Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Production of Contrastive Focus in Mandarin-Speaking Children

Sichen Zhang[1], *Aijun Li*[2], *Jun Gao*[3]

[1]University of Chinese Academy of Social Sciences
[2]Corpus and Computational Linguistics Center, Institute of Linguistics, Chinese Academy of Social Sciences
[3]Institute of Linguistics, Chinese Academy of Social Sciences
zhangsichen@ucass.edu.cn, liaj@cass.org.cn, gao-jun@cass.org.cn

Abstract

This study investigated Mandarin-speaking children's spontaneous speech production of contrastive focus in five-syllable SVO declaratives (constructed as Adjective1+Noun1+Verb+Adjective2+Noun2). Ten children aged 4;7-6;0 participated in a picture-guessing game with their parents, where they were prompted to produce the target structure with contrastive focus at different positions. The results showed a generally accurate placement of the focal accent, which improved with age. Nevertheless, 10.2% of the total contrastive foci were misplaced in wrong prosodic words, and 8.3% in correct prosodic words but on a wrong syllable. Additionally, 37.5% of contrastive foci on Noun2 were realized at incorrect positions, while very few on Verb were misplaced. Misaccenting Noun2 was the most common error, followed by Adjective1 and Adjective2. Acoustic analysis revealed children's awareness of using prosodic features including pitch range expansion, durational adjustment, and post-focus compression to encode contrastive focus. However, the performance varied across individuals and developed gradually. Furthermore, hesitation pauses were found in a third of the utterances. The findings suggest that the adult-like mastery of contrastive focus realization is still developing by age 6, which is closely related the complex interaction of lexical and phrase-level prosody in Mandarin Chinese.

Index Terms: prosodic development, contrastive focus, spontaneous production, Mandarin Chinese

1. Introduction

Intonation, as proposed by [1], is the linguistically structured modulation of suprasegmental phonetic features including fundamental frequency (F0), duration, and intensity that conveys pragmatic meanings at the post-lexical level such as sentence type, speech act, and information structure. Focus, as a universal information-structural category, facilitates the updates of the common ground with new information indicating the relevant alternatives for a given assertion [2-5]. Despite a lack of consensus in the classification of focus types and the variation of focus marking strategies across languages, contrastive focus has been generally recognized as a more marked subtype of narrow focus [6-8]. A contrastively focused constituent is often realized by means of extra prosodic prominence and particular types of pitch accent [1,7-9]. Pragmatically, contrastive focus is used to contradict or correct an assertion in the preceding discourse. The pragmatic use of contrastive focus is closely related to a number of cognitive faculties such as short-term memory and attention [5,10].

In the framework of intonation phonology [1,11-14], a feature-based model proposes that Chinese intonation consists of two key components: focal accent and boundary tone [15,16], corresponding tonal events to pitch accent and boundary tone as in languages like English. Focal accent is associated with narrow focus and is marked with prosodic prominence. Since F0 is an acoustic correlate shared by lexical and phrase-level prosody in Mandarin Chinese, it expands in range on the focused constituent but maintains the basic contour of lexical tone [15-17]. Consequently, Mandarin focal accent has a significantly higher F0 peak in Tone 1 (HH), Tone 2 (LH), and Tone 4 (HL) and a relatively lower trough in Tone 3 (LL/LLH), respectively labeled as [+RaiseH] and [+LowerL]. The duration and intensity are commonly found to increase on focused syllables, nonetheless, a complex pattern of duration adjustment of corrective focus is reported in [18]. Moreover, while the prosodic features of pre-focus syllables are almost intact, post-focus syllables are typically realized with lowered and compressed F0 range, i.e., post-focus compression (PFC) [19].

