Published in final edited form as:
Hear Res. 2009 October ; 256(1-2): 75–84. doi:10.1016/j.heares.2009.07.001.

Concurrent-vowel and Tone Recognition by Mandarin-speaking Cochlear Implant Users

Xin Luo[1], **Qian-Jie Fu**[2], **Hung-Pin Wu**[3,4], and **Chuan-Jen Hsu**[5]

[1]Department of Speech, Language, and Hearing Sciences, Purdue University

[2]Division of Communication and Auditory Neuroscience, House Ear Institute

[3]Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University, Taipei, Taiwan

[4]Department of Otolaryngology, Buddhist Tzuchi General Hospital, Taipei, Taiwan

[5]Department of Otolaryngology, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

Abstract

In Mandarin Chinese, tonal patterns are lexically meaningful. In a multi-talker environment, competing tones may create interference in addition to competing vowels and consonants. The present study measured Mandarin-speaking cochlear implant (CI) users' ability to recognize concurrent vowels, tones, and syllables in a concurrent-syllable recognition test. Concurrent syllables were constructed by summing either one Chinese syllable each from one male and one female talker or two syllables from the same male talker. Each talker produced 16 different syllables (4 vowels combined with 4 tones); all syllables were normalized to have the same overall duration and amplitude. Both single- and concurrent-syllable recognition were measured in 4 adolescent and 4 adult CI subjects, using their clinically assigned speech processors. The results showed no significant difference in performance between the adolescent and adult CI subjects. With single syllables, mean vowel recognition was 90% correct, while tone and syllable recognition were only 63 and 57% correct, respectively. With concurrent syllables, vowel, tone, and syllable recognition scores dropped by 40-60 percentage points. Concurrent-syllable performance was significantly correlated with single-syllable performance. Concurrent-vowel and syllable recognition were not significantly different between the same- and different-talker conditions, while concurrent-tone recognition was significantly better with the same-talker condition. Vowel and tone recognition were better when concurrent syllables contained the same vowels or tones, respectively. Across the different vowel pairs, tone recognition was less variable than vowel recognition; across the different tone pairs, vowel recognition was less variable than tone recognition. Thepresent results suggest that interference between concurrent tones may contribute to Mandarin-speaking CI users' susceptibility to competing-talker backgrounds.

Send correspondence to: Xin Luo, Ph.D., Department of Speech, Language, and Hearing Sciences, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907, Phone: (765) 496-7267 FAX: (765) 494-0771, luo5@purdue.edu.

Keywords

concurrent-vowel recognition; concurrent-tone recognition; cochlear implants; Mandarin Chinese

I. INTRODUCTION

Cochlear implant (CI) technology has advanced markedly over the last few decades, and currently provides many CI users good speech understanding in optimal listening conditions. Multi-channel CI devices are successful because speech signals contain redundant cues that aid in recognition of phonemes, words, and sentences. Thus, a limited number of frequency channels can support robust speech recognition, at least in quiet (e.g., Wilson et al., 1991; Shannon et al., 1995). However, contemporary CI devices do not convey pitch information very well, and CI users have difficulty in pitch-related listening tasks such as music perception (see McDermott, 2004 for a review), competing speech (e.g., Stickney et al., 2004), and tonal language recognition (e.g., Fu et al., 2004; Wei et al., 2004). Normal hearing (NH) listeners extract pitch information from the place of excitation along the basilar membrane ("place cues") and the temporal pattern of auditory nerve responses ("rate cues"; Licklider, 1951). Electric stimulation in CIs is limited to a fixed number of electrode locations, which is insufficient to resolve fundamental frequency (F0) and its harmonics. Consequently, CI users receive limited pitch information from place cues. Alternatively, CI users may extract pitch from the temporal patterns of electric stimulation (rate or envelope; e.g., McKay and Carlyon, 1999). Given these limited place and rate cues, CI users achieve only limited recognition performance for vocal emotion (Luo et al., 2007), speech intonation (Peng et al., 2008), and Chinese tone recognition (e.g., Fu et al., 2004; Wei et al., 2004).

For CI users who speak tonal languages, poor pitch perception poses a special challenge. For example, Mandarin Chinese uses tonal patterns to convey lexical meaning within syllables. There are four lexical tones (i.e., pitch contours) in Mandarin Chinese: high-flat (Tone 1), rising (Tone 2), falling-rising (Tone 3), and falling (Tone 4). Although CI users may also utilize vowel duration and amplitude envelope cues to recognize Chinese tones (e.g., Fu et al., 1998; Luo and Fu, 2004), the pitch contour is the primary cue for tone recognition. Therefore, tone recognition may be more challenging for CI users than vowel recognition, which relies more on gross spectral envelope cues. In a recent study with Mandarin-speaking CI users (Luo et al., 2008), mean tone recognition was only 61% correct (25% chance level), while mean vowel recognition was 69% correct (8.3% chance level).

Pitch information is also critical to complex listening tasks such as auditory scene analysis (Bregman, 1990), in which multiple sound sources are presented at the same time. To identify each sound source, listeners must segregate auditory components that arise from different sound sources in the acoustic mixture, according to different acoustic properties. For example, the different voice pitch of competing talkers allows listeners to segregate and stream different talkers. Concurrent-vowel recognition, although not a common task in everyday life, provides a useful tool to explore the role of F0 information in segregating and identifying two simultaneously presented vowels. NH listeners' concurrent-vowel recognition has been shown to be significantly better when there are larger differences in F0 between the two vowels (e.g., Scheffers, 1983; Assmann and Summerfield, 1990). Harmonic misalignment and/or periodic asynchrony related to the different vowel F0s may have facilitated vowel segregation and identification. Previous studies have also shown that frequency modulation contributes to the perceptual prominence and recognition accuracy of a target vowel in the presence of competing vowels (e.g., Marin and McAdams, 1991; Culling and Summerfield, 1995). Specifically, Chalikia and Bregman (1989) found that NH listeners' concurrent-vowel recognition was

significantly better when frequency components of the two vowels were modulated in opposite direction (up vs. down) rather than in parallel direction (both up or both down).

Acoustic CI simulations (e.g., Shannon et al., 1995) have been used to investigate the effects of CI speech processing on concurrent-vowel and tone recognition. Using synthesized vowel-like stimuli, Qin and Oxenham (2005) found that when listening to acoustic CI simulations (even with 24 channels), NH listeners' concurrent English vowel recognition significantly worsened, relative to unprocessed stimuli. Increasing the F0 separation between the concurrent vowels did not improve vowel recognition in the CI simulations. Recently, Luo and Fu (2009) extended these observations by measuring NH subjects' recognition of concurrent Chinese syllables while listening to unprocessed speech, 8- or 4-channel CI simulations. One purpose of Luo and Fu (2009) was to investigate if competing tonal patterns (i.e., pitch contours) may aid in the segregation and identification of concurrent vowels. Similar to the results in Qin and Oxenham (2005), concurrent Chinese vowel, tone, and syllable recognition were significantly poorer with the CI simulations than with unprocessed speech. There was a small but significant effect of talker F0 separation for both unprocessed speech and the 8-channel CI simulation, but not for the 4-channel CI simulation. Concurrent-tone and syllable recognition with unprocessed speech were better when the two component syllables were produced by a male and a female talker rather than by the same male talker. However, concurrent-tone and syllable recognition with the 8-channel CI simulation showed an opposite pattern, i.e., were better with the same-talker condition than with the different-talker condition. Luo and Fu (2009) also found that with the CI simulations, concurrent-vowel and tone recognition were independent of each other, as suggested by the different error patterns across the various vowel or tone pairs. Concurrent-vowel recognition was quite variable across the different vowel pairs, while concurrent-tone recognition remained largely unchanged. Concurrent-tone recognition was significantly better when both syllables had the same tone, while concurrent-vowel recognition was not significantly affected by the different tone pairs. The poor pitch coding in the acoustic CI simulations may explain why the large F0 separations and different F0 contours did not aid in concurrent-syllable recognition.

