



Voice Pitch Elicited Frequency Following Response in Chinese Elderlies

Shuo Wang^{1*}, Jiong Hu², Ruijuan Dong¹, Dongxin Liu¹, Jing Chen¹, Gabriella Musacchia² and Bo Liu^{1*}

¹ Otolaryngology – Head & Neck Surgery, Beijing Tongren Hospital, Beijing Institute of Otolaryngology, Capital Medical University, Beijing, China, ² Department of Speech-Language Pathology and Audiology, University of the Pacific, San Francisco, CA, USA

Background: Perceptual and electrophysiological studies have found reduced speech discrimination in quiet and noisy environment, delayed neural timing, decreased neural synchrony, and decreased temporal processing ability in elderlies, even those with normal hearing. However, recent studies have also demonstrated that language experience and auditory training enhance the temporal dynamics of sound encoding in the auditory brainstem response (ABR). The purpose of this study was to explore the pitch processing ability at the brainstem level in an aging population that has a tonal language background.

Method: Mandarin speaking younger ($n = 12$) and older ($n = 12$) adults were recruited for this study. All participants had normal audiometric test results and normal suprathreshold click-evoked ABR. To record frequency following responses (FFRs) elicited by Mandarin lexical tones, two Mandarin Chinese syllables with different fundamental frequency pitch contours (Flat Tone and Falling Tone) were presented at 70 dB SPL. Fundamental frequencies (f_0) of both the stimulus and the responses were extracted and compared to individual brainstem responses. Two indices were used to examine different aspects of pitch processing ability at the brainstem level: Pitch Strength and Pitch Correlation.

Results: Lexical tone elicited FFR were overall weaker in the older adult group compared to their younger adult counterpart. Measured by Pitch Strength and Pitch Correlation, statistically significant group differences were only found when the tone with a falling f_0 (Falling Tone) were used as the stimulus.

Conclusion: Results of this study demonstrated that in a tonal language speaking population, pitch processing ability at the brainstem level of older adults are not as strong and robust as their younger counterparts. Findings of this study are consistent with previous reports on brainstem responses of older adults whose native language is English. On the other hand, lexical tone elicited FFRs have been shown to correlate with the length of language exposure. Older adults' degraded responses in our study may also be due to that, the Mandarin speaking older adults' long term exposure somewhat counteracted the negative impact on aging and helped maintain, or at least reduced, the degradation rate in their temporal processing capacity at the brainstem level.

Keywords: aging-related, frequency following responses, Mandarin lexical tone, pitch processing, temporal processing

OPEN ACCESS

Edited by:

Rodrigo Orlando Kuljiš,
University of Miami School of
Medicine, USA

Reviewed by:

Kristina Arousseau,
Centre de Recherche de l'Institut
Universitaire de Gériatrie de Montréal
(CRIUGM), Canada
Jingwen Niu,
Temple University, USA

*Correspondence:

Shuo Wang
shannonwsh@allyun.com
Bo Liu
trliubo@139.com

Received: 15 March 2016

Accepted: 14 November 2016

Published: 29 November 2016

Citation:

Wang S, Hu J, Dong R, Liu D,
Chen J, Musacchia G and Liu B
(2016) Voice Pitch Elicited Frequency
Following Response in Chinese
Elderlies.

Front. Aging Neurosci. 8:286.
doi: 10.3389/fnagi.2016.00286

INTRODUCTION

Many perceptual studies have shown that older adults with and without hearing loss may have difficulty with speech identification, especially in the presence of background noise (Dubno et al., 1984; Souza and Boike, 2006). Psychoacoustic studies have also found that in comparison with younger listeners, older adults with normal hearing may have more difficulty in using time-varying acoustic cues due to the deficits in temporal resolution (Moore and Glasberg, 1988; Moore et al., 1992; Tremblay et al., 2002, 2003; Harkrider et al., 2005). Temporal cues, including temporal envelope and fine structure, play dominant roles in speech perception and pitch perception (Rosen, 1992; Shannon et al., 1995; Smith et al., 2002; Xu and Pfingst, 2003; Wang S. et al., 2011). Thus, it has been proposed that speech identification difficulties suffered by older adults even with normal hearing may be attributed to their degraded temporal processing ability.