The acquisition of focus prosody has been cross-linguistically reported to be a gradual process that continues to develop by early adolescence [20,21]. Previous studies on Mandarin-speaking children's focus-prosody mapping have found that in terms of online processing, children over 5;0 use contrastive focus to facilitate information structure processing, but are not able to use it to predict the upcoming relevant alternatives [22,23]. 3;0-5;0 children rely on word order more than prosodic cues to identify the positions of narrow focus in a behavioral procedure [24]. Regarding production, [25-28] investigates the use of tonal and durational cues in the prosodic realization of broad, narrow, and contrastive foci by children of different ages. These studies reveal a developmental path where pitch-related features are not acquired as well and early as duration-related cues in distinguishing different focus types. The less transparency of focus-pitch mapping of Mandarin Chinese due to the interplay of lexical tone and intonation could explain this difficulty. Furthermore, [29] examines the production of contrastive and narrow focus at the sentence medial position. The lack of accuracy across age groups based on perceptual judgment was due to the lack of PFC, whereas the pitch and duration properties on focused syllables were already adult-like among 4;0-6;0 children.

In light of the contributions to our understanding of the development of the prosody-information structure interface and the foundation for experimental design and acoustic analysis established by the previous studies, the present study takes a further step to identify and explain the successes and failures in children's communicative use of contrastive focus prosody at more sentence positions in three dimensions, phonological

10.21437/SpeechProsody.2024-38

realization, phonetic implementation, and pragmatic use. Therefore, this paper investigates 1) Mandarin-speaking children's phonological realization of focus types and positions to correct an assertion in conversation, and 2) their phonetic implementation of prosodic properties in marking contrastive focus, with due consideration of the interaction between Mandarin lexical and post-lexical prosody.

2. Methods

2.1. Participants

Ten preschool children aged between 4;7-6;0 (6 females and 4 males, mean age: 5;5) as well as their caretakers, all of whom were parents, participated in this study. All participants are native Mandarin Chinese speakers and live in Beijing. None of them have been diagnosed with any speech/hearing disorders.

2.2. Materials

The target structure is a five-syllable SVO sentence, constructed by a disyllabic subject, a monosyllabic verb, and a disyllabic object. The subject and object consist of an adjectival modifier and a noun. Hence, the target structure can be represented as Adjective1 (A1) + Noun1 (N1) + Verb (V) + Ajective2 (A2) + Noun2 (N2), for example, *Hēi*$_{A1}$*māo*$_{N1}$ *qīn*$_{V}$ *xiǎo*$_{A2}$*yā*$_{N2}$ (黑猫亲小鸭, "The black cat kisses the little duck."). A contrastive focus is set on A1, V, A2, or N2 in each sentence, to investigate the effect of focus in the sentence-initial, -medial, or -final positions. Particularly, both A2 and N2 are set to bear the contrastive focus to further observe the interaction between focal accent and boundary tone. The focus-bearing syllables at each position cover four lexical tones in Mandarin Chinese. An extra all-Tone1 sentence is designed for the contrastive focus on the A1, V, and A2 conditions.

2.3. Procedure

A picture-guessing game was designed to elicit spontaneous production of contrastive focus referring to the experimental design in [24-26]. After a word-naming task familiarizing the participants with the target words and structures, parents were presented with two pictures contrasting in one element, that is, one word in the target sentence, as seen in Figure 1(a). Children were presented with only one picture contrasting with parents' pictures in the same element, as seen in Figure 1(b). Parents were instructed to arbitrarily choose one picture as their guess and ask a yes/no question. Given that the children's picture was different from either of the parents' pictures, the parents' guess would always be wrong. Thus, children were expected to produce a corrective answer to their parents' question with a contrastive focus on a certain element. An example of the question-answer dialogue is given in example (1). The participants were divided into two groups. One eliciting 10 target sentences with the contrastive focus on A1 or N2, the other eliciting 11 target sentences with the contrastive focus on V or A2. In order to make the game more natural, 3 fillers were inserted randomly among the trials, in which parents were prompted for the correct answer by texts on their picture.

Recording was conducted in a sound-proof room (3.25 m*3.5 m*2.4 m) decorated as a child-friendly playroom. Participants were asked to sit face-to-face, ask and answer questions about the presented pictures according to the instructions. Their conversation was collected separately by two desktop standing microphones (AKG C1000S). Speech data were sampled at 44.1 kHz with a 16-bit resolution in stereo two channels (children's speech in the right channel and parents' in the left).

