of the subject's speech reception threshold (SRT): signal-to-noise ratio (SNR) at which they score 50% correct. In test situations, a steady-state masking noise is often used. However, fluctuating backgrounds are much more common in real life, and most often speech is heard against a background of other speech. Although in normal-hearing subjects the SRT decreases when the masker has temporal fluctuations, most hearing-impaired subjects show very little difference for a fluctuating and steady-state masker (Drullman and Bronkhorst, 2004; Duquesnoy, 1983; Festen and Plomp, 1990; Hawley *et al.*, 2004; Miller, 1947; Peters *et al.*, 1998; Summers and Molis, 2004; Wagener and Brand, 2005), including those using a cochlear implant (Zeng *et al.*, 2005). Cochlear-implant users may even show negative effects of modulated maskers (Nelson *et al.*, 2003). Normal-hearing subjects appear to be able to take advantage of listening in the gaps which occur when the level of the competing speech is low, for example in pauses between words, or during the production of low-energy sounds like m, n, k, or p (Peters *et al.*, 1998). This allows brief glimpses of the target speech and leads to improved SRTs. Hearing-impaired subjects are usually unable to utilize glimpsing; this discrepancy is not due to inaudibility (Summers and Molis, 2004) or subject age (Festen and Plomp, 1990). Suprathreshold differences like reduced frequency selectivity may be involved (Peters *et al.*, 1998). It has been suggested that cochlear-implant users' difficulty understanding speech in modulated noise may be related to reduced spectral information (Fu *et al.*, 1998) and problems fusing auditory information across temporal gaps (Nelson and Jin, 2004).

Normal-hearing subjects can also use differences in the voice fundamental frequency (F0) between target and masker to help segregate competing voices, resulting in better speech recognition when the F0 of the target voice differs from that of the masker voice (Brokx and Nooteboom, 1982; Brungart, 2001; Brungart *et al.*, 2001; Drullman and Bronkhorst, 2004). No such effect has been seen in cochlear-implant users (Stickney *et al.*, 2007; Stickney *et al.*, 2004) or normal-hearing subjects using a cochlear-implant simulation (Qin and Oxenham, 2003; Qin and Oxenham, 2005; Stickney *et al.*, 2007; Stickney *et al.*, 2004). The speech processing method only weakly encodes the F0, so it may be difficult to segregate voices on this basis, despite reasonably good F0 difference limens (around one semitone or less) in cochlear-

[a)]Portions of this work were presented in "Two's company; three's a crowd. Speech recognition with competing talkers: normally-hearing, cochlear implant and CI simulation subjects," American Auditory Society meeting, Arizona, March 2006.

[b)]Author to whom correspondence should be addressed. Electronic mail: hculling@uci.edu

implant simulation subjects with access to eight or more spectral bands (Carroll and Zeng, 2007; Qin and Oxenham, 2005).

Masking can be broadly divided into two types. *Energetic masking* results from competition between target and masker at the auditory periphery, i.e., overlapping excitation patterns in the cochlea or auditory nerve. *Informational masking* can be defined as the elevation of threshold caused by stimulus uncertainty (Durlach *et al.*, 2003). In the case of a speech target, this would suggest that the interfering talker is intelligible and so similar to the target speech that it becomes difficult for the subject to disentangle target and interfering speech. Energetic masking is believed to be a purely peripheral phenomenon; informational masking is thought to be related to central or attentional mechanisms (Durlach *et al.*, 2003; Watson and Kelly, 1981). The effects of purely energetic masking are well documented, and can be predicted using models such as the Speech Intelligibility Index or Articulation Index (French and Steinberg, 1947).

Informational masking is difficult to predict and document. When other talkers mask speech, there is probably a combination of energetic and informational masking occurring. One method that has been used to separate the two types of masking is reversed speech. When speech is time reversed, its long-term spectral content and the F0 remain unchanged; however, it contains no linguistic information above the phoneme level and thus should cause limited informational masking. Reversal of the temporal envelope though may increase forward masking due to the abrupt offsets (Rhebergen *et al.*, 2005). Some studies have shown that speech recognition is better in the presence of reversed compared with forward maskers (Rhebergen *et al.*, 2005; Summers and Molis, 2004; Trammell and Speaks, 1970). Duquesnoy (1983), however, found negligible difference. Another method to study informational masking is to minimize spectral overlap between the signal and masker, thus eliminating energetic masking. This can be done by presenting speech stimuli into nonoverlapping bands (Arbogast *et al.*, 2002). Brungart and colleagues specifically examined the role of informational masking in speech recognition in normal-hearing subjects. Significant differences in performance were found between talker maskers and noise maskers leading to the conclusion that, although energetic masking occurred, informational masking dominated performance (Brungart, 2001; Brungart *et al.*, 2001). Drullman and Bronkhorst (2004) had assumed that informational masking would reduce with more interfering talkers, until the SRT approached that for steady-state noise. This hypothesis was based on the idea that the spectral and temporal modulations in the masking signal would diminish with increasing numbers of talkers, and eventually approach the dynamics of steady-state noise. However, even with eight interfering talkers, they found poorer SRTs than for steady-state noise. Carhart *et al.* (1975) found that even 64 competing talkers gave more masking that steady-state noise, although informational masking was at its maximum with three competing talkers and thereafter decreased.

The aim of the current research was to investigate the performance of cochlear-implant users in real-life listening situations, in comparison to normal-hearing subjects. Speech recognition was measured in the presence of background talkers as a function of the number and characteristics of the competing voices. Target and maskers originated from the same location so that spatial release from masking was not considered; Arbogast *et al.* (2005) are among several researchers who have conducted work in this field. In addition, most cochlear-implant users listen with just one ear and therefore may be unable to exploit spatial release from masking. Three experiments were performed. The aim of the first experiment was to assess to what extent cochlear-implant users can obtain release from masking due to temporal and spectral fluctuations in the masker. This was done by examining the influence of masker type on the SRT using combinations of female, male, and child talkers, and steady-state noise as maskers. Normal-hearing and cochlear-implant simulation subjects were also evaluated as a control. The simulation subjects were included in an attempt to compensate for the disparity in age and other characteristics between the normal-hearing and cochlear-implant subjects. It is acknowledged, however, that a simulation does not exactly mimic the performance of cochlear-implant subjects, due to inherent differences between acoustic and electric stimulation. Results therefore should be viewed in terms of trends, rather than a quantitative estimate of cochlear-implant performance (Throckmorton and Collins, 2002). Additional results were collected on one implant subject who has virtually normal hearing in the contralateral ear. Comparison of his results between ears reflects only hearing capabilities and removes the effect of subject characteristics. In order to assess the influence of informational masking on the SRT, a second experiment was performed whereby normal-hearing subjects were tested with one and two talker maskers using both forward and time-reversed masker sentences. This was done in an attempt to resolve the conflicting results obtained by previous authors (Duquesnoy, 1983; Rhebergen *et al.*, 2005; Summers and Molis, 2004; Trammell and Speaks, 1970). The third experiment investigated further the masking effectiveness of a child's voice in normal-hearing subjects. Although previous research has used children as subjects in informational masking of speech experiments (Hall *et al.*, 2002; Johnstone and Litovsky, 2006), results have not been reported using children's voices as maskers. Results were examined in relation to F0, syllabic rate, and temporal modulation rate of the talkers.