Neurophysiological and electrophysiological studies have also found evidence of delayed neural timing, decreased neural synchrony, and decreased temporal processing ability with age (Walton et al., 1998; Wang M. et al., 2011; Parbery-Clark et al., 2012). Tremblay et al. (2003) measured the cortical auditory evoked potentials (CAEPs) and observed that the N1 and P2 peak latencies recorded from older listeners were prolonged when they listened to the synthetic speech tokens. They proposed that the auditory neurons in older adults may need more time to recover from the initial excitation before firing again, so older adults may have decreased neural timing during transmission and processing auditory information. Age-related neural synchrony changes were not only observed by the use of auditory cortical potentials, but also reported by measuring auditory brainstem responses (ABRs) (Jeger and Hall, 1980; Konrad-Martin et al., 2012) and speech evoked ABRs (cABR) (Anderson et al., 2012). These studies reported common findings that the aging process reduces the amplitudes and increases the latency of ABR waves, suggesting degraded time-locked neural activity in the auditory brainstem of older listeners. Furthermore, Anderson et al. (2012) recorded ABRs to a speech syllable /da/ in both normal-hearing younger and older adults. They found that older adults had decreased phase-locking and smaller response magnitudes in comparison to younger adults, indicating that older adults have already had the deficits in temporal precision in the subcortical encoding of a speech sound.

The frequency following response (FFR) is a scalp-recorded measure that is dependent on the phase-locked brainstem activity in a population of neurons that phase-lock to the periodicity features in a sound (Krishnan and Parkinson, 2000; Krishnan, 2002; Krishnan et al., 2004; Aiken and Picton, 2008; Clinard et al., 2010). In a study, Krishnan et al. (2004) recorded human FFRs elicited by different Mandarin lexical tones. Mandarin lexical tones are characterized by syllable-level fundamental frequency (f_0) contour patterns, including changes in pitch height and direction of the pitch contour. It was clearly shown that the phase-locked FFR activity carrying pitch-relevant information follows the pitch changes presented

in each stimulus, suggesting that the representation of pitch changes in Mandarin tones are based on the temporal discharge patterns of phase-locked neural activity in the brainstem pathways. As older adults, even those with normal hearing, have shown degraded temporal processing with age-related distorted cortical neural activity, it will be interesting to see how the neural representation of pitch, as reflected by Mandarin tone elicited FFR, in older listeners preserve the pitch-relevant information in comparison with their younger counterparts. By this date, few studies have investigated how aging may impact the neural representation of pitch in the brainstem using the FFR recording measure, especially in populations that speak a tonal language. Clinard et al. (2010) examined the relationship between behavioral performance, as measured by a frequency discrimination task, and physiological performance, as measured by FFRs, using the same 500 and 1000 Hz tone bursts. They found that as the age of listeners increased, pitch discrimination ability and neural representation of frequency both decreased, but no significant relationship was found between each other.

The present study was aimed at investigating the neural representation of Mandarin lexical tones in the brainstem of Mandarin-speaking older adults with normal hearing using the FFR technique. It is hypothesized that in comparison with the younger adults, older adults may have less faithful representations of the f_0 contours in lexical tones and decreased robustness of neural phase-locking.

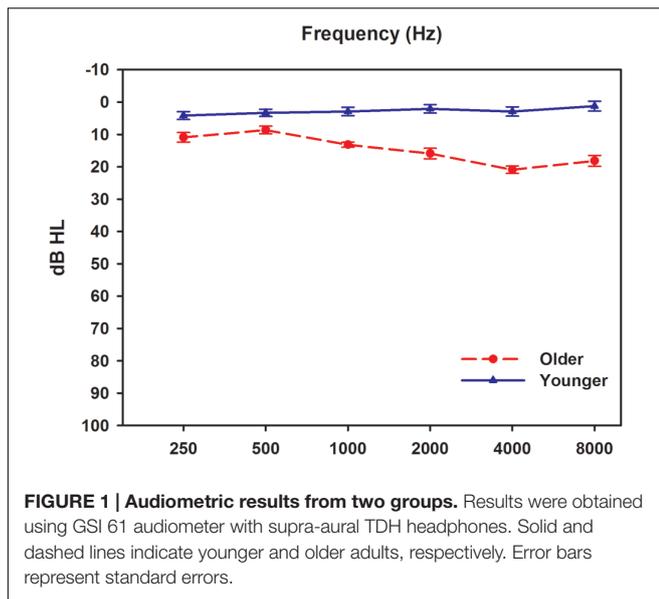
MATERIALS AND METHODS

Ethics Statement

Experimental protocols and procedures used in this study were approved by the Beijing Institute of Otolaryngology Institutional Review Board.

