Tone Perception by Putonghua-Learning Preschool Children in South Xinjiang Uyghur Autonomous Region

Aijun Li[1,2](✉), Zhiwei Wang[3], and Jun Gao[1]

[1] Institute of Linguistics, China Academy of Social Sciences (CASS), Beijing, China
{liaj,gao-jun}@cass.org.cn
[2] Corpus and Computational Linguistics Center, Institute of Linguistics, CASS, Beijing, China
[3] University of China Academy of Social Sciences, Beijing, China
wangzhiwei1@ucass.edu.cn

Abstract. Boundary tones are pitch patterns conveying pragmatic and paralinguistic information. In Putonghua (PTH), they are the results of the interaction between tone and intonation and can be categorized into two types: simultaneous addition boundary tone (SiABT) and successive addition boundary tone (SuABT). PTH differentiates word meanings using tonal contrasts, challenging learners whose native languages lack such contrasts. This study examines whether PTH-learning preschoolers (PTH learners) from south Xinjiang Uyghur Autonomous region can correctly decode the lexical tone categories intertwined with SiABT and SuAB. The correct decoding must be based on the integration of segmental, tonal, and semantic information. Nineteen children aged 3;9–5;3 with about 7 mouths of PTH learning in kindergarten, participated in a word recognition task with two conditions. In Condition 1, 24 familiar monosyllabic words in 12 minimal pairs contrasting in lexical tones and conditioned with SiABT or falling SuABT were used. The results indicated that for the PTH learners, the SuABT facilitated lexical tone decoding, especially for boys. However, compared to younger PTH-speaking children (1;6–2;8), the PTH learners had weaker tonal category decoding ability. In Condition 2, 8 familiar monosyllabic words in 4 pairs without phonological contrast were used. The accuracy rates for PTH learners were similar to those of PTH speakers, except that the SuABT didn't improve the accuracy rate for PTH learners, which indicated that the learners may depend primarily on segmental information and face challenges in integrating the tonal and pragmatic functions of intonation. The results support the automatic selective perception (ASP) model for L2 learners and suggest that tone acquisition is challenging for PTH learners, offering an effective method for teaching tones to preschool learners in the Xinjiang Uyghur Autonomous Region.

Keywords: tone perception · boundary tones · Putonghua-learning preschoolers · south Xinjiang Uyghur Autonomous Region · Automatic Selective Perception (ASP) model

Z. Ling et al. (Eds.): NCMMSC 2024, CCIS 2312, pp. 250–260, 2025.
https://doi.org/10.1007/978-981-96-1045-7_21

1 Introduction

Boundary tones are pitch patterns employed at the end of spoken phrases or sentences to convey both linguistic and paralinguistic information. In tone languages, the phonetic realization of boundary tones reflects a mixed relationship of tone and intonation. Chao [1] proposed that in the case of Putonghua (PTH), this relationship can be manifested through simultaneous addition or successive addition of tonal and intonational features at the intonation boundary. Simultaneous addition boundary tone (SiABT) is achieved as the algebraic sum of tonal and intonational features (a superimposed relation). It alters the phonetic realization without changing the phonological features of the original tone. Lin and Li [2] recognized three patterns of SiABT based on the acoustic analysis of PTH tones in declarative and interrogative intonations. Successive addition boundary tone (SuABT), on the other hand, is achieved by adding a falling or a rising tonal "tail" in a row to the lexical tone without changing the lexical meaning, but the surface pitch contour. Although both boundary tones carry the function of expressing different modalities and moods [2], the relationship between tone and intonation is more complex for SuABT than SiABT. When we examined the interactive dialogues in the longitudinal Child speech corpus, Cass_Child_Word, we found that both children and their caregivers frequently used SuABT to convey a wealth of pragmatic and paralinguistic information, and we summarized 19 functions such as 'confirmation, emphasizing, calling, questioning, negation, request, complaint, surprise, anger, happiness, correcting, poetry recitation…etc.' [3].

The Automatic Selective Perception (ASP) model [4] posits that Second Language (L2) learners process the auditory speech stream in one of two modes, phonological or phonetic, depending on a variety of factors including the learners' linguistic knowledge, the nature of the stimuli and task demands. The online processing of L2 sounds, especially for late L2 learners, demands a greater cognitive effort to distinguish contrasts compared to native language processing. The acquisition of non-native (L2) lexical tones is influenced by both segmental and semantic dimensions, with L2 tone learning being hindered for listeners who are provided with semantic information during training [5].

