INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

On the Duration of Mandarin Tones

Jing Yang[1], Yu Zhang[2], Aijun Li[3], Li Xu[2]

[1]Communication Sciences and Disorders, University of Central Arkansas, USA
[2]Communication Sciences and Disorders, Ohio University, USA
[3]Chinese Academy of Social Sciences, China
jyang@uca.edu, yz137808@ohio.edu, liaj@cass.org.cn, xul@ohio.edu

Abstract

The present study compared the duration of Mandarin tones in three types of speech contexts: isolated monosyllables, formal text-reading passages, and casual conversations. A total of 156 adult speakers was recruited. The speech materials included 44 monosyllables recorded from each of 121 participants, 18 passages read by 2 participants, and 20 conversations conducted by 33 participants. The duration pattern of the four lexical tones in the isolated monosyllables was consistent with the pattern described in previous literature. However, the duration of the four lexical tones became much shorter and tended to converge to that of the neutral tone (i.e., tone 0) in the text-reading and conversational speech. The maximum-likelihood estimator revealed that the durational cue contributed to tone recognition in the isolated monosyllables. With a single speaker, the average tone recognition based on duration alone could reach approximately 65% correct. As the number of speakers increased (e.g., ≥ 4), tone recognition performance dropped to approximately 45% correct. In conversational speech, the maximum likelihood estimation of tones based on duration cues was only 23% correct. The tone duration provided little useful cue to differentiate Mandarin tonal identity in everyday situations.

Index Terms: tone recognition, Mandarin Chinese, duration cues, maximum-likelihood estimator

1. Introduction

Mandarin (in the present study, we refer to Standard Chinese or Putonghua) is a tonal language in which four types of F0 (fundamental frequency) contours are adopted to differentiate lexical meanings for a given stressed isolated monosyllable. These four categories of lexical tones are labeled as tone 1, tone 2, tone 3, and tone 4 with pitch contours being high-level, mid-rising, low-dipping, and high-falling, respectively. For example, the stressed syllable "ba" (/pa/ in IPA) may mean "eight", "pull", "target", and "dad" for tone 1 through tone 4, respectively. In addition to the four lexical tones in stressed isolated monosyllables, there is a fifth tone, usually called a neutral tone or tone 0. It normally occurs in unstressed syllables in multisyllabic words or connected speech.

So far, a large number of studies have shown that the main acoustic correlate of Mandarin tones is F0 [1-5]. In addition to F0 patterns, previous studies revealed that Mandarin tones also vary in intrinsic duration and overall amplitude contour of the vowels [6-9]. However, controversy surrounds which part of a syllable carries the tone. Some researchers [1,10] pointed out that tone is carried by the voiced portion of a syllable in which the voiced initial consonant, if any, should be included. Some other researchers [11] noted that tonal identity was only related to the syllable nucleus, i.e., vowel. The initial voiced consonants and nasal ending should be excluded. In the present study, we followed Howie [7] who suggested that syllable rhyme including vowel and/or nasal endings carry the tones. Howie [7] also suggested that the voiced consonant preceding the nucleus could be viewed as a tonal transition.

In Mandarin words, the rhymes that carry individual tones vary in duration. Among the four lexical tones, many studies have reported that tone 3 has the longest duration; tone 4 has the shortest duration and tones 1 and 2 fall in between [2, 9]. In comparison to isolated monosyllables, the durational features of Mandarin lexical tones were less consistent and less distinguishable in connected speech. Tseng reported that while tone 3 was still the longest, tone 4 was no longer the shortest in connected spontaneous speech [12]. Deng et al. suggested that the durations of four lexical tones in the sentence medial and sentence-final position followed such a pattern: tone 2 > tone 1 > tone 3 > tone 4 [13]. Chang measured the duration of Mandarin lexical tones in sentence-final positions and found that tone 2 was the longest followed by tone1, tone 4, and tone 3 [14].

