

Relative Contributions of Temporal Envelope and Fine Structure Cues to Lexical Tone Recognition in Hearing-Impaired Listeners

SHUO WANG^{1,2,3}, LI XU⁴, AND ROBERT MANNELL^{2,3}

¹*Beijing Institute of Otolaryngology, Key Laboratory of Otolaryngology Head and Neck Surgery (Capital Medical University), Ministry of Education, Beijing Tongren Hospital, Capital Medical University, Beijing, People's Republic of China*

²*Department of Linguistics, Macquarie University, Sydney, Australia*

³*The HEARing Cooperative Research Center, Sydney, Australia*

⁴*School of Rehabilitation and Communication Sciences, Ohio University, Athens, OH 45701, USA*

Received: 7 April 2010; Accepted: 25 July 2011; Online publication: 11 August 2011

ABSTRACT

It has been reported that normal-hearing Chinese speakers base their lexical tone recognition on fine structure regardless of temporal envelope cues. However, a few psychoacoustic and perceptual studies have demonstrated that listeners with sensorineural hearing impairment may have an impaired ability to use fine structure information, whereas their ability to use temporal envelope information is close to normal. The purpose of this study is to investigate the relative contributions of temporal envelope and fine structure cues to lexical tone recognition in normal-hearing and hearing-impaired native Mandarin Chinese speakers. Twenty-two normal-hearing subjects and 31 subjects with various degrees of sensorineural hearing loss participated in the study. Sixteen sets of Mandarin monosyllables with four tone patterns for each were processed through a “chimeric synthesizer” in which temporal envelope from a monosyllabic word of one tone was paired with fine structure from the same monosyllable of other tones. The chimeric tokens were generated in the three channel conditions (4, 8, and 16 channels). Results showed that differences in tone responses among the three channel conditions were minor. On average, 90.9%, 70.9%, 57.5%, and 38.2% of tone responses were consistent with fine

structure for normal-hearing, moderate, moderate to severe, and severely hearing-impaired groups respectively, whereas 6.8%, 21.1%, 31.4%, and 44.7% of tone responses were consistent with temporal envelope cues for the above-mentioned groups. Tone responses that were consistent neither with temporal envelope nor fine structure had averages of 2.3%, 8.0%, 11.1%, and 17.1% for the above-mentioned groups of subjects. Pure-tone average thresholds were negatively correlated with tone responses that were consistent with fine structure, but were positively correlated with tone responses that were based on the temporal envelope cues. Consistent with the idea that the spectral resolvability is responsible for fine structure coding, these results demonstrated that, as hearing loss becomes more severe, lexical tone recognition relies increasingly on temporal envelope rather than fine structure cues due to the widened auditory filters.

Keywords: Mandarin lexical tone, fine structure, temporal envelope, auditory chimera, sensorineural hearing loss

INTRODUCTION

Mandarin Chinese is a tone language with four phonologically distinctive tones, characterized by syllable-level fundamental frequency (F0) contour patterns. These pitch contours are commonly

Correspondence to: Li Xu · School of Rehabilitation and Communication Sciences · Ohio University · Athens, OH 45701, USA. Telephone: +1-740-5930310; email: xul@ohio.edu

described as high–level (tone 1), high–rising (tone 2), falling–rising (tone 3), and high–falling (tone 4) (Lin 1988). Liu et al. (2000) observed that Mandarin tone recognition performance in mild to moderate hearing-impaired subjects was only slightly impaired, with 94%–98% correct, when audibility of the stimulus was compensated for.

Smith et al. (2002) constructed a set of acoustic stimuli, called “auditory chimera”, which provided a way to study the relative importance of temporal envelope and fine structure in speech and pitch perception. Based on the Hilbert transform, fine structure is defined as the instantaneous phase information in the signals. Temporal envelopes are defined as amplitude contours of the signal (see Kong and Zeng 2006; Moore 2008). As fine structure can be related to harmonic resolvability or to fine temporal coding of unresolved harmonics, a more neutral term, “fine structure”, is used in the present study. Xu and Pfingst (2003) divided Mandarin monosyllables into 4, 8, or 16 frequency channels, and exchanged fine structure and temporal envelope of one tone pattern with those of another tone pattern of the same monosyllable to create chimeric tokens. These tokens, containing conflicting cues, were used to investigate the relative contributions of temporal envelope and fine structure cues to Mandarin tone recognition in normal-hearing listeners. They reported that fine structure dominated tone recognition in all three channel conditions.

Questions still remain regarding how sensorineural hearing-impaired listeners utilize temporal envelope and fine structure cues for tone recognition. A few studies have reported that sensorineural hearing-impaired listeners might still have the ability to use temporal envelope cues for speech perception that is equivalent to that of normal-hearing listeners (e.g., Bacon and Gleitman 1992; Turner et al. 1995, 1999; Lorenzi et al. 2006; Xu et al. 2008). However, a number of psychoacoustic and perceptual studies have revealed that listeners with sensorineural hearing loss (SNHL) have a markedly impaired ability to use fine structure information (Moore and Moore 2003; Buss et al. 2004; Lorenzi et al. 2006, 2009; Hopkins and Moore 2007). Hopkins et al. (2008) measured the speech reception thresholds (SRTs) for sensorineural hearing-impaired listeners using processed signals with variable amounts of fine structure information. They reported that as fine structure information increased, listeners with moderate hearing loss benefited less than normal-hearing subjects for speech perception. However, they did not find any correlation between the benefits gained from the addition of fine structure information and pure-tone thresholds of the hearing-impaired listeners, perhaps due to the limited number of hearing-impaired subjects in their study.

Based on the existing evidence, we hypothesized that listeners with SNHL may not be able to perceive lexical tone in the same way as normal-hearing listeners do, as their ability to process fine structure information may be degraded. In addition, a clear relationship between the ability to process fine structure information and the degree of hearing loss has not been established. The present study was designed to determine the relative importance of temporal envelope and fine structure cues in tone recognition for three groups of listeners with SNHL from moderate to severe degree, in order to address the above-mentioned two research questions.

The “auditory chimera” technique was used in the present study. Two technical issues with the chimerizing algorithm have been addressed by a few studies (Zeng et al. 2004; Gilbert and Lorenzi 2006). First, temporal envelope cues can be reconstructed from the fine structure at the output of auditory filters when the input signals are contained in a few wide frequency bands (e.g., one or two channels), leading to a potential over-interpretation of the role of fine structure. The other issue was related to the use of large number of bands (e.g., 32–64), where the impulse responses of the narrow filters produce ringing at the center frequencies of the filters. This artifact would produce misleading results that the temporal envelope dominates pitch perception in chimera processing (Zeng et al. 2004). Therefore, in order to avoid the effects of reconstructed envelope and the ringing artifact, chimeric stimuli were generated in three channel conditions (4, 8, and 16) in the present study.

We hypothesized that listeners with SNHL may, with increasing hearing loss, rely more and more on temporal envelope, and less and less on fine structure to perceive lexical tones in comparison with normal-hearing listeners. The effect of different numbers of channels on tone recognition responses may not be dramatic for both normal-hearing and hearing-impaired listeners.