(a) (b)

Figure 1: *Schematic diagram of speech production.*

Parent: *(Nǐ de túpiàn shì) hēimāo qīn xiǎoyáng*$_{N2}$*?* (1)
（你的图片是）黑猫 亲 小羊？
(Your picture is) the black cat kisses the little sheep?
Child: *(Búduì, wǒ de túpiàn shì) hēimāo qīn xiǎoyā*$_{N2}$*.*
（不对，我的图片是）黑猫 亲 小鸭。
(Wrong, my picture is) the black cat kisses the little duck.

2.4. Data analysis

The data were converted from stereo channels to mono using Praat [30], and then segmented by an automatic alignment program to generate syllable- and phone-level transcriptions [31]. In addition to manually checking the segmentation and transcriptions, two professional annotators labeled the prosodic structure, prosodic stress, and miscellaneous events based on C-ToBI [32,33].

F0 values were taken at ten points in equal intervals of the syllable finals using Praat scripts, excluding creaky and breathy voices. The extracted F0 values were firstly converted into semitones (St) from Herz (Hz) using the formula as seen in (2), then normalized into Z-score scale values for each speaker to eliminate individual differences, using the formula in (3).

$$St_i = 12 * log_2(\frac{f0_i}{f0_{ref}}) \quad (2)$$

Here St_i is the i^{th} F0 value in Hz, $f0_i$ the original i^{th} F0 value in Hz, and $f0_{ref}$ the minimum F0 value of each speaker.

$$Z_i = \frac{f0_i - \mu}{\sigma} \quad (3)$$

Here Z_i is the z-scored i^{th} F0 value, $f0_i$ the i^{th} F0 value, μ the mean F0 value, and σ the standard deviation of the F0 values of the individual speaker, all in St. This formula also applies to the z-scored normalization of the duration of syllable finals.

3. Results

3.1. Phonological realization of focus types and positions based on perceptual judgments

The production of contrastive focus was taken under the perceptual judgments by three well-trained experimenters (including the two annotators) in terms of focus types and focus positions. The judgment of focus types only distinguished between contrastive focus and broad focus. The latter rarely occurred in children's production, only when the child was required to repeat her/his answers. The positions of contrastive focus were identified as the syllable bearing focal accent, i.e., the sentence stress. The results were grouped by the correct and incorrect accent placement in Table 1, by the lexical tone categories of the syllable receiving contrastive focus in Table 2, and by children in Figure 2.

The initial column in Table 1 specifies the intended position for the contrastive focus, as per the experimental design. Columns under "wrong positions" present the misplacement of the focal accent, including different positions in the five-syllable utterances (A1, N1, V, A2, and N2) and one instance of multiple accents due to prolonged hesitation. The percentage figures accompanying the integers in Table 1 and Table 2 denote the proportion taken by the integer in the total 108 utterances (in black), each position group (in blue), or the errors in each position group (in red).

Table 1: *Focal accent placement by positions.*

Positions	**Total #**	**Error #**	**Wrong positions**					
			A1	N1	V	A2	N2	Multiple
A1	24	4 16.7%		2 50%		2 50%		
V	30	2 6.7%				2 100%		
A2	38	8 21.1%	1 12.5%		1 12.5%		5 62.5%	1 12.5%
N2	16	6 37.5%	3 50%	1 16.7%	2 33.3%			
Total	*108*	*20* *18.5%*	*4* 20%	*3* 15%	*1* 15%	*4* 20%	*5* 25%	*1* 5%

Table 2: *Focal accent placement by tone categories.*

Tone category	**Total #**	**Error #**
T1	37	7 18.9%
T2	20	3 15.0%
T3	31	8 25.8%
T4	20	2 10.0%

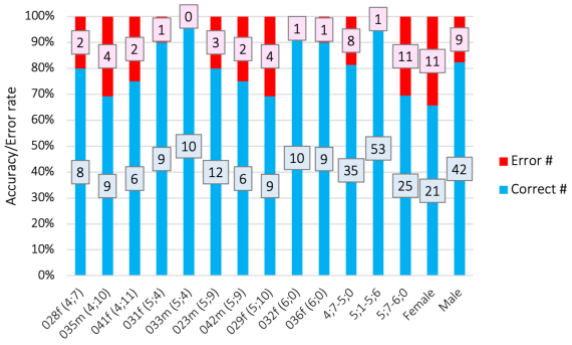

Figure 2: *Focal accent placement by children, age groups, and gender groups.*

According to Table 1, the majority (81.5%) of contrastive foci were realized with focal accent on the correct syllable. However, when N2 and A2 were supposed to carry contrastive focus, 37.5% and 21.1% of the focal accents were misplaced. Very few errors on the V condition were found. Moreover, misaccenting tends to occur on N2, followed by A1 and A2. There were 10.2% of the total contrastive foci were realized in wrong prosodic words, and 8.3% in the correct prosodic word but on a wrong syllable.