The aforementioned studies measured the durations of target words located at different places of designed sentences. However, how the durational features of Mandarin lexical tones are reflected in a more natural setting, such as conversational speech, remains vague. In addition, as connected speech contains a large number of words with neutral tones (i.e., tone 0), except for a general consensus of shorter duration for tone 0 than for the other four tones, few studies have reported data on the extent to which the durational feature of tone 0 differs from the four lexical tones in different speech contexts. More importantly, as pitch information is inadequately coded in modern multichannel cochlear implants, researchers turned to examine the contribution of the secondary cues of tone duration and amplitude to tone recognition in CI recipients [15-24]. To date, almost all tone perception tests used monosyllables produced by a small number (usually one or two) of speakers. If the pattern of tone durations in other types of contexts from a larger number of speakers differs from that in monosyllables by individual speakers, caution should be exerted when generalizing previous findings of tone perception in CI users.

To address these issues, the present study aimed to provide a detailed investigation of tone duration by comparing the duration of Mandarin tones from a relatively large number of speakers in both isolated citation forms and connected speech that includes formal text-reading and casual conversational speech. In addition, a maximum-likelihood estimator was adopted as an ideal observer for tone recognition based on the durational features.

http://dx.doi.org/10.21437/Interspeech.2017-29

2. Method

2.1. Participants

A total of 156 Mandarin-speaking adults was recruited for 3 different speech production tasks: (1) 121 speakers (46 males and 75 females) between 18 to 34 years of age (Mean=24, SD=3) were recruited for the production of isolated monosyllables; (2) 33 speakers (15 males and 18 females) between 20 and 50 years of age were recruited for the production of conversational speech; and (3) 2 speakers (one 24-year-old male and one 23-year-old female) were recruited to read text. None of the participants reported having any speech-language or hearing impairments.

2.2. Speech Materials

For the isolated monosyllables, 44 tokens composed of 11 syllables in all four tones were recorded from each of the 121 speakers. The consonant and vowel contexts across the 11 syllables were not controlled in the present study. For the connected speech, a total of 12,085 syllables were obtained from 18 passages of read text by the 2 speakers and 21,945 syllables were obtained from 20 conversations among the 33 speakers. Because the speech contexts in both text-reading passages and conversations were not designed, the consonant and vowel contexts for the tone productions were not controlled either.

2.3. Recording and Duration Measurement

For the production of monosyllables and text-reading passages, all participants were recorded separately in a quiet room or sound-attenuated booth. For monosyllables, each speaker was provided a list of monosyllables and was asked to produce tones 1 through 4 for each syllable. For text-reading speech, each speaker was provided a number of randomly selected passages and was required to read each passage with no breaks. All productions were recorded to a digital recorder with a 44.1-kHz sampling rate and a 16-bit quantization. The productions of conversational speech were selected from the customer service phone recordings in the corpus of Discourse-CASS, which were recorded with a 16-kHz sampling rate and a 16-bit quantization. The recorded sound files were annotated with Chinese characters and Pinyin. An automatic segmentation program was then applied to determine the locations of individual syllables, onset and rhyme of each syllable. Praat was finally used to manually check the boundary locations of the phonetic units. The duration of each tone was measured as the time interval between the onset of the vowel and the offset of the syllable.

2.4. Maximum-Likelihood Estimator

This study adopted a computational model, namely maximum-likelihood estimator [25] to determine the optimal outcome of tone recognition on the basis of tone durations alone. In this procedure, based on the mean and standard deviation of the duration data for each tone, a normal probability density function from 0 to 1000 ms was derived. The probability density values at every 10-ms step were compared among the 4 tones in the monosyllabic speech or 5 tones in the connected speech. The maximum values of the probability density at all of the 10-ms steps were then summed and the sum was divided by the total probability density between 0 and 1000 ms of all the 4 or 5 probability density functions to derive a percent-correct value for the ideal observer.

