Hearing Research 256 (2009) 75–84

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research papers

Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users

Xin Luo [a,*], Qian-Jie Fu [b], Hung-Pin Wu [c,d], Chuan-Jen Hsu [e]

[a] Department of Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, IN 47907, USA
[b] Division of Communication and Auditory Neuroscience, House Ear Institute, 2100 West 3rd Street, Los Angeles, CA 90057, USA
[c] Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University, Taipei, Taiwan
[d] Department of Otolaryngology, Buddhist Tzuchi General Hospital, Taipei, Taiwan
[e] Department of Otolaryngology, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 11 December 2008
Received in revised form 25 June 2009
Accepted 8 July 2009
Available online 10 July 2009

Keywords:
Concurrent-vowel recognition
Concurrent-tone recognition
Cochlear implants
Mandarin Chinese

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. The present results suggest that interference between concurrent tones may contribute to Mandarin-speaking CI users' susceptibility to competing-talker backgrounds.

1. 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

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

* Corresponding author. Tel.: +1 765 496 7267; fax: +1 765 494 0771.
E-mail address: luo5@purdue.edu (X. Luo).

0378-5955/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2009.07.001

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.

2. Methods

2.1. Subjects

Four adult and 4 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.

2.2. 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