



Research paper

Lexical tone recognition with spectrally mismatched envelopes

Ning Zhou, Li Xu *

School of Hearing, Speech and Language Sciences, Ohio University, Athens, OH 45701, USA

ARTICLE INFO

Article history:

Received 29 April 2008

Received in revised form 16 September 2008

Accepted 17 September 2008

Available online 25 September 2008

Keywords:

Mandarin Chinese tones
Tone recognition
Spectral shift
Shallow insertion
Frequency compression
Cochlear implants

ABSTRACT

It has been shown that frequency-place mismatch has detrimental effects on English speech recognition. The present study investigated the effects of mismatched spectral distribution of envelopes on Mandarin Chinese tone recognition using a noise-excited vocoder. In Experiment 1, speech samples were processed to simulate a cochlear implant with various insertion depths. The carrier bands were shifted basally relative to the analysis bands by 1–7 mm in the cochlea. Nine normal-hearing Mandarin Chinese listeners participated in this experiment. Basal shift of the carriers only slightly affected tone recognition. The resistance of tone recognition to spectral shift can be attributed to the overall amplitude contour cues that are independent from spectral manipulations. Experiment 2 examined the effects of frequency compression, where widened analysis bands by 2, 6, and 10 mm were compressively allocated to narrower carrier bands. Five of the 9 subjects participated in Experiment 2. It appears that the expanded frequency information especially on the low frequency end can compensate for the distortion from frequency compression. Thus, spectral shift might not pose a severe problem for tone recognition, and allocation of wider frequency range to include more low frequency information might be beneficial for tone recognition.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Speech recognition is robust and resistant to many forms of distortion or reduction of information (e.g., Remez et al., 1981; Van Tassel et al., 1987; ter Keurs et al., 1992, 1993; Baer and Moore, 1993, 1994). The channel vocoder has been a useful tool to study and quantify the effects of distortion or degradation of signal on speech recognition (e.g., Hill et al., 1968; Zollner, 1979; Shannon et al., 1995) and to provide estimates for performance of the cochlear implant users (e.g., Fishman et al., 1997; Friesen et al., 2001). With channel vocoder, it has been shown that degraded spectral resolution to as few as 4 frequency channels can still maintain good speech recognition in quiet (Shannon et al., 1995). However, speech recognition is severely affected if speech information in the channels does not match tonotopically to the places in the cochlea (Dorman et al., 1997; Shannon et al., 1998; Fu and Shannon, 1999). Research has also shown that human brains have extraordinary abilities to adapt to frequency-place mismatch with training (e.g., Rosen et al., 1999; Fu et al., 2002; Faulkner, 2006).

In cochlear implant systems, a number of forms of frequency-place mismatch occur as a result of the pathology of hearing loss of the implant patients, shallow insertion, or frequency mapping of the device. One type of frequency-place mismatch happens when the patients have localized losses of auditory neurons, which

results in “holes” in hearing. In this case, elevated electrical thresholds of the corresponding electrodes will be needed for those bands of information to be received. The increased signal level will likely result in spread of electric current to neural fibers that are not intended to be activated, producing frequency warping around the “holes” in the cochlea. Frequency-place warping has been studied using acoustical simulations with noise-excited vocoders and in cochlear implant patients. Shannon and colleagues (Shannon et al., 2002; Baskent and Shannon, 2006) found that holes in the apical regions affected speech recognition more than holes in the middle or basal regions did. Further, redistribution of the missing frequency content around the hole regions did not improve speech recognition.

Another type of frequency-place mismatch involves an overall shift of spectrum as a result of shallow insertion of a cochlear implant. Consider the case where the implant electrode array is not fully inserted into the cochlea so that the electrodes do not match the places corresponding to the frequency map of the speech processor. Typically, the output of a low frequency analysis channel is delivered to the electrode that rests at a higher frequency place, resulting in a basal shift of the spectrum. Dorman et al. (1997) used a 5-channel tone vocoder to simulate various insertion depths of a cochlear implant that resulted in different degrees of basal shift. They found that the basal shift of spectrum progressively deteriorated the recognition of sentences, vowels, and consonants. Compared to vowel and sentence recognition, consonant recognition was less affected by the shift. The feature of place of articulation

* Corresponding author. Tel.: +1 740 593 0310; fax: +1 740 593 0287.
E-mail address: XuL@ohio.edu (L. Xu).

of consonants was, however, transmitted particularly poorly (Dorman et al., 1997). Fu and Shannon (1999) further investigated the effects of frequency-place mismatch on vowel recognition with 4-, 8-, or 16-channel processors. They first fixed the analysis bands of the input signal while changing the simulated insertion depth of the implant. They found that vowel recognition scores decreased significantly when the tonotopic places of the carrier bands were shifted by 3 mm or more. In their second experiment, Fu and Shannon (1999) simulated a fixed electrode array positioned at two relatively shallow places and varied the frequency allocation of the analysis bands. They found that the best performance occurred when the analysis bands transmitted more low frequency information with a small amount of mismatch to the carrier bands. The effects of mismatch between the analysis band frequencies and the electrode locations have also been evaluated in patients with cochlear implants (Fu and Shannon, 1999). Similar results were found. That is, the best performance struck a balance between the loss of low frequency coverage as a result of the basal shift of analysis bands and the degree of the mismatch between the analysis and carrier bands.

Other studies investigated the effects of compression of the entire speech spectrum to the limited stimulation range of the electrode array that typically results from shallow insertion (Pfungst et al., 2001; Baskent and Shannon, 2003, 2004, 2005). In the study of Baskent and Shannon (2005), the most apical electrode was turned off consecutively to simulate partially inserted implants while the analysis frequency range was kept the same or reduced to map the stimulation frequency range. The compressed stimulation avoided truncating low frequency for partially inserted implants but resulted in a distorted frequency-place mapping. Baskent and Shannon (2005) found that while tonotopic mapping started to lose its advantage when more apical electrodes were turned off, compressing a wider analysis frequency to the few basal electrodes contributed to better speech recognition. In addition to the basal electrodes being compressed with an entire speech spectrum, Pfingst et al. (2001) also tested central and apical electrodes and found that the effects of compression were the least when the central electrodes were used.

