10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

An Eye-tracking Study on Mandarin Tone Perception of Children

Xifan ZHANG, Ting WANG

School of Foreign Languages, Tongji University
1995cven@tongji.edu.cn, 2011ting_wang@tongji.edu.cn

Abstract

In Mandarin, tones have the same function as phonemes, which can distinguish a word from others. Lexical tone processing in speech involves acoustic perception and semantic integration. However, the mechanisms that underpin the interaction between acoustic and semantic information in tonal processing are not fully understood. This study investigated RSD (Relative Staring Duration) under the conditions of different image stimuli (the pitch-highlighted image, the semantic-highlighted image) and sound stimuli (real words, pseudo-words, and filtered words) in 11 typically developing children using an SMI eye tracker. Results indicated that Mandarin-speaking children performed better on pitch perception in speech than in non-speech, but showed a comparable performance on real words and pseudo-words. Moreover, results showed that the tone perception of Mandarin-speaking children was not affected by tone types. These findings suggest that Mandarin tone perception of children cannot rely solely on pitch processing. Meanwhile, though the pitch perception of children is not influenced by semantic information in speech, Mandarin-speaking children still have the ability of distinguishing acoustic features from linguistic meanings of high-dimensional speech signals.

Index Terms: Mandarin tone perception, children, eye-tracking

1. Introduction

Pitch plays a significant role in the perception of speech. In non-tonal languages, pitch conveys information from word stress and utterance accent to pragmatic stance and speaker emotion [1]. However, in tonal languages, pitch is used to differentiate lexical meaning of words particularly [2]. According to Chao [3], Mandarin phonemically distinguishes four tones, which are the high-level tone (Tone 1), the high-rising tone (Tone 2), the low-dipping tone (Tone 3) and the high-falling tone (Tone 4). The same segmental information */ma/* in Mandarin, for example, can represent the meaning of “mother”, “hemp”, “horse” and “scold” with these four tones respectively.

There are two phases in the cognitive process of tones, namely local processing and global integration. In local processing (low-dimensional cognitive processing), speakers identify tones from acoustic cues, such as pitch primarily. Then, speakers combine acoustic features with linguistic meaning in global integration phase (high-dimensional cognitive processing) [4]. In tonal languages, the boundary between low and high dimensions is not clear-cut. It is found that individuals whose native language are tone languages perform better in both producing and perceiving musical pitch [5]. Besides, native speakers of Mandarin have a preeminent categorical perception of high-dimensional speech signal compared with those of English. The categorical perception of low-dimensional speech signal for Mandarin speakers is comparable to that of high-dimensional one, but stronger than English speakers, whose categorical perception of low-dimensional speech signal is better than that of high-dimensional one [6].

The research summarized above supports the notion that the fine-tuning caused by language acquisition is demonstrated in the processing of critical auditory dimensions in the speech signal and can be conveyed over into nonlinguistic domains like cognitive perception [5]. In the early stage of language acquisition, the tone consciousness is superior to the vowel and consonant consciousness [7]. Therefore, in this study, we focus on the tone perception of Mandarin-speaking children and explore its characteristics.

Recent works on Mandarin tone perception of children mainly concentrate on special populations since it is conducive to their diagnosis and intervention. It is shown that autistic children have the advantage of local processing compared with typically developing children [8, 9, 10, 11, 12], while having defects in global integration [13, 14, 15, 16, 17, 18]. Cochlear-implanted children are easy to discriminate Tone 1 and Tone 4, while difficult to separate Tone 2 from Tone 3, which is similar to typically developing children. However, different SNRs may affect the tone perception significantly [19]. Contrary to typically developing children, the tone consciousness of children with Down’s syndrome is irrelevant to the reading ability [20]. As for the research centering on typically developing children, Zhang and Wang [21] find that a positive correlation exists between their categorical perception of tone and reading ability, and the categorical perception of children is as alike as adults [22]. Previous studies analyze Mandarin tone perception of children based on behavioral experiments, while in this study, we use SMI, an eye tracker, to explore the mechanism of Mandarin tone perception of children and expect to answer the following questions:

1. Is children’s perception of pitch change more sensitive to speech sounds or non-speech sounds;
2. Is tone perception of Mandarin-speaking children dependent on acoustic cues or integrated information in speech;
3. Is the pitch perception of children influenced by semantic information in speech?