II. EXPERIMENT 1: EFFECT OF MASKER TYPE ON THE SRT IN NORMAL-HEARING, COCHLEAR-IMPLANT, AND COCHLEAR-IMPLANT SIMULATION SUBJECTS

A. Methods

1. Test material

In all three experiments, the target material consisted of sentences drawn from the HINT database, spoken by a male talker. These comprise 25 phonemically balanced lists of ten sentences, with each sentence containing between three and seven words (mean=5.3, mode=5 words) (Nilsson *et al.*, 1994). The HINT sentences were designed to be scored as

correct if the subject repeats all of the words exactly correct, with the exceptions of article confusion (a/the/an) and tense for the verbs "to be" or "to have" (is/was, has/had, etc.). However, preliminary investigation with cochlear-implant subjects showed that most did not repeat the exact sentence word for word, even if they appeared to have clearly understood it. This may be a function of age or hearing impairment. Therefore, a loose keyword scoring method was adopted. The HINT sentences were developed from the Bamford-Kowal-Bench (BKB) sentences (Bench *et al.*, 1979), and most of the sentences are almost identical between the two databases. The BKB sentences are commonly scored for keywords (Bench and Bamford, 1979); it was therefore relatively easy to designate keywords for the HINT sentences. Three to five keywords (mean=3.3, mode=3 keywords) were identified for each sentence. In common with criteria often used for BKB sentences, if two or more keywords were repeated correctly, the sentence was considered correct (Blandy and Lutman, 2005). Loose keyword scoring was used, meaning that if the subject repeated the root of the keyword correctly, this would be considered correct; precise inflexion or word ending were not required. Loose keyword scoring is easier to apply, especially if there are difficulties understanding precisely the speech of the test subject (Foster *et al.*, 1993). No target sentence lists were repeated during the test session, as recommended to avoid familiarity (Foster *et al.*, 1993; Wagener and Brand, 2005). All test material was digitized with a sampling rate of 44.1 kHz, and comprised mono 16 bit resolution wav files.

Twenty different maskers were available to compete with the target talker; they were selected to represent various real-life competing talker situations. These are shown in Table I. The names of the masker conditions were abbreviated to represent the constituent talkers, for example, "m2f2" represented two males and two females. The abbreviations are listed in Table I. Different combinations of maskers were used in each experiment. The talker maskers comprised various combinations of ten different voices: two females, two males, and six children. The first female talker and both the male talkers were obtained from the IEEE sentence material (IEEE, 1969) (used with permission from the Sensory Communication Group of the Research Laboratory of Electronics at MIT). Each spoke 40 different sentences. The IEEE sentences are typically longer, and use more complex language than the HINT sentences. The second female talker spoke 30 of the IEEE sentences; this recording was obtained from and used with permission from Ruth Litovsky at University of Wisconsin, Madison. The third female and male talkers (used only for condition m3f3) were the same as the first female and male talkers; as sentence choice was randomized, they were very unlikely to speak the same sentence.

The child talkers were obtained from the Carnegie Mellon University (CMU) Kids Corpus (Eskenazi, 1996; Eskenazi and Mostow, 1997); this is a large database of sentences read aloud by children. Six child talkers were included; they were labeled child-A to child-F. The details of the children used are shown in Table II. Although the database contains hundreds of sentences, only those spoken fluently, without hesitation, mistakes, or extraneous noise were included. This meant that for some children very few sentences were available.

TABLE I. Masking material used in the three experiments, including the abbreviations utilized in this paper. Each experiment used only a subset of the maskers, due to the limited target material available; the conditions used are indicated by 'X'. All talker maskers spoke sentences. The adults spoke IEEE sentences; the children spoke sentences from the CMU Kids Corpus.

No. of talkers	Masker	Abbreviation	Expt 1	Expt 2	Expt 3
1	female	f	×	×	
	6 different children	child-A to child-F	child-E only		×
	male	m	×	×	
2	2 females	f2	×	×	
	1 male and 1 female	m1f1	×	×	
	2 males	m2	×	×	
	2 children	2ch			×
3	1 male and 2 females	m1f2	×		
	2 males and 1 female	m2f1	×		
	3 children	3ch			×
4	2 males and 2 females	m2f2	×		
	4 children	4ch			×
6	3 males and 3 females	m3f3			×
	6 children	6ch			×
—	steady-state noise	noise	×		

The sentences used for the single-talker maskers were selected such that they would have greater duration than the longest HINT sentence, ensuring that no part of the target sentence would be presented in quiet. All sentence material (including target HINT sentences) was edited digitally so that there were minimal silent periods at the start and end of each sentence.

An eight-channel sine-carrier cochlear-implant simulation was implemented in MATLAB® (The MathWorks, Inc.). The signal was first split into eight logarithmically spaced frequency bands from 80 to 8000 Hz, using eighth-order Butterworth bandpass filters. The amplitude envelope was extracted from each band by full-wave rectification and low-pass filtering with a cutoff frequency of 160 Hz. The enve-

TABLE II. Details of child maskers. Child talkers were obtained from the CMU Kids Corpus. They read aloud from grade-appropriate Weekly Reader Stories. Only sentences spoken fluently without mistakes, hesitation, or background noise were included.

Masker	Sex	Age (years)	School grade	No. sentences used
child-A	female	8	3	8
child-B	male	8	2	41
child-C	female	8	2	9
child-D	male	8	2	15
child-E	female	9	3	13
child-F	female	7	1	10

lope in each band was used to modulate a sine wave carrier whose frequency was equal to the band's center frequency. The modulated signal was filtered again using the original analysis filters to ensure that the amplitude-modulated signal had the same bandwidth. The bands were then summed to produce an eight-channel cochlear-implant simulation (Shannon *et al.*, 1995).

In Experiment 1, ten masker conditions were used: f, child-E, m, f2, m1f1, m2, m1f2, m2f1, m2f2, and steady-state noise. The steady-state noise was a 3 s sample spectrally matched to the average long-term spectrum of the HINT sentences (Nilsson *et al.*, 1994), ensuring that on average the SNR was approximately equal at all frequencies. For the cochlear-implant simulation testing, all maskers were preprocessed with the same eight-channel sine-carrier simulation program used for the target sentences. The target and masker materials were individually processed and then added, to allow real-time variation of signal-to-noise ratio during the test. Although in real-life situations the target and noise mix and are then processed together by the cochlear implant, this method was not used with the simulation as it would have introduced a few seconds processing delay before each stimulus.

