Importance of tonal envelope cues in Chinese speech recognition

Qian-Jie Fu[a)]
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057 and Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089

Fan-Gang Zeng
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057 and Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089

Robert V. Shannon
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057 and Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089

Sigfrid D. Soli
House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 8 May 1997; revised 21 January 1998; accepted 31 March 1998)

Recent studies have shown that temporal waveform envelope cues can provide significant information for English speech recognition. This study investigated the use of temporal envelope cues in a tonal language: Mandarin Chinese. In this study, the speech was divided into several frequency analysis bands; the amplitude envelope was extracted from each band by half-wave rectification and low-pass filtering and was used to modulate a noise of the same bandwidth as the analysis band. These manipulations preserved temporal and amplitude cues in each frequency band, but removed the spectral detail within each band. Chinese vowels, consonants, tones and sentences were identified by 12 native Chinese-speaking listeners with 1, 2, 3, and 4 noise bands. The results showed that the recognition score of vowels, consonants, and sentences increased monotonically with the number of bands, a pattern similar to that observed in English speech recognition. In contrast, tones were consistently recognized at about 80% correct level, independent of the number of bands. This high level of tone recognition produced a significant difference in the open-set sentence recognition between Chinese (11.0%) and English (2.9%) for the one-band condition where no spectral information was available. The data also revealed that, with primarily temporal cues, the falling–rising tone (tone 3) and the falling tone (tone 4) were more easily recognized than the flat tone (tone 1) and the rising tone (tone 2). This differential pattern in tone recognition resulted in a similar pattern in word recognition: words having either tone 3 or 4 were more likely to be recognized while words having tone 1 and 2 were not. The quantitative role of tones in Chinese speech recognition was further explored using a power-function model and found to play a significant role in relating phoneme recognition to sentence recognition. [S0001-4966(98)02607-1]

PACS numbers: 43.71.Es, 43.71.Hw [WS]

INTRODUCTION

Traditionally, spectral information in speech sounds has been regarded as the most important cue for speech recognition, while the temporal waveform envelope of the speech sounds has been considered largely a redundant cue. The view that the temporal envelope plays an insignificant role in speech recognition may originate in Licklider's classic experiment, in which speech sounds were infinitely clipped in amplitude, resulting in a flat temporal envelope, but still maintained a high degree of intelligibility (Licklider and Pollack, 1948). However, the importance of temporal information has not always been neglected; for example, Fletcher (1995) attributed high intelligibility of whispered speech sounds, at least partially, to their ''manner of starting and stopping.'' More recent studies have found that, in the absence of spectral cues, temporal envelope cues alone can produce significant consonant recognition (Rosen, 1992; Van Tasell *et al.*, 1987). Shannon *et al.* (1995) systematically studied the trade-off between the spectral and temporal envelope cues. In their study, temporal envelopes of speech sounds were extracted from one to four broad frequency bands, and used to modulate noises of the same bandwidths. This manipulation preserved temporal envelope cues in each band but restricted the listener to severely degraded spectral information. Shannon *et al.* found that identification of consonants, vowels, and sentences improved monotonically as the number of bands was increased, and near perfect speech recognition was achieved with only four bands of modulated noise.

[a)]Electronic mail: qfu@hei.org

 0001-4966/98/104(1)/505/6/$15.00

The present study extends the Shannon *et al.* study to a tonal language, i.e., Mandarin Chinese speech. Tones are important in Chinese speech recognition because the tonality of a monosyllable is lexically meaningful (e.g., Liang, 1963; Wang, 1989; Lin, 1988). There are four tonal patterns in Mandarin Chinese, which are characterized by their fundamental frequency ($F0$) contours. Tone 1 has a flat $F0$ pattern, tone 2 has a rising $F0$ pattern, tone 3 has a falling–rising $F0$ pattern, and tone 4 has a falling $F0$ pattern. For example, the same syllable /ma/ can mean ''mother,'' ''linen,'' ''horse,'' or ''scold'' for the tone pattern 1, 2, 3, or 4, respectively.