3. Results

Pre-analysis of the tone durations yielded subtle differences between male and female speakers (the gender difference on tone duration was less than 10 ms for the four lexical tones and less than 20 ms for tone 0). Given that the gender difference on tone duration was subtle and not of interest in the present study, the durational data was collapsed across gender for the subsequent analyses. Table 1 presents the means and standard deviations of tone duration in the isolated monosyllables, text-reading speech, and conversational speech. Note that the isolated monosyllables only contained 4 lexical tones while connected speech contained 5 tones. Normal probability density functions were used to estimate the overall distributional patterns of the Mandarin tones in the three contexts (see Figure 1). It was shown that the durations of the four lexical tones were longer and more widely dispersed in the isolated monosyllables than in the text-reading speech or conversational speech. In the text-reading and conversational speech, the four lexical tones were not only significantly shorter than those in isolation but also narrowly distributed within a much smaller range. In addition, tone 0 was shorter than the four lexical tones in the text-reading speech. In conversational speech, the difference between tone 0 and the four lexical tones tended to diminish. A one-way repeated-measures ANOVA was used to compare the subject-mean durations of the four lexical tones in the isolated monosyllables and the five tones in the conversational speech, respectively. When a significant effect of tone was yielded, pair-wise sample t tests with adjustment for multiple comparisons were conducted to compare the differences of duration between individual tone pairs. Consistent with our observation for the isolated monosyllables, the four lexical tones showed significantly different durations in citation form ($F(3,360)=449.466$, $p<0.0001$). All tone pairs were significantly different on duration except for tone 1 and tone 2. In conversational speech, the five tones also showed significantly different durations ($F(4,128)=7.272$, $p=0.001$). In particular, tone 0 was significantly shorter than tone 1, tone 3, and tone 4, but no significant difference was found among the four lexical tones. Note that there were only two speakers in the text-reading speech, the repeated-measures ANOVA was not applied to the speech samples in this condition.

Regarding the durational pattern of the four lexical tones, in the monosyllables, tone 3 has the longest duration while tone 4 has the shortest duration. Tone 1 and tone 2 showed similar durations in between tone 3 and tone 4. This pattern was consistent with the results reported in earlier studies [5,6,10]. In the text-reading speech, tone 2 was the longest and tone 1 was the shortest, which was different from the pattern in the isolated monosyllables. In conversational speech, all five tones tended to converge on the durational feature even though tone 0 was still significantly shorter than the four lexical tones. The durations of the four lexical tones were similar. Close comparison of the four lexical tones revealed that tone 1 had the longest duration and tone 2 had the shortest duration. This pattern was different from that in the text-reading speech or isolated monosyllables. These observations suggested that Mandarin tones' durations tend to converge as the speech style changed from citation form to casual speech. In addition, the durational pattern normally described in isolated monosyllables does not apply to other contexts.

In the following analysis, we adopted a maximum-likelihood estimator as an ideal observer to estimate to what extent the Mandarin tones can be distinguished solely based on

Table 1. *The mean and standard deviation (in ms) of the durations of the four Mandarin lexical tones in the isolated monosyllables, text-reading speech, and conversational speech.*

	Tone 0		Tone1		Tone2		Tone3		Tone4	
	M	SD	M	SD	M	SD	M	SD	M	SD
Monosyllables			416.2	90	417.7	87.7	484	97.7	307.8	92.4
Text-reading	106.2	59.6	149.2	69.3	169.6	65.6	157.3	62.6	152.8	65.7
Conversations	96.9	49.6	115.6	56.4	107.6	51	112.3	54	111.2	60.1

durational features in isolated monosyllables or conversational speech. Figure 2 displays the maximum-likelihood recognition accuracy of Mandarin tones in isolated monosyllables and conversational speech as a function of increasing number of speakers. The number of speakers ranged from 1 to 15. For each number of speakers, the tone duration data of the desired number of speakers were chosen and fed into the maximum-likelihood estimator. This process repeated 30 times. Each time different sets of speakers were chosen randomly with replacement. Therefore, 30 maximum-likelihood recognition scores were obtained for each number of speakers. As shown in Figure 2 for the four lexical tones in the isolated monosyllables, the average maximum-likelihood recognition was approximately 65% correct when there was only one speaker.