2. Subjects

Normal-hearing subjects were undergraduates at UC Irvine. They chose to participate in the experiment in order to receive course credit. All subjects had hearing threshold levels within normal limits (≤20 dB HL re ANSI-1996 for octave frequencies between 0.25 and 8 kHz), reported no history of hearing problems, and stated that English was their native language. The subjects were naïve subjects for the HINT sentences. Each subject was included in only one experiment; a total of 38 normal-hearing people participated in the three experiments. All subjects signed an informed consent. The study protocol was approved by the UC Irvine Institutional Review Board.

a. Normal-hearing subjects. Six females and one male with ages ranging from 18 to 21 years (mean=20 years) participated in Experiment 1.

b. Cochlear-implant subjects. Five females and two males with ages ranging from 49 to 80 years (mean =69 years) participated. They were all regular participants in experiments in our laboratory and others; they had all most likely been exposed to the HINT sentences on some or many occasions. They received payment for their participation and had their travel expenses reimbursed. All subjects were postlingually deafened and were experienced users of the Nucleus® or Advanced Bionics Clarion® cochlear-implant device (five had Nucleus 24, one had Nucleus 22, one had CII). They listened with their usual speech processor (SPrint, ESPrit 3G, Spectra 22, or Auria), without a contralateral hearing aid.

c. Simulation subjects. Four female and three male normal-hearing subjects with ages ranging from 18 to 22 years (mean=20 years) participated.

d. Subject CINH001. This subject has a Clarion® HiRes 90k cochlear-implant in his right ear, and virtually normal hearing in his left ear (≤20 dB HL re ANSI-1996 for octave frequencies between 0.25 and 8 kHz, except 35 dB HL at 4 kHz). He received an implant due to intractable tinnitus and is 46 years old.

3. Procedure

A MATLAB® program, developed by the first author, was used to present and score the sentences. Testing was done in quiet or in noise, with a choice of 20 maskers (as shown in Table I). The target and masker were added digitally, and the root mean square (rms) level in dB sound pressure level was adjusted so that all maskers were at a constant intensity regardless of the number of talkers involved. Testing took place in a sound-treated audiometric booth, with the subject sitting approximately 1 m from a loudspeaker placed at 0° azimuth. The operator was also inside the booth at a computer terminal, scoring the subject's responses and running the test. All test material was presented in the sound field except for testing for subject CINH001, as described later. Testing took approximately 1 h, with an option for a break if required.

Testing began with at least one list of HINT sentences presented in quiet at a rms level of 60 dB(A). The quiet testing allowed the subject to become accustomed to the sound of the target talker's voice. The rms level of the target remained at 60 dB(A) throughout the testing; the masker intensity was adjusted to create the appropriate SNR. Using a fixed target level avoids presenting target stimuli at intensities where compression occurs; the cochlear-implant device has a limited input dynamic range (Stickney *et al.*, 2004). Wagener and Brand (2005) found no significant difference in the SRT for normal-hearing or hearing-impaired subjects whether the target level was held constant and the masker level varied or vice versa, although their experiment used only steady-state noise. Two sentence lists (20 sentences total) were used for each masking condition; the pair of lists used was selected at random.

A one-up, one-down adaptive procedure was used to estimate the subject's SRT. The initial SNR was −5 dB for normal-hearing subjects, and +5 dB for cochlear-implant users. This procedure, first described by Levitt and Rabiner (1967), is commonly used to ensure observations are concentrated in the region of interest. Initially, the same target sentence was presented repeatedly and the SNR was increased by 8 dB until the subject correctly repeated the sentence; this allowed the program to quickly find the approximate SRT. Once this occurred, the step size was reduced to 4 dB, and the adaptive procedure began, with the SNR decreasing by 4 dB when the subject answered correctly, and increasing by 4 dB when the response was erroneous. The SRT (in dB) was calculated as the mean of the last six reversals. Although the usual HINT step size is 2 dB, it was found that with only 20 sentences presented, cochlear-implant users would produce insufficient reversals with this step size. The masker segment was demonstrated to the subject at the beginning of each condition; they were told to ignore this voice or these voices and listen only to the target male talker. The masker sentence began approximately 0.45 s before the target; both target and masker were presented from the same speaker. An onset difference between masker and target has been used by other

authors (Drullman and Bronkhorst, 2004; Festen and Plomp, 1990; Freyman *et al.*, 2004; Wagener and Brand, 2005); it provides a basis for attending to the target, although in this experiment the subjects were not instructed as such.

Due to the relatively small sample size, the univariate (mixed-model) approach was used for statistical analyses in Experiments 1 and 3, using the Greenhouse–Geiser ε adjustment to control for Type I errors. Experiment 2 had more subjects and therefore used multivariate analysis of variance. In the case of multiple planned comparisons, the observed p value was compared to a critical p value of 0.05/C (where C is the number of planned comparisons) in order to maintain the familywise alpha level at 0.05 using the Bonferroni approach.

Subject CINH001 was tested in two ways: listening with his normal-hearing ear in the sound field while not wearing his cochlear implant, and by direct connection to his cochlear-implant speech processor (to prohibit the use of his normal-hearing ear). For the direct cochlear-implant connection, the level was adjusted to a comfortable listening level.

B. Results and discussion

Five normal-hearing subjects scored 100% for sentences in quiet; two missed one word. Sentence scores in quiet for cochlear-implant users varied from 65 to 100%, with a mean of 84%. Although initially it was considered appropriate only to test in noise if scores in quiet exceeded 90%, the score in quiet was not found to be a good predictor of performance in noise, so noise testing was performed on all subjects. The simulation subjects obtained scores in quiet ranging from 80 to 100% correct, with a mean of 90%. Subject CINH001 scored 100% correct for sentences in quiet with his near normal-hearing ear and 70% correct with his cochlear-implant ear.

1. Effect of masker type

Figure 1 shows the mean SRT for each masker condition for the three groups of subjects and for subject CINH001. Lines joining the points are purely for clarity, and are not suggesting a functional relation, due to the maskers being categorically different. Three initial observations are noteworthy. First, normal-hearing subjects performed vastly better than the implant users on all conditions. The mean discrepancy was 8 dB for steady-state noise, but as much as 24 dB difference with a female masker. Second, the standard deviation across the individual cochlear-implant users' SRTs was generally higher than those for the normal-hearing subjects, reflecting a larger variation in listening performance. Third, the simulation provides a very comparable result to cochlear-implant users for steady-state noise, but there is a large discrepancy for the talker maskers. This eight-channel simulation may not provide a fair representation of implant users' performance in the presence of competing talkers.

A repeated-measures analysis of variance (ANOVA) (with Greenhouse–Geiser adjustment where appropriate) showed a highly significant main effect of group (normal-hearing, cochlear-implant, and cochlear-implant simulation) ($F(2,18)=102.8$, $p<0.0005$), and a highly significant main

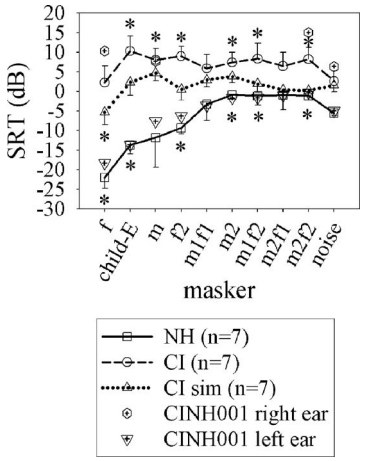

FIG. 1. Mean SRT as a function of masker type in seven normal-hearing, seven cochlear-implant subjects, and seven normal-hearing subjects using an eight-channel sine-carrier cochlear-implant simulation. Target material was HINT sentences spoken by a male. The squares represent the mean SRT for normal-hearing subjects. The circles represent the mean SRT for cochlear-implant subjects. The triangles represent the mean SRT for cochlear-implant simulation subjects. The hexagons and inverted triangles represent the SRTs for the right and left ears, respectively, for subject CINH001, who has a cochlear implant in the right ear and virtually normal hearing in the left ear. Error bars represent one standard deviation. For clarity, only the upward bar is shown for the cochlear-implant users, and only the downward bar for the normal-hearing and simulation subjects. Lines joining the points are purely for clarity, and are not suggesting a functional relation, due to the maskers being categorically different. The asterisks represent a SRT value significantly different from the SRT with a steady-state noise masker.

effect of masker type ($F(4.2,75.8)=36.3$, $p<0.0005$). There was also a highly significant interaction between group and masker type ($F(8.4,75.8)=16.3$, $p<0.0005$), suggesting that the effect of masker type differs in the three groups. Further analysis was therefore performed separately for the three subject groups.