In our previous studies [7], it has been found that for children whose native language is PTH, the falling SuABT, caused by pragmatic functions of emphasizing the final word in child-directed speech, does not hinder the decoding of lexical tones. Contrarily, the falling SuABT assists children in the recognition of the words, especially for younger children (under the age of 32 months). As a result, in word teaching, a counterpart contrasting minimally with lexical tones and with various intonation forms embodying the pragmatic functions will promote the establishment of the mental lexicon.

This study focus on how the situation is for the preschool PTH learners from the Uygur Ethnic Group in south Xinjiang Uyghur Autonomous Region. The Children from the Uygur Ethnic Group began to PTH after entering kindergarten.

Uyghur is a stress-timed language without lexical tones. Each word generally has one primary stress on one of its syllables. The stress in Uyghur words is predictable based on the word's syllable structure and morphology. Due to the influence of their native language, Uyghur children's acquisition of lexical tones is supposed to differ from that of PTH-speaking children. There are two research questions. The first is how Uyghur children decode lexical tone categories (lexical prosody) SuABT (post-lexical prosody) that convey pragmatic information. The second is what are the similarities and differences compared to PTH-speaking children?

To answer the questions, this study adopted a word-recognition paradigm to test whether the falling SuABT added to the lexical tones will influence the PTH learners' word recognition. We hypothesized that the learners may face challenges in integrating the pragmatic functions of intonation in word recognition. Compared with PTH-speaking children, PTH learners are expected to have weaker ability to decode tonal categories based on the ASP model, for the L2 learners must employ greater attentional resources in order to extract sufficient information to differentiate phonological contrasts that do not occur in their native language.

2 Method

2.1 Participants

The recruited participants included 19 children (9 boys and 10 girls) from the first year of a kindergarten in south Xinjiang Uyghur Autonomous Region. None of the children had a history of hearing impairment or intellectual disabilities. The children started learning PTH after entering kindergarten and have learned for about 7 months. The participants' ages ranged from 3;9 to 5;3, with an average age of 4;10.

2.2 Materials

The key target words were all monosyllabic familiar words. They were of two conditions, the tonal-minimal (TM) target words and the non-tonal-minimal (non-TM) target words. There were 24 TM target words in four lexical tones (T1, high level; T2, high rising; T3, low dipping; T4, high falling), making 12 tonal minimal pairs, T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4 (see Table 1). There were 8 non-TM target words also in four lexical tones, making 4 pairs. In each pair, the two members were of the same lexical tones but different segments (see Table 2). Each target word was produced in two intonation types: SiABT and falling SuABT (See Fig. 1). We suppose the falling SuABT is used by parents to emphasize the last word. There were also 30 familiar disyllabic words in 10 pairs tested as fillers and 6 familiar monosyllabic words in 3 pairs for the pretest phase. A middle-aged female native PTH speaker recorded the speech stimuli in a sound-proof chamber. All the test words were embodied in colorful pictures.

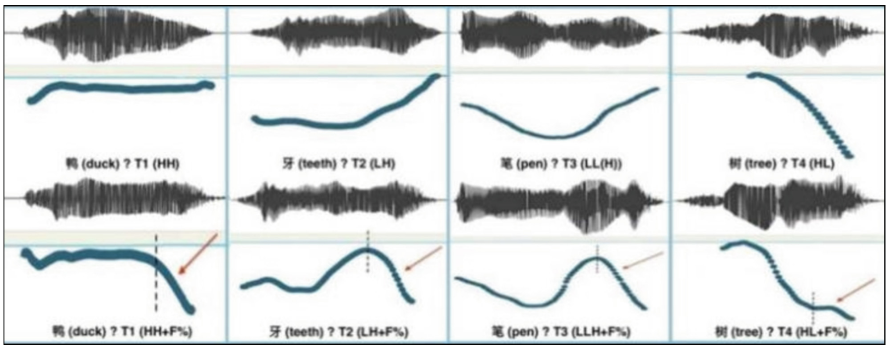

Fig. 1. Two forms of intonation boundary tones (upper panel: SiABT; lower Panel: falling SuABT) for four lexical tones. The falling tail of SuABT is indicated by the arrow. Duration is normalized for boundary tones in the figure.