Participants

Twelve younger participants and 12 older participants were recruited in this study. The younger participants were aged 22–25 years old (mean \pm SD = 22.9 \pm 0.8), while the older participants were aged 61–69 years old (mean \pm SD = 62.3 \pm 2.8). All participants, 13 males and 11 females, were native speakers of Mandarin Chinese. Before each test session, all participants underwent a complete audiological assessment evaluation to fulfill the requirement of hearing sensitivity \leq 25 dB HL at octave frequencies from 125 to 8000 Hz (**Figure 1**). The audiological assessments included otoscopy, pure tone audiometry (using a GSI-61 clinical audiometer), tympanometry (using a GSI 33 Middle ear analyzer), and ABRs test. The ABRs were recorded using the Smart EP module in the IHS System. The click stimuli were presented monaurally at an intensity starting at 80 dB nHL with a rate of 11.1/s. A total of 2,048 sweeps were recorded using two polarities, rarefaction and condensation. All participants had within normal range wave V latencies (5.70–6.08 ms; 5.88 \pm 0.16 ms) in the click-evoked ABR tests. Experimental protocols and procedures used in this study were approved by the Beijing Tongren Hospital Institutional Review



Board and the Ethical Committee. All participants were informed of the purpose and content of the study, and signed an informed consent form. All data collection was conducted in a sound-proof booth at the China Rehabilitation Research Center for Deaf Children.

Recordings

Two monosyllabic Mandarin Chinese syllables with contrast pitch contours (Flat Tone, ‘clothes’ and Falling Tone, ‘meaning’) were prepared. These voice samples then underwent a temporal normalization to a unified duration of 250 ms with a rising/falling time of 10 ms. The final stimulus tokens of the two Mandarin syllables had the frequency ranges containing their f_0 contours ranging from 163 to 180 Hz for Flat Tone, and 105–156 Hz for Falling Tone. Stimulus tokens were controlled and delivered via Stim2 system. They were presented monaurally to the participant’s right ear through a custom-built electromagnetically shielded insert earphone modified from an ER3-A earphone (Biologic Systems, Corp., Mundelein, IL, USA) at 70 dB SPL. Each token was presented for blocks of 2000 accepted sweeps, where artifact rejection criteria was set at $\pm 25 \mu\text{V}$. The inter-stimulus interval between adjacent stimulus tokens was set at 45 ms and the tokens were presented in a random order within and across participants. A control condition (sound tube occluded and removed from the participant’s ear) was conducted at the end of each recording session to ensure that the stimulus artifact was eliminated from the recordings.

Three gold-plated recording electrodes were applied to each participant on the forehead (Fpz, non-inverting), right mastoid (M2, inverting), and left mastoid (M1, ground). All electrode impedances were maintained under 3 kOhms at 10 Hz. Continuous recordings, collected by NeuroScan ACQUIRE 4.4., were amplified using the SynAmp 2 (Compumedics, Compumedics, VIC, Australia) system with a 24-bit resolution and a least significant bit of 0.15 nV.

Analyses

All the data was analyzed using MatLab (Mathworks, Natick, MA, USA) and EEGLab (Swartz Center for Computational Neuroscience, San Diego, CA, USA). Recorded EEGs were filtered through a 100–1500 Hz band-pass filter with a linear-phase of 500 poles. The artifact rejection criteria was set at $\pm 25 \mu\text{V}$. After averaging the remaining EEG sweeps, a cross-correlation between the stimulus and recorded waveforms was carried out to identify the time shift data point that had the maximum cross-correlation value between the 2–10 ms response window. Starting from this data point, a segment of 250 ms was extracted from the averaged data.

A fundamental frequency extraction algorithm based on narrowband spectrogram (Jeng et al., 2011a) was used to extract the pitch information of the sampled EEG signal. In short, all stimulus tokens and recordings were first segmented using a 50 ms sliding Hanning window with a step size of 1 ms. Within each windowed segment, the frequency corresponding to the maximal peak of the spectral amplitudes within a predefined frequency range is determined as the f_0 frequency for that segment. This procedure was repeated for all windowed segments.

