Identification of temporal envelope cues in Chinese tone recognition

Qian-Jie Fu
House Ear Institute, Los Angeles, CA, USA

Fan-Gang Zeng
University of Maryland, College Park, MD, USA

Abstract
For tonal languages such as Mandarin Chinese, tone recognition is important for understanding the meaning of words, phrases or sentences. While fundamental frequency carries the most distinctive information for tone recognition, waveform temporal envelope cues can also produce a high level of tone recognition. This study attempts to identify what types of temporal envelope cues contribute to tone recognition and whether these temporal envelope cues are dependent on speakers and vowel contexts. Several signal-correlated-noise stimuli were generated to separate the contribution of three major temporal envelope cues – duration, amplitude contour, and periodicity – to tone recognition. Perceptual results show that the duration cue contributed mostly to discrimination of Tone-3, the amplitude cue contributed mostly to Tone-3 and Tone-4 discrimination, and the periodicity cue contributed to recognition of all tones. However, tone recognition based on temporal envelope cues was highly variable across speakers and vowel contexts. Acoustic analysis of these temporal envelope cues revealed that this variability in tone recognition is directly related to the acoustic variability between the amplitude contour and fundamental frequency contour.

Key words: Chinese tone recognition, temporal envelope cues, duration, amplitude, periodicity

Introduction

As a distinctive feature in tonal languages, the tonality of a monosyllable is lexically meaningful (Liang, 1963; Lin, 1988; Wang, 1989). In Mandarin Chinese, there are four lexical tones, which can be described as: high-level, mid-rising, mid-falling-rising, and high-falling based on their fundamental frequency (F_0) variation patterns (Howie, 1976; Wang, 1989). These tones are traditionally termed: Tone-1, Tone-2, Tone-3, and Tone-4, respectively. For example, the syllable /ma/ expressed with Tone-1, -2, -3, and -4 would mean 'mother', 'hemp', 'horse', and 'reproach', respectively.

Although the principal phonetic feature for tone recognition is in the frequency domain (Liang, 1963; Abramson, 1978; Lin, 1988), other acoustic characteristics

have been found to co-vary with F0 changes in isolated Chinese words. One such acoustic cue is vowel duration, which differs for different tones. The falling-rising tone (Tone-3) is longest in vowel duration, while the falling tone (Tone-4) is shortest (Howie, 1974). Another acoustic cue is the amplitude contour. The falling-rising tone (Tone-3) generally is produced with the lowest amplitude, while the falling tone (Tone-4) is the highest. Additionally, a correlation has been observed between the amplitude contour and F0 contour (Garding et al., 1986; Sagart, 1986; Whalen and Xu, 1992).

Several studies have investigated the effects of vowel duration and amplitude contour on the perception of Mandarin Chinese tones in the presence of F0 information. Lin (1988) directly manipulated vowel duration but preserved F0 information; he found that the score of tone recognition was only 3% lower for stimuli without the duration cue than for stimuli with the duration cue. Similarly, Lin (1988) directly manipulated the amplitude contour to remove the correlation between the amplitude contour and F0 contour and found that the recognition score dropped by only 1.2%. Lin suggested that the contribution of these acoustic features to tone recognition was negligible in the presence of F0 information.

On the other hand, several studies have attempted to identify the contribution of these temporal envelope cues to tone recognition when the F0 information was partially or totally removed. Liang (1963) presented high-pass filtered (300 Hz) speech sounds and reported 94.6% correct tone recognition. This high recognition score was probably due to the residual pitch, which was easily extracted from the harmonic fine structure (Schouten et al., 1962). Liang (1963) also found that tone recognition score dropped to 64% correct when F0 information and the harmonic fine structure were both removed in whispered speech. Two more recent studies used signal-correlated noise stimuli to remove both the F0 pattern and the harmonic fine structure and found that about 80% correct of tone recognition could be achieved (Whalen and Xu, 1992; Fu et al., 1998). This high level of tone recognition at the syllable level also contributed significantly to Chinese sentence recognition (Fu et al., 1998).

Similarly to Rosen (1992), the present study defines temporal envelope cues as the following three main acoustic cues: duration, amplitude contour and periodicity. The duration cue refers to the overall duration of the single-vowel syllable. The amplitude contour refers to fluctuations in the overall amplitude at a rate between 2 Hz and 50 Hz. Periodicity refers to fluctuations in the overall amplitude at a rate between 50 Hz and 500 Hz, which is directly correlated to the change in fundamental frequency. Both acoustic and perceptual studies were conducted in an attempt to isolate the contribution of these three temporal envelope cues to tone recognition. The acoustic study measured the distribution probability of the duration of different tones and the distribution probability of the correlation coefficients between the amplitude contour and F0 contour. The perceptual study measured tone recognition as a function of stimulus duration, amplitude contour and periodicity. Finally, the acoustic variability was related to perceptual variability as a function of speakers and vowel contexts.