To summarize, recognition of English phonemes, words, or sentences has been studied in terms of its relationship with spectral shift (e.g., Dorman et al., 1997; Shannon et al., 1998; Fu and Shannon, 1999) or spectral distortion (e.g., Pfingst et al., 2001; Baskent and Shannon, 2003, 2004, 2005). The effect of spectral distortion on lexical tone recognition is a topic that has not been closely studied. There is no study, to our knowledge, that has simulated these effects on lexical tone recognition using normal-hearing subjects. The two studies that have examined the effects of frequency compression on tone recognition in cochlear implant patients have reported mixed findings. Chu et al. (2005) tested Cantonese tone recognition in three patients implanted with short electrodes compressively assigned with a normal speech spectrum. In comparison with performance of tonotopically matched condition, they did not find an effect of compression. Liu et al. (2004), however, showed that Mandarin tone recognition in six patients decreased as a result of limiting the number of active electrodes. The reason for the discrepancy between the two studies is not clear, as neither study described the frequency range allocated to the electrodes. Further, there are many factors other than the variable under study that could influence the performance in those cochlear implant patients (e.g., etiology, neural survival, implant device and speech processing strategy, current spread etc.). Therefore, in the present study, we adopted a noise-excited vocoder to simulate the effects of two forms of spectral distortion on Mandarin Chinese tone recognition. We examined the acute effects of basal spectral shift and frequency compression in an attempt to provide reasonable implications for cochlear implants.

Lexical tone is quite different from English phonemes in terms of its recognition mechanisms (Xu and Pfingst, 2008, 2003). The most important acoustic feature for Mandarin Chinese tones is the fundamental frequency (F0). The Mandarin Chinese tones 1–4 demonstrate F0 patterns of (1) flat; (2) rising; (3) low and dipping; and (4) falling, respectively. Temporal information that co-varies with F0 contours, including vowel duration and envelope amplitude also contributes as a secondary cue (Liang, 1963; Lin, 1988; Whalen and Xu, 1992; Xu et al., 2002). Compared with English phoneme recognition, vocoder studies have shown that Mandarin tone recognition saturates with higher frequency resolution (Xu et al., 2002, 2005). Only with very detailed spectral resolution provided by more than 30 spectral channels does the tone recognition performance improve to close to that of unprocessed stimuli (Kong and Zeng, 2006). Kong and Zeng also found that 500 Hz envelope information carried by only one channel provided better tone recognition in quiet than 8 channels with 50 Hz envelope information. Evidence from these studies showed that, when place pitch can not be resolved, tone recognition depends on temporal envelope information that contains the periodicity more than English phoneme recognition does [see Xu and Pfingst (2008) for a review]. Fu et al. (2004) also suggested that speech processing strategies that use high stimulation rates favors tone recognition, probably because more detailed temporal envelope information could be represented with higher stimulation rates. Given these differences, frequency-place mismatch may exert different effects on Mandarin Chinese tone recognition than English phoneme recognition. The effects of basal shift of the spectrum on Mandarin Chinese tone recognition were investigated in Experiment 1, where an implant with varying insertion depths was simulated. Experiment 2 examined the effects of frequency compression with wider frequency ranges allocated to narrower frequency bands.

2. Methods

2.1. Speech material and signal processing

One female and one male native Mandarin Chinese speaker produced the following ten syllables in each of the four tones: /fu/, /ji/, /ma/, /qi/, /wan/, /xi/, /xian/, /yan/, /yang/, and /yi/. Four of the syllables end with nasal consonants while the other six are open syllables. Care was taken so that the four tones produced for the same syllable were equal in duration (see Xu et al., 2002). The rms values of all tone tokens were equalized to control for the overall loudness differences. The amplitude contour cue was intact. The recordings of the 80 tone tokens (2 speakers \times 10 syllables \times 4 tones) were stored at a sampling rate of 22050 Hz with a 16-bit resolution.

Signal processing for the acoustic simulations was performed in MATLAB (MathWorks, Natick, MA). Experiment 1 measured tone recognition in basal shift conditions. The basal shifts of the spectrum simulated an implant with varying insertion depths (Fig. 1A). The raw tone tokens were pre-emphasized by high-pass filtering at 1200 Hz and were bandpass filtered into 4, 8, 12, or 16 frequency bands. The frequency range of the analysis bands was 269–3283 Hz, corresponding to a tonotopic place between 28 mm and 13 mm from the basal end of the cochlea. The bandwidth and corner frequencies of the analysis bands were determined using the formula from Greenwood (1990) that estimates equal spacing on the basilar membrane of the cochlea, $F = 165.4(10^{0.06x} - 1)$, where x is the distance in mm of the estimated place from the apex end of the cochlea, assuming the length of the basilar membrane to be 35 mm. The temporal envelope of each band was extracted by half-wave rectification and then low-pass filtering at 160 Hz (2nd order Butterworth, 12 dB/octave).

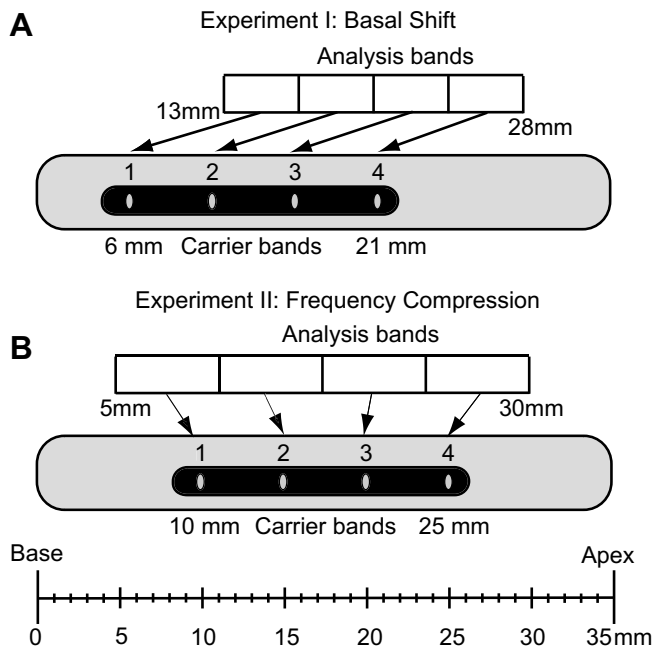


Fig. 1. Schematic representation of the tonotopic location of the electrodes in a 4-band processor (i.e., carrier bands) and their frequency allocations (i.e., analysis bands). The upper panel (A) depicts an example of Experiment 1, where the analysis bands are fixed and the carrier bands are shifted basally by 7 mm. The lower panel (B) depicts an example of Experiment 2, where the analysis bands are widened by 5 mm on each frequency end and are compressively assigned to the narrower carriers.