Two indices, namely Pitch Correlation and Pitch Strength, were used to quantify the pitch-tracking accuracy and phase-locking magnitude of the responses. They each represent a different aspect of pitch processing in the brainstem. Pitch Correlation describes how close the f_0 of the responses are correlated to that of the stimulus. Using the above mentioned f_0 extracting technique, the f_0 of both the responses and stimulus were extracted and the correlation coefficient between them was denoted as the Pitch Correlation index. It describes how precise the brainstem’s response tracks the pitch change in the stimulus. Pitch Strength was used to quantify the phase-locking magnitude of the responses (Krishnan et al., 2005; Jeng et al., 2010; Xu and Jeng, 2011). This index was derived from an autocorrelation function that allowed the measurement of overall periodicity of a sampled signal. Specifically, each recording (i.e., the entire 250 ms of a recording) was multiplied by a duplication of itself with increasing time shifts. For each time shift, an autocorrelation value was calculated and expressed between -1 and 1 . F_0 was calculated using the output of the autocorrelation function by finding the time shift that yielded the maximum autocorrelation value and taking the inverse of that time shift. Because the f_0 contour of the stimulus token used in this study fell within the frequency range of 100–200 Hz, the time shifts were limited to 5–10 ms when searching for the location of the maximum peak in the autocorrelation output. Pitch Strength was calculated using the autocorrelation function by finding the peak-to-trough amplitude starting from the maximum positive peak (within the 5–10 ms time shifts) to the following negative trough in the normalized autocorrelation output. Between the two indices, Pitch Correlations emphasize how faithful the brainstem response is to the stimulus, where Pitch Strength focuses on the magnitude of the response itself. A Student’s t -test was carried on each index for comparisons between the younger and older adult groups for statistical analysis.

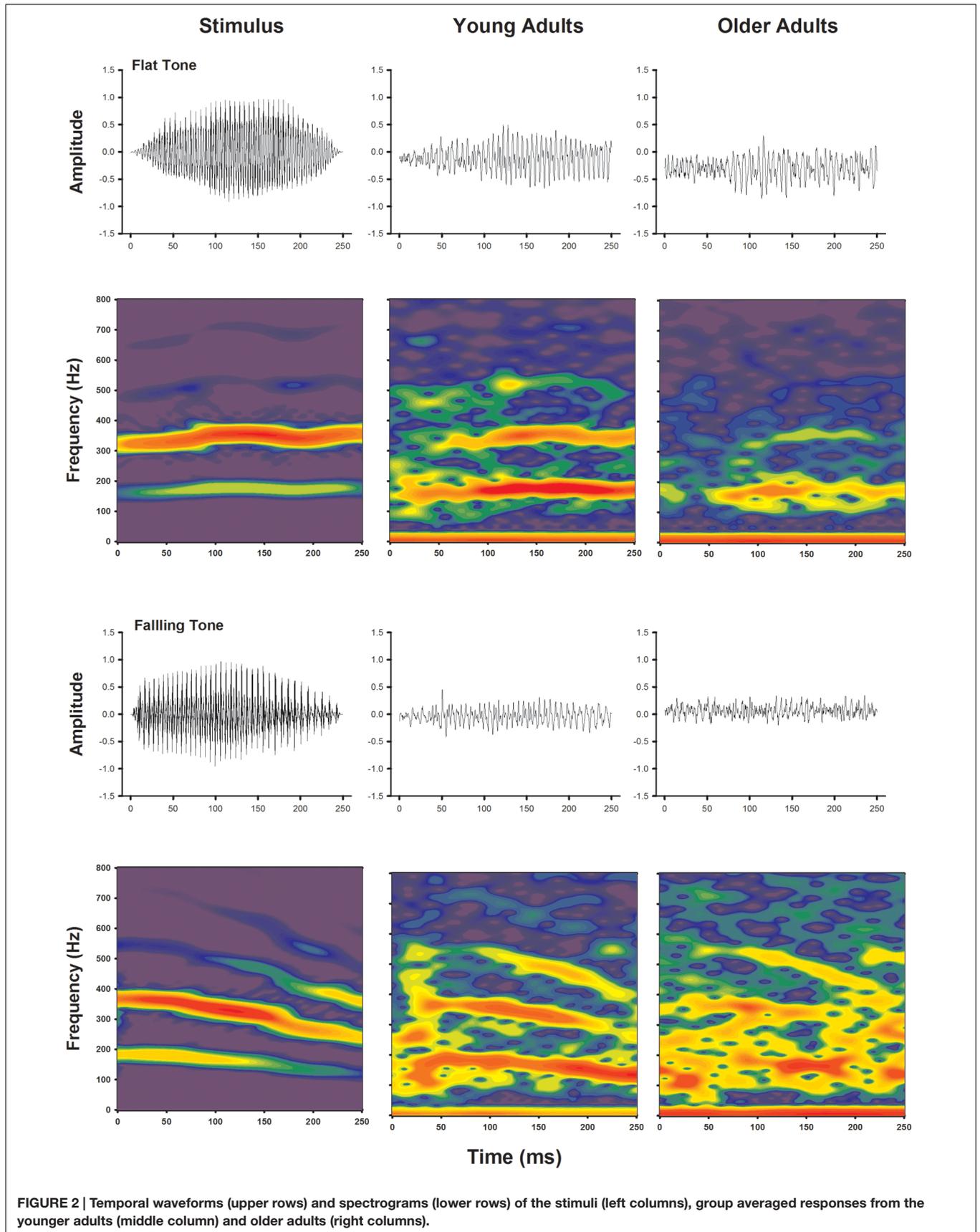
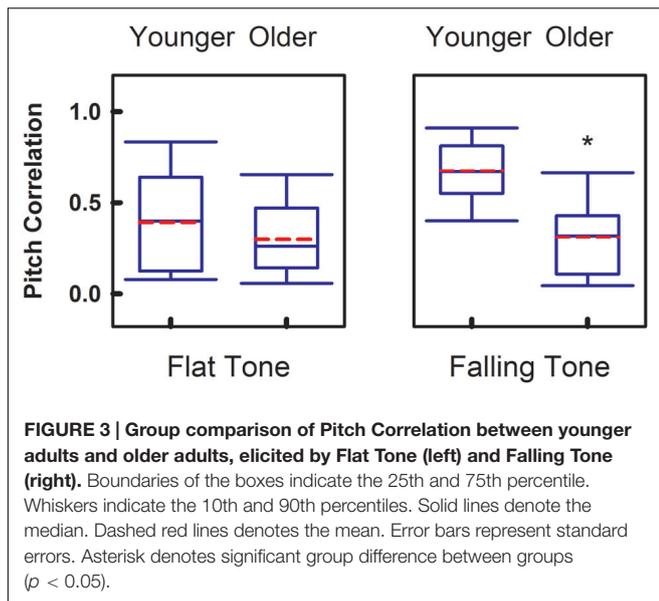


