MANDARIN TONE PERCEPTION

Mandarin tone recognition in acoustic and electric hearing

YING-YEE KONG, Department of Cognitive Sciences and Department of Otolaryngology, University of California, Irvine, CA 92697, USA
FAN-GANG ZENG, Department of Cognitive Sciences, Department of Otolaryngology and Department of Biomedical Engineering, University of California, Irvine, CA 92697, USA

Objectives

Mandarin tone recognition can be achieved using either spectral cues or temporal cues. Although the pitch contour is the most salient cue for voice pitch perception × Liang, 1963; Abramson, 1978), temporal envelope cues such as amplitude contour, periodicity and duration also contribute significantly to Mandarin tone recognition (Liang, 1963; Fu et al., 1998; Fu and Zeng, 2000). Fu and colleagues (1998, 2000) showed that normal-hearing listeners performed approximately 80% tone identification with only temporal envelope cues. However, no studies have been done on whether reliable Mandarin tone recognition could be achieved in the presence of noise. The purpose of this study was to investigate the relative contribution of envelope and fine-structure cues in normal-hearing and cochlear-implanted listeners to Mandarin tone recognition in quiet and in noise. We hypothesized that while the temporal envelope can support tone recognition in quiet, it is not adequate in noisy conditions. We further hypothesized that the temporal fine structure is required for tone recognition in noise.

Materials and methods

Speech materials consisted of a list of 100 words (25 syllables × 4 tones) recorded from one male and one female native Beijing Mandarin speaker. Original stimuli were then processed to extract either the slow-varying amplitude modulation alone (AM-only) from a number of frequency bands (1–32 bands) using the processing scheme of the cochlear implant (Shannon et al., 1995) or the AM with additional slow-varying frequency modulation (AM+FM) using a signal processing algorithm developed in our laboratory (Nie et al., 2003).

One native Mandarin-speaking cochlear-implanted and four normal-hearing listeners participated in the experiment. The cochlear-implanted listener had substantial low-frequency residual hearing in the non-implanted ear. His aided thresholds were obtained at 40 to 75 dB HL from 125 to 2000 Hz. Normal-hearing listeners were tested with both original and processed stimuli with and without fine structure cues. The cochlear-implanted listener was tested with only original stimuli under three listening conditions: cochlear implant alone, hearing aid alone, and binaurally combined acoustic and electric hearing. Stimuli were presented in quiet and in speech-shaped noise. For the normal-hearing listeners, the target speech was presented monaurally via headphones at 65 dBA while the noise levels were varied to produce six signal-to-noise ratio (SNR) conditions. For the cochlear-implanted listener, speech stimuli were presented via a loudspeaker at a comfortable level for each listening condition.

Results

For the normal-hearing listeners, on average, tone recognition with original stimuli maintained at 98 to 100% correct from quiet to –5 dB SNR. Performance decreased by 25 percentage points when the SNR was further reduced to –10 dB, and finally approached the chance level at –15 dB SNR. In the one-band AM-only condition, tone recognition performance decreased monotonically from 79% correct in quiet to 51% at 0 dB SNR and approached the chance level at –5 dB SNR. With envelope cues alone, 32 bands of frequency information were required to produce tone recognition performance similar to the original stimuli. Tone recognition with 32 bands decreased from 99% correct in quiet to 62% at –5 dB SNR and finally to 38% at –10 dB SNR. However, the performance difference between the original and the 32-band AM-only stimuli was still large (p <0.01), especially in noise. The difference was 5, 15, 38 and 36 percentage points for 5, 0, –5 and –10 dB SNRs, respectively. In contrast, only four bands were necessary for both quiet and noise conditions with the additional fine-structure cues. Performance difference between the four-band AM+FM and the original stimuli was only 13 percentage points at –5 dB SNR and 18 percentage points for –10 dB SNR. The difference between the AM+FM and the AM-only condition was also significant (p <0.05) in quiet and at all SNRs independent of the number of frequency bands. Performance with 32 bands in AM-only resembled that obtained with only two bands in the AM+FM condition. The cochlear-implant listener showed better tone recognition with hearing aid alone than with cochlear implant alone, but showed similar performance with hearing aid alone and combined hearing in quiet and at all SNRs. His performance was similar to a one-band AM-only simulation in quiet and in noise with cochlear implant alone, but equivalent to an eight-band AM-only simulation with hearing aid alone or with combined hearing. The difference between cochlear implant alone and hearing aid alone was greater in noise than in quiet, approximately 20 percentage points from –5 to 10 dB SNRs but only nine percentage points in quiet.