Although acoustic CI simulations provide a reasonable estimate of CI users' performance trends (e.g., Friesen et al., 2001), several factors may limit the accuracy of noise-band vocoders in modeling CI users' pitch perception (Laneau et al., 2006). In typical acoustic CI simulations, NH listeners may receive more place pitch cues than do CI listeners due to less spectral smearing or spectral mismatch. However, noise-band CI simulations may transmit fewer temporal pitch cues compared with the CI case, due to the absence of envelope compression/expansion and the potential interference between the speech envelope and the noisy carrier envelope. CI performance also has substantial inter-subject variability, possibly due to differences among CI users' etiologies, neural survival patterns, and peripheral auditory processing abilities (e.g., Zeng, 2004). In light of these differences between real and simulated electric hearing, the present study measured concurrent-vowel and tone recognition by Mandarin-speaking CI users via their clinically assigned speech processors. Three talker conditions were tested: single talker (including a male and a female talker), concurrent talkers (the male talker combined with himself or with the female talker). CI performance across different vowel and tone pairs was analyzed and compared to the previous study's simulation results (Luo and Fu, 2009) to shed light on similarities and differences between CI users' and NH listeners' perception of acoustic cues. Luo and Fu (2009) found significant effects of talker conditions for the 8-channel CI simulation but not for the 4-channel CI simulation. The present study further tested if increasing talker F0 separation would improve CI users' concurrent-vowel and tone recognition. Assuming that CI users' functional spectral resolution was reasonably simulated using 4 and 8 channels in Luo and Fu (2009), CI users' single-vowel recognition was hypothesized to be similar to that of the previous CI simulations. However, if CI users were able to make use of the presumably more salient temporal pitch cues (Laneau et

al., 2006), their single-tone recognition and concurrent-syllable recognition would be expected to be slightly better (although still very limited) than the previous simulation results.

II. METHODS

A. Subjects

Four adult and four adolescent native Mandarin-speaking CI users (3 males and 5 females) participated in the present study. Table 1 shows the demographic information for the subjects. All subjects except for S1 used the Continuous Interleaved Sampling (CIS) strategy; S1 used the Hi-Resolution (HiRes) strategy. The four adult subjects (S1 - S4) were post-lingually deafened, although S1 and S4 lost their hearing when still young. The four adolescent subjects (S5 - S8) were pre-lingually deafened. All subjects were paid for their participation.

B. Stimuli and Speech Processing

The same speech stimuli used in Luo and Fu (2009) were used in the present study. Single-syllable recognition was measured using single-vowel stimuli drawn from the Chinese Standard Database (Wang, 1993). Four Mandarin Chinese single-vowels (/a/, /e/, /u/, /i/ in Pinyin) were produced by one male and one female talker according to four lexical tones, resulting in a total of 32 single-vowel syllables (4 vowels × 4 tones × 2 talkers). These vowels were selected because their productions were relatively stable and they were located at the corners of the Chinese single-vowel space (based on the first and second formant frequencies). These single-vowel stimuli were recorded with a sampling rate of 16 kHz and a quantization resolution of 16 bits. No high-frequency pre-emphasis was applied. After recording, these single-vowel syllables were normalized to have the same long-term root-mean-square (RMS) amplitude (65 dB) and the same time duration (405 ms for vowel segments, with no silence preceding or following the vowel). The durations of the recorded vowels were stretched or compressed to 405 ms using an algorithm in Adobe Audition, without changing their pitch and formant frequencies. The F0 contours of all single-vowel stimuli are shown in Figure 1.

Concurrent syllables were constructed by summing either one syllable each from the male and female talkers (different-talker condition) or two syllables from the male talker (same-talker condition). Because of the normalized amplitude and duration, the two component vowels in the concurrent syllables were matched in overall level and their onsets and offsets were both aligned. The long-term RMS amplitude of the concurrent syllables was normalized to 65 dB after summation. There was a total of 256 concurrent syllables (16 single-vowel syllables from the male talker × 16 single-vowel syllables from the female or the same male talker) in both the same- and different-talker conditions. In the same-talker condition, each single-vowel syllable from the male talker was paired with itself once, and each pair of different single-vowel syllables was repeated once.

C. Procedures

CI users were tested with their clinically assigned speech processors. During the tests, the microphone sensitivity and volume levels were set for conversational speech; once set, these levels were not changed during testing. Subjects were seated in a sound-treated booth and listened to the original, unprocessed speech presented in sound field over a single loudspeaker at a comfortably loud level. The experimental talker conditions were tested in a random order for each subject. Subjects were informed whether they were to be presented with single- or concurrent-syllables, but they did not know whether the concurrent talkers were the same or different. Both single- and concurrent-syllable recognition were measured using a 16-alternative, forced-choice paradigm. In each trial, a stimulus was randomly selected from the stimulus set (without replacement) and presented to the subject. 16 response choices were displayed on a computer screen: a1, a2, a3, a4, e1, e2, e3, e4, u1, u2, u3, u4, i1, i2, i3, i4. The

letters in the response labels refer to the Chinese vowels and the numbers refer to the Chinese tones (1 - high, flat, 2 - rising, 3 - falling and rising, 4 - falling). In the single-syllable recognition tasks, subjects were instructed to click on the response choice corresponding to the syllable they heard. Responses were collected and scored as the percentage of correctly identified Chinese syllables, vowels, and tones. Each subject completed two runs of single-syllable recognition, and the recognition scores were averaged across both runs (a total of 64 trials). In the concurrent-syllable recognition tasks, subjects were instructed to identify the two Chinese syllables by making two consecutive choices; the order of choices was not important for scoring. Subjects were allowed to make only one choice and skip to the next stimulus if they believed they only heard one syllable, which actually happened for 5% of the concurrent-syllable trials; such single choices were incorrect for concurrent recognition tasks, except for in the same-talker condition where a single-vowel syllable was paired with itself. Responses were collected and scored as the percentage that both syllables, both vowels, or both tones were correctly identified. In all tests, no preview or feedback was provided.

III. RESULTS

A. CI vs. NH Performance

Figure 2 shows Chinese syllable, vowel, and tone recognition scores for the different talker conditions. The black bars show results for 8 CI subjects in the present study. For comparison, results for 6 NH subjects (from Luo and Fu, 2009) are presented; the gray bars show performance with the acoustic CI simulations and the white bars show performance with unprocessed speech. The horizontal lines show chance performance levels for specific recognition tasks and talker conditions. A series of one-way analyses of variance (ANOVAs) were performed on the single-talker data shown in each panel, with subject group as the factor; CI and NH subjects listening to unprocessed speech, as well as NH subjects listening to the 4- or 8-channel CI simulation were treated as four levels within the subject group factor. *Post hoc* Bonferroni *t*-tests showed that for single-talker vowel recognition, CI performance was not significantly different from either the 4- or 8-channel simulation results. In contrast, for single-talker tone recognition, CI performance was significantly poorer than both the 4- and 8-channel simulation results ($p < 0.002$). For single-talker syllable recognition, CI performance was not significantly different from the 4-channel simulation results, but was significantly poorer than the 8-channel simulation results ($p < 0.001$). For all three single-talker recognition measures, CI performance was significantly poorer than NH performance with unprocessed speech ($p < 0.04$).

Figure 3 shows CI subjects' vowel confusion matrix with single syllables (left panel) as compared to that of the 4-channel simulation (right panel; data from Luo and Fu, 2009). Only the 4-channel simulation results are presented because they more closely approximate real CI performance than the 8-channel simulation results. While having similar overall single-talker vowel recognition performance, CI subjects showed different vowel confusion patterns than did NH subjects listening to the 4-channel simulation. For example, CI subjects misidentified /a/ as /e/ 8% of the time, while NH subjects had much more such confusions (27% of the time) with the 4-channel simulation. Interestingly, with the 8-channel simulation, NH subjects responded with /e/ when presented with /a/ only 4% of the time. Thus, the frequency channel distribution used in the 4-channel simulation may have contributed to the confusions between /a/ and /e/. Figure 4 shows CI subjects' tone confusion matrix with single syllables (left panel) as compared to that of the 4-channel simulation (right panel; data from Luo and Fu, 2009). As can be seen, the poorer CI tone recognition with single syllables (relative to the 4-channel simulation results) was mostly due to the confusions between Tones 2 and 3. The present CI subjects seemed to be unable to make full use of amplitude envelope cues, which have been

found to contribute strongly to NH listeners' identification of Tone 3 in the absence of salient pitch cues (Luo and Fu, 2004).