METHODS

Subjects

Two groups of subjects, all native Mandarin Chinese speakers, were recruited, including 22 normal-hearing (12 females and ten males) and 31 sensorineural hearing-impaired subjects (eight moderate, 13 moderate to severe, and ten severe). Subjects in the normal-hearing group were graduate students at Macquarie University, Australia. They were aged from 23 to 34 years old (mean=26.1, SD=2.5). Their hearing threshold levels were ≤ 15 dB HL at each octave frequency from 0.25 to 8 kHz. Subjects with

SNHL were all recruited from the Clinical Audiology Center in Beijing Tongren Hospital, China. Most of them had relatively symmetric and flat hearing loss in both ears. The demographic and hearing threshold data of these hearing-impaired subjects are described in Table 1. No remarkable history was present that suggested auditory processing disorder. The use of human subjects was reviewed and approved by the Institutional Review Boards of Macquarie University, Australia and Beijing Tongren Hospital, China.

Original materials and processed stimuli

The original materials were composed of 16 sets of Chinese monosyllables: /bai/, /chou/, /di/, /fei/, /guo/, /hu/, /jia/, /liu/, /ma/, /qü/, /she/, /tan/, /xie/, /yin/, /zao/, and /zhi/. Each of them had four tone patterns, resulting in a total of 64 combinations. These 64 monosyllables all represent real words in Chinese. These original materials were recorded twice in an acoustically treated booth by one adult male and one adult female native Beijing Mandarin speaker. The fundamental frequency (F0) of the male speaker's voice was around 160 Hz, and the F0 of the female speaker was around 280 Hz. Speech signals were captured through an M-Audio Delta 64 PCI digital recording interface attached to a computer. Recordings were made at a 44.1-kHz sampling rate and 16-bit quantization. With natural speech production, there is some variation in the durations of each syllable for different tones. On average for both speakers and each of the monosyllables, the duration of tones 1 to 4 is 683.3 ± 95.2 ms, 706.3 ± 84.5 ms, 780.1 ± 87.3 ms, and 602.0 ± 97.9 ms respectively. Consonant–vowel boundaries were time-aligned for all original speech tokens before creating the chimera. This ensured that the vowel onsets of the tokens contributing to the chimera were time-aligned and that the highly salient tone onsets were aligned across all tokens.

The chimeric tokens were generated in the three conditions with 4, 8, and 16 channels in this study. A number of FIR band-pass filters with nearly rectangular

response were equally spaced on a cochlear frequency map (Greenwood 1990). The overall frequency range for chimera synthesis was between 80 and 8820 Hz. For instance, the cutoff frequencies of filter channels for the condition of eight channels were 80, 205, 405, 724, 1236, 2055, 3366, 5463, and 8820 Hz. The transition over which adjacent filters overlap was 25% of the bandwidth of the narrowest filter in the bank of 4, 8, or 16 channels. Firstly, two tokens of the same syllable but with different tone patterns (e.g., “ma” with tone 1 and “ma” with tone 2) were passed through a number of band-pass filters to split each sound into several channels (i.e., 4, 8, or 16). The output of each filter was then divided into its temporal envelope and fine structure using a Hilbert transform. Next, the temporal envelope of the output in each filter band was exchanged with the temporal envelope in that band for the other token to produce the single-band chimeric wave. Finally, the single-band chimeric waves were summed across all channels to generate two chimeric stimuli (e.g., one with temporal envelope of “ma tone 1” and fine structure of “ma tone 2”, the other with temporal envelope of “ma tone 2” and fine structure of “ma tone 1”). Therefore, for each set of monosyllables with four tone patterns, there were 12 chimeric combinations created in each channel condition for each voice. The duration of each chimeric token is always identical to the duration of the token donating the temporal envelope cues. In addition, the “intact tokens” having fine structure and temporal envelope from the same syllable and same tone were also used to measure how well hearing-impaired listeners could perceive the tone in the original speech. Thus, a total of 1,536 tokens were prepared (i.e., 16 combinations \times 16 sets of syllables \times 3 channel conditions \times 2 voices). Signal processing was performed in MATLAB (Mathworks, Natick, MA, USA) environment.

Additionally, in order to investigate how reconstructed envelope from fine structure at the output of cochlear filters might potentially impact on tone response for these chimeric tokens, a tone-excited

TABLE 1

Age and hearing thresholds (mean \pm SD) at each frequency from 0.25 to 4 kHz for the moderate, moderate to severe, and severely hearing-impaired listeners							
	Age (years)	Ear	0.25 kHz	0.5 kHz	1 kHz	2 kHz	4 kHz
Moderate (5M, 3F)	19.1 \pm 6.3	R	37.5 \pm 5.3	42.5 \pm 5.3	51.3 \pm 5.2	56.3 \pm 6.4	55.6 \pm 6.8
		L	35.6 \pm 5.6	41.3 \pm 3.5	51.3 \pm 4.4	55.6 \pm 3.2	56.9 \pm 5.9
Moderate to severe (7M, 6F)	25.2 \pm 15.0	R	52.3 \pm 9.9	55.8 \pm 7.6	62.7 \pm 6.3	66.2 \pm 4.6	64.6 \pm 5.6
		L	50.0 \pm 8.7	53.8 \pm 7.9	63.1 \pm 4.8	64.6 \pm 6.3	66.2 \pm 5.8
Severe (8M, 2F)	18.1 \pm 4.3	R	68.0 \pm 6.7	72.5 \pm 3.5	73.0 \pm 7.1	79.0 \pm 10.7	78.0 \pm 11.4
		L	66.5 \pm 9.7	69.5 \pm 7.6	75.0 \pm 5.8	80.5 \pm 8.6	76.0 \pm 9.4

vocoder, described in Gilbert and Lorenzi (2006) and Gnansia et al. (2009), was used to extract reconstructed envelope from the chimeric tokens. Each chimeric token was passed through a bank of 32, fourth-order gammatone filters, each 1-ERB wide with center frequencies ranging from 80 to 8600 Hz. Temporal envelope cues were extracted using full-wave rectification and lowpass filtering (sixth-order Butterworth filter) at $ERB_N/2$ for analysis filters whose bandwidths were greater than 128 Hz. For those filters whose bandwidths were smaller than 128 Hz, the envelopes were lowpass filtered at 64 Hz. The extracted envelopes were then used to amplitude modulate sine-wave carriers with the same frequencies as the center frequencies of the gammatone filters. The resulting modulated signals were summed across the 32 frequency bands to create the synthesised tokens. A total of 1,152 tokens were generated (i.e., 12 chimeric combinations \times 16 sets of syllables \times 3 channel conditions \times 2 voices), and four of the normal-hearing subjects were tested in this subset of experiment.