Table 1. Familiar word (FW) pairs with minimal tonal contrasts (TM target words).

Tone contrasts	FW-FW	IPA	Pinyin
T1-T2	鸭-牙 duck-tooth	/ja1/-/ja2/	ya1-ya2
T1-T2	窗-床 window-bed	/tʂʰuɑŋ1/-/ tʂʰuɑŋ2/	chuang1-chuang2
T1-T3	三-伞 three-umbrella	/san1/-/san3/	san1-san3
T1-T3	冰-饼 ice-pancake	/piŋ1/-/ piŋ3/	bing1-bing3
T1-T4	书-树 book-tree	/ʂu1/-/ʂu4/	shu1-shu4
T1-T4	花-画 flower-painting	/xuɑ1/-/xuɑ4/	hua1-hua4
T2-T3	鼻-笔 nose-pen	/pi2/-/pi3/	bi2-bi3
T2-T3	鱼-雨 fish-rain	/y2/-/y3/	yu2-yu3
T2-T4	狼-浪 wolf-wave	/lɑŋ2/-/lɑŋ4/	lang2-lang4
T2-T4	驴-绿 donkey-green	/ly2/-/ly4/	lv2-lv4
T3-T4	土-兔 soil-rabbit	/tʰu3/-/tʰu4/	tu3-tu4
T3-T4	耳-二 ear-two	/ɚ3/-/ɚ4/	er3-er4

Table 2. Familiar Word pairs with same tones (non-TM target words).

Tones	FW-FW	IPA	Pinyin
T1-T1	虾-山 shrimp-mountain	/ɕia1/-/ʂan1/	xia1-shan1
T2-T2	梨-球 pear-ball	/li2/- /tɕʰiou2/	li2-qiu2
T3-T3	手-脚 hand-foot	/ʂəu3/-/tɕiaʊ3/	shou3-jiao3
T4-T4	面-蛋 noodle-egg	/miæn4/-/ tan4/	mian4-dan4

2.3 Procedure

A word-recognition paradigm was employed. The child was tested individually in a relatively quiet room in the kindergarten. During the experiment, the child sat in front of a touch screen which was a display expansion to a laptop put behind it. The X-perception computer program [6] was used to run the experiment on the laptop and to present images and speech stimuli. In each trial, a pair of pictures was first presented simultaneously side by side on the screen (see Figs. 2 and 3), one being the target picture and the other the distractor. After 1s silence, the carrier question “哪个是” (Which one is) of 817 ms was played and then after 350 ms silence the target word of the name of one of the pictures began. The speech stimuli were played through the speaker of laptop. In response to the prompt question, the child was asked to click on the picture on the touch screen that he/she thought was the right picture representing the meaning of the target word. Once the child clicked on the screen, the on-going trial ended immediately, and the next trial started right away. The computer program recorded whether the click was on the right or the left picture. During the experiment, the experimenter kept a record of the invalid trials where the child made a click at will. And the data of the invalid trials would be excluded.

Fig. 2. Familiar word pairs with minimal tonal contrast (TM).

Fig. 3. Word pairs without tonal or segmental contrast (non-TM).

At the beginning of the experiment, a training phase with 3 trials was conducted to familiarize the child with the procedure. The experiment focused on two conditions of words: a tonal minimal condition, and a non-tonal-minimal condition. The children were tested with 12 TM trials (2 target words $\times$ 3 pairs $\times$ 2 forms), 8 non-TM trials (the two forms of one member of the 4 non-TM pairs) and 10 filler trials. Trials order was quasi-randomized in which key trials were separated by filler trials and adjacent trials did not contain the same objects or test the same target word. The position of the pictures was balanced across trials. Eight kinds of trial settings were designed. The participants were assigned randomly to one of the eight experimental settings.

Please note that, the recognition task of choosing images of tone words requires the children to contend with multiple dimensions of information, including segmental, tonal, and semantic. In the TM condition, where the target words are minimal tone contrasts, PTH learning children needed to decode the segments and lexical tone of the target word in two kinds of boundary tones, and then recognized one out of the lexical tonal minimal pairs. In the non-TM condition, where the target word have different segments but same tones, the PTH learning children may decode the segments and lexical tone of the target word in two types of boundary tones as well, however, without requiring to pay more attention to tone information in word recognition.