Methods

Subjects

Four young adult listeners participated in this study and were paid for their service. All subjects had pure tone thresholds less than 15dB HL bilaterally at octave fre-

quencies from 250 Hz to 8000 Hz. Subjects were instructed to listen to the tonality of each token, regardless of the vowel and speaker identity.

Stimuli

The stimuli were derived from the 'Chinese Standard Database' (Wang, 1993). Five male and five female speakers produced four tones for each of six single-vowel syllables in Mandarin Chinese, resulting in a total of 240 tokens. The six vowels were a[a], o[o], e[ɣ], i[i], u[u], ü[y]. These stimuli were digitized using a 16-bit A/D converter at a 16-kHz sampling rate without high frequency pre-emphasis.

Seven stimulus sets were processed to produce temporal waveform envelope acoustic cues containing only duration (D), or amplitude contour (A), or periodicity (P), or a combination of these cues. Table 1 describes the seven stimulus sets.

Table 1: *Seven stimulus sets containing different temporal cues that may contribute to tone recognition*

Stimuli conditions	APD	AP	AD	PD	A	P	D
Duration cue	✓		✓	✓			✓
Amplitude contour	✓	✓	✓		✓		
Periodicity cue	✓	✓		✓		✓	

Notes: A = amplitude contour cue, P = periodicity cue, D = duration cue. The term APD refers to a stimulus set whose tokens contain all three non-spectral cues: duration (D), amplitude contour (A), and periodicity (P).

(1) *APD-stimulus*: The envelope of the speech signal was derived by half-wave rectification and then digitally low-pass filtered with a cut-off frequency of 500 Hz. The sign of each envelope sample was then randomly flipped to produce the 'signal-correlated-noise' stimulus (Schroeder, 1968). These processed stimuli are referred to as 'APD-stimulus' because they contain three non-spectral cues: amplitude contour, periodicity and duration.

(2) *AP-stimulus*: The AP-stimulus set was generated by modifying the APD-stimuli to have the same 400-ms duration for all vowels by a linear interpolation method. For example, assume that the original length of one token is L-ms. For a given sample location n, the output $y(n)$ is equal to: $x(m) + \alpha[x(m+1)-x(m)]$, where m is the integral part of the product $n*L/400$ and α is the remainder. Note that this linear manipulation may change the range of F0 variation within the syllable, but the percentile F0 change remains the same.

(3) *AD-stimulus*: The generation of the AD-stimulus was similar to the APD-stimulus, except that a low-pass envelope filter with a cut-off frequency of 50 Hz (96 dB/octave) was used to eliminate the periodicity information.

(4) *PD-stimulus*: The F0 of the speech signal was derived by the short-term autocorrelation method (Rabiner, 1977), and used to 100% amplitude modulate a wideband noise (white noise with a bandwidth of 8 kHz) having the same duration as the original signal.

(5) *A-stimulus*: The A-stimulus set was generated by modifying the AD-stimulus to

have the 400-ms duration for all vowels by the linear interpolation method as described in (2).

(6) *P-stimulus*: The P-stimulus set was generated by modifying the PD-stimulus to have the same 400-ms duration for all vowels by the linear interpolation method as described in (2).

(7) *D-stimulus*: The speech signal was replaced by a wideband noise (white noise with a bandwidth of 8 kHz), but with the naturally varying vowel duration preserved.

Procedure

A stimulus token was randomly chosen from all 240 tokens, and was processed by a Dell Pentium-90 PC computer on line. The level of each stimulus set was expressed as the level of a steady-state random noise with root-mean-square (RMS) amplitude equal to the average RMS amplitude of all stimuli in the set. The nominal stimulus presentation level was 70 dB SPL, as measured with a TDH-49 headphone mounted in an MX41/AR cushion with an NBS-9A coupler. The subject was forced to identify the token from four alternative choices. All subjects had participated in the previous experiments (Fu et al., 1998) and were familiar with the signal-correlated-noise speech tokens. Subjects were trained on sample conditions before formal testing was begun. No feedback was provided. For all tests, subjects were seated in a double-walled sound-treated booth and tested individually.

For all subjects, the unprocessed speech condition was tested first. All subjects achieved 98% or better tone recognition scores. The order of the experimental conditions was randomized and counterbalanced across subjects. Data were stored in the form of a 4*4 confusion matrix for each different stimulus set and each subject. A total of 28 matrices were collected (4 subjects * 7 conditions). For each subject, each experimental condition was presented four times and the final confusion matrix was formed by summing the matrices from the four runs.