Although the $F0$ pattern is the dominant cue for tone recognition, other acoustic cues can contribute to tone recognition. For example, Liang (1963) found that 94.6% correct tone recognition can be achieved with the speech high-pass filtered at 300 Hz. He argued that this high level of tone recognition in the high-pass filtered speech is due to the residue pitch, extracted from the harmonic information and termed as the ''phenomenon of the missing fundamental'' (Schouten *et al.*, 1962). Liang also found that 64.0% tone recognition can be achieved in whispered speech in which neither fundamental frequency nor the harmonic fine structure was present. The whispered speech results indicated that the temporal envelope could also encode information for tone recognition. However, other studies found that tonal contrasts were not well preserved in whispered speech (Abramson, 1972; Howie, 1976). Whalen and Xu (1992) used signal-correlated-noise stimuli (Schroeder, 1968) to further study the contribution of both amplitude contour and duration to tone recognition. To produce signal-correlated-noise stimuli, a speech signal is digitized and the sign of approximately half of the samples, chosen at random, is flipped, resulting in a new stimulus that has a flat spectrum but exactly the same temporal envelope as the original speech signal. The original $F0$ pattern, spectral fine structure, and the spectral envelope are completely absent in the signal-correlated-noise stimulus. Under these conditions, Whalen and Xu found that about 70% recognition level of tones can be achieved. These results clearly demonstrated that the $F0$ pattern is not the only cue in tone recognition; other acoustic cues, including amplitude contour and duration, can also play a significant role.

The major goal of this study was to investigate the use of temporal envelope cues in a tonal language, specifically in Mandarin Chinese at word and sentence levels. The same approach as in the Shannon *et al.* study was used to manipulate the relative distribution of temporal and spectral information. In the present study, the amount of spectral information in speech was increased systematically by changing the number of bands from one, two, and three to four. The amount of temporal envelope information in speech was manipulated by changing the cutoff frequency of envelope extraction filters. Rosen (1992) defined three main temporal features: envelope (2–50 Hz), periodicity (50–500 Hz), and fine-structure (500–10 000 Hz). Two low-pass filters with cutoff frequencies of 50 and 500 Hz were used for envelope extraction to evaluate the relative contribution of both temporal envelope and periodicity information to speech recognition. Recognition of Chinese vowels, consonants, and sentences was measured in 12 native Chinese-speaking listeners as a function of the number of noise bands. In addition, recognition of four tones in standard Chinese was measured and correlated to recognition of Chinese sentence recognition.

I. METHODS

A. Subjects

Twelve native Chinese-speaking listeners, including seven men and five women, ranging in age from 25 to 35 years old, participated in this study. All listeners were recruited from the University of Southern California and were paid for their services. All listeners had pure-tone thresholds better than 15 dB HL at octave frequencies from 250 to 4000 Hz in both ears.

B. Stimuli

The tape-recorded test data were derived from the ''Chinese minimal auditory capability test'' developed by Beijing Union Hospital in P. R. China (Zhang *et al.*, 1988). The test stimuli were spoken by an adult male speaker. The test materials included 21 initial consonants, 35 final vowels, 4 tones, and 200 daily-life sentences. The letter and its associate IPA for all 21 initial consonants and 35 final vowels used in the present study are listed in Appendix A. All materials were divided into four test groups with each containing vowel, consonant, tone, and sentence recognition. Similar to the study of Shannon *et al.* (1995), the tape-played sound was digitized at a 10-kHz sampling rate and passed through a pre-emphasis filter to whiten the spectrum (low pass below 1200 Hz, -6 dB/octave). Then the signal was split into several analysis frequency bands (24 dB/octave, elliptical bandpass filter) and the amplitude envelope from each band was extracted by half-wave rectification and low-pass filtering (elliptical IIR filters with cutoff frequencies of 50 and 500 Hz, -6 dB/octave). The speech envelope was used to amplitude modulate a wide-band white noise, which was then spectrally limited by the same bandpass filter as in the original analysis band. These manipulations preserved band-specific temporal envelope cues, but removed totally the spectral details within each band. The resulting modulated noises from each band were summed, amplified (CROWN D75), and then presented to the listeners through TDH-49 headphones. The overall levels were calibrated for each combination of parameters to produce an average A-weighted output level of 75 dB for continuous speech. All these manipulations were implemented on a real-time signal processing system. In the present study, the one, two, three, or four analysis bands, each combined with the 50- or 500-Hz envelope filters, produced a total of eight conditions. The total bandwidth was from 100 to 4000 Hz. The corner frequencies for the one-band processor were 100 and 4000 Hz. The corner frequencies for the two-band processor were 100, 1500, and 4000 Hz. The corner frequencies for the three-band processor were 100, 800, 1500, and 4000 Hz. The corner frequencies of the four-band processor were 100, 800, 1500, 2500, and 4000 Hz.