FIGURE 2 | Temporal waveforms (upper rows) and spectrograms (lower rows) of the stimuli (left columns), group averaged responses from the younger adults (middle column) and older adults (right columns).



RESULTS

Frequency following response to voice pitch was visualized by plotting waveforms and spectral energies of the recordings as a function of time. Waveforms and spectrograms of the voice pitch elicited FFR obtained from grand averages of both younger and older adults are shown in **Figure 2**, for Flat Tone and Falling Tone, respectively. Waveforms obtained from both age groups demonstrated clear periodicity and change of periodicity that mimics the fundamental frequencies of the stimuli and their temporal contours. Similarly, spectrograms of the recordings of both age groups also showed clear FFR energy that followed that of the stimulus. The FFR energy following the harmonics was not as clear.

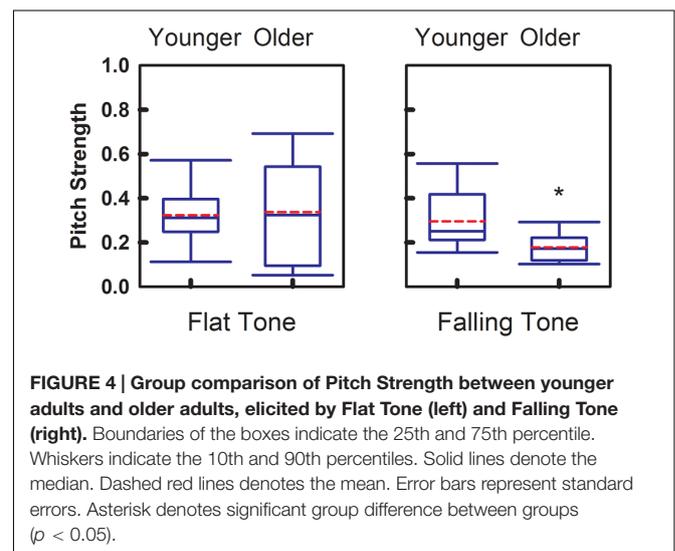
To compare the overall faithfulness of the neural responses in the brainstem to the stimulus, **Figure 3** demonstrates the group comparisons of Pitch Correlations between age groups for both Flat Tone and Falling Tone. Significant group difference ($p < 0.05$) of Pitch Correlation was found in Falling Tone, which has a falling pitch contour, between the younger ($r = 0.67 \pm 0.16$) and the older group ($r = 0.31 \pm 0.21$). Pitch Correlations obtained from the younger group ($r = 0.39 \pm 0.27$) were greater, yet not significantly different, from that of the older group ($r = 0.30 \pm 0.20$) when Flat Tone was used as the stimulus.