The temporal envelope was used to modulate a wideband noise. The modulated signal was then frequency-limited in what is thereafter referred to as the carrier band. The frequency allocation of the carrier bands was varied to simulate various insertion depths of the electrode array. The estimation of the frequency allocation of the carrier bands was also based on the Greenwood formula (1990). The analysis bands and the carrier bands did not necessarily match, which resulted in a tonotopic shift. Matched analysis bands and carrier bands simulated a fully inserted electrode array. The electrode location was manipulated to shift from full insertion (i.e., 28 mm) to 21 mm into the cochlea in a step size of 1 mm. The frequency allocation for the carrier bands is provided in Table 1. The shifting procedure was repeated for all four channel conditions (i.e., 4, 8, 12, and 16). The outputs of all bands were summed up for acoustic presentations.

Experiment 2 measured tone recognition under the condition of frequency compression (Fig. 1B). Different from Experiment 1, the

Table 2
Compressed frequency allocation for a 4-band processor

Compression size (mm)	Cochlear location of analysis bands (mm)	Frequency allocation for analysis bands (4 channels)				
		1	2	3	4	
0 (matched)	25-10	492	938	1686	2943	5053
2	26-9	407	864	1686	3165	5826
6	28-7	269	731	1686	3659	7732
10	30-5	164	616	1686	4225	10246

carrier bands in Experiment 2 were fixed in the frequency range of 492–5053 Hz to simulate an implant located between 25 mm and 10 mm from the base. The widened analysis bands (i.e., 407–5826 Hz, 269–7732 Hz, and 164–10246 Hz) were compressively assigned to the relatively narrower carrier bands (i.e., 492–5053 Hz). The analysis bands were widened evenly on both low and high frequency sides. The amount of widening was equivalent to 1, 3, and 5 mm on each side, thus 2, 6, and 10 mm in total (Greenwood, 1990). Frequency compression was repeated for all four channel conditions. The frequency allocations of the analysis bands for the three compression conditions are provided in Table 2.

2.2. Subjects and procedure

Nine normal-hearing, Mandarin Chinese native speakers (five males and four females, ages 28.4 ± 6.7, mean ± SD) participated in the tone recognition tests. All subjects participated in Experiment 1 and five of them continued with Experiment 2. As will be reported in results, the sample size of Experiment 2 appeared to be sufficient to detect statistical significance between the experimental conditions. All subjects were screened for pure tone thresholds lower than 20 dB HL for octave frequencies between 250 Hz and 8000 Hz. The use of human subjects was approved by the Ohio University Institutional Review Board.

Tone stimuli were presented at a comfortable level to the left ear of the listeners via a circumaural headphone (Sennheiser, HD 265) in an IAC double-walled sound booth. A custom graphical user interface (GUI) was developed in MATLAB to present the tone stimuli and to collect the listeners' responses. In order to avoid floor effects in real tests, all subjects received training and were required to reach an averaged performance of 70% percent correct with spectrally matched tone stimuli processed in a noise-excited vocoder. Each training session contained 1600 stimuli (10 syllables × 4 tones × 2 speakers × 4 channel conditions × 5 repetitions). Each training session started with stimuli processed in a 16-band processor and continued with stimuli processed in progressively fewer

Table 1
Corner frequencies of the carrier bands for 8 insertion conditions in a 16-band processor

Insertion depth (mm from base)	Carrier bands																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
28	269	329	397	475	564	664	779	910	1058	1227	1420	1639	1888	2172	2496	2864	3282
27	333	402	481	570	672	787	919	1069	1239	1434	1655	1906	2193	2519	2890	3312	3793
26	407	486	576	679	795	928	1079	1252	1447	1670	1924	2213	2542	2916	3342	3827	4379
25	492	583	686	804	938	1090	1264	1461	1686	1942	2234	2565	2943	3373	3862	4419	5053
24	589	693	812	947	1101	1276	1476	1702	1961	2255	2589	2970	3403	3897	4459	5098	5826
23	701	821	957	1112	1289	1490	1719	1979	2276	2613	2997	3434	3932	4499	5144	5878	6713
22	829	967	1123	1301	1504	1735	1998	2297	2637	3025	3466	3968	4539	5190	5930	6773	7732
21	977	1134	1314	1519	1751	2016	2318	2661	3052	3497	4004	4580	5236	5983	6833	7801	8902

The most apical edge of the electrode array varies from 28 mm to 21 mm from the base. The frequency allocations of the 8-band and 4-band processors can be derived from this table by combining adjacent 2 and 4 bands, respectively.

band processors. Eight subjects took 2 training sessions that lasted 3 h long to reach the criteria. Only one subject needed an extra training session. Tone recognition was measured in a 4-alternative forced-choice paradigm. Each stimulus was presented for only once in both training and the test. The subjects were instructed to take their best guesses even if they were not sure about the response. Subjects responded by pressing one of the four GUI buttons, each labeled with one of the four possible answers. Feedback was provided after each response during training, while no feedback was provided in the test. The test for Experiment 1 consisted of 5120 stimuli (10 syllables × 4 tones × 2 speakers × 8 insertion depths × 4 channel conditions × 2 repetitions) presented in random order and required about 5.5 h for each subject to complete. Experiment 2 contained 2560 randomized stimuli (10 syllables × 4 tones × 2 speakers × 4 compression conditions × 4 channel conditions × 2 repetitions) and took about 2.5 h for each subject to complete. The experiments were scheduled in blocks of 1–2 h. The listeners were encouraged to take breaks within each test session. The training and testing spanned on average 2–3 weeks.

3. Results

3.1. Experiment 1: Tone recognition with basal shift

Fig. 2 plots the percent correct scores as a function of simulated insertion depth for different numbers of channel conditions. As revealed by a two-way repeated-measure ANOVA, the effects of insertion depth ($F(7, 56) = 22.30, p < 0.00001$) and number of channels ($F(3, 24) = 15.58, p = 0.00007$) were both significant. Poorer performance was associated with fewer numbers of chan-

nels or shallower insertion depths. The greatest basal shift (i.e., 21 mm) caused the tone recognition performance to decrease by approximately 10 percentage points from the performance of unshifted condition (i.e., 28 mm). Data of vowel recognition from Fu and Shannon (1999) are re-plotted in Fig. 2 for comparison.