3 Results

There were 1048 valid responses. Based on the valid responses, the accuracy rates of word recognition for the two types of boundary tones in TM condition and non-TM condition were calculated. The recognition rate was defined as the ratio of the number of trials correctly recognized by the children to the total number of relevant trials.

Moreover, given that children in south Xinjiang Uyghur Autonomous Region are PTH learners, we compared the results in present study with previous studies on PTH-speaking children in Beijing. Both groups were tested with the same test trials, but the PTH speakers were younger in age (18–32 m). For further information, please consult our earlier research [7].

The analysis revealed that none of the recognition rate results conformed to a normal distribution ($p < 0.1$). As a result, the data of two groups were compared using non-parametric statistics, particularly median and interquartile range Mann-Whitney U Tests.

3.1 Recognition Rate for all Word Pairs

PTH-learning children typically achieved a recognition rate of around 77.17% across two types of stimuli (SiABT and SuABT). The distributions of overall recognition rate for different gender of PTH-learning children were depicted in Fig. 4. Specifically, although girls (82.46%) demonstrated a higher accuracy rate compared to boys (77.85%), the difference was not significant ($p = 0.43 > 0.05$). Girls' accuracy rates were more consistent and clustered, while boys showed more variability. Further examination of accuracy rates in the two boundary tone conditions is presented in Fig. 5. On average, the accuracy rates of PTH-learning children in both boundary tone conditions did not differ significantly (SiABT:77.3%, SuABT:77.01%, $p = 0.87 > 0.05$). Girls demonstrated higher recognition rate in SiABT condition (81.6%), whereas boys (68.87%) performed better in SuABT condition than in SiABT. In both follow-up boundary tone conditions, girls consistently outperformed boys, with the difference being more noticeable in the SiABT condition.

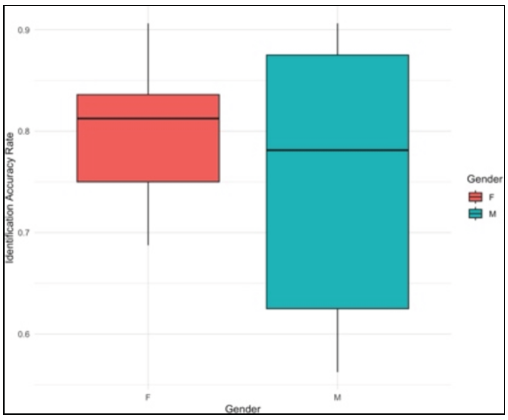

Fig. 4. The distribution of overall recognition rate for PTH-learning children.

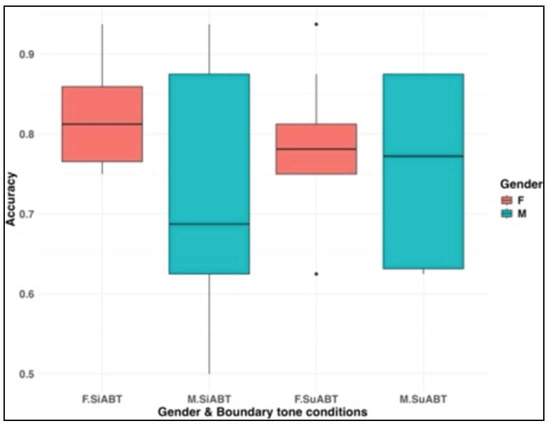

Fig. 5. Recognition rate in the two boundary tone conditions for PTH-learning children.

3.2 Words Recognition Results in TM Condition

The results for TM word pairs of PTH-learning children are displayed in Fig. 6. The average recognition rate is 58.8% in SuABT conditions, and 61.4% in SiABT conditions. There is no significant difference in the average recognition rates of PTH learners across the two boundary tone conditions ($p = 0.07 > 0.05$). The existence of SuABT did not significantly reduce the recognition rate, except in cases involving T2 as the target word and T2 as the distractor in the T1-T2 minimal pairs. Regarding the TM word pairs, pragmatic function of SuABT notably enhanced accuracy rates in most tone pairs compared with SiABT ($ps < 0.05$). This didn't echo our hypothesis that SuABT would hinder the lexical tone encoding of PTH learners.