A series of two-way ANOVAs were performed on the concurrent-talker data in Figure 2, with subject group and talker condition (same- vs. different-talker) as factors. The results showed that concurrent recognition performance was significantly different across subject groups [$F(3,44)=568.92$, $p<0.001$ for syllables, $F(3,44)=175.35$, $p<0.001$ for vowels, and $F(3,44)=347.12$, $p<0.001$ for tones]. There was no significant difference in concurrent recognition performance between the same- and different-talker conditions [$F(1,44)=0.11$, $p=0.75$ for syllables, $F(1,44)=0.01$, $p=0.94$ for vowels, and $F(1,44)=0.74$, $p=0.40$ for tones]. No significant interaction between subject groups and talker conditions was found for syllables [$F(3,44)=1.26$, $p=0.30$], vowels [$F(3,44)=0.82$, $p=0.49$], or tones [$F(3,44)=1.36$, $p=0.27$]. *Post hoc* Bonferroni analyses showed that for all three concurrent-talker recognition measures, CI performance was significantly poorer than NH performance with unprocessed speech, 8- or 4-channel CI simulation ($p<0.01$). Overall, the performance gap between CI and NH subjects listening to unprocessed speech was much larger with concurrent syllables than with single syllables, and for tone recognition than for vowel recognition. Note though that, NH performance with unprocessed speech, particularly for vowel recognition, was subject to ceiling effects.

B. Effects of Talker F0 Separation

To further test if talker F0 separation affected CI users' concurrent-syllable recognition, paired *t*-tests were conducted to compare CI performance between the same- and different-talker conditions. For vowel and syllable recognition, CI performance was not significantly affected by talker F0 separation. However, CI subjects had slightly (~5%) but significantly better tone recognition with the same-talker condition than with the different-talker condition ($t=3.21$, $p=0.02$), which is consistent with the 8-channel CI simulation results in Luo and Fu (2009). In contrast, Luo and Fu (2009) found that with unprocessed speech, NH subjects' tone and syllable recognition were significantly better with the different-talker condition than with the same-talker condition (as revealed by repeated measures ANOVAs).

C. Concurrent- vs. Single-talker Performance, Adult vs. Adolescent CI Performance

Figure 5 shows performance with concurrent syllables (averaged across the same- and different-talker conditions) as a function of performance with single syllables, for individual adult (filled symbols) and adolescent (open symbols) CI subjects; the solid lines show the linear regressions between concurrent- and single-talker performance. There was a significant correlation between concurrent- and single-talker performance for syllables ($r^2=0.66$, $p=0.02$) and tones ($r^2=0.81$, $p=0.002$), but not for vowels ($r^2=0.25$, $p=0.21$). Among the adolescent subjects, the best performer with single syllables (S6) did not perform the best with concurrent syllables. While S6 performed nearly the best among the adolescent subjects in terms of concurrent-vowel and tone recognition, this subject sometimes exchanged the tones of the two component vowels. Performance for the adult CI subjects was generally better than that for the adolescent CI subjects (except for S4). However, there was no significant difference between the adult and adolescent CI subjects for any of the performance measures ($p>0.05$ in *t*-tests).

D. CI Performance Across Different Vowel Pairs

Figure 6 shows CI subjects' concurrent Chinese syllable, vowel, and tone recognition scores, as a function of vowel pairs in the concurrent syllables. The detailed performance patterns with the same- and different-talker conditions are shown in the left and right panels, respectively. Two-way repeated measures (RM) ANOVAs (with vowel pair and talker condition as factors) showed that recognition performance with concurrent syllables was significantly affected by the vowel pairs [$F(9,63)=14.07$, $p<0.001$ for syllables, $F(9,63)=13.74$, $p<0.001$ for vowels,

and $F(9,63)=2.14$, $p=0.04$ for tones]. Concurrent-vowel and syllable recognition were not significantly different between the talker conditions [$F(1,63)=0.57$, $p=0.48$ for syllables and $F(1,63)=1.16$, $p=0.32$ for vowels], while concurrent-tone recognition was significantly better with the same-talker condition than with the different-talker condition [$F(1,63)=9.91$, $p=0.02$]. There was no significant interaction between vowel pairs and talker conditions [$F(9,63)=1.06$, $p=0.40$ for syllables, $F(9,63)=0.62$, $p=0.78$ for vowels, and $F(9,63)=0.57$, $p=0.82$ for tones]. *Post hoc* Bonferroni *t*-tests showed that concurrent-syllable and vowel recognition were significantly better when the concurrent syllables contained the same vowel (except for /e/-/e/) rather than different vowels ($p<0.05$). *Post hoc* analyses also showed that concurrent-tone recognition was significantly different only between vowel pairs /i/-/i/ and /a/-/e/ ($p=0.04$).

Figure 7 shows the distribution of CI subjects' vowel responses for the different vowel pairs used in the concurrent-syllable recognition tests; the left and right panelsshow the results for the same- and different-talker conditions, respectively. Although this figure does not unambiguously represent CI subjects' vowel pair confusions (e.g., show the number of /a/-/i/ recognized as /a/-/e/) or correspond to CI vowel recognition scores [e.g., 50% /a/ and 50% /i/ responses for the vowel pair /a/-/i/ could arise from all /a/-/i/ responses (100% correct) or half /a/-/a/ and half /i/-/i/ responses (0% correct)], it still reflects the trends of vowel perception in the concurrent-syllable context. When the vowel pairs consisted of the same vowel, subjects tended to correctly identify the target vowels, although this tendency was less marked for /e/-/e/. When /u/ was combined with /i/ or /a/, subjects responded less often with /u/ than with /i/ or /a/. For vowel pairs /a/-/u/ and /a/-/i/, subjects often responded with /e/, which was not one of the component vowels. These perceptual confusions contributed to the poorer recognition performance for vowel pairs consisting of different vowels.

E. CI Performance Across Different Tone Pairs

Figure 8 shows CI subjects' concurrent Chinese syllable, vowel, and tone recognition scores, as a function of tone pairs in the concurrent syllables. The detailed performance patterns with the same- and different-talker conditions are shown in the left and right panels, respectively. Two-way RM ANOVAs (with tone pair and talker condition as factors) showed that recognition performance with concurrent syllables was significantly affected by the tone pairs [$F(9,63)=3.31$, $p=0.002$ for syllables,$F(9,63)=2.82$, $p=0.008$ for vowels, and $F(9,63)=3.38$, $p=0.002$ for tones]. Again, concurrent-vowel and syllable recognition were not significantly different between the talker conditions [$F(1,63)=1.50$, $p=0.26$ for syllables and $F(1,63)=1.28$, $p=0.30$ for vowels], while concurrent-tone recognition was significantly better with the same-talker condition [$F(1,63)=12.73$, $p=0.01$]. There was no significant interaction between tone pairs and talker conditions for syllables [$F(9,63)=1.42$, $p=0.20$] and vowels [$F(9,63)=0.97$, $p=0.47$]. However, for concurrent-tone recognition, tone pairs and talker conditions significantly interacted with each other [$F(9,63)=3.01$, $p=0.005$]. *Post hoc* Bonferroni *t*-tests showed that concurrent-syllable recognition was significantly better for tone pair 1-1 than for tone pairs 1-2, 1-3, and 2-4 ($p<0.05$). Concurrent-vowel recognition was significantly better for tone pair 3-4 than for tone pairs 3-3 and 2-2 ($p<0.05$). *Post hoc* analyses also showed that concurrent-tone recognition with the same-talker condition was significantly better (relative to the different-talker condition) only for tone pairs 1-1, 4-4, and 2-3 ($p<0.03$). In the same-talker condition, concurrent-tone recognition was significantly better for tone pairs 1-1 and 4-4 than for tone pairs 1-2, 1-3, 2-4, and 3-4 ($p<0.02$). However, in the different-talker condition, there was no significant difference in concurrent-tone recognition between any tone pairs.