Subject CINH001 was tested over two sessions separated by several months to avoid duplication of target material. At the time of the second test, he had not been using his speech processor for several days and obtained only 30% correct in quiet. At this session, SRTs from 15 to 25 dB were obtained with his implanted right ear for maskers child-E, f2, m1f1, m2, m1f2, and m2f1. Therefore for his implanted ear, only those masker conditions tested in the first session were plotted due to the device nonuse prior to the second session. The SRT results from his normal-hearing ear are almost all within one standard deviation of the mean of those from the normal-hearing young adults. Limited results from his implanted ear fall close to or outside one standard deviation of the mean for the implant subjects. These results suggest that age-related cognitive differences between the normal-hearing and cochlear-implant subject groups are not responsible for the vast differences in the SRT. It is acknowledged, however, that cognitive effects may play a part in some elderly subjects; previous work demonstrated that performance in noise worsened significantly with increasing age (Souza *et al.*, 2007).

A repeated-measures ANOVA was performed to assess nine planned comparisons: the difference between the SRT for the nine masker conditions (f, child-E, m, f2, m1f1, m2, m1f2, m2f1, m2f2) and the SRT for steady-state noise. The

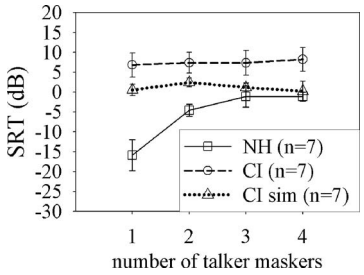

FIG. 2. Mean SRT as a function of number of talker maskers in seven normal-hearing, seven cochlear-implant subjects, and seven normal-hearing subjects using an eight-channel sine-carrier cochlear-implant simulation. Target material was HINT sentences spoken by a male. The squares represent the mean SRT for normal-hearing subjects. The circles represent the mean SRT for cochlear-implant subjects. The triangles represent the mean SRT for cochlear-implant simulation subjects. Error bars represent ± one standard deviation.

analysis was performed separately for each subject group. Those results found to be significantly different from steady-state noise are indicated with an asterisk in Fig. 1. A significant p value was considered to be 0.006, using the Bonferroni correction for multiple comparisons. In addition, for each subject, the SRTs were averaged across conditions f, child-E, and m to obtain a measure of the SRT for a one-talker masker; over conditions f2, m1f1, and m2 for a two-talker condition; over conditions m1f2 and m2f1 for a three-talker condition; and condition m2f2 was used for the four-talker condition. Figure 2 shows mean SRT results as a function of number of interfering talkers. Three planned comparisons were performed to assess the effect of changing from one to two, two to three, and three to four interfering talkers. The p values were compared to 0.017, using the Bonferroni correction for multiple planned comparisons.

a. Normal-Hearing Subjects. All masker conditions gave significantly different SRTs from steady-state noise, except conditions m, m1f1, and m2f1. The SRT for these conditions had the largest standard deviations, probably due to the subject inadvertently following the wrong male talker. This confusion usually happened when these conditions occurred near the beginning of the test, and the subject was unaccustomed to the target talker's voice. The single-talker conditions female and child produced significantly better SRT results than steady-state noise. For two interfering talkers, only f2 gave better SRT results, while m1f1 provided a comparable SRT to steady-state noise, and m2 was worse. With more than two interfering talkers, the SRT was significantly worse than for steady-state noise, except for m2f1 where the SRT was comparable. Considering Fig. 2, there was a significant increase in the mean SRT as the number of interfering talkers changed from one to two ($F(1,6)=45.6$, $p=0.001$) and from two to three ($F(1,6)=18.5$, $p=0.005$). However, as the number of interfering talkers increased from three to four, there was not a significant change in the mean SRT ($F(1,6)<0.0005$, $p=0.991$). This suggests that once there are three interfering talkers, the inclusion of one additional talker did not influence speech recognition, although, as discussed later, there must eventually be a drop in the SRT to the level of that with a noise masker, as more and more talkers are included.

The normal-hearing subjects in this study appeared to be able to use temporal and spectral fluctuations in the interferers to obtain release from masking with one or two interfering talkers. With fewer maskers, the subject can take advantage of favorable SNRs in the temporal gaps; as more maskers are introduced, these gaps are filled in and energetic masking increases. However, as the number of talkers increases, informational masking also increases; with three or four interfering talkers, the SRT was generally worse than for steady-state noise. These results agree with those found by Brungart *et al.* (2001) who showed that, at a negative SNR, performance is worse for two or three interfering talkers than for only one. In common with Drullman and Bronkhorst (2004) the authors believe that when there are only one or two interfering talkers, the interfering speech is still intelligible, so grammatical and semantic information in the masker help the subject to pick out the target speech. However, when there are multiple interfering talkers, the subject is able to hear words in the maskers, but because they cannot fully perceive the linguistic structure of the masker, they are unable to take advantage of this to decide whether the words were spoken by target or masker. The SRTs seem to reach a plateau at three talker maskers, suggesting that additional background talkers would not affect the SRT. An early study by Miller (1947) using target words against continuous discourse maskers and very different methodology had shown an increase in masking from two to four masker voices, but no further increase from four to six or six to eight. However, in this study the SRT for four interfering talkers (m2f2) is significantly worse than that for steady-state noise, and if steady-state noise is considered to be the sum of an infinite number of talkers, it appears that at some point the effect of informational masking would decline, and the SRT would decrease again. Carhart *et al.* (1975) though still demonstrated informational masking with 64 competing talkers.

b. Cochlear-Implant Subjects. The maskers f, m1f1, and m2f1 produced comparable SRTs to steady-state noise; all other maskers gave significantly higher SRTs. As shown in Fig. 2, there was no significant difference in the mean SRT related to number of interfering talkers ($F(2.6,15.4)=1.6$, $p=0.213$). The best SRTs were obtained for one female masker and for steady-state noise. Most of the multiple talker maskers gave worse SRTs than steady-state noise. This may be a result of informational masking, or some kind of modulation interference. The cochlear implant users are clearly unable to take advantage of temporal glimpsing. Assessing whether the subjects confused the masker words with the target would indicate the extent of informational masking; error analysis was not done in this study. In common with Qin and Oxenham (2003) these results suggest that testing implant users in steady-state noise may underestimate the difficulties they experience in everyday life.

c. Simulation Subjects. The only masker condition that gave a significantly different SRT from steady-state noise was the female talker; its mean SRT was lower than that of steady-state noise ($F(1,6)=21.4$, $p=0.004$). Considering Fig. 2, as the number of interfering talkers increased from one to two, there was a significant increase in the mean SRT ($F(1,6)=26.2$, $p=0.002$). Significant changes were not seen for two vs three talkers ($F(1,6)=7.0$, $p=0.038$), or three vs four talkers ($F(1,6)=0.9$, $p=0.377$).