C. Procedures

For the consonant, vowel, and tone tests, a four-alternative, forced-choice procedure was used in which both the pinyin[1] and its corresponding Chinese character were shown on the choice list. One test session consisted of 4 blocks, each of which had 10 trials for consonant, vowel, tone recognition, and 20 trials for sentence recognition. A sample block is shown in Appendix B. The stimulus in consonant, vowel and tone tests was a single syllable, consisting of an initial consonant and a following vowel with a tone. In each trial of the consonant test, four-alternative syllables, which had the same following vowel and tone, were included in the choice list. The listener was asked to identify the consonant by marking the syllable containing the consonant that was presented. Because not all combinations of the initial consonant and the following vowel were lexically meaningful, the vowel and tone varied from trial to trial to accommodate the meaning of Chinese words. The same was true for consonants and tones in the vowel recognition test, and for consonants and vowels in the tone recognition test. In the daily-life ''open-set'' sentence recognition test, the listener was asked to write down as many words as were recognized in each sentence. Each test block included 20 sentences, and each sentence had 2–8 key words, resulting in a total of 100 key words. Sentences were presented without repetition. The recognition score was calculated based on the percentage of the total number of key words correctly recognized. All subjects received extensive training, including familiarization with the testing environment and the speech quality of all eight experimental conditions via 15-min casual conversation with experimenter through the real-time processing system, and informal tests of sample materials. The sample materials were not included in the formal tests. The sequence of these eight experimental conditions were randomized and counterbalanced across subjects. No feedback was provided regarding the correct answer in any test.

II. RESULTS AND DISCUSSION

Figure 1 shows the averaged recognition results from 12 subjects as a function of the number of noise bands. Results from four different tasks, including consonant, vowel, tone, and sentence recognition, are presented in panel A, B, C, and D, respectively. The two temporal envelope filter cutoff frequencies, 50 and 500 Hz, are represented by the short dashed line and the solid line, respectively. As a comparison, panels A, B, and D also show our previous results of English vowel, consonant, and sentence recognition obtained with the envelope filter frequency at 500 Hz (the long dashed line, Shannon *et al.*, 1995). Figure 1A shows that, as the number of bands was increased from one to four, the recognition score of Chinese consonants increased monotonically from 50.1% to 84.7% for the 500-Hz envelope filtering condition and increased from 45.0% to 79.8% for the 50-Hz filtering condition. A two-way, repeated-measures analysis of variance (ANOVA) was performed, with the number of bands and the low-pass envelope filters as within-subjects factors. This analysis showed a significant effect of the number of bands

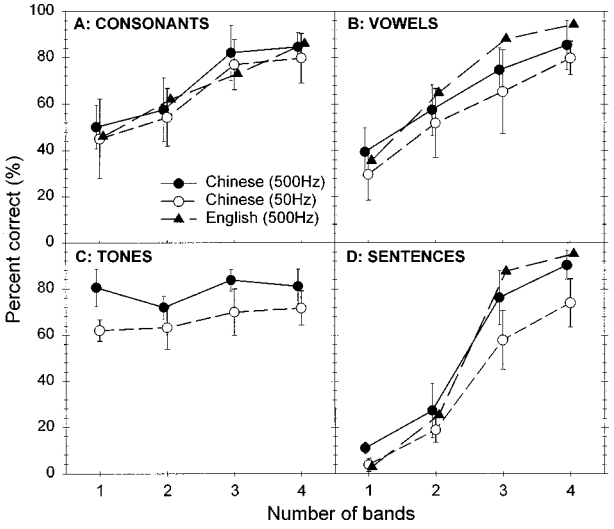

FIG. 1. Speech recognition as a function of the number of noise bands. A: consonants; B: vowels; C: tones; D: sentences. The conditions included English with the envelope filter frequency at 500 Hz (▲), Chinese with the envelope filter frequency at 500 Hz (●) and at 50 Hz (○).