Conclusions

Our results first suggest that while envelope cues alone can produce 80 to 90% tone recognition in quiet, it cannot support robust voice pitch perception in noise. With the envelope cues alone, high spectral resolution with at least 32 bands was needed for tone recognition. However, such fine spectral detail still does not produce the level of tone recognition in noise that is comparable to the original stimuli. Tone recognition in noise improved significantly when additional fine structure cues were provided. These findings were further supported by the tone recognition in the cochlear-implanted listener who performed considerably better in noise with the low-frequency fine-structure information in the non-implanted ear compared to the cochlear-implanted ear. The present study suggests the importance and urgency of accurately encoding the low-frequency fine-structure cues for voice pitch perception in cochlear implants.

References

Abramson AS (1978) Static and dynamic acoustic cues in distinctive tones. Lang Speech 21(4): 319–325.

Fu QJ, Zeng FG (2000) Identification of temporal envelope cues in Chinese tone recognition. Asia Pacific Journal of Speech, Language and Hearing 5: 45–57.

Fu QJ, Zeng FG, Shannon RV, Soli SD (1998) Importance of tonal envelope cues in Chinese speech recognition. Journal of the Acoustical Society of America 104(1): 505–510.

Liang Z. (1963) The auditory basis of tone recognition in Standard Chinese. Acta Physioloica Sinica 26: 85–90.

Nie K, Stickney G, Zeng FG (2003) Encoding frequency modulation to improve cochlear implant performance in noise. IEEE (submitted).
Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270(5234): 303–304.
Address correspondence to: Ying-Yee Kong, 364 Med Surge II, University of California, Irvine, Irvine, CA 92697, USA.

Factors contributing to tone perception performance: implications for designing paediatric assessment instruments

KATHY YUET-SHEUNG LEE, CHARLES ANDREW VAN HASSELT, Division of Otorhinolaryngology, Department of Surgery, Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong, SAR

Objective

To study the factors contributing to the tone perception performance of Cantonese-speaking children.

Material and methods

Two hundred and twenty-five Cantonese-speaking children (100 girls and 125 boys), aged 2;09 to 3;03, with normal hearing and language development were recruited by stratified random sampling in Hong Kong.

The three basic tones of high-level (T1), high-rising (T2) and low-falling (T4) tones were used for experimental testing. Represented by line drawings, a total of 32 tonal pairs on T1/T2, T2/T4, T1/T4 were constructed. Subjects were required to point to the named picture after live voice presentation. Detailed description of the subject selection and the testing procedures can be found in Lee et al. (2002a,b).

Background information including the children characteristics, the language input, the family and schooling status were obtained by parent interview. The Cantonese Receptive Vocabulary Test (Lee et al., 1996), a test standardized on 609 Cantonese-speaking children, was performed on each subject to assess the level of vocabulary knowledge. The familiarity and the frequency of occurrences of the test words used were measured by ratings given by parents. All the possible contributing factors to the tone-perception score were then explored by logistic regression.

Results

Table 1 shows the results of logistic regression. Five significant factors and six insignificant factors to the tone-perception score were found.

Discussion

Five factors were found to be significant in contributing to the tone perception scores. The first factor is age, which suggests that tone-perception ability is still in the process of development in children within the age range of 2;09 to 3;03. The factor of vocabulary knowledge, which may be related to the age factor, was also found to be significant. With increasing age, one's vocabulary knowledge improves. The more solid one's language