Here we compare the above TM results with that of the PTH-speaking group, aged 18-32m, shown in Fig. 7 [7]. The average recognition rate of PTH-speaking group is 95.24% in SuABT condition, and 93.33% in SiABT condition. PTH-learning children exhibited significantly lower recognition rates than PTH-speaking children in both boundary tone conditions ($p = 0.035 < 0.05$). And there is no significant difference in the average recognition rates of PTH-speaking children across the two boundary tone conditions ($p = 0.87 > 0.05$). But for PTH-learning children, recognition rates in the form of SuABT was higher than those in the form of SiABT. For further discussions on PTH-speaking children, refer to our prior research [7].

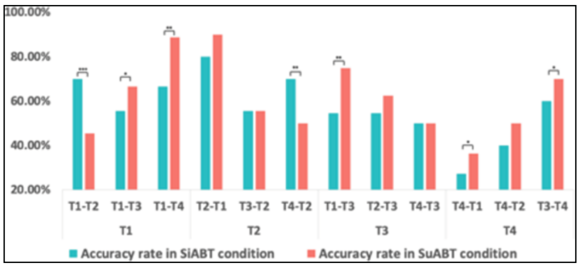

Fig. 6. The accuracy rates for TM word pairs of PTH-learning children.

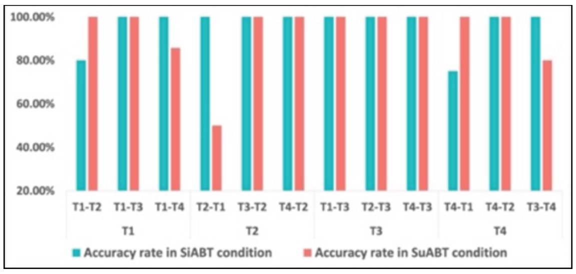

Fig. 7. The accuracy rates for TM word pairs of PTH-speaking children (18–32 m) [7].

3.3 Words Recognition Results in Non-TM Condition

As shown in Fig. 8, the results for non-TM word pairs of PTH-learning children suggest that, overall, there is little difference in recognition rates across the two boundary tone conditions. The children's word recognition rates for all four lexical tones exceeded 80%.

Here we also compared the above non-TM results with those of the PTH-speaking group, aged 18–32 m, shown in Fig. 9 [7]. It can be observed that the recognition rates between two groups of children are quite similar in this task. This indicates that PTH-learning children being exposured to PTH for about 7 months, reached the level of 18–32 m-old PTH-speaking children in non-TM condition of word recognition. But the biggest difference between the two groups is that PTH-speaking children can use SuABT to improve word recognition, while the learners cannot.

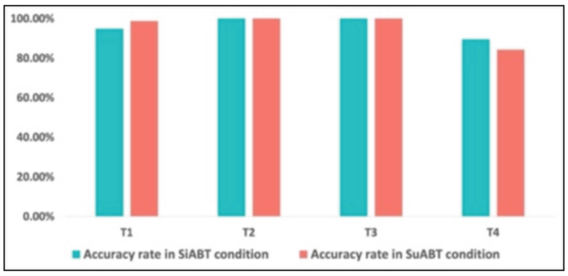

Fig. 8. The accuracy rates grouped by tones for non-TM word pairs of PTH-learning children.

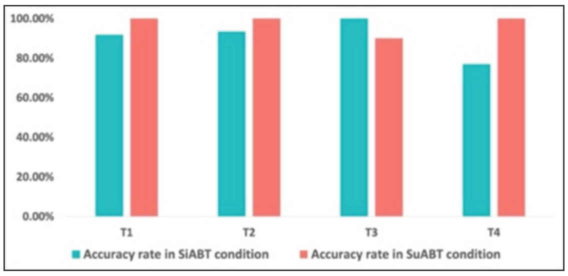

Fig. 9. The accuracy rates grouped by tones for non-TM word pairs of PTH-speaking children (18–32 m) [7].

4 Discussion

4.1 Overall Recognition Rates

The overall recognition rate for PTH-learning children is approximately 77.17%. This value indicates a lower level of proficiency, compared with PTH-speaking children (94.30%). For PTH-learning children, girls achieved higher recognition rates compared to boys. Girls displayed more consistent and clustered recognition rates, while boys exhibited more variability.