Figure 9 shows the distribution of CI subjects' tone responses for the different tone pairs used in the concurrent-syllable recognition tests; the left and right panels show the results for the same- and different-talker conditions, respectively. Note that Figure 9 does not show exact tone pair confusions or tone recognition scores, and represents only tone perception trends with

concurrent syllables. Compared to the vowel response patterns as shown in Figure 7, subjects' tone responses were more broadly distributed across the four tone choices (see tone pairs 1-3, 2-4, and 3-4 for examples). For tone pairs 1-1 and 4-4 in the same-talker condition, subjects tended to correctly identify the target tones, resulting in significantly better tone recognition than for other tone pairs in the same-talker condition or the same tone pairs in the different-talker condition. For tone pairs 2-2 and 3-3, subjects tended to confuse Tones 2 and 3.

F. Responses With the Same Vowel or Tone Within a Pair

Finally, CI subjects had a strong bias towards responding with the same vowel or tone for concurrent syllables within a pair. Averaged across the same- and different-talker conditions, CI subjects made a single choice (i.e., only heard one syllable) for 5% of the concurrent-syllable trials (range across subjects: 1 - 16%); the percentage of single-choice responses was not significantly different between the same and different-talker conditions (4 vs. 6%). In addition, the percentages of responses with the same vowel (60% in average, range: 33 - 89%) or tone (40% in average, range: 9 - 90%) within a pair were 1.6 - 2.4 times the percentages of concurrent syllables actually consisting of the same vowel or tone (25% each). The percentage of responses with the same vowel within a pair was not significantly different between the same- and different-talker conditions (62 vs. 58%). However, subjects made significantly more (paired *t*-test: p=0.03) responses with the same tone within a pair in the same-talker condition (43%) than in the different-talker condition (36%).

IV. DISCUSSION

The present CI data (Figure 2, black bars) show that while clinical CI devices and strategies are able to transmit segmental speech information (e.g., vowel cues in quiet), they are able to deliver only limited supra-segmental speech information (e.g., tonal cues in quiet). In the present study, CI subjects' single-syllable vowel recognition (mean: 90% correct) was much better than their single-syllable tone recognition (mean: 63% correct). This difference between vowel and tone recognition was much larger than that in our previous CI study (Luo et al., 2008), in which mean vowel and tone recognition were 69 and 61% correct, respectively. In that study, the two poorest-performing CI subjects exhibited better tone recognition than vowel recognition, whereas the reverse was true for the remaining eight subjects, resulting in similar overall vowel and tone scores. Without the two poorest-performing subjects, mean vowel and tone recognition in Luo et al. (2008) were 77 and 64% correct respectively, showing a trend of better vowel recognition than tone recognition. Some differences between the present and previous studies may explain the difference in vowel results. In the present study, vowel and tone recognition data were extracted from the syllable recognition data; as such, chance performance levels differed among the single-syllable measures (6.25% correct for syllable recognition, 25% correct for both vowel and tone recognition). In Luo et al. (2008), vowel recognition was measured independently with 12 vowels; chance level was 8.33% correct. Thus, differences in the test methods, number of stimuli, and chance performance levels may have contributed to the differences in vowel recognition scores between studies. Although tone recognition performance was comparable between the present and previous CI studies (e.g., Wei et al., 2004;Fu et al., 2004;Luo et al., 2008), the present CI subjects may have relied more exclusively on pitch cues for tone recognition than did the previous studies' CI subjects. In the present study, duration was normalized across the stimuli; thus, subjects could not access duration cues that co-vary with tonal patterns (Fu and Zeng, 2000). Besides, the present CI subjects may not have learned to optimally use amplitude envelope cues associated with Chinese tones, as revealed by their tone confusion matrix (Figure 4).

The performance gap between NH listeners (Figure 2, white bars) and CI subjects (Figure 2, black bars) was much larger with concurrent syllables than with single syllables, partially

because ceiling effects were observed in NH performance with both concurrent and single syllables. To recognize concurrent syllables, listeners first must segregate the two competing syllables, and then identify each component syllable (Assmann and Summerfield, 1990). The difference between the single- and concurrent-syllable data suggests that for CI users, the segregation step is much more challenging than the identification step, most likely due to the poor pitch coding in CI speech processors. Significant correlations were found between CI subjects' single- and concurrent-talker performance (except for vowel recognition). CI subjects with better single-talker performance were more likely to have better concurrent-talker performance. The weaker correlation for vowel recognition was likely due to ceiling effects in CI subjects' single-talker performance. The relatively easy single-syllable vowel recognition task may not have fully differentiated the spectral processing abilities of different CI subjects. Overall, CI users' performance with single talkers may be a good predictor of performance with competing talkers, although most CI users have only limited ability to handle complex auditory scene analysis.

Unlike previous studies with Mandarin-speaking CI users, which only tested either post-lingually deafened adults (e.g., Luo et al., 2008; Wei et al., 2004) or prelingually deafened children (e.g., Peng et al., 2004; Fu et al., 2004), the present study provided an opportunity to directly compare speech performance between the two subject groups. The post-lingually deafened adult CI subjects in the present study were relatively late-implanted and had shorter duration of implant use, compared with the pre-lingually deafened adolescent CI subjects (Table 1). Note that the best performer in this study was subject S3, who was deafened post-lingually and had the shortest duration of deafness. Although she was most recently implanted among the subjects, her duration of implant use appeared to be long enough for her to adapt to the "novel" CI stimuli. While adult performance was generally better than adolescent performance, there was no significant difference in performance between these two subject groups, and more subjects should be tested before drawing any strong conclusions. The number of pre-lingually deafened adolescent CI subjects (4) was too small to confirm the significant negative correlation between speech performance and age at implantation found in previous pediatric CI studies (e.g., Kirk et al., 2002; Zwolan et al., 2004; Connor et al., 2006; Wu et al., 2006). While there were too few subjects to fairly compare speech processing strategies, performance for subject S1 (the only HiRes user) was comparable to that of the other subjects who used the CIS strategy.

Previous studies have shown that NH listeners' concurrent-vowel recognition is significantly better when the competing vowels have large separations in F0 (e.g., Scheffers, 1983; Assmann and Summerfield, 1990) or when the competing vowels have pitch contours that move in different directions (e.g., Chalikia and Bregman, 1989). In the present study, CI subjects' vowel and syllable recognition were similar between the same- and different-talker conditions (Figure 2). However, consistent with the previous 8-channel simulation results (Luo and Fu, 2009), CI subjects had significantly better tone recognition with the same-talker condition than with the different-talker condition. CI subjects' better tone recognition in the same-talker condition mainly derived from their better performance for tone pairs 1-1 and 4-4. In the same-talker condition, concurrent syllables consisting of two exactly same single-vowels may have introduced minimal interference between the component syllables. Also, CI subjects had a stronger bias towards responding with the same tone within a pair in the same-talker condition, which may have resulted in their better recognition of tone pairs 1-1 and 4-4. Contrary to the results of Chalikia and Bregman (1989), CI performance worsened rather than improved for concurrent syllables consisting of different tones (Figure 8). Similar to NH subjects listening to the CI simulations (Luo and Fu, 2009), CI subjects were unable to take full advantage of the available pitch cues (as well as formant transitions and temporal asynchronies) in the acoustic input signals.

With the clinically assigned speech processors, CI subjects' single-syllable vowel recognition was not significantly different from that of NH subjects listening to the 4- or 8-channel CI simulation in Luo and Fu (2009). In contrast, CI subjects' single-syllable tone recognition and concurrent-syllable recognition were significantly poorer than the 4- or 8-channel simulation results. Although the number of electrodes in the CI subjects was more than the number of channels in the CI simulations, the available place pitch cues may have been limited by the interactions between electrodes and the frequency-to-electrode mismatch. Inconsistent with our hypothesis, the present CI subjects did not seem to perceive more salient temporal pitch cues than NH subjects listening to the CI simulations. Alternatively, the poor single-syllable tone recognition of the CI subjects may have been due to their inability to use amplitude envelope cues associated with Chinese tones (see their tone confusion matrix in Figure 4). Targeted auditory training may help CI users associate amplitude envelope cues with Chinese tones and better identify Tones 2 and 3.