$[F(3,57)=19.45,\ p<0.001]$ but no significant difference between the 500- and 50-Hz condition $[F(1,57)=1.34,\ p>0.05]$.

Figure 1B shows Chinese vowel recognition as a function of the number of bands. The recognition score of Chinese vowels increased monotonically from 39.4% in the one-band condition to 85.6% in the four-band condition for the 500-Hz condition and increased from 29.6% to 79.9% for the 50-Hz condition. A similar two-way ANOVA showed a significant effect of the number of bands $[F(3,57)=35.09,\ p<0.001]$ but no significant effect of the envelope filters $[F(1,57)=3.84,\ p>0.05]$.

Figure 1C shows Chinese tone recognition as a function of the number of bands. In contrast to the Chinese consonant and vowel results, the two-way ANOVA showed no significant effect of the number of bands $[F(3,57)=2.15,\ p>0.05]$ but a significant effect of the envelope filters $[F(1,57)=26.54,\ p<0.001]$. The percent correct score of tone recognition averaged across four band conditions was 80.8% in the 500-Hz filtering condition and 66.9% in the 50-Hz filtering condition.

Figure 1D shows Chinese sentence recognition as a function of the number of bands. Similar to consonant and vowel recognition patterns, the recognition score of Chinese sentences increased monotonically from 11.0% in the one-band condition to 90.4% in the four-band condition for the 500-Hz condition and increased from 3.8% to 74.0% accordingly for the 50-Hz condition. The two-way ANOVA showed both significant effects of the number of bands $[F(3,55)=17.08,\ p<0.001]$ and the envelope filters $[F(1,55)=138.49,\ p<0.001]$.

The present results showed several interesting similarities as well as differences between Chinese and English speech recognition with primarily temporal cues. First, both Chinese and English consonant, vowel, and sentence recognition improved monotonically as a function of the number of bands. No significant effect was observed between the

50-Hz and 500-Hz cutoff frequencies of the envelope filter in Chinese consonant and vowel recognition, similar to English consonant and vowel recognition (Shannon *et al.*, 1995). Second, different from consonant, vowel, and sentence results, tones were consistently recognized at an 80.8% correct level in the 500-Hz filtering condition and 66.9% correct level in the 50-Hz filtering condition, independent of the number of bands. Third, a significant effect of the envelope filters was observed in Chinese sentence recognition but not in English sentence recognition. In particular, for the one-band, 500-Hz filtering condition, where no spectral information was available, Chinese produced a significantly higher score in sentence recognition (11.0%) than English (2.9%) [$t(17)=4.71$, $p<0.001$].

The similar effect of the envelope filter cutoff frequency on Chinese tone and sentence recognition suggested a possibly important role for tonal envelope cues in Chinese speech recognition. Two approaches were used to address the possible relationship between Chinese tone and sentence recognition. First, a qualitative relationship between tone recognition and sentence recognition was investigated based on the results from the 500-Hz condition. Results showed that different tones had different recognition scores. As shown in the solid line of Fig. 2(A), the recognition scores of the falling-rising tone (tone 3) and the falling tone (tone 4) were almost twice as high as those of the flat tone (tone 1) and the rising tone (tone 2) in the one-band condition. This result was consistent with an earlier finding (Whalen and Xu, 1992) and might be explained by the differences in the amplitude contour (Fu *et al.*, 1995). Our acoustic analysis showed that the amplitude envelope was highly correlated with $F0$ contour for the falling–rising tone and falling tone, and this correlation was likely responsible for the high recognition score of these two tones. However, tone 1 and tone 2 did not seem to have amplitude envelopes that were highly correlated to their respective $F0$ contours. If tone recognition had played an important role in sentence recognition, then words with tone 3 and tone 4 would be more easily recognized than words with tone 1 and tone 2. The dashed line in Fig. 2(A) shows the distribution patterns of tones (right y axis) for correctly recognized words in sentence recognition for the one-band condition. Indeed, most of the words that were correctly recognized in sentence recognition had either tone 3 or tone 4. Figure 2(B) shows the recognition score of the individual tones and the distribution tonal patterns of the correctly recognized words in sentence recognition for the four-band condition. Although the difference was smaller between tones for the four-band condition than the one-band condition, the recognition scores of tone 3 and tone 4 were still significantly higher than those of tone 1 and tone 2, resulting in a similar tonal distribution pattern of the correctly recognized words. These results indicated that tone played a greater role in sentence recognition when no spectral information was available, and became less important when more spectral information was available.