Magnitude of the phase-locking mechanism at the brainstem level is represented by the Pitch Strength index. **Figure 4** demonstrates the group comparisons of Pitch Strength obtained from the two age groups. Similar to Pitch Correlation, Falling Tone yielded significantly greater ($p = 0.017$) Pitch Strength in the younger adult group ($r = 0.30 \pm 0.14$) than that obtained from the older group ($r = 0.18 \pm 0.07$). Pitch Strength obtained from the younger group ($r = 0.32 \pm 0.14$) for Flat Tone was not significantly different from that from the older group ($r = 0.33 \pm 0.25$).

DISCUSSION

Overall, results of this study demonstrated that in a tonal language speaking population, the pitch processing ability at the brainstem level in older adults are not as strong and robust as their younger counterparts. This is not surprising given that the temporal processing capacity of older adults are generally degraded due to aging (Wingfield et al., 1999; Tremblay et al., 2002). Previous electrophysiological reports have also provided objective evidence that the aging process affects the neural processing functions along the auditory pathway (Jerger and Hall, 1980). For example, compared to younger adults, older adults have prolonged latencies in click evoked ABR when a higher presentation rate was used (Burkard and Sims, 2001). As to speech evoked brainstem responses, such as Consonant-Vowel (CV) evoked cABR, evidence has found that normal hearing older adults have diminished neural precision. (Anderson et al., 2012, 2013a,b).

In this study, Mandarin voice pitch elicited FFR obtained from younger Mandarin speaking adults were significantly stronger than that of older adults in Falling Tone, but not Flat Tone, which has a flat f_0 contour. This is consistent with previous Mandarin tone elicited FFR reports (Krishnan et al., 2004; Jeng et al., 2010, 2011b) where in normal hearing young adults, Falling Tone usually elicited stronger FFR than Flat Tone. It was suggested that the change of f_0 contour of Falling Tone, which started around 180 Hz and falls to around 120 Hz in just 250 ms, requires additional neural encoding in the brainstem area to track such rapid pitch change. The results from our study may suggest that such ability to encode rapid spectral change in voice pitch, e.g., Mandarin Falling Tone, may have degraded in the older group, while the ability to detect and follow relatively flat pitch, e.g., Flat Tone, remained fairly stable in the older group, when compared to their younger counterparts. This finding is also consistent with previous reports on cABR in older populations that did not speak tonal languages (Anderson et al., 2012, 2013a,b). In those studies, older adults' responses were notably poorer during rapid spectral



transitions, such as transition between a transient consonant and a steady-state vowel, which is similar to the results of our study where change of spectral slope made it harder for older adults to track.

On the other hand, auditory exposure and trainings have been proven to have a positive impact on behavioral auditory processing capacity (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Strait and Kraus, 2014), as well as enhance auditory electrophysiological responses, including brainstem and cortical responses (Marques et al., 2007; Musacchia et al., 2007, 2008; Marie et al., 2010; Chobert et al., 2011). Specifically for scalp reported FFR, reports have shown that due to neural plasticity at the brainstem level, short-term and long-term auditory exposure, including music training and language exposure, could enhance FFR responses in different age groups with different language backgrounds (Krishnan, 2002; Krishnan et al., 2004; Musacchia et al., 2007; Jeng et al., 2010, 2011b). In this study, as our older adult groups are on average nearly 40 years older than their younger counterparts, the length of their exposure to a tonal language are notably longer too.

From this perspective, since both aging and length of language exposure play a part in reshaping the neural pathway, it may be difficult to discern which one is the dominating factor here. However, reports from cABR in older adults have shown that without lifelong exposure to a language that is acoustically rich in pitch information, older adults almost always showed significantly declined brainstem responses to pitch (Anderson et al., 2012, 2013b). In our study, Falling Tone elicited significantly stronger responses in the younger adults than that of older adults, while Flat Tone did not. Thus, it may be possible that the Mandarin speaking older adults' long-term exposure somewhat counteracted the negative impact on aging and helped maintain, or at least reduced, the degradation rate in their temporal processing capacity at the brainstem level.