Simulation subjects showed little difference in the mean SRT across the masker types except a significantly better result with a female masker. They did not show a pronounced effect of informational masking and seemed unable to make much use of amplitude minima in the temporal

structure of the maskers. This does not agree with simulation results from Qin and Oxenham (2003) who found that simulation subjects performed significantly worse with male or female single-talker maskers than steady-state noise. This simulation used narrowband noise carriers, and the present study used sine carriers, so it is possible that this caused the discrepancy. The current data do show that the SRT for the male masker was worse than that for steady-state noise, and if a p value of 0.05 were used (as used by Qin and Oxenham), then this difference would be significant. As discussed later, the female masker used in Qin and Oxenham's study had an abnormally low F0, so cannot be compared to that used in the current research.

Five subjects from each group were asked to identify the number and gender of talkers in each of the maskers played separately without the target. Although the normal-hearing subjects were able to accurately specify the one, two, and three talker conditions, none was correct for the four-talker condition. All stated that they heard three voices: two male and one female, when in fact there were four voices (two male and two female). (One subject stated that he heard one male talker as two males; this is believed to have been a concentration error). The plateau for masking effectiveness coincided with the limit of talker number that the subjects were able to identify. The cochlear-implant and simulation subject results were variable, but worse than those for normal-hearing subjects. Although all implant and most simulation subjects were able to identify one female or one male talker, only one implant user and three simulation subjects could identify the child's voice. The others reported the child talker to be the female.

Cochlear-implant users clearly performed much worse than normal-hearing subjects when identifying talkers. Voice gender perception is dependent on accurate pitch information, which is lacking in cochlear-implant speech processing. The implant users were able to accurately identify one female or one male talker as found by Fu *et al.* (2005), but identification of the child's voice was very poor.

2. Effect of voice pitch on the SRT

Further analysis was conducted using only the single-talker maskers, in order to evaluate the effect of voice pitch on the SRT. The hypothesis was that more separation between talker and masker F0 would lead to lower (better) SRTs. A modification of the MATLAB® program STRAIGHT was used to extract the F0 for voiced parts of the speech (Kawahara *et al.*, 1999); these were averaged across each sentence. The mean for each sentence was then averaged over all the sentences (250 for the target, 40 for the adult maskers, 13 for the child) providing a single mean value for the target and each of the maskers. These are shown in Fig. 3. The child's voice clearly has the highest F0, followed by the female voice, followed by the two male voices, as expected. Three planned paired comparisons were assessed to evaluate differences in the SRTs for the single-talker maskers, and the observed p value was compared to 0.017 using the Bonferroni correction for multiple comparisons.

For all three groups (normal-hearing, cochlear-implant, and simulation), the mean SRT for the female masker was significantly better than the mean SRT for the child masker, and significantly better than the mean SRT for the male masker. There was not a significant difference between the mean SRT for the child and male maskers. Table III contains the t and p values for the comparisons.

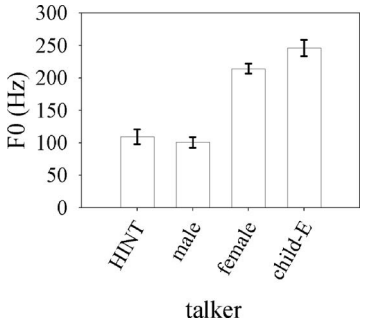

FIG. 3. Mean fundamental frequency (F0) for the target and each of the masker voices. The target HINT sentences were spoken by a male. Error bars represent ± one standard deviation. As expected, the child has the highest F0, followed by the female, then the two males.

The three subject groups behaved in a similar manner on this comparison; they were all able to take advantage of differences in voice F0 to separate the female from the male and child talkers. This result does not agree with other researchers who found that implant (Stickney *et al.*, 2007; Stickney *et al.*, 2004) and simulation users (Qin and Oxenham, 2003; Stickney *et al.*, 2007; Stickney *et al.*, 2004) showed no difference in speech recognition when the gender of the masker was changed. The simulations used were different in two of the studies (Qin and Oxenham, 2003; Stickney *et al.*, 2004): these used noise carriers, but this should not affect the result. The previous studies had more generous low-pass cutoff frequencies (300 and 500 Hz, respectively, for Qin and Oxenham and Stickney *et al.*) compared to the current study's 160 Hz. Qin and Oxenham (2003) felt that one likely reason for their null effect was that their female voice had an atypically low mean F0 of 129 Hz. The current study uses a female masker with F0 of 214 Hz, which is much closer to the average value of 220 Hz (Hillenbrand *et al.*, 1995). In Stickney *et al.*'s 2004 study, although the cochlear-implant subjects performed well in quiet (78%–92% on IEEE sentences), their performance in noise was poor, with the average reaching only around 50% even at 20 dB SNR. The performance of the subjects in the later

TABLE III. t and p values from the paired t tests performed on the mean SRT with female, child, and male masker in normal-hearing, cochlear-implant, and cochlear-implant simulation subjects. Values shown are two tailed, with six degrees of freedom. The p values were compared to 0.017 using the Bonferroni correction for multiple comparisons. All three groups showed a significantly better mean SRT for the female masker than both the child and male maskers. There was no significant difference between the mean SRT for the child and male maskers.

	Normal-hearing	Cochlear-implant	Cochlear-implant simulation
Female/child	t=−21.8	t=−6.0	t=−4.3
	p<0.0005	p=0.001	p=0.005
Female/male	t=−4.8	t=−4.0	*t*=−6.6
	p=0.003	p=0.007	p=0.001
Child/male	t=−8	t=1.7	t=−1.4
	p=0.465	p=0.135	p=0.204

study was even worse (Stickney *et al.*, 2007). This may explain their null finding. The cochlear-implant users in this study performed much better in noise. It is possible that the subjects in the current study are using some characteristic other than the F0 in order to segregate the voices, for example, speaking rate or coarse spectral differences. However, it is shown later that syllabic rate is similar between the female and male maskers.

Considering that previous research has shown that voices are easier to segregate if there is a larger separation between their pitches, one would expect that the child's voice would be the least effective masker. This was not found. In fact no significant difference was seen in the masking effectiveness between the child and the male masker, although their F0 difference is 145 Hz. Brungart *et al.* (2001) suggested that a particularly salient masker could cause the subject's attention to be drawn away from the target phrase; this appears to be the situation here. The characteristics of the child's voice make it more difficult to ignore than would be expected given its spectral qualities; this is further explored in Experiment 3.

III. EXPERIMENT 2: EFFECT OF REVERSED SPEECH MASKERS ON THE SRT IN NORMAL-HEARING SUBJECTS

A. Methods

1. Test material

The HINT sentences were the target material, as described in Experiment 1. Only the single- and two-talker adult maskers were used: f, m, f2, m1f1, and m2. They were played either forward or reversed, making ten different maskers. In order for target material not to be repeated, only these conditions were involved.