The CI simulations in Luo and Fu (2009) somewhat over-estimated the present "real" CI results with concurrent syllables. The present CI results were compared to the closest CI simulation results (with 4 channels) in terms of response and error patterns. As shown in Figure 6, concurrent-vowel recognition varied significantly across the different vowel pairs, while tone recognition was largely unaffected by the vowel pairs, similar to the previous 4-channel simulation results (Luo and Fu, 2009). Note that CI tone recognition was generally poor, and thus may have been subject to floor performance effects that might not have been sensitive to the different vowel pairs. Also consistent with the previous 4-channel simulation results, the different tone pairs affected CI subjects' concurrent-tone recognition much more than vowel recognition (Figure 8). These independent effects on concurrent-vowel and tone recognition suggest that the spectral envelope cues used for vowel recognition and the temporal envelope cues used for tone recognition may not cooperate in CI users' sound source segregation, maybe because of the weak salience for either cue. On the other hand, Green et al. (2004) found that dynamic spectral variations even obscured temporal pitch cues in CI users. Theimproved vowel and tone recognition for concurrent syllables consisting of the same vowel or tone, respectively, may have reflected CI subjects' bias towards responding with the same vowel or tone for concurrent syllables within a pair.

There were some common response errors for the previous simulation study (Luo and Fu, 2009) and the present CI study, in terms of concurrent-vowel and tone recognition. For example, the poorer recognition performance for vowel pairs /a/-/u/ and /i/-/u/ was likely the result of /u/ being masked by /a/ or /i/. Similar to the simulation results, CI subjects' perception of vowel pairs /a/-/u/ or /i/-/u/ was dominated by the higher formant frequencies of /a/ or /i/. The better recognition scores for vowel pair /e/-/i/ may have been because /i/ contains the largest ratio between the first formant (F1) and second formant (F2) frequencies among the four Chinese vowels; this large formant ratio may have produced minimal interaction with /e/. Also similar to the simulation results, CI subjects often responded with /e/ when presented with vowel pair /a/-/i/. According to Luo and Fu (2009), the combination of F1 for /i/ with F2 for /a/ may have produced the perception of /e/. Finally, CI subjects often selected Tone 1 for tone pair 2-4, even though Tone 1 was not presented. The rising pitch contour (and amplitude envelope) of Tone 2 and the falling pitch contour (and amplitude envelope) of Tone 4 may have cancelled each other out, resulting in the perception of Tone 1 (flat).

There were also different response errors between the previous simulation study (Luo and Fu, 2009) and the present CI study, in terms of concurrent-vowel recognition, especially for some vowel pairs consisting of the same vowel. For example, the present CI subjects had significantly better vowel recognition scores when the vowel /a/ was paired with /a/ rather than /e/, /u/, or /i/. With the 4-channel CI simulation, NH listeners often heard /e/ (25% of responses) when presented with /a/-/a/, resulting in poorer performance than that of CI subjects for the /a/

pairings. This discrepancy in concurrent-vowel recognition between CI and NH subjects can be interpreted by comparing their vowel confusion matrices with single syllables. As shown in Figure 3, NH listeners often responded with /e/ when presented with /a/ in the 4-channel CI simulation; CI subjects made far fewer errors when presented with /a/.

There were also different response errors between the previous simulation study (Luo and Fu, 2009) and the present CI study, in terms of concurrent-tone recognition, especially for certain tone pairs consisting of the same tone. For example, NH subjects listening to the 4-channel CI simulation had significantly better tone recognition scores with tone pairs 2-2 and 3-3 than with the other tone pairs. In contrast, CI performance with tone pairs 2-2 and 3-3 was poorer than the 4-channel simulation results, as CI subjects often confused Tone 2 with Tone 3 (Figure 9). Similar to vowel recognition, this discrepancy in concurrent-tone recognition between CI and NH subjects is consistent with their single-syllable tone confusion patterns. With single syllables, CI subjects also more often confused Tones 2 and 3 than did NH subjects listening to the 4-channel CI simulation (Figure 4).

V. CONCLUSIONS

Recognition of Chinese syllables, vowels, and tones was measured for single- and concurrent-syllables in both post-lingually deafened adult and pre-lingually deafened adolescent CI subjects. There were no significant differences in single- or concurrent-syllable recognition scores between the adolescent and adult CI subjects, possibly due to the small number of subjects. Using clinically assigned speech processors, CI subjects performed only slightly above chance for concurrent syllables. CI subjects' concurrent-syllable performance was significantly correlated with single-syllable performance. CI subjects' response patterns for vowel pairs with the same vowel or tone pairs with the same tone were consistent with single-vowel or tone confusion matrices, respectively. CI subjects' concurrent-vowel and syllable recognition were not significantly different between the same- and different-talker conditions, while concurrent-tone recognition was significantly better with the same-talker condition (mainly for tone pairs 1-1 and 4-4). Vowel and tone recognition were better when concurrent syllables contained the same vowels or tones, respectively. Across the different vowel pairs, tone recognition was less variable than vowel recognition; across the different tone pairs, vowel recognition was less variable than tone recognition. The results suggest that for CI users who speak tonal languages, different tones presented simultaneously are not useful for sound source segregation, but rather pose special challenges for speech recognition with competing talkers, due to the lack of spectro-temporal details provided by CI devices.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in the present study. We thank John J. Galvin III for assistance in editing the manuscript. We would also like to thank two anonymous reviewers for their constructive comments on an earlier version of this paper. Research was supported in part by NIH grants R03-DC-008192 and R01-DC-004993.

Abbreviations

CI, cochlear implant; NH, normal hearing; F0, fundamental frequency; CIS, continuous interleaved sampling; HiRes, Hi-Resolution; RMS, root mean square..

REFERENCES

Assmann PF, Summerfield Q. Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. J. Acoust. Soc. Am 1990;88(2):680–697. [PubMed: 2212292]

Bregman, AS. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press; Cambridge, MA: 1990.

Chalikia MH, Bregman AS. The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation. Percept. Psychophys 1989;46(5):487–496. [PubMed: 2813035]

Connor CM, Craig HK, Raudenbush SW, Heavner K, Zwolan TA. The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: is there an added value for early implantation? Ear Hear 2006;27(6):628–644. [PubMed: 17086075]

Culling JF, Summerfield Q. The role of frequency modulation in the perceptual segregation of concurrent vowels. J. Acoust. Soc. Am 1995;98(2):837–846. [PubMed: 7642822]

Friesen LM, Shannon RV, Baskent D, Wang X-S. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am 2001;110(2):1150–1163. [PubMed: 11519582]

Fu Q-J, Hsu C-J, Horng M-J. Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant users. Ear Hear 2004;25(5):501–508. [PubMed: 15599196]

Fu Q-J, Zeng F-G. Effects of envelope cues on Mandarin Chinese tone recognition. Asia-Pacific Journal of Speech, Language and Hearing 2000;5(1):45–57.

Fu Q-J, Zeng F-G, Shannon RV, Soli SD. Importance of tonal envelope cues in Chinese speech recognition. J. Acoust. Soc. Am 1998;104(1):505–510. [PubMed: 9670541]

Green T, Faulkner A, Rosen S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J. Acoust. Soc. Am 2004;116(4):2298–2310. [PubMed: 15532661]

Kirk KI, Miyamoto RT, Lento CL, Ying E, O'Neill T, Fears B. Effects of age at implantation in young children. Ann. Otol. Rhinol. Laryngol. Suppl 2002;189:69–73. [PubMed: 12018353]

Laneau J, Moonen M, Wouters J. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. J. Acoust. Soc. Am 2006;119(1):491–506. [PubMed: 16454303]

Licklider JCR. A duplex theory of pitch perception. Experimentia 1951;7:128–134.

