

# The Relative Contributions of Temporal Envelope and Fine Structure to Mandarin Lexical Tone Perception in Auditory Neuropathy Spectrum Disorder

Shuo Wang, Ruijuan Dong, Dongxin Liu, Luo Zhang and Li Xu

**Abstract** Previous studies have demonstrated that temporal envelope (E) is sufficient for speech perception, while fine structure (FS) is important for pitch perception for normal-hearing (NH) listeners. Listeners with sensorineural hearing loss (SNHL) have an impaired ability to use FS in lexical tone perception due to the reduced frequency resolution. Listeners with auditory neuropathy spectrum disorder (ANSD) may have deficits in temporal resolution. Little is known about how such deficits may impact their ability to use E and FS to perceive lexical tone, and whether it is the deficit in temporal resolution or frequency resolution that may lead to more detrimental effects on FS processing in pitch perception. Three experiments were conducted in the present study. Experiment I used the “auditory chimera” technique to investigate how SNHL and ANSD listeners would achieve lexical tone recognition using either the E or the FS cues. Experiment II tested their frequency resolution as measured with their psychophysical tuning curves (PTCs). Experiment III tested their temporal resolution as measured with the temporal gap detection (TGD) threshold. The results showed that the SNHL listeners had reduced frequency selectivity, but intact temporal resolution ability, while the ANSD

---

\*Parts of this paper may be published elsewhere.

---

S. Wang (✉) · R. Dong · D. Liu · L. Zhang  
Otolaryngology—Head & Neck Surgery, Beijing Tongren Hospital,  
Beijing Institute of Otolaryngology, Capital Medical University, Beijing, China  
e-mail: shannonwsh@aliyun.com

R. Dong  
e-mail: drj\_789@163.com

D. Liu  
e-mail: liudx891209@163.com

L. Zhang  
e-mail: dr.luozhang@139.com

L. Xu  
Communication Sciences and Disorders, Ohio University, Athens, OH, USA  
e-mail: xul@ohio.edu

© The Author(s) 2016

P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Advances in Experimental Medicine and Biology 894, DOI 10.1007/978-3-319-25474-6\_25

listeners had degraded temporal resolution ability, but intact frequency selectivity. In comparison with the SNHL listeners, the ANSD listeners had severely degraded ability to process the FS cues and thus their ability to perceive lexical tone mainly depended on the ability to use the E cues. These results suggested that, in comparison with the detrimental impact of the reduced frequency selectivity, the impaired temporal resolution may lead to more degraded FS processing in pitch perception.

**Keywords** Auditory neuropathy spectrum disorder (ANSD) · Sensorineural hearing loss (SNHL) · Auditory chimera · Lexical tone · Pitch · Temporal cues · Fine structure · Temporal envelope · Temporal resolution · Frequency resolution

## 1 Introduction

Acoustic signals can be decomposed into temporal envelope (E) and fine structure (FS). Previous studies have demonstrated that E is sufficient for speech perception (Shannon et al. 1995; Apoux et al. 2013), while FS is important for pitch perception (Smith et al. 2002) and lexical tone perception (Xu and Pfingst 2003). Mandarin Chinese is a tone language with four phonologically distinctive tones that are characterized by syllable-level fundamental frequency (F0) contour patterns. These pitch contours are described as high-level (tone 1), rising (tone 2), falling-rising (tone 3), and falling (tone 4). It has been demonstrated that FS is also important for lexical tone perception in listeners with sensorineural hearing loss (SNHL), but as their hearing loss becomes more severe, their ability to use FS becomes degraded so that their lexical tone perception relies increasingly on E rather than FS cues (Wang et al. 2011). The reduced frequency selectivity may underlie the impaired ability to process FS information for SNHL listeners. Auditory neuropathy spectrum disorder (ANSD) is an auditory disorder characterized by dys-synchrony of the auditory nerve firing but normal cochlear amplification function. Several studies have demonstrated that ANSD listeners have a dramatically impaired ability for processing temporal information (Zeng et al. 1999; Narné 2013). It is important to assess the ability of lexical tone perception for ANSD listeners and to investigate how they use FS and E cues to perceive lexical tone, as these results may shed light on the effects of frequency resolution and temporal resolution on FS processing in pitch perception.