An interaction between the two main factors (i.e., insertion depth and spectral resolution) was also significant ($F(21, 168) = 2.73, p = 0.0003$). A post-hoc Cicchetti's test often used for un-confounded comparisons of interaction was performed. Comparisons of insertion depth conditions nested in the factor of number of channels showed that the effects of insertion depth were greater for stimuli with greater numbers of channels than with fewer ones. None of the 28 comparisons between insertion depth conditions for 4 channels were significant, indicating that there were no shallow insertion effects at all for 4 channels. For 8 channels, only scores for insertion depths of 22 and 21 mm differed from that of unshifted condition ($p < 0.05$). Similarly, for 12 and 16 channels, the effects of shallow insertion did not appear till 22 mm ($p < 0.05$). In addition, scores for insertion depth of 21 mm also differed from those for 27, 26, and 24 mm for 12 channels and differed from those for 23–27 mm for 16 channels ($p < 0.05$). Comparisons of the channel conditions nested in the factor of insertion depth revealed that the effects of number of channels diminished as the simulated insertion depth became shallower and eventually disappeared when insertion depth became shallower than 24 mm ($p > 0.05$).

Pattern of performance across different tones was similar for different channel conditions but different for syllable types, the types being open syllables and syllables with a nasal coda. Performance patterns of individual syllables were consistent within syllable types (i.e., open syllables or syllables with a nasal coda). Hence in Fig. 3 we show the mean scores for the two types of syllables as a function of simulated insertion depth. For the open syllables (i.e., /fu/, /ji/, /ma/, /qi/, /xi/, and /yi/), the scores for the four tones were relatively consistent across insertion conditions

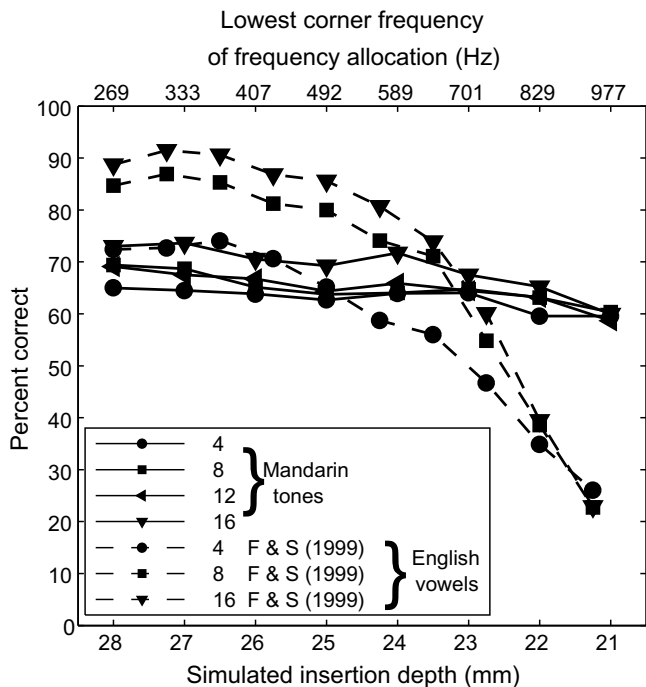


Fig. 2. Group mean tone recognition performance as a function of simulated insertion depth. Tone recognition performance is plotted for 4, 8, 12, and 16 channel conditions in solid lines with different symbols. The lowest corner frequency of frequency allocations for the carriers is noted for each simulated insertion depth. Data of vowel recognition from Fu and Shannon (abbreviated as F & S in the legend) (1999) are re-plotted with permission from the Acoustical Society of America. Simulated insertion depth of 28 mm corresponds to a full insertion or tonotopically matched condition.

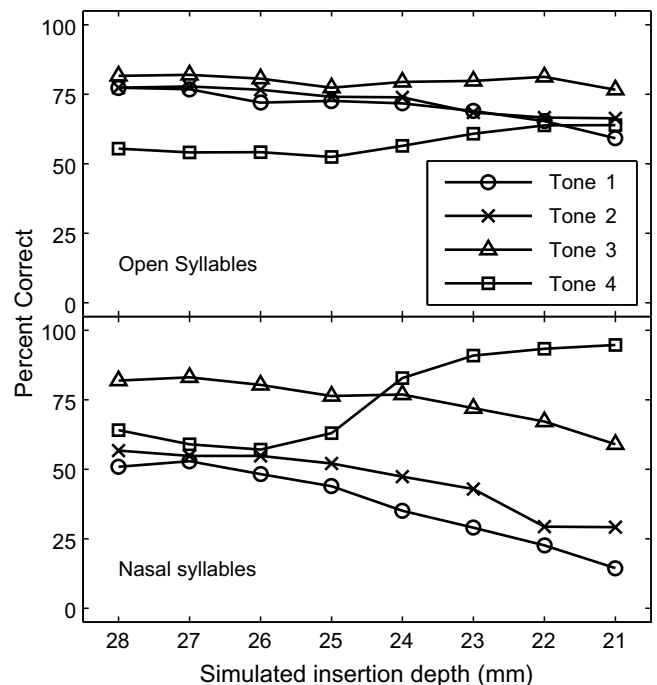


Fig. 3. Group mean performance of individual tones with spectral shift for two types of syllables. Tones 1–4 are plotted with circles, crosses, triangles, and squares, respectively. Simulated insertion depth of 28 mm corresponds to a full insertion or tonotopically matched condition.

(Fig. 3, upper panel). However, for the syllables with a nasal coda (i.e., /wan/, /xian/, /yan/, and /yang/), a seemingly increase in performance as a function of insertion depth was seen for tone 4. In contrast, the performance of the other three tones showed a declining pattern as the insertion depth became shallower (Fig. 3, lower panel). We will discuss in detail that there was a trend in responses as tone 4 that led to the increased accuracy as well as false detection of tone 4.