A quantitative relationship among Chinese vowel, consonant, tone, and sentence recognition scores was also assessed using a power-function model (Boothroyd and Nittrouer, 1988; Rabinowitz *et al.*, 1992). The power-function

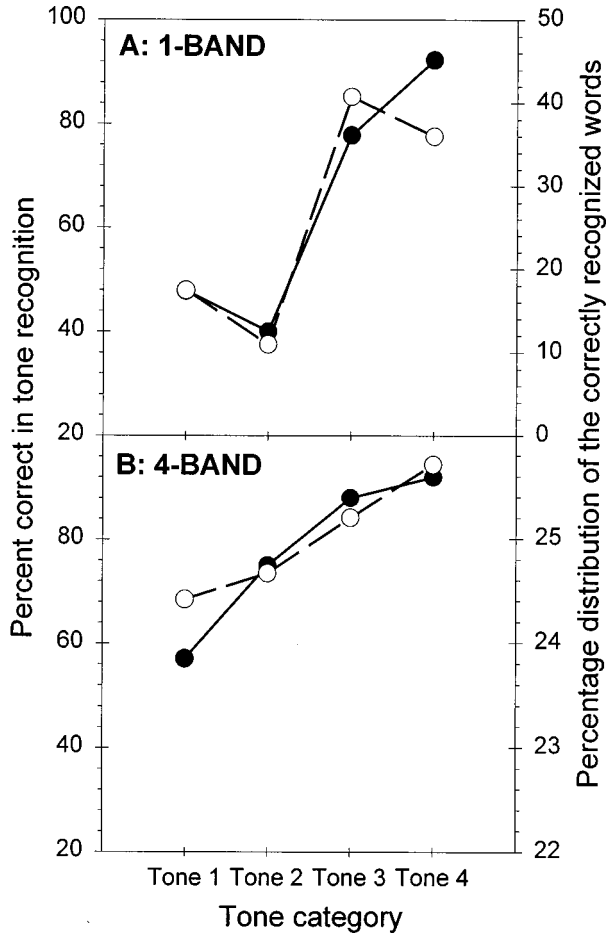

FIG. 2. (A) Percent correct of tone recognition in the one-band condition (●). The tonal distribution pattern of the correctly recognized words in sentence recognition for the one-band condition (○). (B) Percent correct of tone recognition in the four-band condition (●). The tonal distribution pattern of the correctly recognized words in sentence recognition for the four-band condition (○).

model, based on the original work by Fletcher (1995), could account for the benefit of sentence context (factor k) and the relation between word and phoneme recognition (factor j). First, the sentence context factor is represented in the following equation:

$$p_s = 1-(1-p_w)^k, \tag{1}$$

where p_s is the recognition probability for words in sentences, p_w is the recognition probability for isolated words, and the exponent k is the sentence context factor.

Second, the recognition probability for an isolated CVC word is represented as follows:

$$p_w = p_p^j, \tag{2}$$

where p_p is the recognition probability for the individual phonemes and p_w is the recognition probability for the isolated words, j has a value between 2 and 3, reflecting the fact that, due to language constraints, only 2–3 phonemes must be recognized for correct recognition of isolated CVC words.

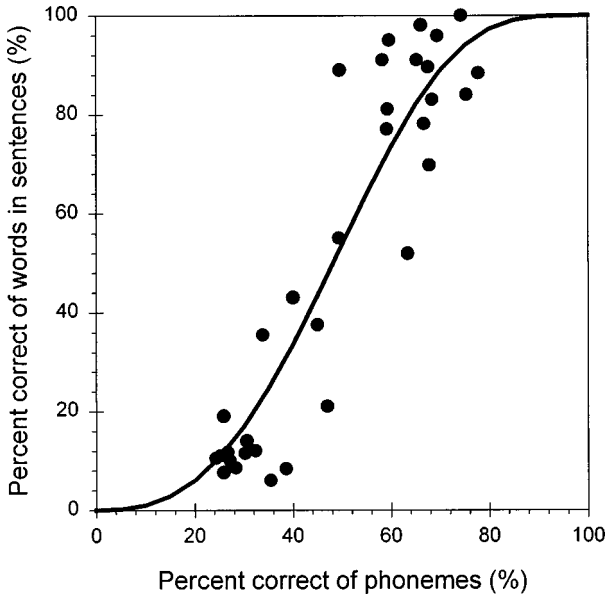

FIG. 3. The relationship between sentence and phoneme recognition based on a power-function model. Symbols showed experimental data and the solid line represented the predicted relationship between the recognition probability of phonemes and sentences.