## 2 Lexical Tone Perception

Experiment I was aimed to investigate the ability of lexical tone perception for both SNHL and ANSD listeners, and to assess the relative contributions of FS and E to their lexical tone perception using the “auditory chimera” technique (Smith et al. 2002; Xu and Pfingst 2003). As described in Wang et al. (2011), the chimeric tokens

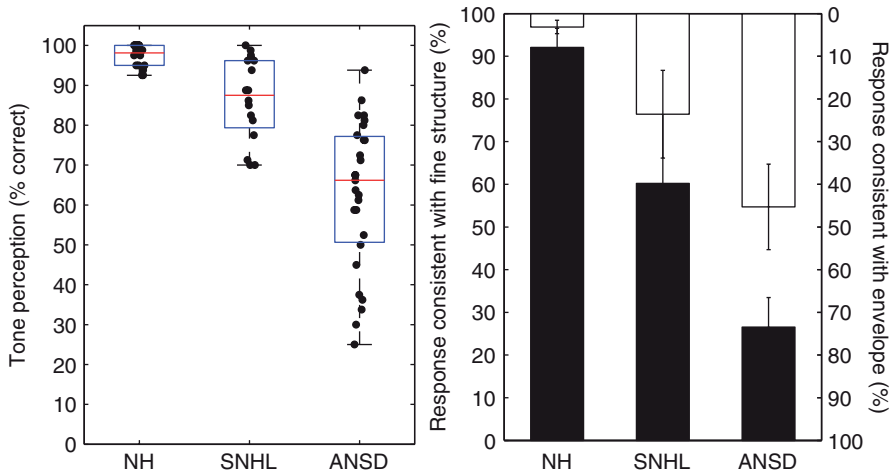
were generated in a condition with 16 channels. These 16 FIR band-pass filters with nearly rectangular response were equally spaced on a cochlear frequency map, ranging from 80 to 8820 Hz. In order to avoid the blurriness between E and FS in the chimeric stimuli, we adopted a lowpass filter (cut-off at 64 Hz) for extraction of the envelopes from the filters. By analysing the tone responses to the chimeric tone tokens, we could understand the competing roles of FS and E in tone perception in various groups of listeners.

## 2.1 *Methods*

Three groups of adult subjects, including 15 NH subjects (8 females and 7 males), 16 patients with SNHL (10 females and 6 males), and 27 patients with ANSD (9 females and 18 males), participated in the present study. The subjects with SNHL had hearing loss ranging from moderate to severe degree, and the ANSD subjects had hearing loss ranging from mild to severe. The original tone materials consisted of 10 sets of Chinese monosyllables with four tone patterns for each, resulting in a total of 40 commonly used Chinese words. These original words were recorded using an adult male and an adult female native Mandarin speaker. The tone tokens in which the durations of the four tones in each monosyllable were within a 5-ms precision were chosen as the original tone tokens. These tone tokens were processed through a 16-channel chimerizer in which the FS and E of any two different tone tokens of the same syllable were swapped. For example, two tokens of the same syllable but with different tone patterns were passed through 16 band-pass filters to split each sound into 16 channels. The output of each filter was then divided into its E and FS using a Hilbert transform. Then, the E of the output in each filter band was exchanged with the E in that band for the other token to produce the single-band chimeric wave. The single-band chimeric waves were summed across all channels to generate two chimeric stimuli. Finally, a total of 320 tokens were used in the tone perception test, including 80 original unprocessed tone tokens and 240 chimeric tone tokens. A four-alternative, forced-choice procedure was used for the tone perception test.