3.2. Experiment 2: spectral compression

Experiment 2 measured tone recognition with compressed frequency allocation for 4 channel conditions (Fig. 4, upper panel). Wider analysis frequency ranges were assigned to relatively narrower carriers (Table 2). As shown by a two-way repeated-measure ANOVA, the main effect of compression was found to be significant ($F(3, 36) = 20.33, p = 0.00005$), but the effects of number of channels were not, ($F(3, 36) = 2.55, p = 0.11$). The interaction between the two factors was also significant ($F(9, 36) = 3.82, p = 0.002$). Data collapsed across channel conditions are plotted in the lower panel of Fig. 4. Post-hoc analysis further showed that scores of 6 and 10 mm compression significantly improved from the tonotopically matched condition ($p < 0.05$). The best score was found for the 6 mm compression condition, but it was not significantly better than that of the 10 mm compression condition ($p > 0.05$). A tonotopically matched condition in this experiment simulated an implant located between 25 and 10 mm from the base (i.e., 492–5053 Hz). Note that it was different from the matched condition in Experiments 1 (e.g., 28–13 mm, or 269–3282 Hz). The recognition score of the matched condition derived from Experiment 1 is plotted in an open circle in the lower panel

of Fig. 4. A paired *t*-test showed that performance for tonotopically matched condition at 25 mm was significantly worse compared to the matched condition at 28 mm obtained from Experiment 1 ($t = 2.91, p = 0.04$).

4. Discussion

4.1. Tone recognition in tonotopically matched condition

Tone recognition was measured in two tonotopically matched conditions with simulated insertion depths of 28 and 25 mm in Experiments 1 and 2, respectively. In either condition, compared to English phoneme recognition measured using similar vocoder simulations, our results showed that tone recognition is poorer even though the chance level for tone recognition (25% correct) is much higher than phoneme recognition (5% correct for consonants and 8.33% correct for vowels). Our findings were consistent with previous vocoder studies (e.g., Xu et al., 2002, 2005) or observations from tone-language-speaking cochlear implant users (e.g., Wei et al., 2000; Ciocca et al., 2002; Lee et al., 2002; Wong and Wong, 2004). The poorer tone recognition was not surprising, as the spectral resolution provided by cochlear implants does not allow the transmission of F0 and the harmonics, while features of phonemes can be transmitted with temporal envelope information. Recognition performance with an insertion depth of 25 mm was lower than that of an insertion depth of 28 mm (Fig. 4, lower panel). This is probably due to more low frequency information being eliminated for the insertion depth of 25 mm.

4.2. Effects of basal shift

Tone recognition was much more resistant to the basal spectral shift compared to English phoneme and sentence recognition (Dorman et al., 1997; Shannon et al., 1998; Fu and Shannon, 1999). The deteriorating effects did not show until the carriers were shifted to almost two octaves higher. Data from Fu and Shannon (1999) provided a clear contrast between the effects of spectral shift on vowels and tones (Fig. 2). At the shallowest insertion depth (i.e., 21 mm), the averaged tone performance across channel conditions only decreased from the unshifted condition by approximately 10 percentage points, whereas a dramatic 60 percentage point drop was demonstrated in vowel recognition (Fu and Shannon, 1999). Although the 10 percentage point decrease in performance was statistically significant, it may not have noticeable effects in communication using lexical tones. However, clinical data related to this are not available. For the 4-channel condition, shifting the spectrum did not affect tone recognition at all. In contrast, Dorman et al. (1997) reported that consonant and vowel recognition with a 5-channel tone vocoder decreased from the unshifted condition by 30 and 60 percentage points, respectively. Given that the subjects used in their study received extensive pre-test training of 12–15 h and the unlimited numbers of presentations that the subjects were allowed to listen to, the effects of spectral shift on tones were remarkably negligible. The effects of shift on tones for other channel conditions were also quite limited, being confined to insertion depths shallower than 23 mm (i.e., 22 mm and 21 mm). The general trend was that stimuli with better spectral resolution were affected by the shifts to a greater degree. This was linked to a diminishing channel effect with greater shift of the spectrum. Such an interaction between the spectral resolution and spectral shift was also evident in the vowel recognition scores from Fu and Shannon (1999) before the scores were normalized. We speculate that a stimulus generated with a greater number of channels contains more speech information, and therefore, the amount of information vulnerable for the frequency-place mismatch is also greater.

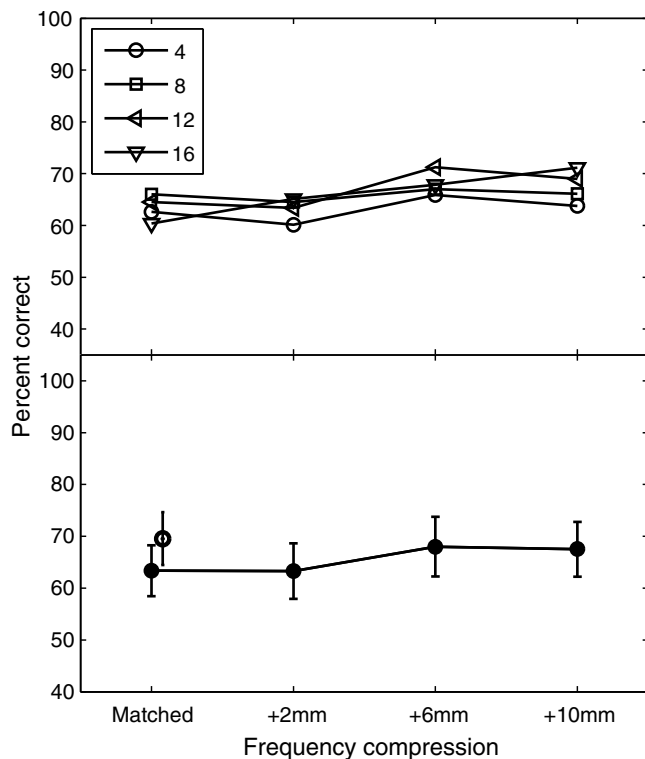


Fig. 4. Group mean tone recognition performance as a function of frequency compression. Upper panel: tone recognition performance is plotted for 4, 8, 12, and 16 channel conditions in solid lines with different symbols. Lower panel: averaged tone recognition scores across channel conditions is plotted in filled circles with error bars (\pm SD). Score of tonotopically matched condition at 28 mm derived from Experiment 1 is plotted in an open circle with error bar (\pm SD) for comparison.