2. Subjects

The normal-hearing subjects were as described previously; results were obtained from 12 females and four males, with ages ranging from 18 to 36 years (mean=21 years). None of these subjects had taken part in Experiment 1.

3. Procedure

The procedure was as described in Experiment 1. A within-subjects crossed design was used, with each subject acting as his or her own control; this had the advantage of requiring fewer subjects, although there was the possible disadvantage of differential carryover effects. Differential carryover effects occur when a subject's participation in one part of the experiment affects their performance on a later condition one way, and on a different condition in another way. In contrast, practice effects affect all treatment conditions equally. Test order was randomized for each subject across both factors in an attempt to avoid differential carryover effects.

B. Results and discussion

All subjects scored 100% for sentences in quiet. All 16 subjects were tested in the single-talker masker-forward and masker-reversed conditions. The final seven (six female, one male, age range 18–23 years, mean=20 years) were also tested in the two-talker masker-forward and masker-reversed conditions.

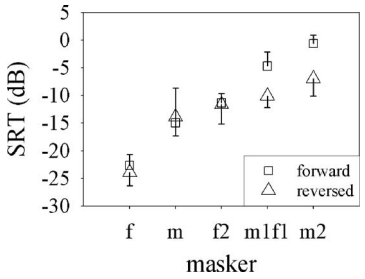

FIG. 4. Mean SRT as a function of masker type and masker reversal in normal-hearing subjects. Target material was HINT sentences spoken by a male. Maskers were one or two female or male talkers played either forward or reversed. The squares represent the mean SRT when masker sentence was played forward. The triangles represent the mean SRT when the masker segment was reversed. Sixteen subjects were evaluated for the single-talker maskers (f and m), seven subjects for the two-talker maskers (f2, m1f1, and m2). Error bars represent one standard deviation. For clarity, only the upward bar is shown for the masker-forward condition, and only the downward bar for the masker-reversed condition.

Mean SRT results are shown in Fig. 4 for the single-talker (16 subjects) and two-talker (seven subjects) masker conditions. For the single-talker maskers no difference is observed between the masker-forward and masker-reversed conditions. (Multivariate analysis of variance $F(1,15)=0.02$, $p=0.896$). For the two-talker conditions, the reversed masker produced better SRTs in conditions m1f1 and m2. A repeated-measures ANOVA showed a significant main effect of reversal ($F(1,6)=25.7$, $p=0.002$). After examining the data, the significance of the difference between m1f1 and m2 for masker-forward or masker-reversed conditions was tested. As this was a post hoc comparison, the critical value used was a multivariate extension of Scheffé's method developed by Roy and Bose (1953). Both comparisons were statistically significant (m1f1: $t=4.6$, $df=6$, $p=0.004$; m2: $t=4.7$, $df=6$, $p=0.003$; critical $t=2.97$). The results in Fig. 4 again show a much larger variance for masker condition m, when the masker was forward. This was presumably because subjects were more likely to pay attention to the wrong talker in this condition.

A significant effect of reversal was only seen for conditions m1f1 and m2: the two-talker conditions that involved a male voice. These results coincided with those of Hawley *et al.* (2004), who used a male target and the same male interferers. Their results did not show a large effect of reversed speech when there was only one interfering talker; however, for two interferers, reversed speech gave consistently better SRTs than forward speech. In common with Drullman and Bronkhorst (2004), it is believed that this occurs because there is minimal informational masking with only one interfering talker as the subject can use the grammatical and semantic information in the masker to segregate it from the target. However, with several interfering talkers, the sentences are not individually intelligible, so the subject cannot take advantage of the grammatical and semantic content. Rhebergen *et al.* (2005) did, however, show a signifi-

cant improvement in the SRT when using a male target and a time-reversed female masker. This result was found in Dutch listeners using Dutch target and masker; differences in the dynamics of this language compared to English may contribute to the discrepancy.

Although time-reversed speech has unchanged spectral content, the temporal envelope is reversed. In forward speech, words usually begin with plosives: quick onset and slow decay. When speech is reversed, it contains abrupt offsets. The auditory system cannot follow these abrupt offsets so accurately, so soft sounds can be masked by a preceding strong signal: forward masking (Rhebergen *et al.*, 2005). When a speech masker is time reversed, one may expect an improvement in the SRT due to the release from informational masking. However, there may also be a decrease in performance due to increased forward masking. The two effects act in opposition, therefore in order to show a release from informational masking using reversed speech, this effect must exceed the opposing increase in forward masking. Rhebergen *et al.* (2005) found the increase in the SRT due to forward masking to be approximately 2 dB. Results from the current study therefore suggest that any release from informational masking that is present with reversed speech in conditions f, m, and f2 is less than or around 2 dB. The release from informational masking for conditions m1f1 and m2 may be approximately 7–8 dB, assuming the 2 dB of forward masking still applies with a multitalker background. The two effects can be separated using a speech masker in a foreign language, thus offering release from informational masking, but not altering the amount of forward masking (Rhebergen *et al.*, 2005).

IV. EXPERIMENT 3: FURTHER INVESTIGATION OF CHILD MASKERS

A. Methods

1. Test material

The target material was the HINT sentences as in Experiments 1 and 2. Subjects were tested with six different child maskers (child-A to child-F), combinations of two, three, four, and six children (2ch, 3ch, 4ch, and 6ch), and six adults (m3f3). Results from Experiment 1 showed that a child's voice had a greater masking effect than expected, given its fundamental frequency. The purpose of Experiment 3 was to further investigate this finding. Six different child talkers were used, in order to rule out the hypothesis that there was an anomalous feature associated with the child's voice used in Experiment 1. Combinations of child maskers were used to replicate and extend the adult talker masker findings shown in Fig. 2. The masking effect of a babble of six child talkers was examined and compared to that of a babble of six adults. It was hypothesized that whatever feature of a child masker that increased its masking effectiveness would disappear once the voices were not individually distinguishable. Although previous work has used children as subjects in masking experiments (Hall *et al.*, 2002), children's voices have not been examined as maskers.

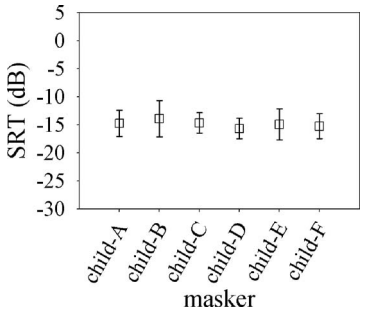

FIG. 5. Mean SRT for six different child maskers in eight normal-hearing subjects. Target material was HINT sentences spoken by a male. Error bars represent ± one standard deviation. Four female and two male children were used, with ages ranging from seven to nine years, with voice fundamental frequencies from 215 to 281 Hz. The mean SRT, however, was approximately the same for all child maskers.

2. Subjects

Eight normal-hearing subjects (five female, three male), with ages ranging from 18 to 23 years (mean=21 years) participated. None of these subjects had taken part in Experiments 1 or 2.

3. Procedure

The procedure was as used in Experiment 2.

B. Results and discussion

Seven subjects scored 100% correct for sentences in quiet; one scored 90% correct for sentences, 98% correct for words.