For PTH-learning children, the average recognition rates under the two boundary tone conditions (SiABT and falling SuABT) did not show a significant difference. This suggests that, overall, boundary tone complexity does not impact PTH-learning children's perception of lexical tones significantly. Regarding the differences in sex, in the condition of SiABT, girls performed better than boys, which indicates that girls may have a better ability to process SiABT. In the condition of a falling SuABT (complex pitch contour), girls' overall tone decoding ability slightly decreased. This suggests that girls may find complex pitch contours more challenging and therefore perform worse when the pitch complexity is increased. In contrast, boys showed a relative improvement with the SuABT boundary tone, suggesting an interesting dynamic where the more complex boundary tone might offer a positive boost for boys to differentiate lexical tones.

4.2 The TM Condition

In the TM condition, after 7-months-long exposure to PTH, PTH-learning children in south Xinjiang Uyghur Autonomous Region, with an average age of 4;10, cannot perceive the lexical tones well enough as PTH-speaking children aged 18–32 months. According to the Automatic Selective Perception (ASP) model, providing semantic information during the acquisition of second language (L2) tonal contrasts may increase the processing load, inhibiting tone perception, even though listeners were not explicitly required to attend to the tonal information [5].

Regarding the pragmatic function of boundary tones, the PTH-learning children can recognize tones with falling SuABT facilitated the tonal recognition. However, their ability to understand the interaction between lexical tone and intonation is still developing and is weaker than even younger PTH-speaking children (18–32 months).

4.3 The Non-TM Condition

According to the non-TM recognition results, complex boundary tones are more beneficial for PTH-speaking children in word recognition, suggesting that they combine tonal and intonational information to recognize words. Although PTH-learning children exhibit a high recognition rate as well, indicating that complex boundary tones do not negatively impact their recognition. This indicates that PTH-learning children may rely primarily on segmental information to recognize words rather than processing tonal and intonational information.

Different from the TM word pairs, the word pairs differ only in segments but not in tones, reduce the processing load for the PTH-learning children. The minimum impact of boundary tones on recognition rates further emphasizes that segmental information is the primary factor in word recognition. This suggests that PTH-learning children in Xinjiang Uyghur Autonomous Region have developed segmental decoding skills better than tone.

Moreover, this suggests that for PTH-learning children whose native language is a non-tone language, they only prime tone recognition when there is no segmental information available at least for early second language (L2) learners. On the other hand, 2.5-year-old PHT-speaking children have relatively strong but still-developing abilities to integrate segmental and tonal information. The results indicates that tone representation may not fully realized for 2.5-year-old PTH-speaking children, and even worse for PTH-learning children aged 3;9–5;3 with a 7-month-long exposure to PTH.

5 Conclusions

This study concluded three key findings:

PTH-learning children (3;9–5;3) from the Uygur Ethnic Group in south Xinjiang Uyghur Autonomous Region have a lower proficiency in tone recognition in two boundary tone conditions. In general, girls performed better than boys. However, the more complex boundary tone (falling SuABT) might offer a positive boost for boys.

Under TM condition, PTH-learning children exhibit weaker tone recognition compared with younger PTH speaking children (18–32 m). Semantic information during the acquisition of L2 tonal contrasts may have increased the processing load, inhibiting tone learning. However, the pragmatic function of emphasizing the target word provided by the falling SuABT significantly enhances tone perception. Under non-TM condition, PTH-learning children exhibit higher proficiency in recognizing words that differ only in segments which reduces the processing load. These results can be explained by ASP model.

These findings inform an effective method for teaching tones to preschool learners in Xinjiang Uyghur Autonomous Region. Teaching only tone without lexical meaning (such as a1, a2, a3 and a4 just reach phonetic level rather phonological level of tone processing), will not enhance tone processing ability. A learning task should be carried out together with lexical meaning and should be in the contexts of different intonations conveying rich pragmatic functions.

For future work, we will verify the results with additional data from bigger sample size, more age groups and more testing conditions involving unfamiliar “wug” words.