2. Methods

2.1. Participants

11 children (M_{age} = 7.64, SD_{age} = 1.9) (7 boys, 4 girls) participated in the present study. All of them speak Mandarin Chinese as their native language and have no history of vision or hearing impairment.

2.2. Stimuli

Sound stimuli were monosyllable real words, pseudo-words and filtered words. Both real words and pseudo-words were recorded by a professional female speaker using a Zoom H6 recorder (Zoom Corporation, Japan) and AKG C544 microphone (Samsung US, US). The recording rate was 44.1KHz,

10.21437/SpeechProsody.2020-61

Finals	Real Words	Pseudo-words
ɑ	蛙，娃，瓦，袜 (wɑ)	ʐɑ1, ʐɑ2, ʐɑ3, ʐɑ4
o	摸，磨，抹，墨 (mo)	no1, no2, no3, no4
ɤ	婀，鹅，恶，饿 (ɤ)	nɤ1, nɤ3
i	衣，姨，椅，亿 (i)	k‘i1, k‘i2, k‘i3, k‘i4
u	猪，竹，煮，柱 (tʂu)	
y	迂，鱼，雨，玉 (y)	ny1, ny2
ɑu	猫，毛，铆，帽 (mɑu)	fɑu1, fɑu2, fɑu3, fɑu4

Figure 1: *Real words and pseudo-words used in sound stimuli*

and the sampling accuracy was 16bit. The real words contained all six monophthongs in Mandarin, namely */ɑ/, /o/, /ɤ/, /i/, /u/, /y/*, and one diphthong */ɑu/*. As for pseudo-words, vowels were as the same as the real words while initials were changed to make them meaningless (see Figure 1). Filtered words were created by Praat using real words as prototypes. Pitch contour of real words was extracted and then resynthesized with sine waves. Therefore, the spectral information was eliminated from the acoustic signals, leaving only the pitch information.

2.3. Procedure

Two perception tasks were conducted in the study: the pitch-highlighted identification task and semantic-highlighted identification task. The pitch-highlighted identification task investigated children's perceptual ability of identifying pitch directions. The semantic-highlighted identification task tested children's perceptual ability of extracting linguistic meanings by integrating pitch and semantic information in acoustic signals.

E-prime software was used to connect to an SMI eye tracker and present image and sound stimuli together. In the pitch-highlighted identification task, the image stimulus, in which the driving path of the car simulated pitch changes, showed four pitch contours of Mandarin tones without any semantic information (See the left picture in Figure 2). Three blocks were designed with the same image stimulus and sound stimuli of real words, pseudo-words, and filtered words respectively. Presented in random order, each block randomly played sound stimuli twice for a total of two rounds, and each trial lasted for 6000ms. Participants were required to stare at the corresponding picture matching the pitch contour of the sound stimulus. However, in the semantic-highlighted identification task, the sound stimuli were only real words and the image stimulus showed the lexical meaning of the words instead of the pitch contour. For example, */tʂu/* can represent meanings of "pig", "bamboo", "boil" and "pillar" with different tones (See the right picture in Figure 2). Since meanings of some words were too abstract to use a picture to represent, only 24 sound stimuli were used in this block. All trials were randomly played twice for two cycles and also lasted for 6000ms. Participants were required to stare at the corresponding picture matching the lexical meaning of the sound stimulus. The two tasks were counterbalanced among subjects.