Procedure

The experiments were conducted in an acoustically treated booth. A GSL-16 clinical audiometer connected with a computer was used to adjust the intensity of chimeric tokens. A custom graphical user interface (GUI) written in MATLAB as described in Xu and Pfingst (2003) was used to present the stimuli and to record the responses during the experiment. Subjects listened to the chimeric speech materials presented through headphones bilaterally. For normal-hearing subjects, the presentation level of each token was fixed at 80 dB SPL. For hearing-impaired subjects, the intensity of the stimuli was set at the most comfortable loudness level. Tone tokens were presented in a fully randomized order. As the stimuli were presented, a set of four Chinese characters with four tone patterns and their 'pinyin' were displayed on the screen (i.e., 妈mā; 麻má; 马mǎ; 骂mà). The subjects were asked to use a computer mouse to select which tone pattern (or what Chinese word) was in the chimeric

token that they had heard. No feedback was provided. Before the formal test, they were all given instructions and undertook a practice session listening to the tokens and using the computer mouse in order to familiarize them with the test. They were encouraged to take their time to practise until they felt ready for the test. On average, the practice session was about 10 minutes, and the entire experiment took approximately 1.5–2 hours for each subject. The subjects could take breaks at any time as they desired.

RESULTS

Overall tone recognition performance

No significant correlation was found between the age of subjects and the percentages of tone responses in each hearing-impaired group (all $p > 0.05$), indicating that our results were not affected by the age of subjects. Likewise, no correlation was found to be significant between the duration of monosyllables and tone responses that were consistent with temporal envelope cues, suggesting that the results were also not affected by the duration of each chimeric token (all $p > 0.05$).

Tone recognition results were pooled across channel conditions, both male and female voices, and all subjects in the normal-hearing group and the three subgroups of sensorineural hearing-impaired listeners. As shown in Table 2, on average 90.9%, 70.9%, 57.5%, and 38.2% of the tone responses were consistent with fine structure for the normal-hearing, moderate, moderate to severe, and severely hearing-impaired groups, whereas 6.8%, 21.1%, 31.4%, and 44.7% respectively of the tone responses were consistent with temporal envelope for the above-mentioned groups. More errors (i.e., tone responses that were neither consistent with temporal envelope nor fine structure) occurred as the hearing loss of subjects became more severe, with averages of 2.3%, 8.0%, 11.1%, and 17.1% of responses for the normal-hearing, moderate, moderate to severe, and severely hearing-impaired listeners respectively. The accuracy of tone recognition to the intact tokens was 99.0%,

TABLE 2

Mean results of tone responses to the intact tokens, responses that were consistent either with fine structure cues or temporal envelope cues, and responses that were neither consistent with fine structure cues nor temporal envelope cues. These results were averaged across three channel conditions and both talkers

%	<i>Intact tokens</i>	<i>Fine structure</i>	<i>Temporal envelope</i>	<i>Error</i>
Normal	99.0 \pm 0.8	90.9 \pm 2.6	6.8 \pm 2.1	2.3 \pm 0.9
Moderate	97.3 \pm 1.0	70.9 \pm 3.8	21.1 \pm 3.2	8.0 \pm 1.8
Moderate to severe	94.5 \pm 4.1	57.5 \pm 6.2	31.4 \pm 4.7	11.1 \pm 3.1
Severe	85.9 \pm 5.3	38.2 \pm 8.3	44.7 \pm 6.8	17.1 \pm 3.6

97.3%, 94.5%, and 85.9% correct for the above-mentioned groups of subjects.

To illustrate the observation that subjects tended to rely more and more on temporal envelope rather than fine structure in tone recognition as their hearing impairment became more severe, Figure 1 shows the mean tone recognition performance for each individual tone for each group of subjects. The four large panels in Figure 1 represent the results from each group or subgroup (i.e., normal-hearing, moderate, moderate to severe, and severely hearing-impaired, from left to right). Each panel consists of 16 bar plots, each representing one of the combinations of fine structure (column-wise) and temporal envelope (row-wise). The mean and standard deviation of the percentages of tone responses for tone 1 through tone 4 are plotted in each bar plot. As shown in the left panel of Figure 1, when normal-hearing subjects listened to chimeric tokens, tone responses were almost always consistent with fine structure cues of the stimuli but rarely consistent with temporal envelope cues. When moving from left to right in each row (i.e., same temporal envelope but different fine structure), the most frequent tone responses moved along from tone 1 to tone 4. However, when moving from top to bottom in each column (i.e., same fine structure but different temporal envelope), the most

frequent tone responses remained unchanged. However, this pattern of tone responses gradually changed as the hearing loss of subjects became more severe. In contrast to the normal-hearing data, subjects with severe SNHL showed an increasing percentage of tone responses that became consistent with temporal envelope cues (Fig. 1, right panel). When moving from top to bottom in each column (i.e., same fine structure but different temporal envelope), a significant proportion of tone responses moved along from tone 1 to tone 4. The tone response patterns for subjects with moderate and moderate to severe hearing loss appeared to be intermediate between those of the subjects with normal hearing and severe hearing loss.

Two sets of three-way ANOVA were performed to examine the effects of the following three factors on tone response results which were consistent with temporal envelope or fine structure: audiometric thresholds (normal-hearing and three groups of hearing-impaired listeners), three channel conditions (four, eight, and 16), and two talkers (male and female). For tone responses that were consistent with fine structure, a significant interaction was found between the group and talker [$F(3, 288) = 23.2, p < 0.001$]. As the hearing loss became more severe, the number of tone responses based on fine structure

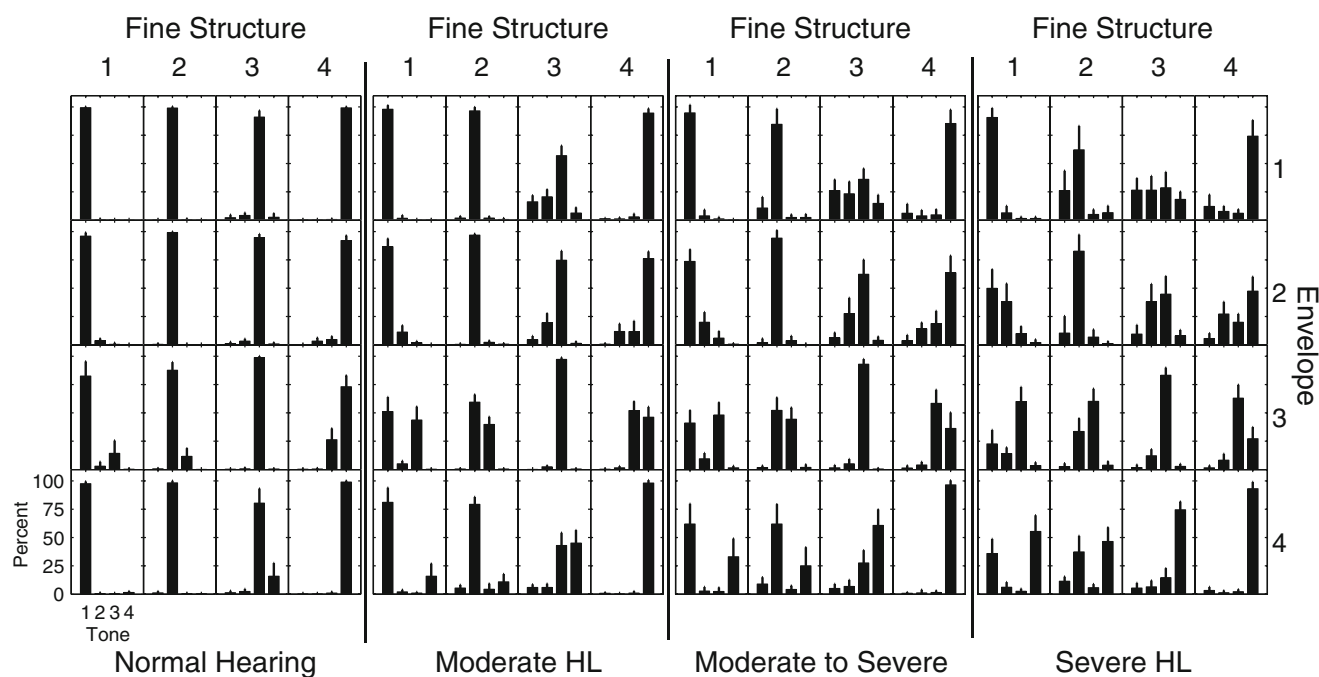


FIG. 1. Overall means and standard deviations of the percentages of the tone recognition responses across channel conditions, both male and female voices, and all subjects in the normal-hearing group and the three subgroups of sensorineural hearing-impaired listeners. The four large panels from left to right represent the data from subjects with normal-hearing, moderate, moderate to severe, and severe hearing loss respectively. In each large panel, small panels