1. Comparison of the SRT in Six Different Child Maskers

Figure 5 shows the mean SRT for each of the child maskers in eight normal-hearing subjects. A repeated-measures ANOVA using the Greenhouse–Geiser correction showed no significant difference ($F(2.9,20.3)=0.655$, $p=0.585$). This confirms that the finding in Experiment 1 (that the child masker produced more masking than expected) was not simply a result of characteristics of that particular child's voice.

2. Effect of number of interfering child talkers

Figure 6 shows the mean SRT for one, two, three, four, and six child maskers. A repeated-measures ANOVA using

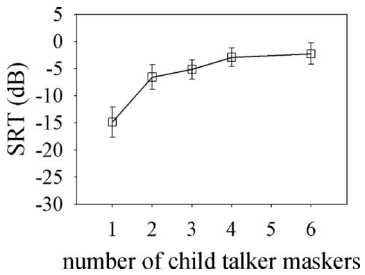

FIG. 6. Mean SRT for one, two, three, four, and six child maskers in eight normal-hearing subjects. Target material was HINT sentences spoken by a male. Error bars represent ± one standard deviation. The mean SRT gradually increases as the number of interfering talkers increases.

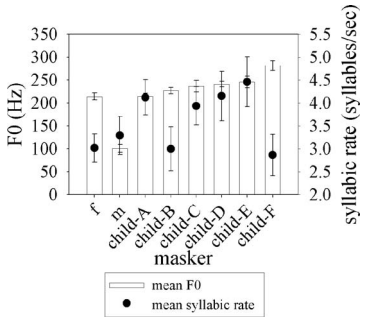

FIG. 7. Mean F0 and syllabic rate for all the single-talker maskers. The bars represent the mean F0 for each talker, using the left-hand axis. The circles represent the means syllabic rate for each talker, using the right-hand axis. Error bars represent ± one standard deviation. The male F0 is significantly lower than all the other maskers. All child maskers except child-A have a significantly higher F0 than the female. There is much variation in syllabic rate, with some children having higher rate than the female, and some comparable.

the Greenhouse–Geiser correction showed a significant effect of the number of talkers ($F(2.8,19.8)=47.3$, $p<0.0005$). Four planned paired comparisons were evaluated to assess the change from one to two, two to three, three to four, and four to six interfering child talkers. The p value was compared to 0.01 using the Bonferroni correction for multiple planned comparisons. The only significant change was from one to two interfering child talkers ($F(1,7)=38.3$, $p<0.0005$); the other changes were too gradual to reach significance. A comparison of the child six-talker SRT to that obtained using steady-state noise in Experiment 1, showed that the child six-talker masker had a significantly greater masking effect than steady-state noise (two-tailed $t=4.3$, $df=13$, $p<0.0005$). This reflects informational masking. The mean SRT with six adults as the masker was −2.7 dB; mean SRT with six child maskers was −2.3 dB. A paired t test showed that this difference was not significant (two-tailed $t=0.424$, $df=7$, $p=0.685$). This demonstrates that the characteristic of a child masker that is providing more masking than expected is not apparent when a babble of child voices is used. A babble of child voices produces essentially the same masking as a babble of adult voices.

3. Effect of child voice pitch, syllabic rate, and temporal modulation on the SRT

The program STRAIGHT implemented in MATLAB® was used to calculate the mean F0 for the six child maskers (Kawahara *et al.*, 1999). The rate of the talkers was also examined. This was simply the mean number of syllables spoken per second; it was evaluated for 15 sentences for female, male, child-B, and child-D, and fewer sentences for the other children (see Table II). Figure 7 shows mean F0 and syllabic rate for all the one-talker maskers. The male F0 is significantly lower than the female (two-tailed $t=63.4$, $df=78$, $p<0.0005$), and all the child maskers except child-A have a significantly higher F0 than the female masker ($p<0.008$ using the Bonferroni correction for multiple comparisons). Even though F0 of the child maskers is higher than the female, the SRT is still significantly worse; this suggests that frequency separation alone is not responsible for the SRT difference. Figure 7 also shows the variation in the syllabic rate of the child talkers. Compared to the female masker, child-A, child-C, child-D, and child-E have a significantly faster syllabic rate ($p<0.008$ using the Bonferroni correction for multiple comparisons). Child-B and child-F have comparable speaking rates to the female talker. It seems that talking rate is not a consistent factor, and therefore is not responsible for the increased masking effectiveness of a child talker.

The authors observed that children's voices often have a "sing-song" characteristic: voice pitch appears to rise and fall more during a sentence than for adult talkers. The F0 calculation using STRAIGHT involves averaging instantaneous F0 across a sentence. In order to evaluate whether the variation in F0 was similar for female, male, and child maskers, the coefficient of variation (standard deviation/mean) was evaluated in each case. The values were 0.16, 0.32, and 0.21 for female, male, and child, respectively. The child's value is between that for the female and male; this rejects the hypothesis that the child's voice is more distracting because of the variation in the F0 across a sentence.

Spectra of the octave band envelope modulations were examined for the target HINT sentences, female, male, and child maskers. Ten sentences of each were concatenated to simulate running speech; no gaps were inserted between the sentences. Following a method described by Payton and Braida (1999) and implemented in MATLAB®, the speech material was first bandpass filtered using octave bandwidth digital Butterworth filters with center frequencies from 125 to 8000 Hz (the 8000 Hz filter was a high-pass filter). The samples were squared, low-pass filtered to extract the intensity envelope, and power spectra were computed. Average spectra were summed to third octave band representations. The square root of this sum was the one third octave modulation spectrum. The modulation index represents depth of modulation. Greater informational masking may occur if temporal modulation properties are similar between target and masker. Figure 8 shows the modulation index difference between each masker and the target as a function of third octave band modulation frequency for female masker, male masker, and child-E. The measurement was made for seven octave bands (125 Hz through 8000 Hz), although the child results at 125 Hz were artifactual and were excluded. Temporal envelope modulations play an important role in speech intelligibility (Shannon *et al.*, 1995). It is difficult to visually assess from these graphs whether any real differences exist in terms of envelope modulation depth, although in the 1000 Hz band it appears that the child masker had slightly more modulation than the other talkers from 3 to 8 Hz. Correlation coefficients between the target and the female, male, and two child maskers (child-B was included because it has a similar syllabic rate to the female) were calculated for each octave band. The values were averaged across frequencies from 250 to 8000 Hz. The mean correlation coefficients were as follows: female 0.709, male 0.670, child-E 0.653, child-B 0.670. These numbers describe how similar the modulation indices of the masker are to the target. There is no consistent trend in these numbers; they do not explain the masking properties of the child maskers. However, this