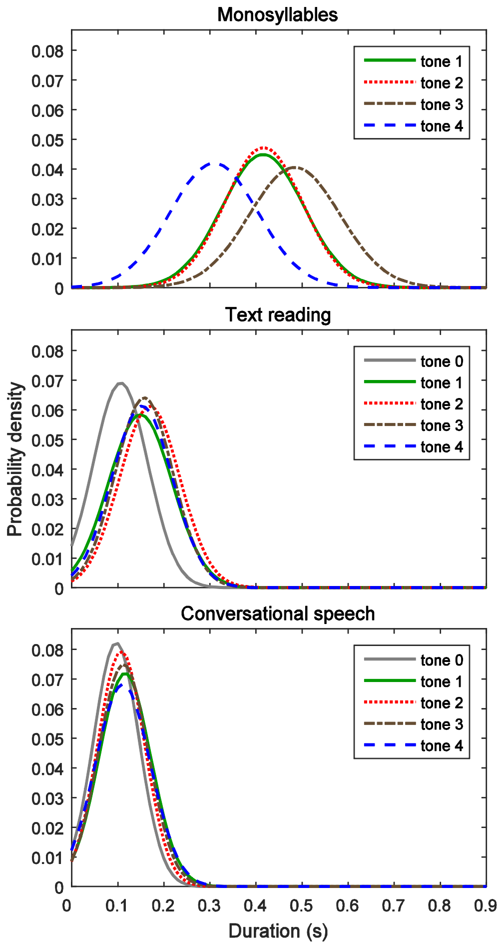

Figure 1. Probability density function of tone durations in the isolated monosyllables (top panel), text-reading speech (middle panel), and conversational speech (bottom panel).

For certain speakers who showed clearly distinct durational features on the four lexical tones, the accuracy could reach as high as 80% correct solely based on tone durations. However, for other speakers who did not show distinctive durational features, the maximum-likelihood recognition accuracy dropped to approximately 45%. As the number of speakers increased, the average maximum-likelihood score gradually decreased from approximately 65% to 45% correct. In addition, the maximum-likelihood recognition scores showed less variability as the number of speakers increased. When the number of speakers increased to 5 or above, the average maximum-likelihood score remained relatively stable. When the number of speakers increased to 10 or above, the average score dropped to 45% correct, which was similar to the score of 42.7% correct for all 121 speakers.

For conversational speech, the highest maximum-likelihood recognition accuracy in one speaker condition only reached 35% correct. The average maximum-likelihood recognition scores, in the range between 24 and 28% correct, showed little change regardless of the number of speakers (1 to 15), which was close to the overall maximum-likelihood recognition score of 23% correct for all 33 speakers (Figure 2). However, the variability of maximum-likelihood recognition scores for the conversational speech was much smaller than that for the isolated monosyllables.

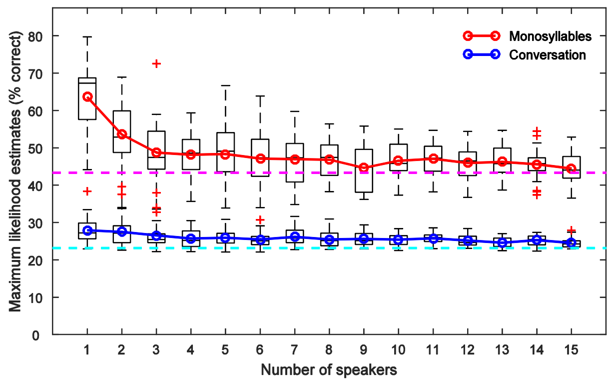

Figure 2. The maximum-likelihood recognition score of Mandarin tones based on tone durations in the isolated monosyllables (red) and conversational speech (blue). For each number of speakers, the maximum-likelihood estimator was implemented 30 times. Each box shows the horizontal lines at the lower quartile, median, and upper quartile values. The whiskers show the range of the data. The circles represent the average recognition scores. The pink dashed line represents the overall maximum-likelihood recognition score of the 121 speakers in the isolated monosyllables while the cyan dashed line represents the overall maximum-likelihood recognition score of the 33 speakers in the conversational speech.