column-wise from left to right represent the fine structure cues from tone 1 to tone 4 in the chimeric stimuli, while those row-wise from top to bottom represent the envelope cues from tone 1 to tone 4 in the stimuli. Each bar and its error bar in the bar plot represents the mean and standard deviation of the percentages of tone responses for tone 1, 2, 3, or 4.

reduced more markedly for the male voice than the female voice. The significant interaction was also observed between the channel and talker [$F(2, 288) = 36.6, p < 0.001$]. As the number of channels increased, tone responses were more often consistent with fine structure for the male voice, while for the female voice, the largest proportion of tone responses that were consistent with fine structure appeared at eight channels. There was no significant interaction between the group and channel [$F(6, 288) = 1.6, p > 0.05$]. Similarly, for tone responses that were consistent with temporal envelope, significant interactions were also found between the group and talker [$F(3, 288) = 17.4, p < 0.001$], and the channel and talker [$F(2, 288) = 44.3, p < 0.001$]. As the hearing loss of subjects became more severe, the number of tone responses based on temporal envelope increased more markedly for the male voice than the female voice. As the number of channels increased, tone responses were less often consistent with temporal envelope for the male voice, while for the female voice, the smallest proportion of tone responses that were consistent with temporal envelope occurred at eight channels. No significant interaction was observed between the group and channel [$F(6, 288) = 1.3, p > 0.05$]. Therefore, these results indicate that the effects of hearing abilities and channel conditions on tone-response performance differed in listening to male and female recorded materials. The effects of hearing abilities on tone-response performance were not affected by the different channel conditions (see below for details).

Effects of audiometric thresholds on tone responses

One of the goals of our study was to determine whether tone responses vary systematically based on the degree of hearing loss. The main effect of the audiometric thresholds on tone responses which were consistent either with fine structure or with temporal envelope were found to be highly significant [fine structure: $F(3, 288) = 862.2, p < 0.001$; temporal envelope: $F(3, 288) = 640.3, p < 0.001$]. Post hoc tests with Bonferroni correction further showed that significant differences were found between any two groups of subjects, suggesting that as audiometric thresholds increase up to different hearing degrees, tone-response performance based on fine structure or temporal envelope significantly changes (see Table 2). The correlation between the average of pure-tone hearing thresholds between 0.5 and 2 kHz ($PTA_{0.5 \text{ to } 2 \text{ kHz}}$) and tone-recognition performance for the hearing-impaired subjects is shown in Figure 2. The upper panel shows the negative correlation between the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ and tone

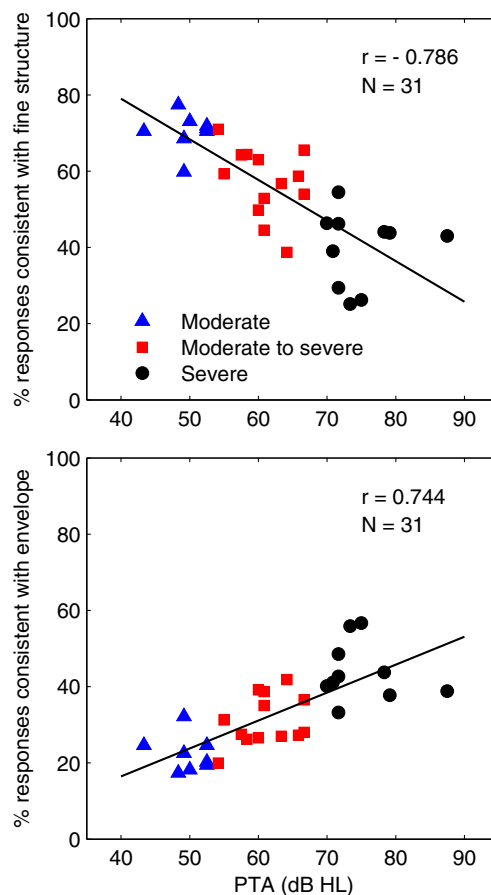


FIG. 2. The correlation between the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ and tone recognition performance across three subgroups of hearing-impaired subjects. The different symbols represent different degrees of hearing loss as indicated in the legend. The upper panel represents the correlation between the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ and tone responses that were consistent with fine structure cues, while the lower panel represents the correlation between the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ and tone responses based on temporal envelope cues.

responses that were consistent with fine structure [$r(31) = -0.786, p < 0.001$], suggesting that as the hearing loss becomes more severe, tone-recognition responses that were based on fine structure progressively decreased. The lower panel shows the positive correlation between the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ and tone responses that relied on temporal envelope [$r(31) = 0.744, p < 0.001$], suggesting that as the hearing loss became more severe, tone recognition responses that were consistent with temporal envelope cues progressively increased.

We chose to use the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ in Figure 2, because the $PTA_{0.5 \text{ to } 2 \text{ kHz}}$ is most commonly used in the clinic. However, when we calculated correlations between pure-tone thresholds of various frequencies and tone responses that were consistent with either fine structure or temporal envelope, it was found that the correlations were the strongest when the lower frequencies (i.e., 0.25 and 0.5 kHz) were used (see Table 3). For fine structure, the strength of the coefficients at the low frequency (i.e., 0.25 or 0.25

Correlation coefficients of between pure-tone thresholds and tone responses consistent with fine structure or temporal envelope. All correlation coefficients are significantly different from zero (z test, $p < 0.001$, $N = 31$)		
Frequency (kHz)	r values	
	Fine structure	Temporal envelope
0.25	-0.839	0.818
0.5	-0.813	0.774
1	-0.697	0.652
2	-0.683	0.644
4	-0.588	0.560
0.25-0.5	-0.840	0.808
1-2	-0.721	0.677
2-4	-0.654	0.620

and 0.5 kHz in combination) was significantly higher than that at the high frequency (i.e., 4 kHz) (two-tailed z test, $p < 0.05$). For temporal envelope, no significant difference was present between the strength of the coefficients at the low frequency and that at the high frequency (two-tailed z test, $p > 0.05$).