Findings from this study could provide insight to current research and clinical applications of auditory training. For example, it is well-established that short-term and long-term auditory training, especially music training, can provide significant enhancement in pitch processing, speech in noise performance and other behavioral responses (Parbery-Clark et al., 2011, 2012; Alain et al., 2014), as evidenced by electrophysiological studies (Musacchia et al., 2007, 2008; Strait et al., 2012; Anderson et al., 2013a). Similar to lifelong music training, language exposure across the life span could be

as analogous to a long-term auditory training program. The maintenance of excitatory and inhibitory subcortical neural networks (Parbery-Clark et al., 2012) for the perception of tonal information is essential for Mandarin speaking populations and may have been executed during this lifelong auditory training. It may suggest that long-term auditory training can improve and enhance neural phase-locking ability at the brainstem level, which will be beneficial for the pitch encoding capacity.

For future directions, we will continue monitoring the status and change of pitch coding mechanism at the brainstem level with the aging process. With a better understanding of the aging factor in neural encoding at brainstem and the associated plasticity characteristics, we can associate it with behavioral studies that focus on temporal processing in the same population and broaden our understanding of the aging process as a whole.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: SW, JH, and BL. Performed the experiments: SW, RD, DL, and JC. Analyzed the data: SW, JH, and GM. Contributed reagents/materials/analysis tools: SW and JH. Wrote the paper: SW, JH, and BL.

FUNDING

This work was funded in part by grants from the National Natural Science Foundation of China (81200754, 81371103), the 2012 Beijing Nova Program (Z121107002512033), Beijing Natural Science Foundation (7122034, 7154190), the Capital Health Research and Development of Special from the Beijing Municipal Health Bureau (2011-1017-04), and the Promotion Grant for High-level Scientific and Technological Elites in Medical Science from the Beijing Municipal Health Bureau (2015-3-012).

ACKNOWLEDGMENTS

We thank all the subjects who participated in this study and our colleagues at the Beijing Tongren Hospital for recruiting the subjects. We also acknowledge the China Rehabilitation Research Center for Deaf Children for providing the sound-proof booth and NeuroScan for collecting data.

REFERENCES

- Aiken, S., and Picton, T. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47. doi: 10.1016/j.heares.2008.08.004
- Alain, C., Zendel, B. R., Hutka, S., and Bidelman, G. M. (2014). Turning down the noise: the benefit of musical training on the aging auditory brain. *Hear. Res.* 308, 162–173. doi: 10.1016/j.heares.2013.06.008
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2012). Aging affects neural precision of speech encoding. *J. Neurosci.* 32, 14156–14164. doi: 10.1523/JNEUROSCI.2176-12.2012
- Anderson, S., White-Schwoch, T., Parbery-Clark, A., and Kraus, N. (2013a). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear. Res.* 300, 18–32. doi: 10.1016/j.heares.2013.03.006
- Anderson, S., White-Schwoch, T., Parbery-Clark, A., and Kraus, N. (2013b). Reversal of age-related neural timing delays with training. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4357–4362. doi: 10.1073/pnas.1213555110
- Burkard, R. F., and Sims, D. (2001). The human auditory brainstem responses to high click rates: aging effects. *Am. J. Audiol.* 10, 53–61. doi: 10.1044/1059-0889(2001/008)
- Chobert, J., Marie, C., François, C., Schön, D., and Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *J. Cogn. Neurosci.* 23, 3874–3887. doi: 10.1162/jocn_a_00088
- Clinard, C. G., Tremblay, K. L., and Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: evidence from