Since the text-reading speech samples were recorded from only two speakers, an overall maximum-likelihood score for these two speakers was calculated. Associated with the significantly shorter duration in tone 0 than in the other four lexical tones, the overall maximum-likelihood recognition score was approximately 27% correct in the text-reading speech, slightly higher than the score in the conversational speech but much lower than the score in the isolated monosyllables.

4. Discussion and Summary

The four lexical tones in the isolated monosyllables showed distinctive durational features, which was consistent with the pattern described previous studies. However, the maximum-likelihood computational model revealed that tone durations did not always provide a reliable cue in tone recognition. As the number of speakers increased from one to four, the average optimal recognition score dropped from approximately 65% to 50% correct. This result suggested that the contribution of the durational cue to tone recognition decreased as a result of the inter-speaker variability of tone durations. It is worth noting that when the number of speakers increased to 10 or above, the optimal recognition score maintained at a relatively stable level of 45% correct despite further increase of the number of speakers. This result illustrated that although the absolute values of tone durations differ in various speakers, the relative durational pattern in the isolated monosyllables may still provide certain perceptual cue for tone recognition.

In comparison to the isolated monosyllables, the duration pattern of the longest tone 3 and the shortest tone 4 does not exist anymore in the text-reading or conversational speech. Some previous studies reported that Mandarin lexical tones showed a different duration pattern in sentence context relative to isolated citation form. The present study revealed that the durations of all four lexical tones became highly similar in connected speech when a large number of tokens (e.g., N > 10,000) were collected. In addition, the tone durations were dramatically shortened from isolated monosyllables to formal text-reading speech and casual conversations. This pattern conforms to the findings of shorter speech units in more casual spontaneous speech than in formal speech [26, 27]. Although the four lexical tones were apparently longer than tone 0 in the formal text-reading speech, the difference between tone 0 and the four lexical tones decreased substantially in the conversation speech. In such a case, the optimal recognition score based on the durational cues alone for the five tones dropped to approximately 23% correct. This result suggested that tone duration was no longer a reliable cue to differentiate Mandarin lexical tones in connected speech.

The results of the present study have important implications in the development of tone recognition tests. To date, several Mandarin tone recognition tests have been developed to evaluate the identification and differentiation of Mandarin lexical tones in hearing-impaired listeners [28-32]. All these tests were based on a limited number of isolated monosyllables produced by only one or two speakers. As illustrated in Figure 2, the optimal tone recognition score in isolated monosyllables based on the durational cues alone varied from 45% to 80% correct with only one speaker. However, when the number of speakers increased to 4 or more, the optimal tone recognition score demonstrated much less variability and stabilized at around 45% correct (i.e., 20 percentage points above chance performance). Therefore, when we develop such tone perception tests, stimuli from multiple speakers (≥ 4 speakers) should be used to reduce the influence of inter-speaker variability on the test outcome and the contribution of the duration cues to the performance should be considered.

In addition, modern multichannel cochlear implants are widely used in people with a severe-to-profound hearing loss. As pitch information was poorly coded in current signal processing strategies, a large number of perceptual studies were conducted to examine the tone perception of cochlear implant recipients. While a few studies normalized the tone duration in their perceptual tests [16-19], many studies did not equalize the tone durations for tone recognition [20-24]. As the four lexical tones exhibited distinctive durational features in isolated monosyllables, it was very likely that when cochlear implants users listened to the speech stimuli, the tone durations might be more weighted in tone recognition in lieu of the poor F0 information provided by the cochlear implant devices [15]. Indeed, the tone-recognition scores in cochlear implant users are generally higher when the tone durations were not equalized than when they were equalized. However, as the durational features are not distinctive in the spontaneous speech (shown in Figures 1 and 2) and the tone durations do not provide a reliable perceptual cue for tone recognition, it is reasonable to predict that the recognition accuracy of lexical tone in natural connected speech will decrease significantly. In such a case, the relatively high recognition accuracy of the four tones in isolated monosyllables does not necessarily represent the actual perceptual ability of lexical tones in everyday situations. Furthermore, if the purpose of the research is to investigate pitch-related perception, the unequal tone duration may function as a confounding factor and should be eliminated. Xu et al. reported in a vocoder study that tone recognition of isolated monosyllables dropped by approximately 15 percentage points with tone duration equalized relative to that with the original tone duration preserved [5].