Effects of number of channels on tone responses

In normal-hearing listeners, the effects of number of channels (from 4 to 16) on chimera tone responses were found to be minimal (Xu and Pfungst 2003). In the present study, such effects in the hearing-impaired listeners were explored. Figure 3 plots the mean percentages and standard deviations of tone-recognition responses that were consistent with fine structure (upper panel) or temporal envelope (lower panel) in each group for four, eight, and 16 channels. Under all three channel conditions, subjects with more severe hearing loss tended to base their identification of tones more and more on temporal envelope cues and less and less on fine structure cues. Although the main effect of the different numbers of channels on tone responses that were either consistent with fine structure or temporal envelope were found to be significant [fine structure: $F(2, 288) = 12.5$, $p < 0.001$; temporal envelope: $F(2, 288) = 13.2$, $p < 0.001$], the differences in tone responses between channel conditions were minor, with only approximately 3.5 percentage points on average (from 0.3 to 7.4 percentage points). Statistical analysis showed a significant interaction between the gender and channel in tone responses. For the female voice, the largest difference in tone responses was found between 8 and 16 channels, with 8.2 percentage points on average, while the smallest difference was found between four and eight channels, with 2.6 percentage points on average. For the male voice, the largest difference in tone responses was found between 4 and 16 channels,

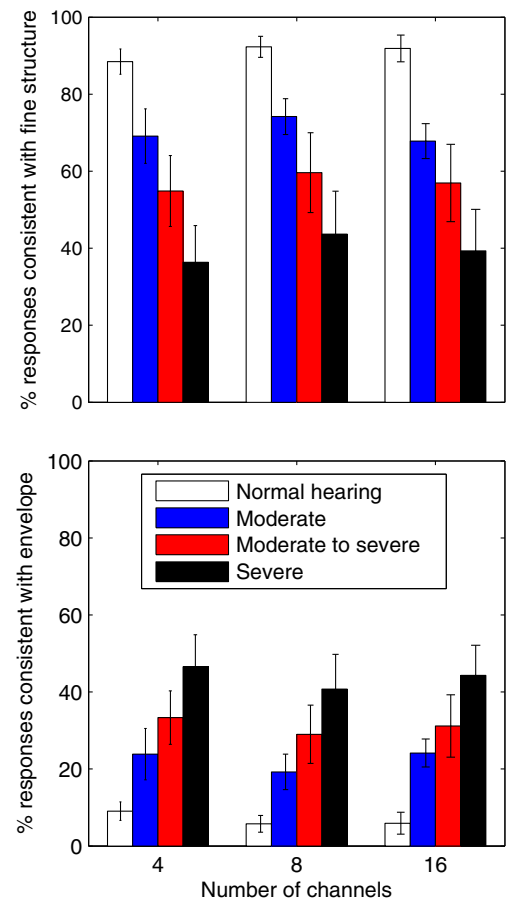


FIG. 3. Mean percentages and standard deviations of tone recognition responses in the normal-hearing group and the three hearing-impaired subgroups using 4, 8, and 16 channels. The upper panel represents the tone responses that were consistent with fine structure of the chimeric stimuli in each group. The lower panel represents the tone responses consistent with temporal envelope of the stimuli in each group.

with 10 percentage points on average, while the smallest difference was found between 8 and 16 channels, with 3.4 percentage points on average.

Effects of talker gender on tone responses

Male and female talkers have respectively relatively low and high voice F_0 . Such a difference in voice F_0 may influence how listeners utilize the fine structure of the stimuli. The main effect of talker gender on tone responses that were either consistent with fine structure or temporal envelope were found to be highly significant [fine structure: $F(1, 288) = 465.0$, $p < 0.001$; temporal envelope: $F(1, 288) = 382.8$, $p < 0.001$]. Eight paired t -tests were further conducted, and showed significant differences in tone recognition that were consistent either with fine structure or temporal envelope between the male and female voices for each group of subjects (all $p < 0.01$ with Bonferroni correction) (see Fig. 4). Subjects achieved

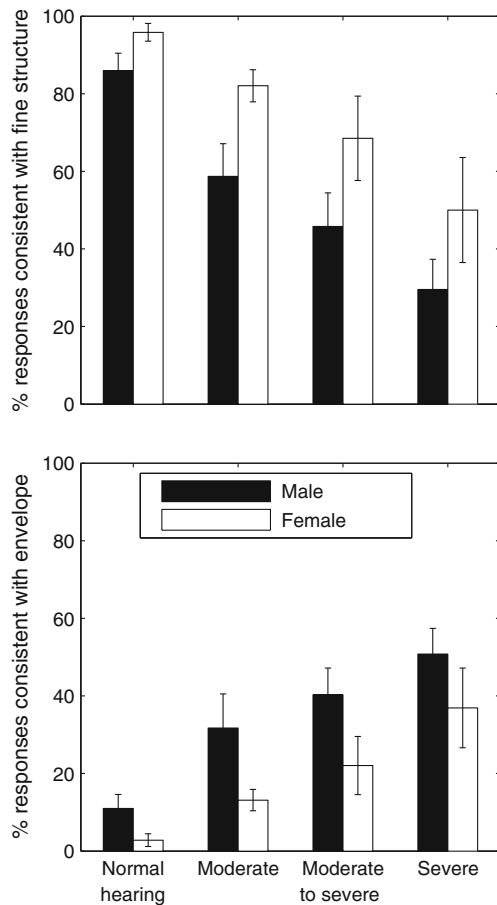


FIG. 4. Mean and standard deviations of the percentages of tone responses consistent with fine structure (*upper panel*) and temporal envelope (*lower panel*) in four groups of subjects. The *filled and open bars* represent the tone recognition performance listening to the male and female voices respectively.

their tone perception with more responses that were consistent with temporal envelope for the male voice relative to the female voice, whereas they relied more on fine structure in tone perception for the female voice than for the male voice. In comparison with the normal-hearing subjects, larger differences of tone responses between the female and male voices were present for hearing-impaired subjects.

Effects of reconstructed envelope from fine structure on tone responses

As mentioned in Introduction, the reconstructed envelope from fine structure may impact on tone response results. The effects of reconstructed envelope on tone responses were examined in an additional experiment. Figure 5 shows the mean tone response scores consistent with fine structure obtained in non-vocoded chimeric tokens (the upper solid line) and those consistent with reconstructed envelope obtained in vocoded chimeric tokens conveying reconstructed enve-

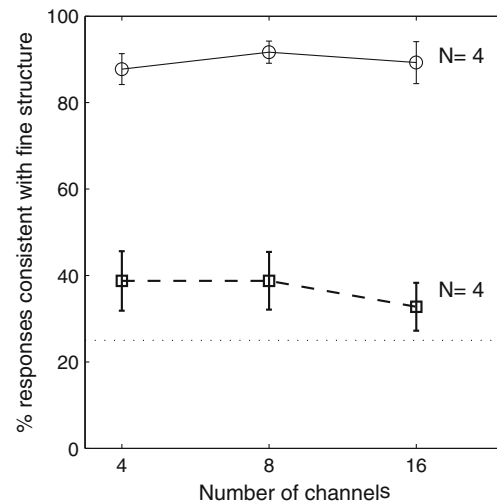


FIG. 5. Mean percentage of tone response consistent with fine structure obtained with the chimeric tokens (the *upper solid line with circles*) and that consistent with reconstructed envelope obtained with the tone-vocoded tokens (the *lower dashed line with squares*) using 4, 8, and 16 channels for four normal-hearing listeners. Chance level corresponded to 25% (the *dotted line*). Each error bar represents ± 1 SD.

lope (the lower dashed line) for four normal-hearing listeners using 4, 8 and 16 channels. These four normal-hearing listeners produced nearly identical results to the group of 22 normal-hearing subjects in responses to non-vocoded chimeric tokens (data in the upper panel of Fig. 3). As described above (Table 2 and Fig. 3), nearly 90% of tone responses were perceived based on fine structure in all three channel conditions. The tone responses based on reconstructed envelope were much lower than those based on fine structure, with approximately 37% in the 4- and 8-channel conditions, and 31% in the 16-channel condition. In addition, to respond vocoded chimeric tokens, approximately 39% of tone responses on average consisted with original temporal envelope cues, and about 25% of tone responses were consistent with neither original envelope nor reconstructed envelope.