Results

Acoustic analysis

The duration of all 240 tokens was extracted to form a duration distribution histogram. A Gaussian distribution function was used to fit the duration distribution histogram. To calculate the correlation coefficient between amplitude contour and F0 contour, the whole vowel token was first divided into 20 even frames. Within each frame, the F0 value was extracted by the short-term autocorrelation method (Rabiner, 1977) and the amplitude was computed by estimating the RMS value within this frame. The correlation coefficient for this token was then obtained by a linear regression analysis on the F0-contour array and amplitude-contour array.

Duration distribution

Figure 1 shows the duration histogram for each individual tone across all six syllables and ten speakers. The solid line represents the fitting curve based on Gaussian distribution. Tone-3 had the longest mean duration across all syllables and speakers (m = 464.3 ms; σ = 66.5 ms) and Tone-4 had the shortest (m = 334.4 ms; σ = 49.5 ms). An ANOVA test showed a significant difference in vowel duration across tones [$F(3,236) = 77.30$, $p < 0.001$]. Post-hoc Tukey HSD tests indicated that Tone-3

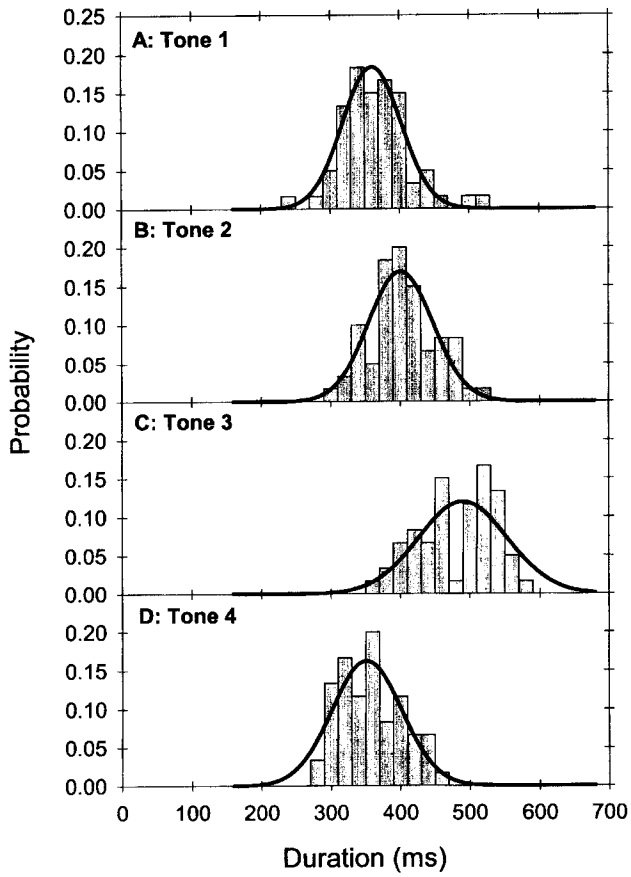

Figure 1: *The duration distribution probabilities of the four tones for all tokens: (A) Tone-1; (B) Tone-2; (C) Tone-3; (D) Tone-4.*
Note: The solid line represents the fitting curve based on the Gaussian distribution.

was significantly longer than the other tones ($p<0.05$); Tone-2 (m = 374.7 ms; σ = 50.7 ms) was significantly longer than Tone-1 (m = 339.5 ms; σ = 48.9 ms) and Tone-4, but that there was no significant difference between Tone-1 and Tone-4.

Correlation between the amplitude contour and F0 contour

Figure 2 shows the mean F0 contour and amplitude contour as a function of normalized time frames. Panel A shows female data and Panel B shows male data. The mean F0 value (240 Hz) for female speakers was significantly higher than for male speakers (131 Hz). The range of F0 variation was highly dependent on tone categories: Tone-4 had the largest F0 range for both female (146 Hz) and male (82 Hz) speakers and Tone-1 had the smallest range for both female (15 Hz) and male (11 Hz) speakers. While the F0 variation range was much greater in female speakers than in male speakers, the percentile F0 change was similar for female (5.3%, 40.0%, 31.5%, and 56.6% for Tone-1, -2, -3, -4, respectively) and male speakers