2.4. Data analysis

The statistical software of SMI, BeGaze, was applied to divide the image stimulus into four areas of interest (AOI), and collect

Figure 2: *Samples for image stimuli: pitch-highlighted one on the left (top left: high-level pitch; top right: high-rising pitch; bottom left: low-dipping pitch; bottom right: high-falling pitch) and semantic-highlighted one on the right (top left: zhu1"pig"; top right: zhu2"bamboo"; bottom left: zhu3"boil"; bottom right: zhu4"pillar"). They are used in different type of tasks.*

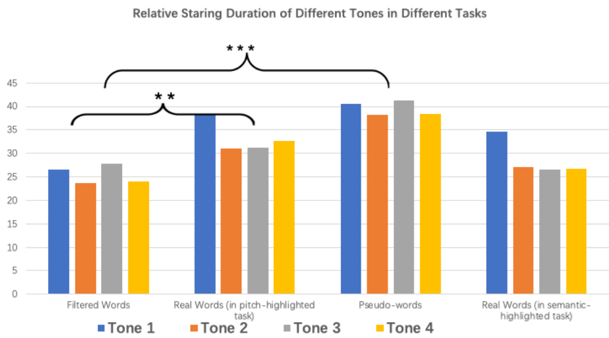

Figure 3: *RSD of Different Tasks in Different Tones*

detailed AOI data. The relative staring duration (RSD) of the target AOI was calculated statistically.

$$\text{RSD} = \frac{\text{Staring Duration on Target AOI}}{\text{Total Staring Duration on four AOI}}$$

Therefore, the higher the RSD, the more accurate the children's perception of the pitch direction or meaning of each tone.

3. Results

Figure 3 presents children's relative staring duration of different tones in both the pitch-highlighted task and semantic-highlighted task. As can be seen, children performed much better when sound stimuli were pseudo-words. In order to further analyze the influence of sound types and tone types on tone perception statistically, a linear mixed model using the *lmer*Test package in R was conducted, where the independent variables were sound stimuli (pseudo-words, filtered words and real words in the pitch-highlighted task; real words in the semantic-highlighted task) and tone types (Tone 1: high level tone, Tone 2: rising tone, Tone 3: low/diping tone and Tone 4: falling tone), and the random factor was participants.

Results showed that there was no significant effect of tone types on the tone perception of children (X^2 = 1.785, df = 3, p = .148), but a main effect of types of sound stimuli (X^2 = 24.969, df = 3, p <.001). No significant two-way interaction was found between sound stimuli and tone types (X^2 = .788, df = 9, p = .627).

To explore the within-group effect in sound stimuli, a *post-hoc* multiple comparison test was conducted using the *mult-comp* package in R. The results, as presented in Table 1, revealed that filtered words motivated a significantly smaller RSD than both real words and pseudo-words in the pitch-highlighted task (RW(P) VS FW: $p < .01$; PW VS FW: $p < .001$), whereas there was no significant difference between real words and pseudo-words ($p = .336$). Besides, the results showed that real words in the semantic-highlighted task did not motivate a greater RSD when compared to other sound stimuli.

Table 1: *The output of multiple comparison analysis for RSD between different sound types (Estimate, coefficient estimates; S.E., standard errors).*

	Estimate	**S.E.**	**z-value**	**p-value**
RW(P) VS FW	11.645	3.428	3.397	0.003^{**}
PW VS FW	16.654	3.632	4.586	$<0.001^{***}$
RW(S) VS FW	8.745	4.008	2.182	0.116
RW(P) VS PW	5.009	3.632	1.379	0.336
RW(S) VS RW(P)	-2.900	4.008	-0.724	0.469
RW(S)VS PW	-7.909	4.168	-1.897	0.173

Notes: **p <0.01; ***p <0.001
RW(P) = Real words (in pitch-highlighted task)
RW(S) = Real words (in semantic-highlighted task)
PW = Pseudo-words
FW = Filtered words

4. Discussion

The present study explored the mechanism of tone perception in Mandarin-speaking children using the SMI eye tracker. Our results indicated that Mandarin-speaking children showed superior performance on the pitch perception of speech instead of non-speech, but comparable performance on real words and pseudo-words in both the pitch-highlighted task and the semantic-highlighted task. Moreover, the tone perception of Mandarin-speaking children was not affected by tone types. These findings suggest that Mandarin tone perception of children cannot rely solely on part of acoustic signals of words in pitch processing. Meanwhile, though the pitch perception of children is not influenced by semantic information in speech, Mandarin-speaking children still have the ability to distinguish acoustic features from linguistic meanings of high-dimensional speech signals.