DISCUSSION

Our results are consistent with the hypothesis that, in contrast to normal-hearing listeners (Xu and Pfingst 2003), hearing-impaired listeners rely less and less on fine structure, but more and more on temporal envelope cues to perceive tones, as their hearing loss progresses. These results are also in agreement with previous studies that demonstrated the preserved ability to use temporal envelope cues for sensorineural hearing-impaired subjects (e.g., Bacon and Gleitman 1992; Turner et al. 1995, 1999; Lorenzi et al. 2006; Xu et al. 2008; Kale and Heinz 2010), and the degraded ability to

use fine structure cues (e.g., Moore and Moore 2003; Buss et al. 2004; Lorenzi et al. 2006, 2009; Hopkins and Moore 2007, 2010; Santurette and Dau 2007; Hopkins et al. 2008; Strelcyk and Dau 2009; Heinz et al. 2010).

Two interesting observations are reported from this study. Firstly, with a large cohort of subjects ($N=31$), we demonstrated that the hearing thresholds were negatively correlated with tone responses consistent with fine structure and positively correlated with tone responses consistent with temporal envelope (see Fig. 2 and Table 3). Listeners with SNHL relied more and more on temporal envelope cues, but less and less on fine structure for tone perception using the speech–speech chimeras. They did not make substantial responses that were consistent with neither fine structure nor temporal envelope. Even severely hearing-impaired listeners primarily made tone responses based on temporal envelope rather than errors. Presumably, the ability to use temporal envelope for tone recognition can compensate for the deficit in the ability to use fine structure to some extent, as hearing loss becomes more severe. Thus, overall tone-recognition performance with intact tone tokens is not dramatically reduced in hearing-impaired listeners in the present study. Secondly, we observed a tendency for the correlation between hearing thresholds and tone responses based on fine structure and temporal envelope to become weaker as the frequency range shifts from the low region (e.g., 0.25–0.5 kHz) towards the high region (e.g., 4 kHz). The difference between the strength of the coefficients at the low frequency and that at the high frequency was found to be significant for fine structure. This observation is in agreement with the finding in Hopkins and Moore (2010) that evaluated the importance of fine structure at different spectral regions contributing to speech perception. They measured the SRTs with a competing talker background for normal-hearing listeners using processed signals with variable amounts of fine structure information in two conditions, adding fine structure information starting either from low-frequency bands or from high-frequency bands. Results indicated that the greatest improvement in speech perception occurred when fine structure information was added to channels with center frequencies below 1000 Hz, consistent with the idea that fine structure information is important for coding F0. Further research is warranted to explore the effects of different frequency regions on fine structure or temporal envelope processing.

Issues in methodology

While “auditory chimera” provided an opportunity to examine perceptual processing when two conflicting cues, i.e., temporal envelope and fine structure,

coexist in the stimuli, the technique suffers certain limitations. The auditory chimera algorithm, as developed by Smith et al. (2002), utilizes the Hilbert transform, which rigorously computes the fine structure and temporal envelope based on the mathematical formula, without any free parameters to limit the temporal envelope extraction (Bedrosian 1963). We examined empirically the waveforms and spectral contents of the extracted envelopes of the speech tokens used in the experiment. In all channel conditions, the periodicity fluctuations between 100 and 300 Hz were present in the envelopes of the analysis bands. This would contribute to the blurring between temporal envelope and fine structure cues in such a frequency range in the chimeric speech tokens. Thus, we caution over-interpretation of a clear-cut separation of temporal envelope and fine structure in the chimera technique. To avoid such a technique limitation, future research with specific controls of the upper frequency limits of the envelopes extracted through Hilbert transform, as suggested by Schimmel and Atlas (2005), is warranted.

The potential contribution of reconstructed envelope to tone responses was examined directly on normal-hearing subjects in this study. As shown in Figure 5, tone responses based on reconstructed envelope were only approximately 12 percentage points above the chance level in the 4- and 8-channel conditions and 6 percentage points above the chance level in the 16-channel condition, indicating that reconstructed envelope did not make a substantial contribution to tone responses using chimeric tokens. Moreover, many studies have demonstrated that the broadening of auditory filters often associated with SNHL may degrade the recovery of envelope cues from fine-structure speech (e.g., Lorenzi et al. 2006; Heinz and Swaminathan 2009). Therefore, it is reasonable to speculate that the reconstructed envelope cues from fine structure through the impaired cochlear filters of listeners should be degraded relative to normal-hearing listeners, and that the contribution of reconstructed envelope to tone responses in sensorineural hearing-impaired listeners should be even less than in normal-hearing listeners. In turn, the reduced reconstructed envelope from fine structure associated with broader auditory filters may induce the degraded ability to use fine structure cues in listeners with SNHL to some extent. However, the degraded ability to use fine-structure speech for hearing-impaired listeners does not result solely from poorer reconstruction of envelope cues from fine-structure speech. Ardoint et al. (2010) found that large differences in speech intelligibility still remained between normal-hearing and hearing-impaired listeners when temporal-envelope reconstruction from fine-structure speech was strongly limited, indicating

genuine deficits in the ability to use fine structure by SNHL listeners.

In this study, spectral shaping was not considered in order to ensure complete audibility of speech cues across the tested frequency range, because Sheft et al. (2008) demonstrated that the ability to use temporal envelope and fine structure cues in a speech-identification task was not significantly affected by large variations in presentation level (40–45 dB vs 80 dB SPL) for normal-hearing listeners. Therefore, the present results were unlikely to be affected by the reduced audibility in the high-frequency range.

Furthermore, in previous perceptual studies using fine-structure speech to examine speech perception, subjects generally required plenty of time for training (i.e., more than ten sessions) (e.g., Lorenzi et al. 2006; Ardoit and Lorenzi 2010). However, this was not the case in the present study. When listening to the chimeric tokens, subjects might hear two conflicting tone patterns (one is from fine structure and the other is from temporal envelope) carried on the syllable. Presumably, they made their choice based on which tone pattern was more perceptually salient to them. Therefore, in the simple four-alternative forced-choice paradigm, our subjects usually took only several minutes for practice.