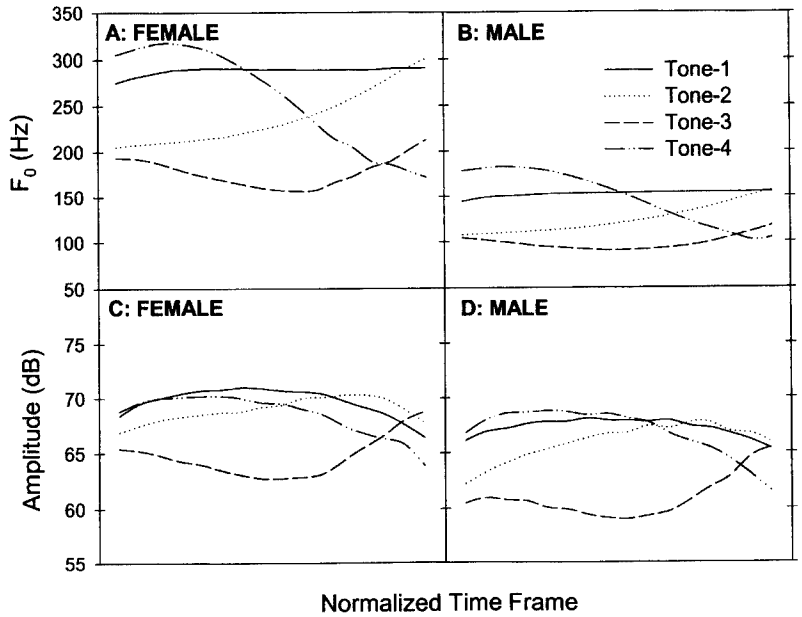

Figure 2: *The averaged F0 contour and amplitude contour for male and female talkers within normalized time frames: (A) F0 contour for female talkers; (B) F0 contour for male talkers; (C) amplitude contour for female talkers; (D) amplitude contour for male talkers.*

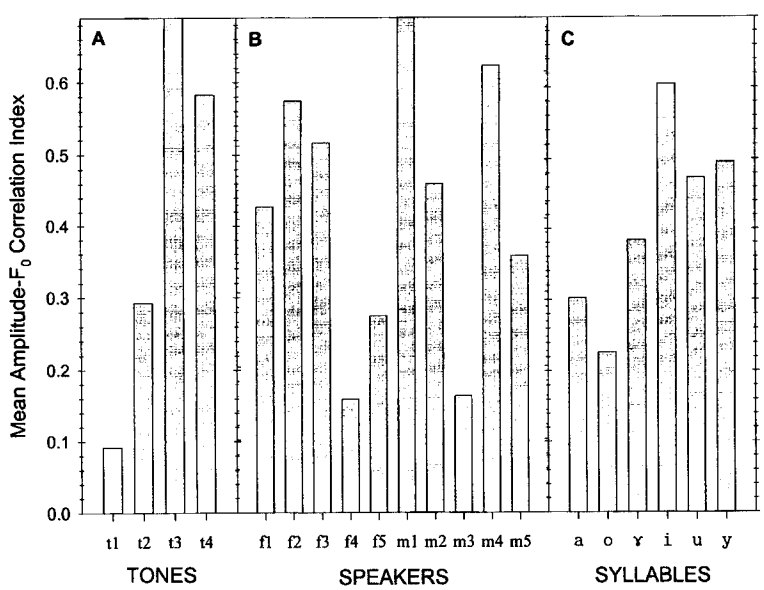

Figure 3: *The mean correlation index between amplitude and F_0 contour as a function of tones, speakers, and syllables: (A) four tones: t1, t2, t3, t4 represent Tone-1, Tone-2, Tone-3, and Tone-4, respectively; (B) 10 speakers: f1–f5 represent female talkers and m1–m5 represent male talkers; (C) six syllables.*

(7.0%, 39.4%, 27.4%, and 55.0% for Tone-1, -2, -3, and -4, respectively). Panel C shows female amplitude contour data and Panel D shows male data. Note the general similarity between the F0 contour and the amplitude contour.

Figure 3 shows the quantitative analysis of this similarity in terms of the mean correlation coefficients between F0 and amplitude contours as a function of tone (A), speaker (B), and syllable (C). Consistent with previous data (Garding et al., 1986; Sagart, 1986; Whalen and Xu, 1992), there was considerable similarity between the amplitude contour and F0 contour. However, there was also large variability across tones, speakers and syllables. The mean correlation coefficients ranged from 0.09 to 0.69 for tones, with Tone-3 yielding the highest correlation coefficient and Tone-1 yielding the lowest. The mean correlation coefficient for different speakers ranged from 0.16 to 0.67, with an average of 0.42. The mean correlation coefficient ranged from 0.23 to 0.61 for different vowels, with /i/ at the highest and /o/ at the lowest.