## 2.2 *Results*

Figure 1 (left panel) plots the accuracy of tone perception to the original tone tokens for the NH, SNHL, and ANSD subjects. The median tone perception performance was 97.2, 86.5, and 62.8% correct, respectively for the three groups. The performance slightly reduced for the subjects with SNHL, while the ANSD subjects had dramatic decreases in their tone perception performance with a very large individual variability. The right panel of Fig. 1 plots the mean percentages and standard deviations of tone perception responses to the chimeric tone tokens. The



**Fig. 1** Tone perception performance. *Left:* Boxplots show the tone perception performance with the original, unprocessed tone tokens for the subjects with NH, SNHL, and ANSD. The box represents the 25th to 75th percentiles and the horizontal line in the box represents the median. The whiskers represent the range of the data. *Right:* Mean and standard deviation of tone perception with the chimeric tone tokens for the subjects with NH, SNHL, and ANSD. *Black bars* represent responses consistent with FS (right ordinate) and *open bars* represent responses consistent with E (left ordinate)

subjects with NH relied on FS nearly entirely to perceive lexical tones regardless of E. In comparison with the NH subjects, the subjects with SNHL depended more on E cues to perceive tones, as they had impaired ability to use FS. In contrast to the finding in the NH and SNHL subjects, only 26.5% of the tone responses were consistent with the FS of the chimeric tone tokens for the ANSD subjects, indicating that the ANSD subjects had more severely impaired ability to use FS in tone perception. On the other hand, the ANSD subjects relied more heavily on E cues in perceiving lexical tones, indicating that the ability of the ANSD subjects to use E cues for lexical tone perception still remained at a reasonable level.

### 3 The Frequency Resolution and Temporal Resolution

Studies have shown that listeners with SNHL may have reduced frequency resolution (Kluk and Moore 2006), but probably normal temporal resolution (Glasberg et al. 1987). In contrast, listeners with ANSD may have a dramatically impaired ability for processing temporal information (Narne 2013), while their frequency selectivity may be close to normal (Vinay and Moore 2007). To clarify whether it is the deficit in temporal resolution or frequency resolution that may lead to more detrimental effects on FS processing in lexical tone perception, we also measured the frequency resolution using the mean  $Q_{10\text{ dB}}$  values of the psychophysical tuning

curves (PTCs), and the temporal resolution using the temporal gap detection (TGD) threshold for the SNHL and ANSD subjects.

### 3.1 *Methods*

Three groups of subjects who were recruited in Experiment I also participated in Experiment II in which we measured the psychophysical tuning curves (PTCs) of the auditory system using a fast method called SWPTC, developed by Sek and Moore (2011). Each subject was required to detect a sinusoidal signal, which was pulsed on and off repeatedly in the presence of a continuous noise masker. The PTCs at 500 and 1000 Hz in each ear were tested separately. The sinusoidal signals were presented at 15 dB sensation level (SL). The masker was a narrowband noise, slowly swept in frequency. The frequency at the tip of the PTC was estimated using a four-point moving average method (4-PMA), and the values of  $Q_{10\text{ dB}}$  (i.e., signal frequency divided by the PTC bandwidth at the point 10 dB above the minimum level) was used to assess the sharpness of the PTC. The greater the  $Q_{10\text{ dB}}$  values, the sharper the PTC.