Underlying mechanisms for the deficits of processing fine structure in SNHL

As the hearing impairment becomes more severe, the frequency selectivity of listeners with SNHL is reduced due to increased bandwidths of the auditory filters. Spectral resolvability in the auditory system plays an important role in processing each frequency component more or less independently. This may imply the possibility that reduced frequency selectivity of sensorineural hearing-impaired listeners may induce the impaired ability to process fine structure information in tone recognition. Moore and Peters (1992) observed a weak correlation between reduced frequency selectivity and reduced F0 discrimination performance for a group of hearing-impaired listeners. In addition, Bernstein and Oxenham (2006) reported that a transition between good and poor F0 difference limens in hearing-impaired listeners corresponded well with the measurement of frequency selectivity, indicating that reduced frequency selectivity may result in the poorer F0 discrimination. Based on this view, we may offer a reasonable interpretation of the results in the present study. Listeners with SNHL may have reduced frequency selectivity to resolve low-order harmonics associated with broader auditory filters, so that they have the degraded ability to process fine structure information in the chimeric tokens. As the hearing impairment becomes greater,

this degradation appears even more severe, so that hearing-impaired listeners rely less and less on fine structure cues in lexical tone perception, and make more errors in their tone responses. On the other hand, we observed that all hearing-impaired listeners produced far more responses consistent with temporal envelope than errors (Table 2). Thus, it seems that the deficit of fine structure processing has been compensated for by the ability to process temporal envelope information in lexical tone perception to some extent, so that even severely hearing-impaired listeners could still achieve a reasonably good tone-perception performance. This in turn also indicated the importance of temporal envelope cues in tone perception in listeners with SNHL. Another possible explanation is that fine structure may be coded in the phase locking across auditory nerve fibers, so the deficits of fine structure processing may be also attributed to the impaired temporal coding (i.e., phase-locking) in sensorineural hearing-impaired listeners (Lorenzi et al. 2006, 2009; Moore et al. 2006). However, the physiological evidence underlying the fine structure coding deficit in hearing impairment continues to be debated.

Implications for cochlear implants

Contemporary cochlear implant technology has provided unfavorable lexical tone perception for tone language speakers (e.g., Xu et al. 2002, 2011; Wei et al. 2004; Xu and Pfungst 2008; Zhou et al. 2008; Han et al. 2009; Wang et al. 2011; see Xu and Zhou 2011 for a review). Previous studies based on the measurement of normal-hearing listeners have demonstrated that fine structure is more critical for pitch perception than temporal envelope (e.g., Smith et al. 2002; Xu and Pfungst 2003; Kong and Zeng 2006). Therefore, transmitting more fine structure through cochlear implants to users might be an effective approach to improving pitch perception for them (e.g., Nie et al. 2005). So far, however, little evidence has shown that any of the fine-structure strategies would make a dramatic improvement for lexical tone perception or music perception in cochlear implant users (e.g., Riss et al. 2008, 2009; Han et al. 2009; Firszt et al. 2009; Schatzer et al. 2010). Many studies have shown that cochlear implant users could only detect differences in pitch for frequencies up to about 300 Hz (e.g., Shannon 1983; Zeng 2002), which is much poorer than that observed in normal-hearing listeners (Carlyon and Deeks 2002). With this limit of temporal pitch, cochlear implant users may not be able to utilize fine structure information that has been transmitted from the novel processing strategies. As summarized by Carlyon et al. (2007), a match between place and rate of stimulation, a credible

phase transition among the peaks of the travelling wave, and the ability to transmit periodicity information may be necessary to improve the encoding of pitch-related information in cochlear implants. Further research is warranted to explore how much fine structure information is available for cochlear implant users in electric hearing to perceive pitch (e.g., lexical tone or music perception).

Talker gender effects

Another notable observation from this study is that subjects achieved their tone perception more consistently with temporal envelope cues for the male voice in relation to the female voice, whereas they relied more on fine structure cues in tone perception for the female voice than the male voice (see Fig. 4). The differences in each group were all significant, while the difference became more remarkable in the hearing-impaired subjects as compared to the normal-hearing subjects. This observation may be explained by the resolvability view as discussed above. The auditory filters are wider as a proportion of their center frequencies in low frequency than high frequency region. Therefore, the harmonics would not be well-resolved for the male voice, as F0 of the male voice is much lower than that of the female voice (160 Hz vs. 280 Hz). As hearing impairment becomes more severe, the auditory filters become even wider in hearing-impaired subjects. Thus, the hearing-impaired listeners showed even greater difference in tone responses between the two voices.

SUMMARY

In summary, our study indicates that the relative importance of fine structure and temporal envelope cues in lexical tone recognition in sensorineural hearing-impaired listeners is different from that in normal-hearing listeners. Hearing-impaired listeners rely more and more on temporal envelope cues to perceive tones, but less and less on fine structure as the hearing loss becomes more and more severe. These findings indicate that sensorineural hearing-impaired listeners have a progressively reduced ability to use fine structure in lexical tone perception, but that the ability to use temporal envelope in tone perception may not be degraded. In addition, there is a strong negative correlation between pure-tone thresholds and lexical tone responses based on fine structure, and a positive correlation between pure-tone thresholds and lexical tone responses based on temporal envelope. The strength of the correlation between hearing thresholds and tone responses consistent with fine structure or temporal envelope tends to become weaker as the

frequency range shifts from the low to high region. Furthermore, hearing-impaired listeners, even listeners with severe SNHL, made many more tone responses based on temporal envelope than erroneous responses. The ability to use temporal envelope for tone recognition can compensate for the deficit in the ability to use fine structure to some extent, as hearing loss becomes more severe.

ACKNOWLEDGMENTS

We thank all the subjects for participating in this study, and all staff at the Clinical Audiology Center in Beijing Tongren Hospital for helping in the recruitment of the hearing-impaired subjects. We are grateful to Dr. Christian Lorenzi for providing their tone-excited vocoder for us to use in the present study, to Drs. Ning Zhou and Heather Schultz for their assistance in the preparation of the manuscript, and to Associate Editor, Dr. Robert Carlyon, and Dr. Christophe Micheyl and two anonymous reviewers for their comments on earlier versions of the manuscript. We acknowledge the support of the HEARing CRC, established and supported under the Australian Government's Cooperative Research Centers Program. This work was funded in part by National Natural Science Foundation of China (81070784 and 81070796), NIH/NIDCD grant (R15 DC009504), and the Linguistics Department, Macquarie University, Australia.