Perceptual results

Because the results were not significantly different for the four subjects, only the averaged overall tone recognition scores as a function of the seven stimulus sets are presented (Figure 4). The dashed line indicates the chance level (25%) for tone identification. Recognition scores for all stimulus sets were above the chance level (p <0.001). The duration only (D-stimulus set) produced the lowest tone recognition score (35.6%), while the set with the full temporal envelope cue (APD-stimulus set) produced the highest tone recognition score (69.3%). One-way ANOVA tests

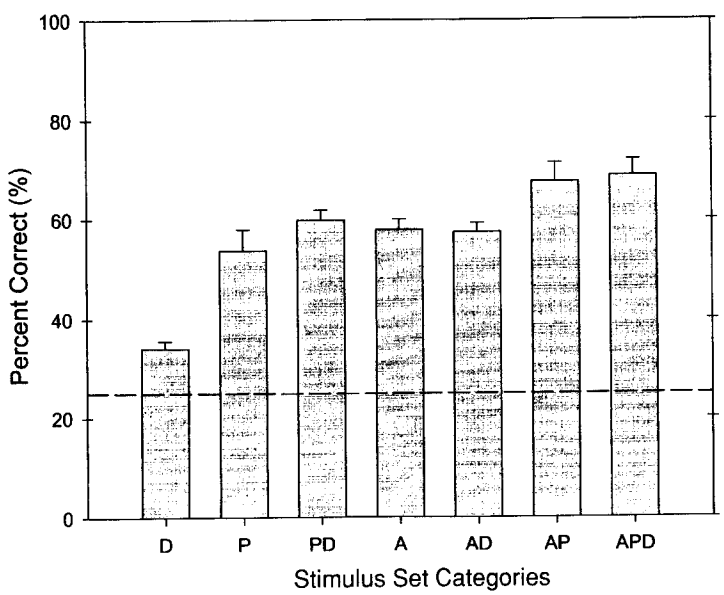

Figure 4: *The percentage correct of tone recognition as a function of all stimuli sets.*
Note: The error bars represent the standard deviation. The dashed line represents chance level (25%).

showed that the stimulus sets produced significantly different scores [$F(6,21)=50.36$, $p < 0.0001$]. The perceptual performance in stimulus sets APD and AP was significantly better than that in the AD-, PD-, A- and P-stimulus sets, which, in turn, were better than the duration only stimulus set (D). There was no significant difference between the APD and AP sets, nor among AD, PD, A and P sets.

Figure 5 shows mean tone recognition scores as a function of speakers, vowels and tones for the APD-stimulus set. The mean recognition score was 69.3% for this stimulus set. The recognition scores ranged from 56% to 81% correct across speakers, from 61% to 75% across syllables and from 52% to 85% across tones. These results indicate that tone identification based on temporal envelope cues is dependent on tones, speakers, and syllables.

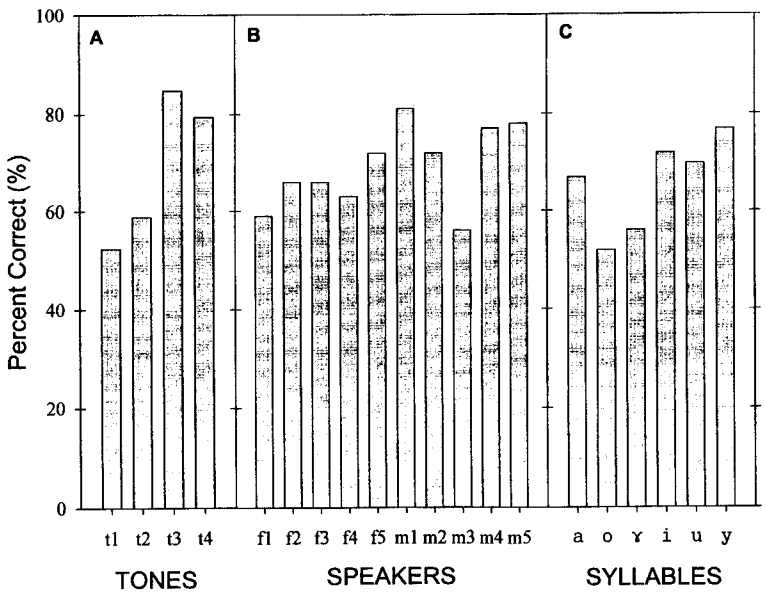

Figure 5: *The percentage correct of tone recognition as a function of tones, speakers, and syllables: (A) four tones: t1, t2, t3, t4 represent Tone-1, Tone-2, Tone-3, and Tone-4, respectively; (B) 10 speakers; f1–f5 represent female talkers and m1–m5 represent male talkers; (C) six syllables.*

Correlational analyses between acoustic analysis and perceptual results

Correlational analyses were undertaken to examine whether perceptual results could be predicted from the measures of acoustic parameters – the duration distribution. To obtain an overall measure of the duration as well as recognition scores, the mean duration and recognition scores were averaged across either ten speakers, or six syllables or four tones. The mean correlation coefficients were then used in correlational analyses with the measures of identification performance. The data are shown as scatter plots in Figure 6. The correlation coefficients for the D-, AD-, PD- and APD-stimulus sets were, respectively, 0.40, 0.41, 0.47 and 0.40. However, none of them was statistically significant.