Luo X, Fu Q-J. Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. J. Acoust. Soc. Am 2004;116(6):3659–3667. [PubMed: 15658716]

Luo X, Fu Q-J. Concurrent vowel and tone recognition in acoustic and simulated electric hearing. J. Acoust. Soc. Am 2009;125(5):3223–3233. [PubMed: 19425665]

Luo X, Fu Q-J, Galvin JJ III. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. Trends Amplif 2007;11(4):301–315. [PubMed: 18003871]

Luo X, Fu Q-J, Wei C-G, Cao K-L. Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. Ear Hear 2008;29(6):957–970. [PubMed: 18818548]

Marin CMH, McAdams S. Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width. J. Acoust. Soc. Am 1991;89(1): 341–351. [PubMed: 2002173]

McDermott HJ. Music perception with cochlear implants: a review. Trends Amplif 2004;8(2):49–82. [PubMed: 15497033]

McKay CM, Carlyon RP. Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains. J. Acoust. Soc. Am 1999;105:347–357. [PubMed: 9921661]

Peng S-C, Tomblin JB, Cheung H, Lin Y-S, Wang L-S. Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. Ear Hear 2004;25(3):251–264. [PubMed: 15179116]

Peng S-C, Tomblin JB, Turner CW. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. Ear Hear 2008;29(3):336–351. [PubMed: 18344873]

Qin MK, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. Ear Hear 2005;26(5):451–460. [PubMed: 16230895]

Scheffers, MTM. Ph.D. thesis. Groningen University; The Netherlands: 1983. Sifting Vowels: Auditory Pitch Analysis and Sound Segregation.

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995;270:303–304. [PubMed: 7569981]

Stickney GS, Zeng F-G, Litovsky RY, Assmann PF. Cochlear implant speech recognition with speech maskers. J. Acoust. Soc. Am 2004;116(2):1081–1091. [PubMed: 15376674]

Wang R-H. The standard Chinese database. University of Science and Technology of China, internal materials. 1993

Wei C-G, Cao K-L, Zeng F-G. Mandarin tone recognition in cochlear-implant subjects. Hear. Res 2004;197:87–95. [PubMed: 15504607]

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. Nature 1991;352:236–238. [PubMed: 1857418]

Wu JL, Lin CY, Yang HM, Lin YH. Effect of age at cochlear implantation on open-set word recognition in Mandarin speaking deaf children. Int. J. Pediatr. Otorhinolaryngol 2006;70(2):207–211. [PubMed: 16043234]

Zeng F-G. Trends in cochlear implants. Trends Amplif 2004;8(1):1–34. [PubMed: 15247993]

Zwolan TA, Ashbaugh CM, Alarfaj A, Kileny PR, Arts HA, El-Kashlan HK, Telian SA. Pediatric cochlear implant patient performance as a function of age at implantation. Otol. Neurotol 2004;25 (2):112–120. [PubMed: 15021769]

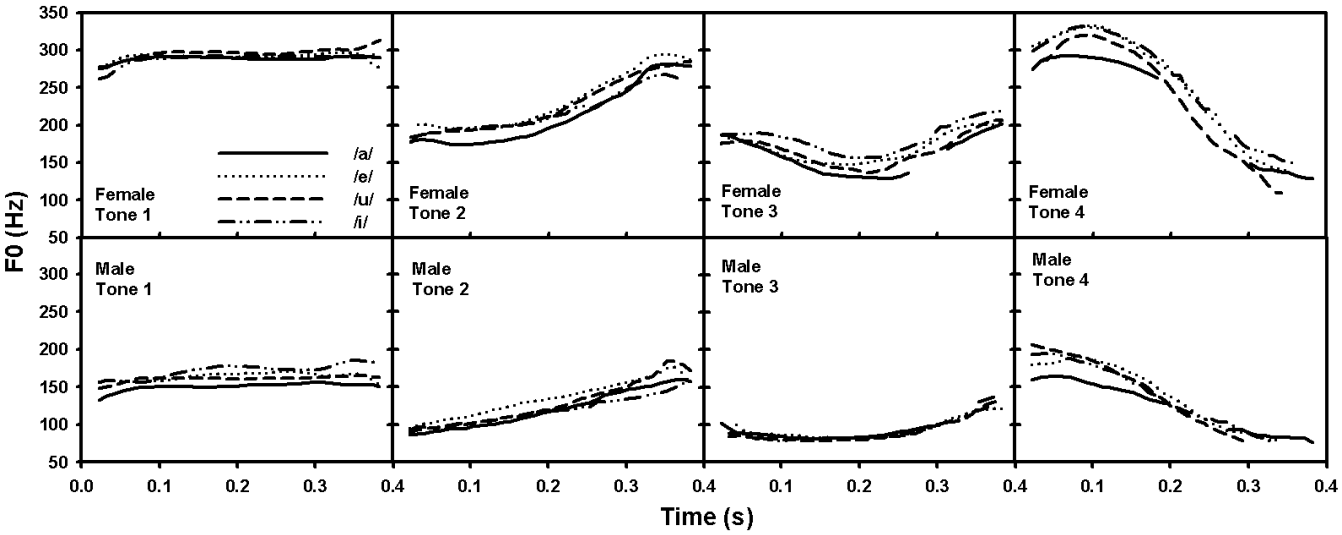

Figure 1.

F0 contours of the Chinese single-vowel stimuli produced by the female (upper panels) and male talkers (lower panels). Different columns represent different tones and different line types represent different vowels.

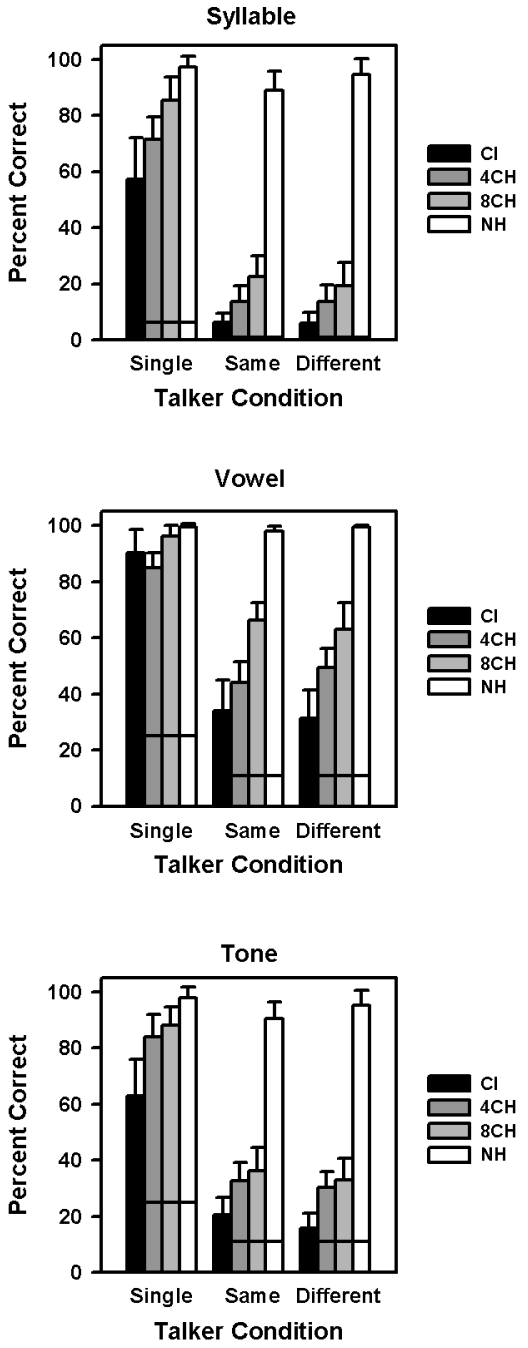

Figure 2.

Mean Chinese syllable (top panel), vowel (middle panel), and tone recognition scores (bottom panel), as a function of talker conditions (Single: single-talker; Same: the male talker combined with himself; Different: the male talker combined with the female talker). CI data from the present study is shown by the black bars. NH performance from a previous CI simulation study (Luo and Fu, 2009) is shown for 4 channels (dark gray bars), 8 channels (light gray bars), and unprocessed speech (white bars). The error bars represent one standard deviation of the mean. The horizontal line within each bar indicates the chance performance level for the specific recognition task.