Sound stimuli as real words and pseudo-words fall into the category of speech sounds, while filtered words belong to non-speech sounds. As for those speech sounds, the acoustic signals of tones are mainly contained in the information of the frequency domain which involves both fine structure and spectral envelope. The fine structure includes fundamental frequency and its harmonics and the spectral envelope refers to the outer envelope structure formed by the formants [23]. Researches showed that fundamental frequency and its harmonics and formants are the most important information in the frequency domain related to the perception of tones [24, 25, 26]. Reserving the fundamental frequency and removing other information, filtered words only contain the pitch contour of real words without any linguistic information. Consistent with previous studies [27, 28], typically developing children's perception of pitch change is more sensitive to speech instead of non-speech, which is in the contrast of children with autism [10, 12, 27]. In [27], children with autism identified and memorized single-note frequencies better than typically developing children. When stimuli were speech-like sounds, the performance between two groups were comparable, while typically developing children performed significantly well compared with non-speech. The result indicated that typically developing children cannot rely solely on part of acoustic signals of words to perceive Mandarin tones. In the present study, when comparing with filtered words, RSD of pseudo-words was relatively higher than that of real words. In the case that real words and pseudo-words both contain the information of the frequency domain, real words also carry the semantic information. The result further corroborated the above-mentioned point of view and indicated that the semantic information of words may interfere with the tone perception of Mandarin-speaking children.

The previous result seemed to turn out that the semantic information of words would hinder children's perception of tones. However, real words and pseudo-words in the pitch-highlighted task showed no significant differences, meaning that the Mandarin tone perception of children was in the same level in different stimuli of speech sounds. Similarly, in [29], when the sound stimuli were real syllables and pseudo-syllables in Cantonese, the performance of typically developing individuals were comparable. Also in [17], the findings did not show a significant difference between real words and nonsense words for either typically developing children and children with autism. In Mandarin tone perception, children perceive tone information through the acoustic cues perception and the semantic information integration of speech sounds. When the sound stimulus is presented as a real word, both two channels would work simultaneously, and when the sound stimulus is presented as a pseudo-word, the channel of semantic information would not be triggered. The experimental results showed that basically, the tone perception of Mandarin-speaking children in speech sounds could not necessarily rely on the channel of semantic information of words but on that of acoustic cues. In other words, children still tend to perceive Mandarin tones based on acoustic features in high-dimensional cognitive processing.

However, it does not mean that children cannot perceive Mandarin tones through linguistic meaning or integrated information in high-dimensional cognitive processing. The present study conducted a semantic-highlighted task to compare with the pitch-highlighted task. Under the circumstance that the sound stimuli were all real words, the image stimuli presented on the screen were divided into two categories, the pitch diagrams and semantic diagrams, as mentioned in the previous section. The pitch diagrams aimed to make the participants focus on pitch changes in the stimuli while semantic diagrams aimed to get the participants to concentrate on the lexical meaning of words. As speech sound, real words require both low-dimensional cognitive processing and high-dimensional cognitive processing. The result indicated that children could perceive Mandarin tones through both pitch and semantic information in high-dimensional cognitive processing since there was no difference between RW(S) and RW(P). Furthermore, it showed that children had the ability to distinguish between acoustic features and linguistic meanings in global integration. In this regard, scholars have explored the brain mechanisms in high-dimensional cognitive processing through hemispheric laterality studies of pitch pattern processing. The functional hypothesis states that lateralization is determined by the function of the sound signal [30], while the acoustic hypothesis states that lateralization is determined by the physical properties of the sound signal [31]. The results of the present study showed that children, like adults, also had the ability to perceive Mandarin tones through acoustic or semantic cues alone. In subsequent studies, EEG technology could be used to further explore the brain mechanisms of children's Mandarin tone perception.