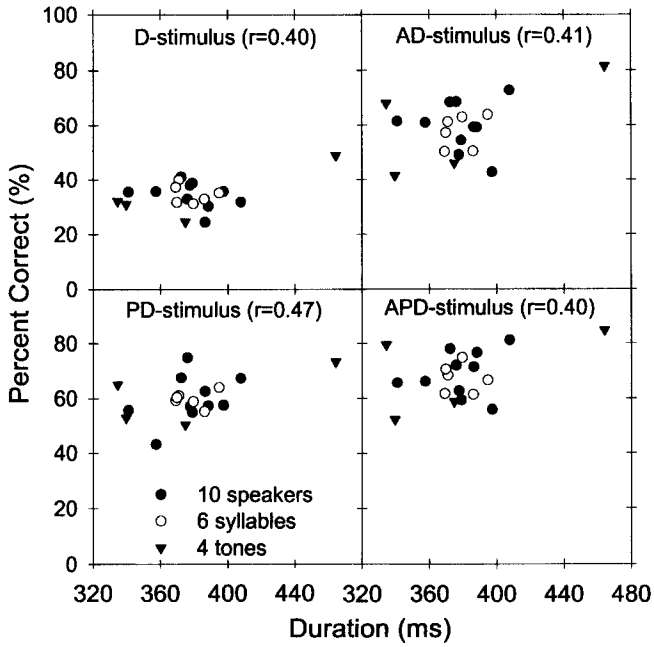

Figure 6: *The correlation between tone recognition score and the duration distribution for four stimulus set conditions (D-, AD-, PD- and APD-stimulus sets).*
Note: Filled circles show the data for 10 speakers. Open circles show the data for six syllables. Filled triangles show the data for four tones.

Correlational analyses were also undertaken to examine whether perceptual results could be predicted from the measures of acoustic parameters – the similarity between the amplitude contour and F0 contour. To obtain an overall measure of the similarity between the amplitude contour and F0 contour, the mean correlation coefficients were averaged across either ten speakers, or six syllables or four tones. The mean correlation coefficients were then used in correlational analyses with the measures of identification performance. The data are shown as scatter plots in Figure 7. The correlation coefficients for the A-, AD-, AP- and APD-stimulus sets were, respectively, 0.85, 0.80, 0.61 and 0.77. All were statistically significant (p <0.001). The correlation between the mean amplitude-F0 correlation coefficients and the perceptual results was also analysed for speakers, syllables or tones, separately. For the speaker category, two stimulus sets had a statistically significant correlation (AD, r = 0.70, p = 0.02; A, r = 0.89, p <0.001). Similarly, the correlation was significant between two stimulus sets for the tone category (APD, r = 0.99, p = 0.01; AD, r = 0.97, p = 0.03). However, none of seven stimulus sets had a significant correlation for the syllable category.

Discussion

The quantitative contribution of the duration cue to tone recognition can be simply evaluated from the D-stimulus set. Given only the duration cue, 34.2% of tones were correctly identified, significantly higher than chance level (25%). The contribution of

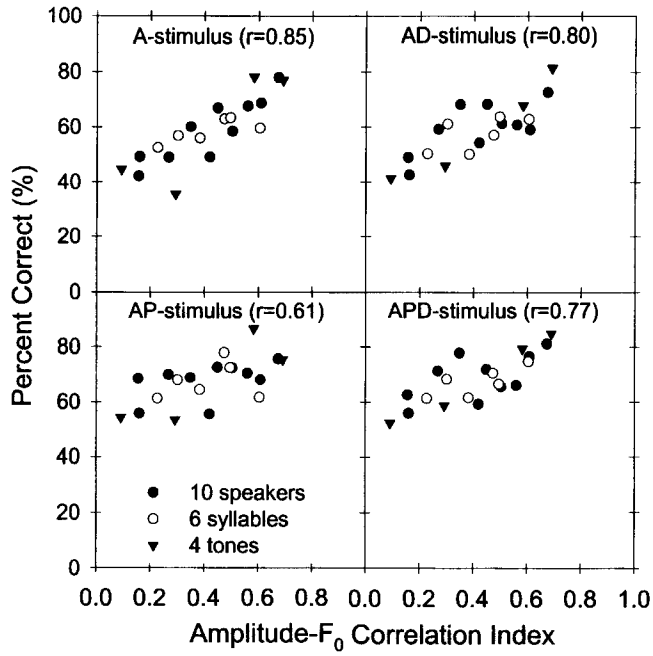

Figure 7: *The correlation between tone recognition score and the mean amplitude-F_0 correlation index for four stimulus set conditions (A-, AD-, AP- and APD-stimulus sets).*
Note: Filled circles show the data for 10 speakers. Open circles show the data for six syllables. Filled triangles show the data for four tones.