Acknowledgements. This work was supported by the Independent Topic Selection Project of Cultural Experts and "Four Batches" of Talents awarded to Aijun Li, and also Key Laboratory of Linguistics, Chinese Academy of Social Sciences (2024SYZH001).

References

1. Chao, Y.: Tone and intonation in Chinese. Bull. Inst. History Philol. **4**, 121–134 (1933)
2. Lin, M., Li, Z.: Focus and boundary tone in Chinese Intonation. In: ICPhS XVII, Hong Kong (2011)
3. Li, A.: A preliminary exploration of the successive addictive boundary tones in Mandarin Children's interactive dialogue, to appear in Interactional linguistics and Chinese Language Studies (Vol. 5), Social Science Literature Publishing House (2024)
4. Strange, W.: Automatic selective perception (ASP) of first and second language speech: a working model. J. Phon. **39**(4), 456–466 (2011)
5. Cooper, A., Wang, Y., Behne, D.M.: Effects of semantic information and segmental familiarity on learning lexical tone. In: Nyvad, A.M., Hejná, M., Højen, A., Jespersen, A.B., Sørensen, M.H. (eds.), A Sound Approach to Language Matters, Honor of Ocke-Schwen Bohn, pp. 211–236. Aarhus University Press, Denmark (2019)
6. Xiong, Z.: xPerception: Yuyin Tingbian Ganzhi Guocheng Jilu Gongju [xSegmenter: A Speech Recognition Perception Tool for Recording]. https://mp.weixin.qq.com/s/kDuviMtajg5NJixUKWuIfg. Accessed 19 Nov 2015
7. Li, A., Gao, J., Wang, Z.: Effect of complex boundary tones on tone recognition: an experimental study with mandarin-speaking preschool children. In: Interspeech, 2024 (under review) (2024)
8. Li, A.: Encoding and decoding of emotional speech: a cross-cultural and multimodal study between Chinese and Japanese. Springer, New York (2015)
9. Li, A.: A preliminary exploration of the successive addictive boundary tones in Mandarin Children's interactive dialogue. In: Interactional linguistics and Chinese Language Studies (Vol. 5), Social Science Literature Publishing House (2024)
10. Woo, H., LeTendre, G.K., Trang, P., Xiong, Y.: The use of social robots in classrooms: a review of field-based studies. Educ. Res. Rev. **33**, 100388 (2021)
11. Xu, R.N., Tang, P., Yuen, I., Gao, L., Demuth, K.: Five-year-olds' acoustic realization of mandarin tone sandhi and lexical tones in context are not yet fully adult-like. Front. Psychol. **9**, 817 (2018)
12. Wong, P.: Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. J. Acoust. Soc. Am. **133**(1), 434–443 (2013)
13. Wong, P., Strange, W.: Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: a perceptual study. PloS one. **12**, e0182337 (2017)
14. Wang, Z.: A Study on Tone and Tonal Realization Patterns of Mandarin Preschool Children. Beijing International Study University, Beijing (2023)
15. Wang, Z., Gao, J., Li, A.: Disyllabic tones in mandarin preschool children and child-directed speech. In: ICPhS 2023, Prague, Czech Republic (2023)
16. Mueller-Liu, P.: Signalling affect in Mandarin Chinese - the role of non-lexical utterance-final edge tones. In: Speech Prosody 2006, paper 048 (2006)
17. Gao, J., Li, A., Xiong, Z.: Mandarin multimedia child speech corpus: Cass_Child. In: 2012 International Conference on Speech Database and Assessments, Macau, China, pp. 7–12 (2012)
18. Gussenhoven. C., Chen, A.: The Oxford Handbook of Language Prosody. (Oxford Handbooks). Oxford University Press (2020)

19. Chen, A., Fikkert, P.: Intonation of early two-word utterances in Dutch. In: J. Trouvain, & W. J. Barry (Eds.), Proceedings of the 16th International Congress of Phonetic Sciences, Priiot, pp. 315–320 (2007)
20. Rhee, N., Chen, A., Kuang, J.: Going beyond F0: the acquisition of mandarin tones. J. Child Lang. **48**(2), 387–398 (2021)
21. Wang, L., Kalashnikova, M., Kager, R., Lai, R., Wang, P.: Lexical and prosodic pitch modifications in Cantonese infant-directed speech. J. Child Lang. **48**, 1–27 (2021)