In terms of the perceptual results in different tones, previous study found that children performed better in perceiving Tone 1 and Tone 4 [19]. In [32, 33], the results showed that the Mandarin Tone 1 and Tone 4 are acquired before Tone 2 and Tone 3. Since the difficulty and confusability of Tone 2 and Tone 3 [32], the perception of Tone 1 and Tone 4 would be easier. However, our study revealed that there was no significant effect of tone types, suggesting that the tone perception of Mandarin-speaking children was not influenced by different tones of words. This observation is not of expectation. Through the post-experiment interview, it was found that most of our participants had learned the Pinyin transcription system and tone system of Mandarin before, which would be the reason why they had a good command of tone perception.

5. Conclusions

Our study analyzed pitch processing in a group of children who speak a tone language using the SMI eye tracker. Results suggest that Mandarin-speaking children are more sensitive to pitch changes in speech than non-speech. Moreover, in speech sounds, they depend more on acoustic cues instead of integrated information to perceive different tones. Meanwhile, though the pitch perception of children is not influenced by semantic information in speech, Mandarin-speaking children still have the ability to distinguish acoustic features from linguistic meanings of high-dimensional speech signals. In addition, the tone perception of Mandarin-speaking children is not affected by tone types.

6. Acknowledgements

Special thanks go to all the lovely children for their participation in the perception experiments. This study was supported in part by the Youth Project of Humanities and Social Sciences Foundation (grant no. 18YJC740103) of the Ministry of Education in China, and the Chenguang Program (grant no. 16CG21) of the Shanghai Education Development Foundation and the Shanghai Municipal Education Commission, which were awarded to the second author.

7. References

[1] A. Cruttenden, *Intonation*. Cambridge University Press, 1997.

[2] M. Yip, *Tone*. Cambridge University Press, 2002.

[3] Y. R. Chao, *Mandarin primer: An intensive course in spoken Chinese*. Harvard University Press, 1948.

[4] A. Schirmer and S. A. Kotz, "Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing," *Trends in cognitive sciences*, vol. 10, no. 1, pp. 24–30, 2006.

[5] P. Q. Pfordresher and S. Brown, "Enhanced production and perception of musical pitch in tone language speakers," *Attention, perception, & psychophysics*, vol. 71, no. 6, pp. 1385–1398, 2009.

[6] Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 1063–1074, 2006.

[7] B. Xu and F. Li, "Experimental study on pre- school children's chinese phonological awareness," *Studies in Preschool Education*, no. 4, pp. 14–18, 2007.

[8] L. Mottron and J. Burack, "Enhanced perceptual functioning in the development of persons with autism," *The development of autism: Perspectives from theory and research*, pp. 131–148, 2001.

[9] L. Mottron, M. Dawson, I. Soulieres, B. Hubert, and J. Burack, "Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception," *Journal of autism and developmental disorders*, vol. 36, no. 1, pp. 27–43, 2006.

[10] A. Bonnel, L. Mottron, I. Peretz, M. Trudel, E. Gallun, and A.-M. Bonnel, "Enhanced pitch sensitivity in individuals with autism: a signal detection analysis," *Journal of cognitive neuroscience*, vol. 15, no. 2, pp. 226–235, 2003.

[11] A. Bonnel, S. McAdams, B. Smith, C. Berthiaume, A. Bertone, V. Ciocca, J. A. Burack, and L. Mottron, "Enhanced pure-tone pitch discrimination among persons with autism but not asperger syndrome," *Neuropsychologia*, vol. 48, no. 9, pp. 2465–2475, 2010.

[12] P. Heaton, "Pitch memory, labelling and disembedding in autism," *Journal of Child Psychology and Psychiatry*, vol. 44, no. 4, pp. 543–551, 2003.