duration cues to tone recognition can also be estimated by comparing the performance of two stimulus sets, such as A- and AD-stimulus, P- and PD-stimulus, or AP- and APD-stimulus. No significant improvement in performance was observed when the duration cue was added to a stimulus set containing either the amplitude contour or periodicity cue. However, the duration cue did show a significant effect on Tone-3 discrimination. Almost 50% of Tone-3 tokens were correctly identified with only the duration cue available, and the recognition score of Tone-3 improved significantly when the duration cue was added to the periodicity cue (P-stimulus vs. PD-stimulus). These results indicate that the major contribution of the duration cue to tone recognition, although relatively minor, is to discriminate Tone-3 from the other tones, consistent with previous findings (Yang, 1989; Blicher et al., 1990).

Theoretically, the acoustic data should be responsible for the contribution of the vowel duration to tone recognition. For example, the significant overlap in the vowel duration across tones, especially Tone-1 and Tone-4, can well explain the overall low level of tone recognition using only the duration cue. The mean duration difference across tones and *a priori* knowledge (e.g. Tone-3 has the longest duration) also provide a reasonable explanation for the relatively higher recognition score for Tone-3 and relatively lower score for the other tones. However, the statistical analyses reveal no significant correlation between perceptual results and the mean duration, even for the D-stimulus condition (the duration cue only), indicating that the duration cue is not the primary factor for Chinese tone recognition owing to the high variability in the vowel duration.

The contribution of amplitude cues to tone recognition can be observed by the 58.5% correct tone identification found in the A-stimulus set condition. Performance also significantly improved when the amplitude contour cue was added to stimulus sets containing duration and/or periodicity cues. For example, when amplitude contour cues were made available, a significant improvement was observed from the D- (34.4%) to AD-stimulus set (59.1%), P- (53.9%) to AP-stimulus set (67.7%), and PD- (59.8%) to APD-stimulus set (69.3%). Similarly to the duration cue, the amplitude contour cue did not contribute equally to recognition of all tones [$F(3,12)=16.68$, $p<0.001$]. For example, in the A-stimulus set, a moderate 44.7% and 35.5% of tokens were correctly identified for Tone-1 and Tone-2; however, a significantly higher 76.9% and 78.1% of tokens were correctly identified for Tone-3 and Tone-4.

The contribution of the amplitude contour cues to tone recognition is clearly illustrated by the significant correlation between the amplitude contour and F0 contour, as reported in previous studies (Garding et al., 1986; Sagart, 1986; Whalen and Xu, 1992). Further statistical analyses also reveal a significant correlation between the perceptual performance and the amplitude-F0 correlation index for those conditions containing the amplitude cue, indicating that the amplitude cue is one of the primary factors in Chinese tone recognition when the fundamental frequency (F0) is not available.

The periodicity cue contributes significantly to tone recognition, as seen by the 53.9% correct tone recognition score in the P-stimulus set. However, unlike the duration and amplitude cues, the periodicity cue contributes to all four tones. For Tone-1, -2, -3, and -4, listeners scored 56.3%, 46.1%, 49.8%, and 63.9% correct, respectively.

It is not surprising that the periodicity cue makes a significant contribution to tone recognition, because the F0 shift in tones represented by a modulation rate change over time does elicit a weak pitch (Burns and Viemeister, 1976). Also, the mean F0 value is 131 Hz for male speakers and 240 Hz for female speakers. However, the percentile F0 change is similar between male and female speakers. Detection of the modulation rate shift would be easier for male speakers than for female speakers because there would be a greater increase in the relative rate difference when the standard modulation rate was increased from 80 Hz to 320 Hz (Patterson et al., 1978; Hanna, 1992). If perception of these tones depends on the detection of the modulation rate change, male speakers should show some advantage over female speakers. The data did show that the mean recognition score of male speakers is about 10 percentage points higher than that of female speakers for the P-stimulus set, and about eight points higher for the APD-stimulus set.