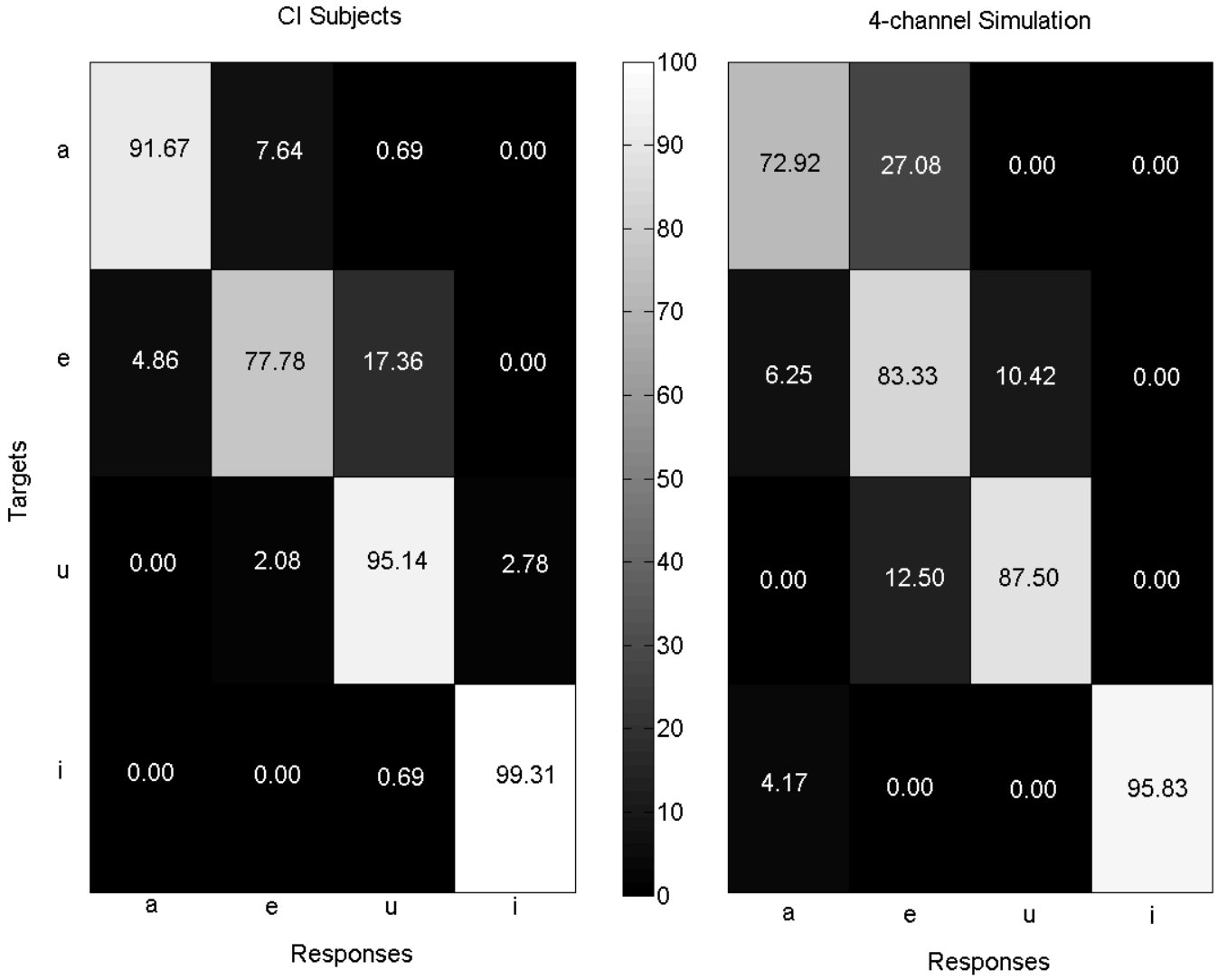

Figure 3.
Vowel confusion matrices for single-syllable vowels for CI subjects (left panel) and NH subjects listening to a 4-channel noise-band CI simulation (right panel; data from Luo and Fu, 2009). The percentages of vowel responses to target vowels are shown in each cell using numbers and shaded according to a grayscale continuum (shown between the panels).

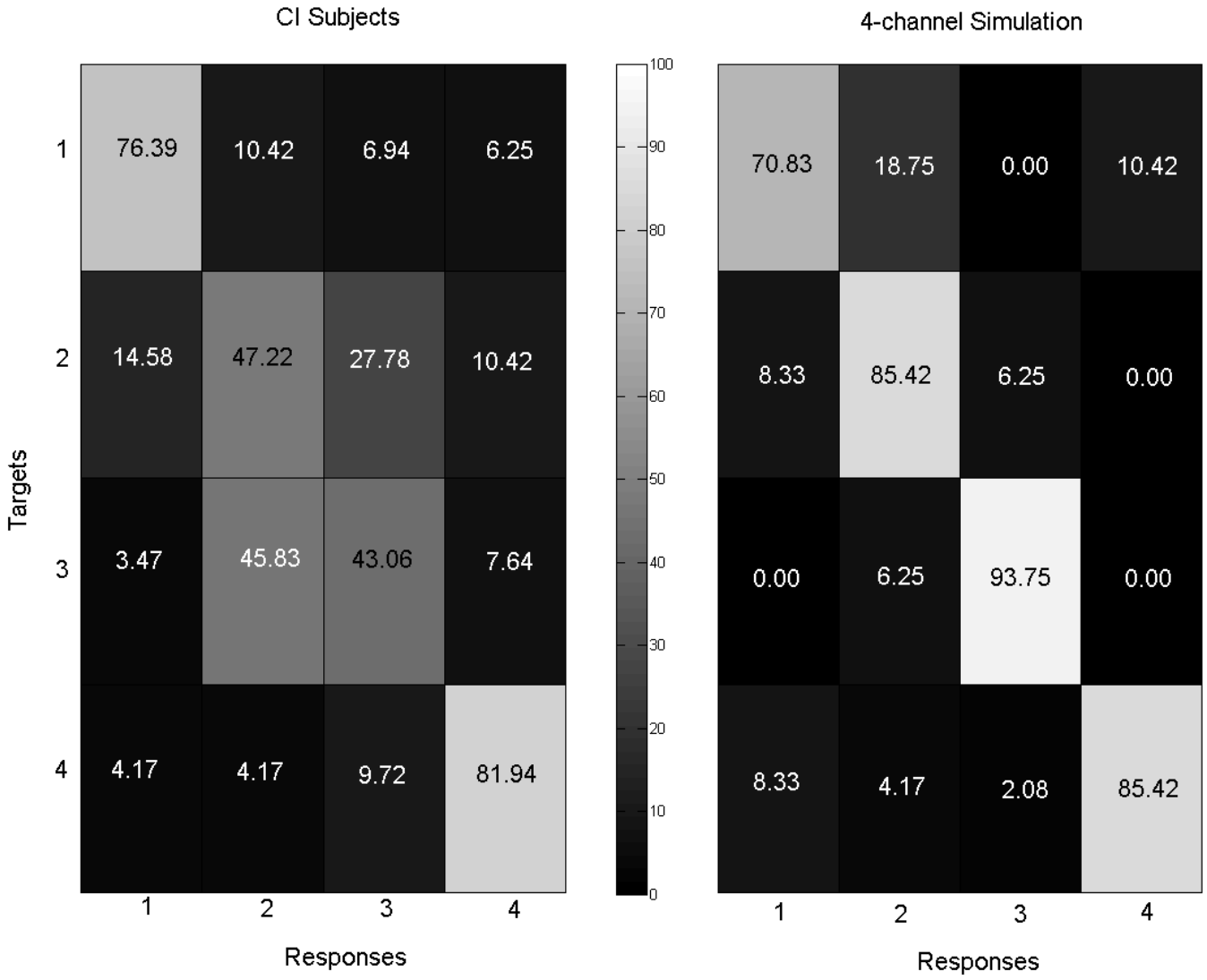

Figure 4.