Finally, we combine Eqs. (1) and (2) to obtain the relation between the recognition probability for words in sentences and phonemes:

$$p_s = 1 - (1 - p_p^j)^k. \tag{3}$$

Because no comparable results were available in Chinese materials, the present study used the same $k(4.5)$ and $j(2.67)$ values as in English materials (Rabinowitz *et al.*, 1992). Different from the previous study, the recognition probability for phonemes was calculated by combing the recognition probability for vowels, consonants, and tones:

$$p_p = p_c^{w_c} p_v^{w_v} p_t^{w_t}, \tag{4}$$

where p_c, p_v, and p_t are the recognition probability of consonants, vowels, and tones, and w_c, w_v, and w_t are weighting parameters. Figure 3 shows the best fitting power-function (the solid line) in relating the recognition probability of phonemes to the recognition probability for words in sentences[2] ($r = 0.92$). The weights for consonants, vowels, and tones were 0.64 (s.e.$=0.13$), 0.71 (s.e.$=0.13$), and 0.79 (s.e.$=0.18$), respectively. A one-way ANOVA showed no significant difference between the weights [$F(2,96) = 0.26$, $p > 0.05$]. These results suggested an approximately equal contribution of consonants, vowels, and tones to Chinese sentence recognition under the present conditions.

III. CONCLUSIONS

High-level performance of Mandarin Chinese speech recognition was achieved with primarily temporal envelope cues, extending previous results obtained with English speech. Due to the tonal nature of Chinese speech, several significant differences were also observed between English and Chinese speech recognition. In contrast to the recognition of consonants, vowels, and sentences, which increased monotonically as a function of the number of bands, tones were recognized at an 80% correct level in the 500-Hz envelope filter condition, independent of the number of bands. This high level of tone recognition produced a significant difference in the open-set sentence recognition between Chinese (11.0%) and English (2.9%) for the one-band condition where no spectral information was available. The analysis of tone recognition and the tone distribution pattern for words in sentence recognition indicated a high correlation between tone recognition and sentence recognition. The present study also used a power-function model to reveal that, with temporal envelope cues, tones contribute to Chinese speech recognition with temporal envelope cues in a similar way to consonants and vowels.

ACKNOWLEDGMENTS

Portions of this paper were presented at the 129th Meeting of the Acoustical Society of America. We are grateful to all subjects for their participation in our experiments and to Alena Wilson for editing the manuscript. We also thank Dr. Winifred Strange, Dr. Adrian Fourcin, Dr. Xu Yi, and an anonymous reviewer for their helpful comments. Research was supported in part by NIH (DC-02267 and DC-01526).

APPENDIX A: A LIST OF CONSONANTS AND VOWELS

A Chinese single syllable consists of an initial consonant and a following vowel with a tone. There are 21 initial consonants and 35 final vowels used in the present study. The letter and its associated IPA symbol are shown as follows:

(1) 21 initial consonants
b[p], p[p^h], m[m], f[f], d[t], t[t^h], n[n], l[l], g[k], k[k^h], h[x], j[tɕ], q[$tɕ^h$], x[ɕ], zh[tʂ], ch[$tʂ^h$], sh[ʂ], r[ʐ], z[ts], c[ts^h], s[s]

(2) 35 final vowels
(a) 6 simple vowels
a[a], o[o], e[ɤ], i[i], u[u], ü[y]
(b) 13 complex vowels
ai[aɪ], ei[eɪ], ao[aʊ], ou[oʊ], ia[ia], ie[iɛ], iao[iɑʊ], iou[iəʊ], ua[ua], uo[uo], uai[uaɪ], uei[ueɪ], üe[yɛ]
(c) 16 compound nasal vowels
an[an], en[ən], ang[ɑŋ], eng[əŋ], ong[oŋ], ian[iɛn], in[in], iang[iɑŋ], ing[iŋ], ion[iʊŋ], uan[uan], uen[uən], uang[uɑŋ], ueng[ʊəŋ], üan[yɛn], ün[yn].