Disrupted frequency-place mapping does not have an equal effect on different aspects of speech perception. Conceivably, vowels are the most prone to spectral shift (Dorman et al., 1997; Shannon et al., 1998; Fu and Shannon, 1999). The most heavily weighted cues for vowel recognition are their formant frequencies. Shifting the spectrum to higher frequencies easily destroys these spectral cues and results in acute decrease in vowel recognition. Consonant recognition, however, is not as vulnerable to spectral shift as vowel recognition is. Information transmitted for place of articulation, which greatly relies on the place of spectrum peaks, was considerably affected as a result of spectral shifts (Dorman et al., 1997). Perception of manner of articulation or voicing distinctions defined by temporal features, such as noise rising time or duration of adjacent vowels (Pickett, 1999), might not be affected by spectral changes as much as that of place of articulation. Tone recognition presumably also relies on temporal information when spectral information is limited to a small number of frequency bands (Xu and Pfungst, 2008), but the reliance is on different aspects of temporal information than phoneme recognition, which includes the periodicity contained in local channel envelopes or in the overall amplitude contours of the signal. As we have shown earlier, tone recognition was the least prone to spectral shift, but there was a difference in the performance for two types of syllables, the open syllables and the syllables with a nasal coda. The differences between the performances for the two types of syllables allowed us to investigate the contributions of the two temporal cues (i.e., envelope cues in local channels and overall amplitude contour cues) to tone recognition in frequency-place mismatched conditions.

4.3. Effects of the overall amplitude contour

Recognition of the four Mandarin tones for open syllables was quite consistent across insertion conditions. In contrast, for syllables with a nasal coda, a trend in responses as tone 4 was observed with increasing shift (Fig. 3, lower panel). This trend in responses as tone 4 caused the instances of hit for tone 4 as well as the false detection of other tones as tone 4 to increase. This response tendency could be explained by the differences between the two types of syllables in their overall amplitude contours. Fig. 5 shows the overall amplitude contours for all syllables. The overall amplitude contours for each of the four tones are plotted on top of the 16-channel vocoded waveforms in the unshifted condition. Note that the difference in overall amplitude contours between the two types of syllables remains, despite the spectral shifts. The amplitude contours of the open syllables demonstrated resemblance to various degrees to the F0 contours of the tones. The amplitude contours of the syllable with a nasal coda, however, were greatly affected by the presence of the nasal coda. Syllables with a nasal coda tend to have low amplitude at the syllable ending, as nasal stops are greatly damped due to the broader band frequency response in the vocal tract (Fujimura, 1962). As a result, the ending of nasal syllables demonstrates a downward movement regardless of its original tone identity. The downward movement in the syllable ending might be perceived as a drop in F0 and might elicit tone 4 responses. Whalen and Xu (1992) demonstrated that responses of the native Mandarin Chinese-speaking listeners primarily depended on the movement of the amplitude segment when no F0 cue was provided.

In conditions of moderate basal shift (i.e., >25 mm), when the local channel envelopes stimulated at moderately shifted places could still be perceived, the overall amplitude contours do not seem to influence tone recognition, possibly because it remains to be a less weighted cue. With increasing shift, however, the mismatch between the envelopes and the places being stimulated becomes increasingly large. In these cases, any periodicity information in the envelopes is coded at places with higher characteristic

frequencies. It is possible that the periodicity information processed in mismatched auditory filters might not be perceived well. In fact, Oxenham and colleagues (2004) demonstrated that disassociation of temporal information from the cochlear places affects temporal pitch coding. They showed that frequency discrimination of transposed tones, which are higher frequency carriers modulated with low frequency F0s, is much worse than frequency discrimination of pure tones. Their findings indicated that tonotopic representation is a necessary element for temporal pitch coding. In our cases, with increasing shift, the periodicity information is delivered to neurons that have mismatched characteristic frequencies and presumably could not be well represented. The derivation of the overall amplitude contour, however, only requires the summation of neuron firing rate across channels and therefore is independent from spectral manipulation. It becomes perceptually more dominant with increasing shift. For syllables with a nasal coda, a trend in responses for tone 4 was therefore introduced and the consequences were the enhanced recognition of tone 4 together with the increased instances of false detection for tone 4. It remains to be tested whether tonotopic mapping is critical to perceive temporal pitch, and whether the overall amplitude contour cue comes into play because of the distorted periodicity.

Our results suggest that the transition in dominance of the two cues occurred at simulated insertion depth of 25 mm (Fig. 3, lower panel). The amplitude contour of tone 3 was somewhat preserved even in the presence of the weak nasal energy (Fig. 3, lower panel). Hence recognition of tone 3 was less biased than tones 1 and 2. The increased scores of tone 4 at shallower depths offset the decreasing recognition accuracy of other tones so that the overall score did not change much with increasing shift. The overall amplitude contours of open syllable tones always provided relatively reliable cues for tone recognition, regardless of how much the envelopes in the local channels were distorted as a result of the spectral shift. Therefore, the performance of open syllable tones did not change with insertion conditions.

Even though the overall amplitude contours are independent from the manipulation of carriers in channels, the above observations suggest that this cue is not as reliable for all Chinese syllables. It is reliable for open syllables, the vocalic portion of which contains only a monophthong vowel. The amplitude contours of tones are prone to changes for other types of syllables depending on the relative energies of the nucleus and the coda. The system of Chinese vowels consists of monophthong and diphthong. Further, nasal or nasal cluster is the only possible syllable coda (Duanmu, 2002). In natural speech, the overall amplitude contours rarely work on their own to provide cues for tone recognition. Therefore, tones with nasal codas should not be more confused than tones carried by other types of syllables. Spectral shift as a result of shallow insertion potentially faced by many cochlear implant users created a unique condition where amplitude contours serve as the primary temporal cue to tone recognition. The results indicate that although not reliable for certain syllable types, the overall amplitude contour cue contributes to the resistance of tone recognition to spectral modifications. Luo and Fu (2004) also showed that tone recognition can be enhanced by modifying the overall amplitude contours according to the F0 contours.