In sum, the present study showed distinctive duration patterns of Mandarin tones in various speech contexts. The durations of the four lexical tones in the isolated monosyllables varied significantly, with tone 3 being the longest and tone 4 being the shortest. In text-reading and natural conversation, the duration of the four lexical tones tended to converge and approximated the duration of the neutral tone. Consistent with different durational patterns in various contexts, the tone recognition rate based on the durational cues alone varied significantly. The durational cue provided reliable cue to the tone recognition in the isolated monosyllables with a small number of speakers, but contributed little, if any, to the tone recognition in connected speech. The results in the present study suggested that speech tokens from multiple speakers (≥ 4) should be used for tone recognition test using monosyllables to avoid the biased recognition outcome caused by inter-speaker variability. In addition, tone duration may be a confounding factor when the purpose of the research is pitch-related perception and such a confounding effect can be avoided with equalized tone durations.

5. Acknowledgements

This study was supported in part by the NIH/NIDCD Grant No. R15-DC014587.

6. References

[1] Y. R. Chao, *A Grammar of Spoken Chinese*. Los Angeles: University of California Press, 1968.

[2] A. T. Ho, "The acoustic variation of Mandarin tones," *Phonetica*, vol. 28, pp. 353–367, 1976.

[3] J. M. Howie, "On the domain of tone in Mandarin," *Phonetica*, vol. 30, pp. 129–148, 1974.

[4] H. Lin and B. H. Repp, "Cues to the perception of Taiwanese tones," *Language and Speech*, vol. 32, pp. 25–44, 1989.

[5] L. Xu, Y. Tsai, and B. E. Pfingst, "Features of stimulation affecting tonal-speech perception implications for cochlear prostheses," *Journal of the Acoustical Society of America*, vol. 112, no. 1, pp. 247–258, 2002.

[6] Q. J. Fu and F. G. Zeng, "Identification of temporal envelope cues in Chinese tone recognition," *Journal of Speech, Language, and Hearing Research*, vol. 5, pp. 45–57, 2000.

[7] J. M. Howie, *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press, 1976.

[8] H.-B. Lin, *Contextual Stability of Taiwanese Tones*. Doctoral Dissertation, The University of Connecticut, 1988.

[9] D. H. Whalen and Y. Xu, "Information for Mandarin tones in the amplitude contour and in brief segments," *Phonetica*, vol. 49, pp. 25–47, 1992.

[10] W. S-Y. Wang, "Phonological features of tones," *International Journal of American Linguistics*, vol. 33, no. 2, pp. 93–105, 1967.

[11] Lin, M.-C. A perceptual study on the domain of tones in Standard Chinese. *Chinese Journal of Acoustics*, vol.14, no.4, pp. 350-357. 1995.

[12] C. Tseng, *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Doctoral Dissertation, Brown University, 1981.

[13] D. Deng, S. Feng, and S. Lu, "The contrast on tone between Putonghua and Taiwan Mandarin," *Sheng Xue Xue Bao* [Acta Acustica], vol. 31, no. 6, pp. 536–541, 2006.

[14] Chang, C. Y. *Dialect Differences in the Production and Perception of Mandarin Chinese Tones*. (Doctoral Thesis) Retrieved from ProQuest Dissertations and Theses. (UMI No. 3435661). 2010.

[15] S.-C. Peng, H.-P. Lu, N. Lu, Y.-S. Lin, M. L. D. Deroche, and M. Chatterjee, "Processing of acoustic cues in lexical tone identification by pediatric cochlear implant recipients," *Journal of Speech, Language and Hearing Research*, vol.60, no. 5, pp. 1223-1235, 2017.

[16] D. Han, B. Liu, N. Zhou, X. Chen, Y. Kong, H. Liu, Y. Zheng, and L. Xu, "Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users," *Ear & Hearing*, vol. 30, pp. 169–177, 2009.