While the periodicity cue makes a significant contribution to tone recognition, the recognition score is still much lower than the perfect scores that would be obtained in the presence of the F0 information. One possible explanation is the dramatic difference between modulation rate discrimination and frequency discrimination. Modulation rate discrimination is much more difficult than frequency discrimination. Patterson et al. (1978) reported that about a 20% rate difference is needed to detect a change in the modulation rate while only about 1% frequency difference is needed for frequency discrimination (Wier et al., 1977). So, given that tones had a percentile change in F0 ranging from 29% to 56%, only 1–2 just noticeable difference (JND) steps were available if the change was based on the rate discrimination; however, about 30–60 JND steps were available if the change was based on the F0 discrimination.

Summary and conclusion

Several signal-correlated-noise stimuli were generated to measure the overall and individual contributions of temporal envelope cues to tone recognition. The results lead to the following conclusions:

(1) High levels of tone recognition can be achieved by temporal envelope cues only. The contribution of the duration cue to tone identification is relatively minor while the amplitude contour and periodicity cues play a major role in tone recognition.
(2) The exclusive contribution of the duration cue is to discriminate Tone-3 from the other tones. The amplitude cue contributes mostly to Tone-3 and Tone-4 discrimination, while the periodicity cue contributes equally to all tones' recognition.
(3) There is a high variability in tone recognition performance across speakers, syllables and tones, but this perceptual variability can be accounted for by the acoustic variation existing in speakers, syllables and tones. Given only temporal envelope cues, a speech token with an amplitude contour similar to its F0 contour is more likely to be identified correctly.

Acknowledgements

The authors would like to thank Professor Wang Ren-Hua for allowing the use of the multi-talker phoneme sets. They also thank the four subjects who participated in this experiment and John J. Galvin III for editing the manuscript. The research was supported in part by NIDCD (DC-03861 and DC-02267).

References

Abramson AS. Static and dynamic acoustic cues in distinctive tones. Lang Speech 1978; 21: 319–25.
Blicher DL, Diehl RL,Cohen LB. Effects of syllable duration on the recognition of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. J Phonet 1990; 18: 37–49.
Burns EM, Viemeister NF. Nonspectral pitch. J Acoust Soc Am 1976; 60: 863–9.
Fu Q-J, Zeng F-G, Shannon RV, Soli S. Importance of tonal envelope cues in Chinese speech recognition, J Acoust Soc Am 1998; 104: 505–10 .
Garding E, Kratochvil P, Svantesson JO, Zhang J. Tone 4 and Tone 3 discrimination in modern Standard Chinese. Lang Speech 1986; 29: 281–93.
Hanna TE. Discrimination and identification of modulation rate using a noise carrier. J Acoust Soc Am 1992; 91: 2122–8 .
Howie JM. On the domain of tone in Mandarin. Phonetica 1974; 30: 129–48.
Howie JM. Acoustical Studies of Mandarin Vowels and Tones. Cambridge: Cambridge University Press, 1976.
Liang ZA. The Auditory Basis of tone recognition in Standard Chinese. Acta Physioloica Sinica 1963; 26: 85–90.
Lin MC. The acoustic characteristics and perceptual cues of tones in Standard Chinese. Chinese Linguistic (Chinese Yuwen) 1988; 204: 182–93.
Patterson RD, Johnson-Davies D, Milroy R. Amplitude-modulated noise: the detection of modulation versus the detection of modulation rate. J Acoust Soc Am 1978; 63: 1901–11.
Rabiner LR. On the use of autocorrelation analysis for pitch detection. IEEE Trans Acoustics Speech Signal Processing 1977; 26: 24–33 .
Rosen S. Temporal Information in speech: acoustic, auditory and linguistic aspects. Phil Trans R Soc Lond B 1992; 36: 367–73.
Sagart L. Tone production in modern standard Chinese: an electromyographic investigation. Cahiers de Linguistique, Asie Orientale, Paris, 1986; 205–21.
Schouten JF, Ritsma RJ, Cardozo BL. Pitch of the residue. J Acoust Soc Am 1962; 34: 1418–24.
Schroeder MR. Reference signal for signal quality studies. J Acoust Soc Am 1968; 44: 1735–6.

Wang RH. Chinese phonetics. In Chen YB, Wang RH (Eds). Speech Signal Processing. University of Science and Technology of China Press, 1989, Ch. 3, pp 37–64.

Wang RH. The Standard Chinese Database. University of Science and Technology of China, internal materials, 1993.

Whalen DH, Xu Y. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica 1992; 49: 25–47.

Wier CC, Jesteadt W, Green DM. Frequency discrimination as a function of frequency and sensation level. J Acoust Soc Am 1977; 61: 178–84.

Yang Y-F. The vowels and the perception of Chinese tones. Acta Psychol Sinica 1989; 34: 29–34.

Address correspondencee to: Qian-Jie Fu, PhD, Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057, USA. Tel: 213-273-8036; fax: 213-413-0950; email: qfu@hei.org

Received October 1999; accepted January 2000