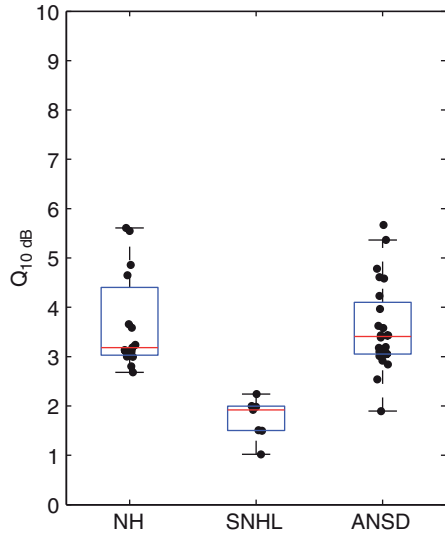
Experiment III tested the temporal gap detection (TGD) threshold for all subjects using a TGD program developed by Zeng et al. (1999). The test stimuli were generated using a broadband (from 20 to 14,000 Hz) white noise. The noise had a duration of 500 ms with 2.5-ms cosine-squared ramps. A silent gap was produced in the center of the target noise. The other two reference noises were uninterrupted. The TGD threshold of left and right ear was measured separately for each subject using a three-alternative, forced-choice procedure. The intensity level of the white noise was set at the most comfortable loudness level for all subjects.

### 3.2 *Results*

The values of  $Q_{10\text{ dB}}$  were determined in all 15 NH subjects, 7 of the 16 SNHL subjects, and 24 of the 27 subjects with ANSD. Note that for 9 of the 16 subjects with SNHL, the PTCs were too broad so that the  $Q_{10\text{ dB}}$  values could not be determined. Since no significant differences were found between the  $Q_{10\text{ dB}}$  values at 500 and 1000 Hz, the  $Q_{10\text{ dB}}$  values averaged across the two test frequencies are plotted in Fig. 2. Statistically significant differences of the  $Q_{10\text{ dB}}$  values were found between the subjects with NH and those with SNHL and between the subjects with ANSD and those with SNHL, while no statistically significant differences were present between the subjects with NH and those with ANSD. This indicates that a majority of the subjects with ANSD showed close-to-normal frequency resolution, whereas the SNHL subjects had poor frequency resolution in the present study.

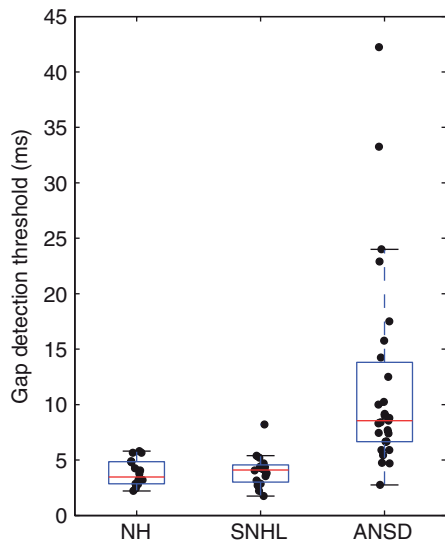
Figure 3 shows the averaged TGD thresholds of both ears for the three groups of subjects. Welch's test in a one-way ANOVA found significant differences of

**Fig. 2** The  $Q_{10\text{dB}}$  values of psychophysical tuning curves for subjects with NH, with SNHL, and with ANSD. The box represents the 25–75th percentiles and the horizontal line in the box represents the median. The whiskers represent the range of the data. Data points outside of the whiskers are outliers. Note that for 9 of the 16 subjects with SNHL, the PTCs were too broad so that the  $Q_{10\text{dB}}$  values could not be determined



TGD thresholds among the three groups of subjects [ $F(2,56)=9.9, p<0.001$ ]. Post hoc Tamhane’ T2 correction test analysis indicated that the mean TGD thresholds for both ears was significantly higher in the listeners with ANSD (11.9 ms) than in the listeners with SNHL (4.0 ms;  $p<0.001$ ) and in the NH (3.9 ms;  $p<0.001$ ) listeners; while the subjects with SNHL had the TGD thresholds close to normal. It is notable that the variability in TGD thresholds for the ANSD subjects was very

**Fig. 3** Temporal gap detection thresholds for subjects with NH, SNHL, and ANSD. The box represents the 25th to 75th percentiles and the horizontal line in the box represents the median. The whiskers represent the range of the data. Data points outside of the whiskers are outliers



large, ranging from normal limits to 10 times the norm. This result indicates that the SNHL subjects had normal temporal resolution as long as the audibility was compensated, but a majority of the subjects with ANSD had poor temporal resolution in the present study.