APPENDIX B: SAMPLE TEST MATERIALS

One test block of vowel recognition is shown.

Trial 1:	ba(3)	ben(3)	bi(3)	biao(3)
	把	本	笔	表
Trial 2:	du(2)	di(2)	de(2)	die(2)
	读	敌	得	迭
Trial 3:	gai(3)	gan(3)	guo(3)	guang(3)
	改	敢	果	广
Trial 4:	ji(1)	jiang(1)	jian(1)	jia(1)
	鸡	将	尖	家
Trial 5:	le(4)	liang(4)	li(4)	lue(4)
	乐	亮	利	略
Trial 6:	men(2)	mian(2)	mi(2)	mao(2)
	门	棉	迷	毛
Trial 7:	she(4)	shu(4)	shi(4)	shang(4)
	射	树	是	上
Trial 8:	hun(1)	hua(1)	he(1)	hei(1)
	昏	花	喝	黑
Trial 9:	ye(3)	yu(3)	yi(3)	yao(3)
	也	雨	以	咬
Trial 10:	rong(2)	ren(2)	ru(2)	ran(2)
	容	人	如	然

[1]Pinyin transcription is a phonemic spelling system for Mandarin. The pinyin of a single syllable consists of an initial consonant and a following vowel with a tone. In this paper, the four tones are represented by the tone mark ''1'' for the high tone, ''2'' for the rising tone, ''3'' for the falling–rising tone, and ''4'' for the falling tone.

[2]The relation between the phoneme recognition probability and the sentence recognition probability can also be predicted by a linear function (slope $=1.65$, $r=0.90$). However, the power function model is based on the combination of the two relations [factor j and k, Eq. (3)]. Each relation can not be well predicted by a linear function. Besides, if there were more data in the two extreme conditions, the relation would be more like a power function instead of a linear function.

Abramson, A. S. (**1972**). ''Tonal experiments with whispered Thai,'' in *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, edited by A. Valdman (Mouton, The Hague), pp. 31–44.

Boothroyd, A., and Nittrouer, S. (**1988**). ''Mathematical treatment of context effect in phoneme and word recognition,'' J. Acoust. Soc. Am. **84**, 101–114.

Fletcher, H. B. (**1995**). ''The speaking mechanism,'' in *Speech and Hearing in Communication*, edited by J. B. Allen (Acoustical Society of America, Woodbury, NY), Chap. 2, p. 16.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. (**1995**). ''Chinese Speech Recognition only Using Amplitude Envelope Cues,'' 1995 Conference on Implantable Auditory Prostheses, abstract booklet, 61.

Howie, J. M. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge U. P., Cambridge, England).

Liang, Z.-A. (**1963**). ''The auditory perception of Mandarin Tones,'' Acta Phys. Sin. **26**, 85–91.

Licklider, J. C. R., and Pollack, I. (**1948**). ''Effects of differentiation, integration, and infinite peak clipping on the intelligibility of speech,'' J. Acoust. Soc. Am. **20**, 42–51.

Lin, M.-C. (**1988**). ''The acoustic characteristics and perceptual cues of tones in Standard Chinese,'' Chinese Yuwen **204**, 182–193.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). ''Relations among different measures of speech reception in subjects using a cochlear implant,'' J. Acoust. Soc. Am. **92**, 1869–1881.

Rosen, S. (**1992**). ''Temporal information in speech: acoustic, auditory and linguistics aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech Recognition with Primarily Temporal Cues,'' Science **270**, 303–304.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.

Wang, R.-H. (**1989**). ''Chinese phonetics,'' in *Speech Signal Processing*, edited by Y. B. Chen and R.-H. Wang (University of Science and Technology of China Press), Chap. 3, pp. 37–64.

Whalen, D. H., and Xu, Y. (**1992**). ''Information for Mandarin Tones in the amplitude contour and in brief segments,'' Phonetica **49**, 25–47.

Zhang, H., Zhao K. L., and Wang, Z. Z. (**1988**). ''MACC: Chinese Minimal Auditory Capability Test,'' Beijing Union Hospital, Beijing, P. R. China.