[13] U. Frith, *Autism: Explaining the enigma*. Blackwell Publishing, 2003.

[14] F. Happé, "Autism: cognitive deficit or cognitive style?" *Trends in cognitive sciences*, vol. 3, no. 6, pp. 216–222, 1999.

[15] U. Frith and F. Happé, "Autism: Beyond "theory of mind"," *Cognition*, vol. 50, no. 1-3, pp. 115–132, 1994.

[16] F. Happé and U. Frith, "The weak coherence account: detail-focused cognitive style in autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 36, no. 1, pp. 5–25, 2006.

[17] P. Heaton, K. Hudry, A. Ludlow, and E. Hill, "Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders," *Cognitive neuropsychology*, vol. 25, no. 6, pp. 771–782, 2008.

[18] A. Järvinen-Pasley and P. Heaton, "Evidence for reduced domain-specificity in auditory processing in autism," *Developmental science*, vol. 10, no. 6, pp. 786–793, 2007.

[19] L. Cui, Z. Huang, and X. Du, "Thetone recognition ofchildren with cochlearim plants and hearing aidsin background noise," *Chinese Scientific Journal of Hearing and Speech Rehabilitation*, no. 2, pp. 39–42, 2011.

[20] C. Chen, "A study on the relationship between phonological awareness and reading skills for chinese children with ds," Master's thesis, Nanjing Normal University, 2014.

[21] Q. Zhang and Y. Wang, "Correlation analysis of children's mandarin tone consciousness and reading ability," *Modern Chinese*, no. 7, pp. 31–33, 2013.

[22] J. Xi, W. Jiang, L. Zhang, and H. Shu, "Categorical perception of vot and lexical tones in chinese and the developmental course," *Acta Psychologica Sinica*, vol. 41, no. 07, pp. 572–579, 2009.

[23] Z. Mou, Z. Chen, W. Yang, and C. Wang, "Research progress and application of mandarin tone perception," *Chinese Journal of Rehabilitation*, vol. 31, no. 6, pp. 418–421, 2016.

[24] 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 and hearing*, vol. 25, no. 3, pp. 251–264, 2004.

[25] Y.-M. Feng, L. Xu, N. Zhou, G. Yang, and S.-K. Yin, "Sine-wave speech recognition in a tonal language," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. EL133–EL138, 2012.

[26] J. Wang, H. Shu, L. Zhang, Z. Liu, and Y. Zhang, "The roles of fundamental frequency contours and sentence context in mandarin chinese speech intelligibility," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. EL91–EL97, 2013.

[27] P. Heaton, B. Hermelin, and L. Pring, "Autism and pitch processing: A precursor for savant musical ability?" *Music Perception: An Interdisciplinary Journal*, vol. 15, no. 3, pp. 291–305, 1998.

[28] C. J. Stevens, P. E. Keller, and M. D. Tyler, "Tonal language background and detecting pitch contour in spoken and musical items," *Psychology of Music*, vol. 41, no. 1, pp. 59–74, 2013.

[29] S. T. Cheng, G. Y. Lam, and C. K. To, "Pitch perception in tone language-speaking adults with and without autism spectrum disorders," *i-Perception*, vol. 8, no. 3, p. 2041669517711200, 2017.

[30] D. Van Lancker, "Cerebral lateralization of pitch cues in the linguistic signal," *Research on Language & Social Interaction*, vol. 13, no. 2, pp. 201–277, 1980.

[31] R. B. Ivry and L. C. Robertson, *The two sides of perception*. MIT press, 1998.

[32] C. N. Li and S. A. Thompson, "The acquisition of tone in mandarin-speaking children," *Journal of Child Language*, vol. 4, no. 2, pp. 185–199, 1977.

[33] Y. R. Chao, "The cantian ideolect: An analysis of the chinese spoken by a twenty-eight-months-old child," *University of California Publications in Semitic Philolology*, vol. 1, pp. 27–44, 1951.