## 4 Discussion and Conclusions

The fine structure (FS) cues in the speech signals play a dominant role in lexical tone perception (Xu and Pfingst, 2003; Wang et al. 2011). The temporal envelope (E) cues contribute to tone perception when the FS cues are not available (Xu et al. 2002; Kong and Zeng 2006; Xu and Zhou 2011). Experiment I of the present study demonstrated that the listeners with ANSD had great deficits in using the FS cues for lexical tone perception, which resulted in great difficulties in perceiving lexical tones for the ANSD listeners. Even though the ability of the listeners with ANSD to use E cues for lexical tone perception remained at a reasonable level, this level of performance was much lower in comparison with that of the NH listeners. The NH listeners could utilize the FS cues and achieve perfect tone perception performance. The tone perception performance in the listeners with ANSD was also lower than that in the listeners with SNHL who were still able to use the FS cues for tone perception to some extent. An earlier report has shown that the ability to use the FS cues for tone perception is negatively correlated with the degree of hearing loss in listeners with SNHL (Wang et al. 2011).

Experiments II and III evaluated the frequency resolution and temporal resolution, respectively for the subjects with SNHL and ANSD. Consistent with previous studies, the listeners with SNHL in the present study had poor frequency resolution whereas a majority of the listeners with ANSD had normal frequency resolution. On the other hand, a majority of the listeners with ANSD had poor temporal resolution whereas the listeners with SNHL had normal temporal resolution. This may imply that poor temporal resolution rather than the frequency resolution exerts the major detrimental effects on FS cue processing for pitch perception.

**Acknowledgements** We thank all the subjects for participating in this study and our colleagues at the Clinical Audiology Center in Beijing Tongren Hospital for helping recruit the hearing-impaired subjects. This work was funded in part by grants from the National Natural Science Foundation of China (81200754), the 2012 Beijing Nova Program (Z121107002512033), the Capital Health Research and Development of Special from the Beijing Municipal Health Bureau (2011-1017-04).

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

## References

- Apoux F, Yoho SE, Youngdahl CL, Healy EW (2013) Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners. *J Acoust Soc Am* 134:2205–2212
- Glasberg BR, Moore BCJ, Bacon SP (1987) Gap detection and masking in hearing-impaired and normal-hearing subjects. *J Acoust Soc Am* 81:1546–1556
- Kluk K, Moore BCJ (2006) Detecting dead regions using psychophysical tuning curves: a comparison of simultaneous and forward masking. *Int J Audiol* 45:463–476
- Kong YY, Zeng FG (2006) Temporal and spectral cues in Mandarin tone recognition. *J Acoust Soc Am* 120:2830–2840
- Name VK (2013) Temporal processing and speech perception in noise by listeners with auditory neuropathy. *Plos ONE* 8:1–11
- Sek A, Moore BCJ (2011) Implementation of a fast method for measuring psychophysical tuning curves. *Int J Audiol* 50:237–242
- Shannon RV, Zeng FG, Wygonski J (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304
- Smith ZM, Delgutte B, Oxenham AJ (2002) Chimeric sounds reveal dichotomies in auditory perception. *Nature* 416:87–90
- Vinay, Moore BCJ (2007) Ten(HL)-test results and psychophysical tuning curves for subjects with auditory neuropathy. *Int J Audiol* 46:39–46
- Wang S, Mannell R, Xu L (2011) Relative contributions of temporal envelope and fine structure cues to lexical tone recognition in hearing-impaired listeners. *J Assoc Res Otolaryngol* 12:783–794
- 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, 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, pp 341–364
- 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
- Zeng FG, Oba S, Grade S, Sininger Y, Starr A (1999) Temporal and speech processing deficits in auditory neuropathy. *Neuroreport* 10:3429–3435