4.4. Effects of frequency compression

Due to the limited length of the cochlear implant electrode array, the frequency range stimulated by a cochlear implant is also limited. Assigning a corresponding tonotopic frequency map to the electrodes would eliminate frequency coverage especially in the low frequency region, if the electrodes are inserted shallowly or short electrodes are used. Clinically-used maps usually com-

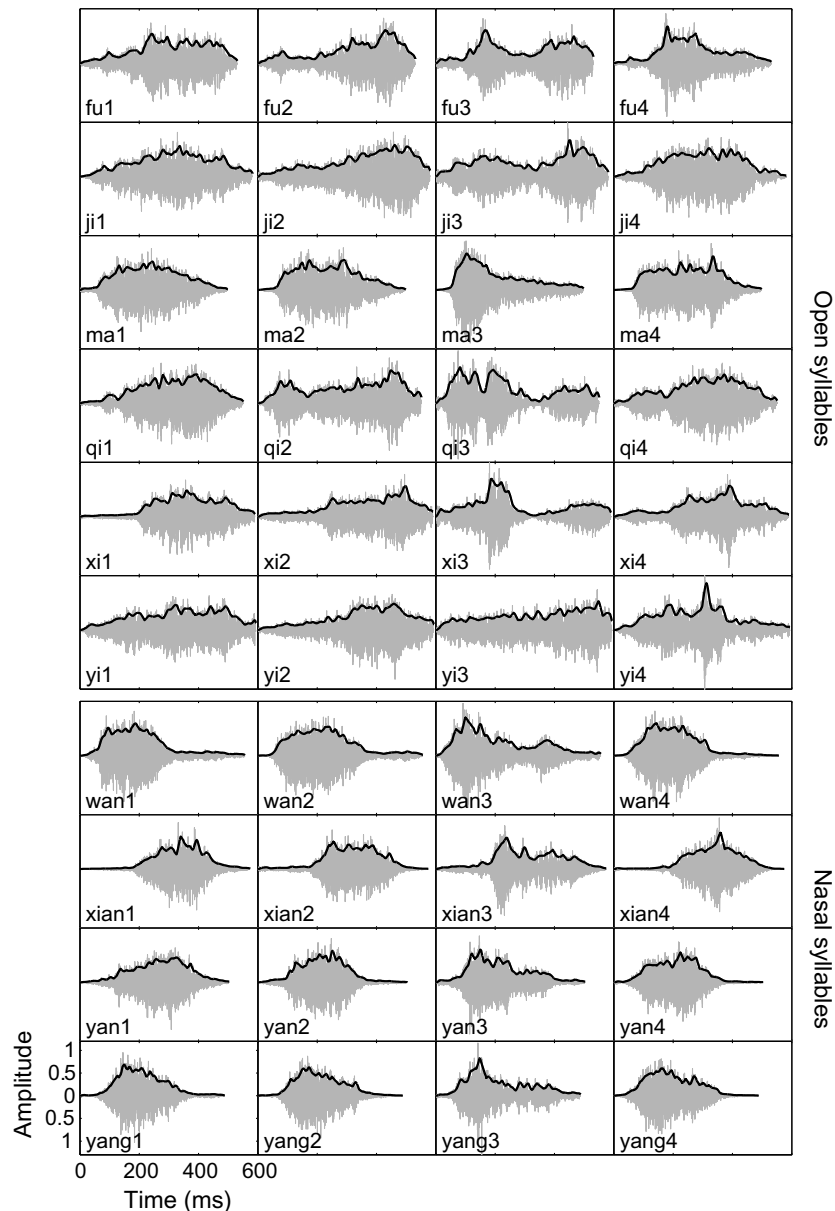


Fig. 5. Overall amplitude contours of all syllables processed in a 16-band vocoder. The original tone tokens were from a male speaker. The overall amplitude contours were extracted by detecting and smoothing the amplitude peaks of the waveform. The overall amplitude contours of four tones for each syllable are plotted in black lines on top of the waveforms that are plotted in grey.

pressively allocate a wider frequency range sufficient for speech understanding to electrodes that cover a narrower cochlear location. Experiment 2 simulated such a scenario. The implant was simulated to have a length of 15 mm and an insertion depth of 25 mm from the base.

Consistent with the previous studies (e.g., Fu and Shannon, 1999; Faulkner et al., 2003) on English phoneme recognition, a tonotopically matched condition resulted in truncation of a fair amount of low frequency information, which in turn resulted in decreased tone recognition performance compared with the performance of the insertion depth of 28 mm obtained from Experiment 1 (Fig. 4, lower panel). Compression of a wider frequency range with 3 mm or 5 mm at both frequency ends produced better performance than that without compression (Fig. 4, lower panel). A small amount of compression (i.e., 1 mm) did not provide an acoustic range wide enough, particularly on the low frequency end, to compensate for the frequency-place distortion as a result of the compression. However, the optimal performance did not oc-

cur with the largest amount of compression either. The best performance was found with a moderate compression that enhanced low frequency information for tone recognition. This was consistent with the findings by Baskent and Shannon (2003, 2005) that in shallow insertion conditions, a moderate amount of compression was better than tonotopic mapping with low frequency truncation for English speech recognition. These results suggest that wider frequency allocation that includes low frequency information critical for pitch perception may benefit tone recognition. However, the degree of compression should also be controlled relative to insertion depth to provide maximum benefit for implant patients.

In conclusion, tone recognition is fairly resistant to spectral mismatch because of its use of the overall amplitude contours independent from spectral alternation. The amplitude contour cue is not reliable for all Chinese syllables, especially for syllables with a nasal coda. Tone recognition with moderate frequency compression is improved as a result of the extended low frequency end provided with the compression. Frequency mapping for modern

cochlear implant devices often involves two forms of frequency-place mismatch examined in the present study. Although technically difficult, it would be interesting to confirm the findings of the present study in cochlear implant users.

Acknowledgements

The authors thank Dr. Qian-jie Fu for providing data from his previous publication (1999, JASA) to be re-plotted in Fig. 2. The study was supported in part by NIH NIDCD Grants R03-DC006161 and R15-DC009504.