Tone confusion matrices for single-syllable tones for CI subjects (left panel) and NH subjects listening to a 4-channel noise-band CI simulation (right panel; data from Luo and Fu, 2009). The percentages of tone responses to target tones are shown in each cell using numbers and shaded according to a grayscale continuum (shown between the panels).

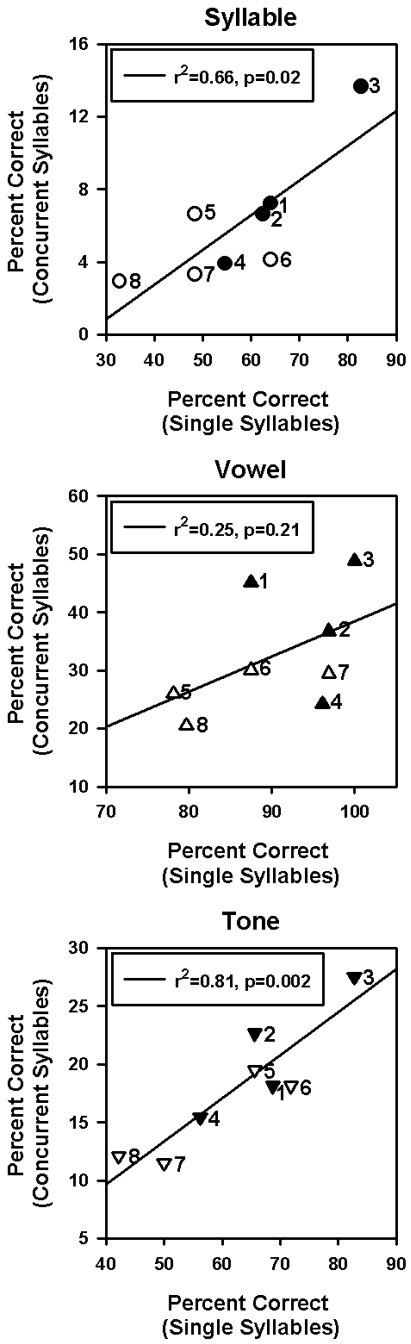

Figure 5.

Chinese syllable (top panel), vowel (middle panel), and tone recognition scores (bottom panel) with concurrent syllables (averaged across the same- and different-talker conditions) for individual adult (filled symbols) and adolescent CI subjects (open symbols), as a function of single-syllable performance. Individual subjects are identified by the numbers to the right of the symbols. The solid lines show the linear regressions between single- and concurrent-syllable performance.

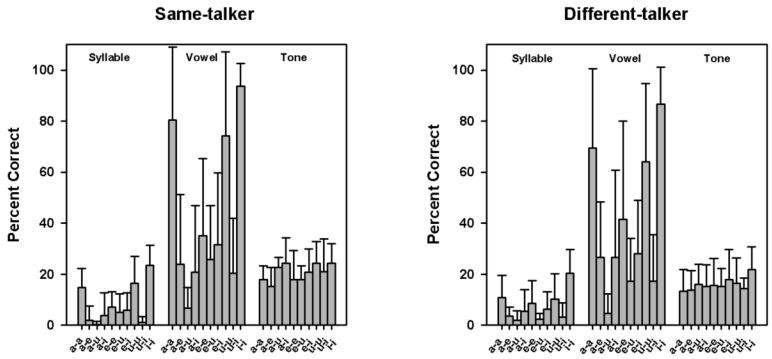

Figure 6.

Concurrent Chinese syllable, vowel, and tone recognition scores for CI subjects, as a function of vowel pairs in the concurrent syllables. The detailed performance patterns with the same- and different-talker conditions are shown in the left and right panels, respectively. The error bars represent one standard deviation of the mean.

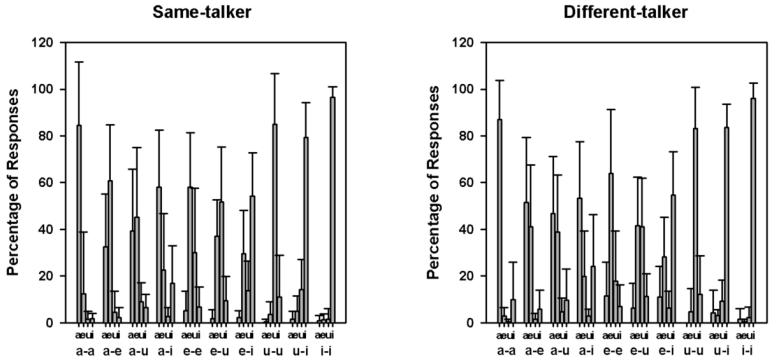

Figure 7.

Distribution of CI subjects' vowel responses (in percentage of responses) for different vowel pairs in the concurrent syllables. The left and right panels show the results for the same- and different-talker conditions, respectively. The error bars represent one standard deviation of the mean.

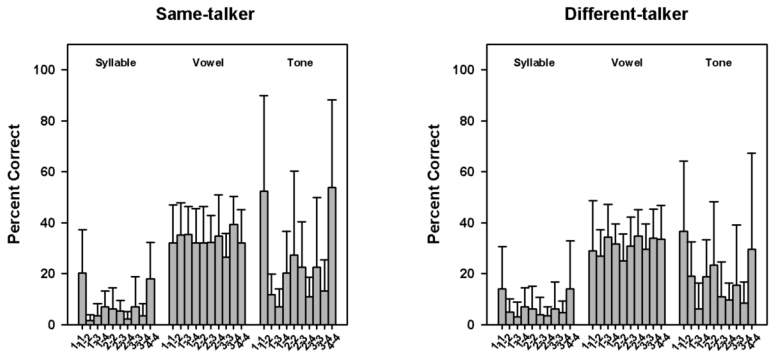

Figure 8.

Concurrent Chinese syllable, vowel, and tone recognition scores for CI subjects, as a function of tone pairs in the concurrent syllables. The detailed performance patterns with the same- and different-talker conditions are shown in the left and right panels, respectively. The error bars represent one standard deviation of the mean.

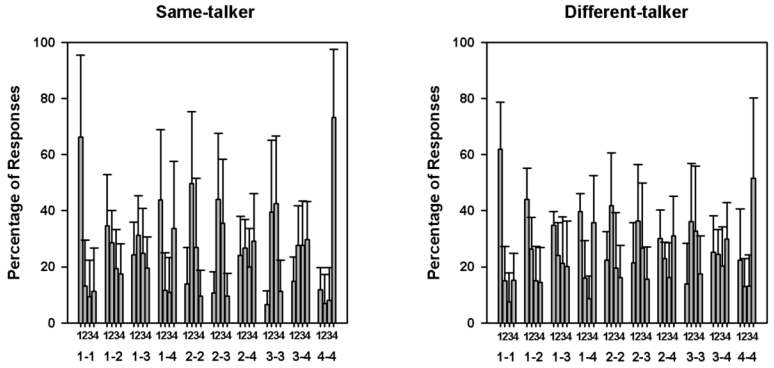

Figure 9.

Distribution of CI subjects' tone responses (in percentage of responses) for different tone pairs in the concurrent syllables. The left and right panels show the results for the same- and different-talker conditions, respectively. The error bars represent one standard deviation of the mean.

Table 1

CI subject demographics

Patients	Age	Gender	Duration of Deafness (years)	Etiology	Device	Strategy	Years with Prosthesis (yy:mm)
S1	32	M	29	Sudden hearing loss	Advanced Bionics Auria	HiRes	2:0
S2	48	F	15	Unknown	Medel Tempo+	CIS	7:8
S3	32	F	4	Unknown	Medel Tempo+	CIS	1:2
S4	58	F	53	Unknown	Medel Tempo+	CIS	5:8
S5	15	F	15	Unknown	Medel Tempo+	CIS	9:10
S6	15	F	15	Unknown	Medel Tempo+	CIS	9:10
S7	12	M	10	Unknown	Medel Tempo+	CIS	9:3
S8	10	M	8	Unknown	Medel Tempo+	CIS	4:7