- human frequency-following response recordings. *Hear. Res.* 264, 48–55. doi: 10.1016/j.heares.2009.11.010
- Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87–96. doi: 10.1121/1.391011
- Harkrider, A. W., Plyler, P. N., and Hedrick, M. S. (2005). Effects of age and spectral shaping on perception and neural representation of stop consonant stimuli. *Clin. Neurophysiol.* 116, 2153–2164. doi: 10.1016/j.clinph.2005.05.016
- Jeng, F. C., Hu, J., Dickman, B. M., Lin, C. Y., and Lin, C. D. (2011a). Evaluation of two algorithms for detecting human frequency-following responses to voice pitch. *Int. J. Audiol.* 50, 14–26. doi: 10.3109/14992027.2011.582164
- Jeng, F. C., Hu, J., Dickman, B. M., Montgomery-Reagan, K., Tong, M., Wu, G., et al. (2011b). Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear. Hear.* 32, 699–707. doi: 10.1097/AUD.0b013e31821cc0df
- Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., and Li, X. (2010). Early maturation of frequency-following responses to voice pitch in infants with normal hearing. *Percept. Mot. Skills.* 111, 765–784. doi: 10.2466/10.22.24.PMS.111.6.765-784
- Jerger, J., and Hall, J. (1980). Effects of age and sex on auditory brainstem response. *Arch. Otolaryngol.* 106, 387–391. doi: 10.1001/archotol.1980.00790310011003
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12, 125–143.
- Konrad-Martin, D., Dille, M. F., McMillan, G., Griest, S., McDermott, D., Fausti, S. A., et al. (2012). Age-related changes in the auditory brainstem response. *J. Am. Acad. Audiol.* 23, 18–35.
- Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hear. Res.* 166, 192–201. doi: 10.1016/S0378-5955(02)00327-1
- Krishnan, A., and Parkinson, J. (2000). Human frequency-following response: representation of tonal sweeps. *Audiol. Neurootol.* 5, 312–321. doi: 10.1159/000013897
- Krishnan, A., Xu, Y., Gandour, J. T., and Cariani, P. A. (2004). Human frequency following response: representation of pitch contours in Chinese tones. *Hear. Res.* 189, 1–12. doi: 10.1016/S0378-5955(03)00402-7
- Krishnan, A., Xu, Y., Gandour, J. T., and Cariani, P. A. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.* 25, 161–168. doi: 10.1016/j.cogbrainres.2005.05.004
- Marie, C., Magne, C., and Besson, M. (2010). Musicians and the metric structures of words. *J. Cogn. Neurosci.* 23, 294–305. doi: 10.1162/jocn.2010.21413
- Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453
- Micéyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47. doi: 10.1016/j.heares.2006.05.004
- Moore, B. C. J., and Glasberg, B. R. (1988). Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *J. Acoust. Soc. Am.* 83, 1093–1101. doi: 10.1121/1.396054
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1992). Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss. *J. Acoust. Soc. Am.* 92:1923. doi: 10.1121/1.405240
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898. doi: 10.1073/pnas.0701498104
- Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42. doi: 10.1016/j.heares.2008.04.013
- Parbery-Clark, A., Anderson, S., Hittner, E., and Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiol. Aging* 33:1483.e1-4. doi: 10.1016/j.neurobiolaging.2011.12.015
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE* 6:e18082. doi: 10.1371/journal.pone.0018082
- Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070
- Shannon, R. V., Zeng, F. G., and Wygonski, J. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90. doi: 10.1038/416087a
- Souza, P. E., and Boike, K. T. (2006). Combining temporal-envelope cues across channels: effects of age and hearing loss. *J. Speech. Lang. Hear. Res.* 49, 138–149. doi: 10.1044/1092-4388(2006/011)
- Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi: 10.1016/j.heares.2013.08.004
- Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang.* 123, 191–201. doi: 10.1016/j.bandl.2012.09.001
- Tremblay, K., Piskosz, M., and Souza, P. E. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin. Neurophysiol.* 114, 1332–1343. doi: 10.1016/S1388-2457(03)00114-7
- Tremblay, K. L., Piskosz, M., and Souza, P. (2002). Aging alters the neural representation of speech cues. *Neuroreport* 13, 1865–1870. doi: 10.1097/00001756-200210280-00007
- Walton, J. P., Frisina, R. D., and O'Neill, W. E. (1998). Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse. *J. Neurosci.* 18, 2764–2776.
- Wang, M., Wu, X., Li, L., and Schneider, B. A. (2011). The effects of age and interaural delay on detecting a change in interaural correlation: the role of temporal jitter. *Hear. Res.* 275, 139–149. doi: 10.1016/j.heares.2010.12.013
- Wang, S., Mannell, R., and 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. doi: 10.1007/s10162-011-0285-0
- Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (1999). Regaining lost time: adult aging and the effect of time restoration on recall of time-compressed speech. *Psychol. Aging* 14, 380–389. doi: 10.1037/0882-7974.14.3.380
- Xu, L., and Jeng, F. C. (2011). Noise tolerance in human frequency following responses to voice pitch. *J. Acoust. Soc. Am.* 129, 21–26.
- Xu, L., and Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception. *J. Acoust. Soc. Am.* 114, 3024–3027. doi: 10.1121/1.1623786

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Wang, Hu, Dong, Liu, Chen, Musacchia and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.