References

- Baer, T., Moore, B.C.J., 1993. Effects of spectral smearing on the intelligibility of sentences in noise. *J. Acoust. Soc. Am.* 94, 1229–1241.
- Baer, T., Moore, B.C.J., 1994. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *J. Acoust. Soc. Am.* 95, 2277–2280.
- Baskent, D., Shannon, R.V., 2003. Speech recognition under conditions of frequency-place compression and expansion. *J. Acoust. Soc. Am.* 113, 2064–2076.
- Baskent, D., Shannon, R.V., 2004. Frequency-place compression and expansion in cochlear implant listeners. *J. Acoust. Soc. Am.* 116, 3130–3140.
- Baskent, D., Shannon, R.V., 2005. Interactions between cochlear implant electrode insertion depth and frequency-place mapping. *J. Acoust. Soc. Am.* 117, 1405–1416.
- Baskent, D., Shannon, R.V., 2006. Frequency transposition around dead regions simulated with a noiseband vocoder. *J. Acoust. Soc. Am.* 119, 1156–1163.
- Chu, K.T.Y., Au, D.K.K., Chow, C.K., Wei, W.I., 2005. Short electrode insertion in cochlear implantation: speech perception performance of Cantonese-speaking subjects. *Acta Otolaryngol.* 125, 718–724.
- Ciocca, V., Francis, A.L., Aisha, R., Wong, L., 2002. The perception of Cantonese lexical tones by early-deafened cochlear implantees. *J. Acoust. Soc. Am.* 111, 2250–2256.
- Dorman, M.F., Loizou, P.C., Rainey, D., 1997. Simulating the effect of cochlear implant electrode insertion depth on speech understating. *J. Acoust. Soc. Am.* 102, 2993–2996.
- Duanmu, S., 2002. *The Phonology of Standard Chinese*. Oxford University Press, Oxford.
- Faulkner, A., 2006. Adaptation to distorted frequency-to-place maps: Implications of simulations in normal listeners for cochlear implants and electroacoustic stimulation. *Audiology & NeuroOtolology Audiol. Neurotol.* 11, 21–26.
- Faulkner, A., Rosen, S., Stanton, D., 2003. Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *J. Acoust. Soc. Am.* 113, 1073–1080.
- Fishman, K.E., Shannon, R.V., Slatery, W.H., 1997. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J. Speech Lang. Hear. Res.* 40, 1201–1215.
- Friesen, L., Shannon, R.V., Baskent, D., Wang, X., 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110, 1150–1163.
- Fu, Q.-J., Hsu, C.-J., Horng, M.-J., 2004. Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant users. *Ear Hearing* 25, 501–508.
- Fu, Q.-J., Shannon, R.V., 1999. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J. Acoust. Soc. Am.* 105, 1889–1990.
- Fu, Q.-J., Shannon, R., Galvin III, J., 2002. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *J. Acoust. Soc. Am.* 112, 1664–1674.
- Fujimura, O., 1962. Analysis of nasal consonants. *J. Acoust. Soc. Am.* 34, 1865–1875.
- Greenwood, D.D., 1990. A cochlear frequency-position function for several species-29 years later. *J. Acoust. Soc. Am.* 87, 2592–2605.
- Hill, F.J., McRae, L.P., McClellan, R.P., 1968. Speech recognition as a function of channel capacity in a discrete set of channels. *J. Acoust. Soc. Am.* 44, 13–18.
- Kong, Y.Y., Zeng, F.G., 2006. Temporal and spectral cues in Mandarin tone recognition. *J. Acoust. Soc. Am.* 120, 2830–2840.
- Lee, K.Y.S., van Hasselt, C.A., Chiu, S.N., Cheung, D.M.C., 2002. Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *Int. J. Pediatr. Otorhinolaryngol.* 63, 137–147.
- Liang, Z.-A., 1963. The auditory perception of Mandarin tones. *Acta Physiologica Sinica* 26, 85–91.
- Lin, M.C., 1988. The acoustical properties and perceptual characteristics of Mandarin tones. *Zhongguo Yuwen* 3, 182–193.
- Liu, T.C., Chen, H.P., Lin, H.C., 2004. Effects of limiting the number of active electrodes on Mandarin tone perception in young children using cochlear implants. *Acta Otolaryngol.* 124, 1–6.
- Luo, X., Fu, Q.-J., 2004. Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *J. Acoust. Soc. Am.* 116, 3659–3667.
- Oxenham, J.A., Bernstein, G.W.J., Penagos, H., 2004. Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences* 101, 1421–1425.
- Pfingst, B.E., Franck, K.H., Xu, L., Bauer, E.M., Zwolan, T.A., 2001. Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses. *JARO* 2, 87–103.
- Pickett, J.M., 1999. *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Boston, Allyn & Bacon.
- Remez, R.E., Rubin, P.E., Pisoni, D.B., Carrell, T.D., 1981. Speech perception without traditional speech cues. *Science* 212, 947–950.
- Rosen, S., Faulkner, A., Wilkinson, L., 1999. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *J. Acoust. Soc. Am.* 106, 3629–3636.
- Shannon, R.V., Galvin III, J.J., Baskent, D., 2002. Holes in hearing. *JARO* 3, 185–199.
- Shannon, R.V., Zeng, F.-G., Wygonski, J., 1998. Speech recognition with altered spectral distribution of envelope cues. *J. Acoust. Soc. Am.* 104, 2467–2476.
- Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- ter Keurs, M., Festen, J.M., Plomp, R., 1992. Effect of spectral envelope smearing on speech reception. I. *J. Acoust. Soc. Am.* 91, 872–2880.
- ter Keurs, M., Festen, J.M., Plomp, R., 1993. Effect of spectral envelope smearing on speech reception. II. *J. Acoust. Soc. Am.* 93, 1547–1552.
- Van Tassel, D.J., Soli, S.D., Kirby, V.M., Widin, G.P., 1987. Speech waveform envelope cues for consonant recognition. *J. Acoust. Soc. Am.* 82, 1152–1161.
- Wei, W.I., Wong, R., Hui, Y., Au, D.K.K., Wong, B.Y.K., Ho, W.K., Tsang, A., Kung, P., Chung, E., 2000. Chinese tonal language rehabilitation following cochlear implantation in children. *Acta Otolaryngol.* 120, 218–221.
- Whalen, D.H., Xu, Y., 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25–47.
- Wong, A.O.C., Wong, L.L.N., 2004. Tone perception of Cantonese-speaking prelingually hearing-impaired children with cochlear implants. *Otolaryngology – Head and Neck Surgery* 130, 751–758.
- Xu, L., Pfingst, B.E., 2008. Spectral and temporal cues for speech recognition: implications for auditory prostheses. *Hear. Res.* 242, 132–140.
- Xu, L., Pfingst, B.E., 2003. Relative importance of the temporal envelope and fine structure in tone perception. *J. Acoust. Soc. Am.* 114, 3024–3027.
- Xu, L., Thompson, C.S., Pfingst, B.E., 2005. Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267.
- Xu, L., Tsai, Y., Pfingst, B.E., 2002. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J. Acoust. Soc. Am.* 112, 247–258.
- Zollner, V.M., 1979. Intelligibility of the speech of a simple vocoder. *Acustica* 43, 271–272.