REFERENCES

- ARDOINT M, LORENZI C (2010) Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hear Res* 260:89–95
- ARDOINT MS, SHEFT S, FLEURIOT P, GARNIER S, LORENZI C (2010) Perception of temporal fine-structure cues in speech with minimal envelope cues for listeners with mild-to-moderate hearing loss. *Int J Audiol* 49:823–831
- BACON SP, GLEITMAN RM (1992) Modulation detection in subjects with relatively flat hearing losses. *J Speech Hear Res* 35:642–653
- BEDROSIAN E (1963) A product theorem for Hilbert transform. *Proc IEEE* 51:868–869
- BERNSTEIN JG, OXENHAM AJ (2006) The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss. *J Acoust Soc Am* 120:3929–3945
- BUSS E, HALL JW, GROSE JH (2004) Temporal fine-structure cues to speech and puretone modulation in observers with sensorineural hearing loss. *Ear Hear* 25:242–250
- CARLYON RP, DEEKS JM (2002) Limitations on rate discrimination. *J Acoust Soc Am* 112:1009–1025
- CARLYON RP, LONG CJ, DEEKS JM, MCKAY CM (2007) Concurrent sound segregation in electric and acoustic hearing. *J Assoc Res Otolaryngol* 8:119–133
- FIRSZT JB, HOLDEN LK, REEDER RM, SKINNER MW (2009) Speech recognition in cochlear implant recipients: comparisons of standard HiRes and HiRes120 sound processing. *Otol Neurotol* 30:146–152
- GILBERT G, LORENZI C (2006) The ability of listeners to use recovered envelope cues from speech fine structure. *J Acoust Soc Am* 119:2438–2444

- GNANSIA D, PEAN V, MEYER B, LORENZI C (2009) Effects of spectral smearing and temporal fine structure degradation on speech masking release. *J Acoust Soc Am* 125:4023–4033
- GREENWOOD DD (1990) A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87:2592–2605
- HAN D, LIU B, ZHOU N, CHEN XQ, KONG Y, LIU HH, ZHENG Y, XU L (2009) Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users. *Ear Hear* 30:169–177
- HEINZ MG, SWAMINATHAN J (2009) Quantifying envelope and fine-structure coding in auditory nerve responses to chimeric speech. *J Assoc Res Otolaryngol* 10:407–423
- HEINZ MG, SWAMINATHAN J, BOLEY JD, KALE S (2010) Across-fiber coding of temporal fine-structure: effects of noise-induced hearing loss on auditory-nerve responses. In: Lopez-Poveda EA, Palmer AR, Meddis R (eds) *The neurophysiological bases of auditory perception*. Springer Science + Business Media, New York, pp 621–630
- HOPKINS K, MOORE BCJ (2007) Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J Acoust Soc Am* 122:1055–1068
- HOPKINS K, MOORE BCJ (2010) The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects. *J Acoust Soc Am* 127:1595–1608
- HOPKINS K, MOORE BCJ, STONE MA (2008) Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J Acoust Soc Am* 123:1140–1153
- KALE S, HEINZ MG (2010) Envelope coding in auditory nerve fibers following noise-induced hearing loss. *J Assoc Res Otolaryngol* 11:657–673
- KONG YY, ZENG FG (2006) Temporal and spectral cues in Mandarin tone recognition. *J Acoust Soc Am* 120:2830–2840
- LIN MC (1988) The acoustical properties and perceptual characteristics of Mandarin tones. *Zhongguo Yuwen* 3:182–193
- LIU TC, HSU CJ, HORNG MJ (2000) Tone detection in Mandarin-speaking hearing-impaired subjects. *Int J Audiol* 39:106–109
- LORENZI C, GILBERT G, CARN H, GARNIER S, MOORE BCJ (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci USA* 103:18866–18869
- LORENZI C, DEBRUILLE L, GARNIER S, FLEURIOT P, MOORE BCJ (2009) Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *J Acoust Soc Am* 125:27–30
- MOORE BCJ (2008) The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J Assoc Res Otolaryngol* 9:399–406
- MOORE BCJ, MOORE GA (2003) Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hear Res* 182:153–163
- MOORE BCJ, PETERS RW (1992) Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J Acoust Soc Am* 91:2881–2893
- MOORE BCJ, GLASBERG BR, HOPKINS K (2006) Frequency discrimination of complex tones by hearing-impaired subjects: evidence for loss of ability to use temporal fine structure. *Hear Res* 222:16–27
- NIE KB, STICKNEY G, ZENG FG (2005) Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Trans Biomed Eng* 52:64–73
- RISS D, ARNOLDNER C, BAUMGARTNER WD, KAIDER A, HAMZAVI JS (2008) A new fine structure speech coding strategy: speech perception at a reduced number of channels. *Otol Neurotol* 29:784–788
- RISS D, ARNOLDNER C, REISS S, BAUMGARTNER WD, HAMZAVI JS (2009) 1-year results using the Opus speech processor with the fine structure speech coding strategy. *Acta Otolaryngol* 129:988–991
- SANTURETTE S, DAU T (2007) Binaural pitch perception in normal-hearing and hearing-impaired listeners. *Hear Res* 223:29–47
- SCHATZER R, KRENMAYR A, AU DK, KALS M, ZIERHOEFER C (2010) Temporal fine structure in cochlear implants: preliminary speech perception results in Cantonese-speaking implant users. *Acta Otolaryngol* 130:1031–1039
- SCHIMMEL SM, ATLAS LE (2005) Coherent envelope detection for modulation filtering of speech. *Proc ICASSP*, 221–224
- SHANNON RV (1983) Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res* 11:157–189
- SHEFT S, ARDOINT M, LORENZI C (2008) Speech identification based on temporal fine structure cues. *J Acoust Soc Am* 124:562–575
- SMITH ZM, DELGUTTE B, OXENHAM AJ (2002) Chimeric sounds reveal dichotomies in auditory perception. *Nature* 416:87–90
- STRELCEK O, DAU T (2009) Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J Acoust Soc Am* 125:3328–3345
- TURNER CW, SOUZA PE, FORGET LN (1995) Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *J Acoust Soc Am* 97:2568–2576
- TURNER CW, CHI SL, FLOCK S (1999) Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *J Speech Lang Hear Res* 42:773–784
- WANG W, ZHOU N, XU L (2011) Musical pitch and lexical tone perception with cochlear implants. *Int J Audiol* 50:270–278
- WEI CG, CAO KL, ZENG FG (2004) Mandarin tone recognition in cochlear implant subjects. *Hear Res* 197:87–95
- XU L, PFINGST BE (2003) Relative importance of temporal envelope and fine structure in lexical-tone perception. *J Acoust Soc Am* 114:3024–3027
- XU L, PFINGST BE (2008) Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hear Res* 242:132–140
- XU L, ZHOU N (2011) Tonal languages and cochlear implants. In: Zeng F-G, Popper AN, Fay RR (eds) *Auditory prostheses: new horizons*. Springer Science + Business Media, LLC, New York, (in press)
- XU L, TSAI Y, PFINGST BE (2002) Features of stimulation affecting tonal-speech perception: implications for cochlear prostheses. *J Acoust Soc Am* 112:247–258
- XU L, ZHOU N, BRASHEARS R, RIFE K (2008) Relative contributions of spectral and temporal cues in speech recognition in patients with sensorineural hearing loss. *J Otol* 3:84–91
- XU L, CHEN XW, LU HY, ZHOU N, WANG S, LIU QY, LI YX, ZHAO X, HAN DM (2011) Tone perception and production in pediatric cochlear implant users. *Acta Otolaryngol* 131:395–398
- ZENG FG (2002) Temporal pitch in electric hearing. *Hear Res* 174:101–106
- ZENG FG, NIE KB, LIU S, STICKNEY G, RIO ED, KONG YY, CHEN HB (2004) On the dichotomy in auditory perception between temporal envelope and fine structure cues. *J Acoust Soc Am* 116:1351–1354
- ZHOU N, ZHANG W, LEE CY, XU L (2008) Lexical tone recognition by an artificial neural network. *Ear Hear* 29:326–335