[17] Y. Mao and L. Xu, "Lexical tone recognition in noise in normal-hearing children and prelingually deafened children with cochlear implants," *International Journal of Audiology, published online: 26 Aug 2016 | http://dx.doi.org/10.1080/14992027.2016.1219073.* 2017

[18] L. Xu, X. Chen, H. Lu, N. Zhou, S. Wang, Q. Liu, Y. Li, X. Zhao, and D. Han, "Tone perception and production in pediatric cochlear implants users," *Acta Oto-Laryngologica*, vol. 131, pp. 395–398, 2011.

[19] N. Zhou, J. Huang, X. Chen, and L. Xu, "Relationship between tone perception and production in prelingually-deafened children with cochlear implants," *Otology & Neurotology*, vol. 34, pp. 499–506, 2013.

[20] Y. Chen, L. L. N. Wong, F. Chen, and X. Xi, "Tone and sentence perception in young Mandarin-speaking children with cochlear implants," *International Journal of Pediatric Otorhinolaryngology*, vol. 78, pp. 1923–1930, 2014.

[21] A. Li, N. Wang, J. Li, J. Zhang, and Z. Liu, "Mandarin lexical tones identification among children with cochlear implants or hearing aids," *International Journal of Pediatric Otorhinolaryngology*, vol. 78, no. 11, pp. 1945–1952, 2014.

[22] S. C. Peng, J. B. Tomblin, H. Cheung, Y. S. Lin, and L. S. Wang, "Perception and production of Mandarin tones in prelingually deaf children with cochlear implants," *Ear & Hearing*, vol. 25, no. 3, pp. 251–264, 2004.

[23] D. Tao, R. Deng, Y. Jiang, J. J. Galvin III, Q.-J. Fu, and B. Chen, "Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users," *Ear & Hearing*, vol. 36, pp. 102–110, 2015.

[24] Y. Zheng, S. D. Soli, Z. Meng, Y. Tao, K. Wang, K. Xu, and H. Zheng, "Assessment of Mandarin-speaking pediatric cochlear implant recipients with the Mandarin Early Speech Perception (MESP) test," *International Journal of Pediatric Otorhinolaryngology*, vol. 74, pp. 920–925, 2010.

[25] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

[26] J. Mcallister, A. Potts, and G. Marchant, "Word duration in monologue and dialogue speech," *Language and Speech,* vol. 37, no. 4, pp.393-405, 1994.

[27] M. Ernestus, and N. Warner, "An introduction to reduced pronunciation variants," *Journal of Phonetics,* vol. 39, no. 3, pp. 253-260, 2011.

[28] A. Krenmayr, B. Qi, B. Liu, H. Liu, X. Chen, D. Han, R. Schatzer, and C. M. Zierhofer, "Development of a Mandarin tone identification test: Sensitivity index d' as a performance measure for individual tones," *International Journal of Audiology*, vol. 50, pp. 155–163, 2011.

[29] B. Liu, B. Qi, A. Krenmayr, X. Chen, S. Wang, R. Schatzer, C. Zierhofer, and D. Han, "Development of Mandarin tonal identification in noise test materials," *Chinese Journal of Otorhinolaryngology Head and Neck Surgery*, vol. 49, no. 9, pp. 733–737, 2014.

[30] K. C. P. Yuen, K.-L. Cao, C.-G. Wei, L. Luan, H. Li, and Z.-Y. Zhang, "Lexical tone and word recognition in noise of Mandarin-speaking children who use cochlear implants and hearing aids in opposite ears," *Cochlear Implants International*, vol. 10, Suppl. 1, pp. 120–129, 2009.

[31] Y. Zheng, Z.-L. Meng, K. Wang, Y. Tao, K. Xu, and S. D. Soli, "Development of the Mandarin early speech perception test: Children with normal hearing and the effects of dialect exposure," *Ear & Hearing*, vol. 30, pp. 600–612, 2009.

[32] S. Zhu, L. L. N. Wong, and F. Chen, "Development and validation of a new Mandarin tone identification test," *International Journal of Pediatric Otorhinolaryngology*, vol. 78, no. 12, pp. 2174–2182, 2014.