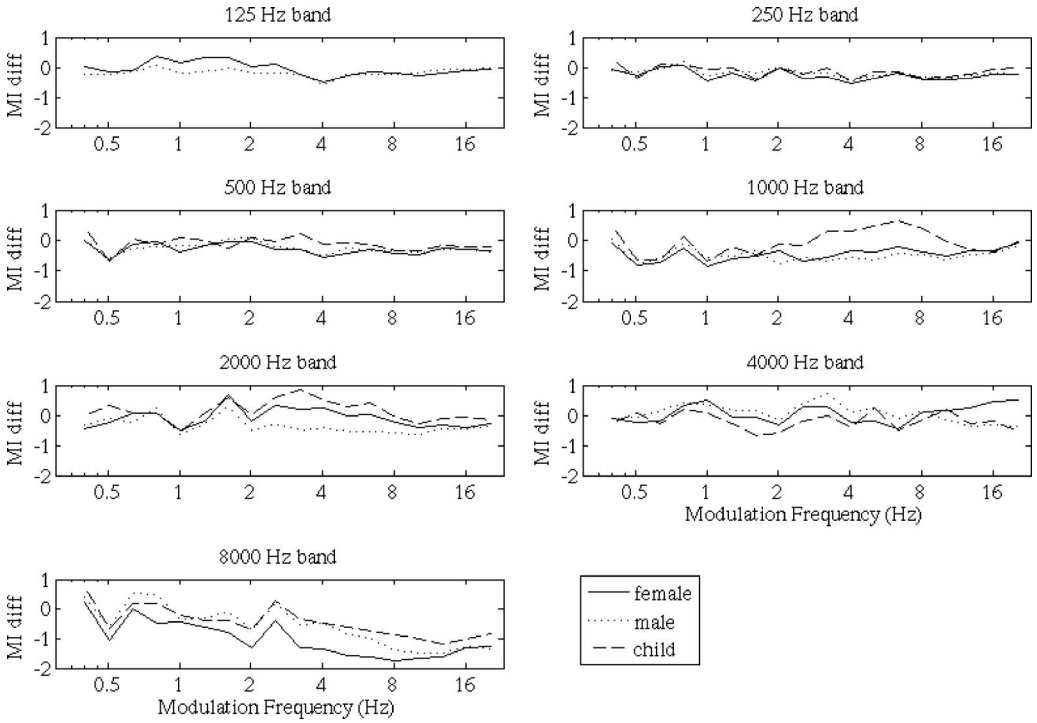

FIG. 8. Temporal envelope modulation data. Graphs show envelope spectra differences between masker and target for seven octave bands, depicted by modulation index difference (MI diff) as a function of third octave band modulation frequency. Data are shown for three maskers: female, male, and child (child-E).

analysis is preliminary, and further research is required. Future work should also examine the speech intelligibility index (SII) of speech masked by a child's voice, using the modification to the SII proposed to predict intelligibility in the presence of a fluctuating masker (Rhebergen and Versfeld, 2005).

V. SUMMARY AND CONCLUSIONS

This research investigated speech recognition in normal-hearing and cochlear-implant subjects, under a variety of talker and noise masker conditions. Normal-hearing subjects performed vastly better than implant users on all conditions; the largest mean discrepancy in the SRT was 24 dB with a female masker. These differences were not caused by age-related cognitive differences in the subject groups. Although an eight-channel sine-carrier cochlear-implant simulation provided an almost identical SRT to cochlear-implant users with a steady-state noise masker, there was a large discrepancy for talker maskers.

Normal-hearing subjects used temporal fluctuations in interferers to obtain release from masking. Cochlear-implant and simulation subjects made much less use of temporal fluctuations. A talker background provides a combination of energetic and informational masking. Results from masker reversal suggested that single-talker maskers produce little informational masking. As the number of talkers increase, both energetic and informational masking increase. Normal-hearing, cochlear-implant, and simulation subjects all showed a significantly better SRT for a female than male masker. Despite the weak representation of voice fundamental frequency in their coding scheme, implant users appeared to use spectral differences in the talkers to segregate the voices.

Although the child maskers had higher voice pitch, all subject groups showed no difference between the mean SRT for the male and child maskers, and a significantly better SRT for the female compared to the child masker. The child maskers possessed greater masking ability than suggested by their spectral qualities; this did not seem to be related to talking rate, variation in the F0 within a sentence, or temporal envelope modulation characteristics.

Clinical cochlear-implant testing generally uses steady-state noise as a masker. The current research suggests that this does not reflect the vast discrepancy between normal-hearing and cochlear-implant subjects in real-life situations with competing talkers. Caution must be exercised when a cochlear-implant simulation is used, as results may reflect implant users' performance in a steady-state noise background, but are discrepant in more realistic listening situations.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic Press, London).

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Ophthamol. **13**, 108–112.

Blandy, S., and Lutman, M. (**2005**). "Hearing threshold levels and speech recognition in noise in 7 year olds," Int. J. Audiol. **44**, 435–443.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.
Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.
Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.
Carhart, R., Johnson, C., and Goodman, J. (**1975**). "Perceptual masking of spondees by combinations of talkers," J. Acoust. Soc. Am. **58**, S35.
Carroll, J., and Zeng, F. G. (**2007**). "Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations," Hear. Res. **231**, 42–53.
Drullman, R., and Bronkhorst, A. W. (**2004**). "Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers," J. Acoust. Soc. Am. **116**, 3090–3098.
Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74** 739–743.
Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.
Eskenazi, M. (**1996**). "KIDS: A database of children's speech," J. Acoust. Soc. Am. **100**(4), 2759.
Eskenazi, M., and Mostow, J. (**1997**). "The CMU KIDS Speech Corpus (LDC97S63)," Linguistic Data Consortium (http://www.ldc.upenn.edu), University of Pennsylvania (viewed 8-27-07).
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., and Rosen, S. (**1993**). "Lip-reading the BKB sentence lists: Corrections for list and practice effects," Br. J. Ophthamol. **27**, 233–246.
French, N., and Steinberg, J. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.
Fu, Q. J., Chinchilla, S., Nogaki, G., and Galvin, J. J., III (**2005**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.
Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**, 159–165.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**, 225–246.
Johnstone, P. M., and Litovsky, R. Y. (**2006**). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," J. Acoust. Soc. Am. **120**, 2177–2189.
Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.
Levitt, H., and Rabiner, L. R. (**1967**). "Use of a sequential strategy in intelligibility testing," J. Acoust. Soc. Am. **42**, 609–612.
Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**, 105–129.
Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Payton, K. L., and Braida, L. D. (**1999**). "A method to determine the speech transmission index from speech waveforms," J. Acoust. Soc. Am. **106**, 3637–3648.
Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am. **118**, 1274–1277.
Roy, S. N., and Bose, R. C. (**1953**). "Simultaneous confidence interval estimation," Ann. Math. Stat. **24**, 513–536.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Souza, P. E., Boike, K. T., Witherell, K., and Tremblay, K. (**2007**). "Prediction of speech recognition from audibility in older listeners with hearing loss: Effects of age, amplification, and background noise," J. Am. Acad. Audiol **18**, 54–65.
Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.
Throckmorton, C. S., and Collins, L. M. (**2002**). "The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model," J. Acoust. Soc. Am. **112**, 285–296.
Trammell, J. L., and Speaks, C. (**1970**). "On the distracting properties of competing speech," J. Speech Hear. Res. **13**, 442–445.
Wagener, K. C., and Brand, T. (**2005**). "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters," Int. J. Audiol. **44**, 144–156.
Watson, C. S., and Kelly, W. J. (**1981**). In *Auditory and Visual Pattern recognition*, D. J. Getty and J. H. Howard, eds. (Erlbaum, Hillsdale, NJ), pp. 37–59.
Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.