Other than the focus placement, lexical tone categories also affect the accuracy of focal accent placement, as seen in Table 2. Errors occurred more commonly when the target syllable receiving the focal accent was in Tone 3, followed by Tone 1 and Tone 2. Tone 4 syllables have the lowest error rate.

Figure 2 illustrates the proportional distribution of correct and erroneous accent placement among each child participant, roughly divided age groups by 6 months, and gender groups. The performance varies largely across the child participants with no continuous developmental trajectory seen. In addition, The differences among age groups (p=0.6939) and between gender groups (p=0.6655) were found insignificant.

3.2. Phonetic implementation of contrastive focus prosody properties based on acoustic analysis

This part analyzes children's implementation of the acoustic parameters, including F0 contour, F0 range, F0 register, and duration, in the focal accents that encoded contrastive foci at correct sentence positions.

Firstly, Figure 3 demonstrates the F0 properties of the target structure grouped by focus-bearing syllable. Figure 4 shows the results of the all-Tone1 sentence for a clearer picture of the adjustment of pitch register.

Figure 3: *F0 properties of contrastive foci realized at different sentence positions.*

Comparing the F0 properties of focused syllables at different sentence positions, the expansion of F0 range was larger in sentence-medial syllables than in sentence-beginning and -final syllables. The [+RaiseH] feature of the focal accent was found the highest in V conditions followed by A1, A2, and then N2. [+LowerL] was found the lowest in the A2, followed by V, N2, and then A1. Pitch register of the focal accent was the highest in the A1 condition, i.e., at sentence-beginning positions, and gradually descending towards the end of the sentence. Given that the post-focus syllable in the A1 and A2 conditions only had an H target, it is not plausible to compare the pitch range of focused and post-focus syllables. Although the PFC phenomenon can be seen in the V condition, the post-focus syllables were at the sentence-final position, which also accounted for the suppression of F0 range. The pre-focus syllables were acknowledged to be barely unchanged in terms of F0 and duration properties. However, in the A2 and N2 conditions, when the pre-focus syllable was in the same prosodic word as the focused syllable, its F0 contours changed with the following syllable, causing an expansion in range. Also, the peak delay of the H target in Tone 3 can be seen in the immediately following Tone1 in A1 condition and in the beginning of Tone 4 carried by the last syllable in the same prosodic word in V condition. The L target in Tone 4 can be seen delayed to the immediately following tones in A1, V, and A2 conditions.

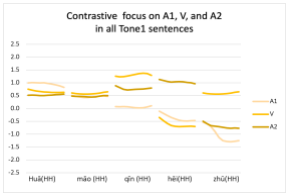

Figure 4: *F0 properties of contrastive foci realized at different sentence positions in all-Tone1 sentences.*

The all-Tone 1 sentences illustrated in Figure 5 visualize a common and obvious lowering of post-focus pitch register. However, the differences between focal register and post-focus register vary according to sentence position.

Secondly, Figure 5 presents the pattern of normalized duration of contrastive foci realized at different sentence positions. Specially, N1 was not intended to carry a contrastive focus, the data presented here were from utterances with accent misplacement on N1.

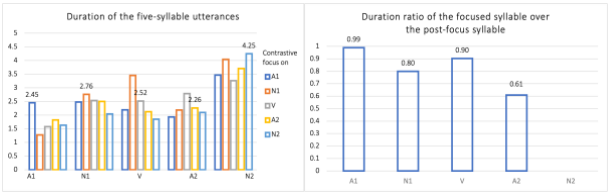

Figure 5: *Duration properties of contrastive foci realized at different sentence positions.*

The duration pattern of the five-syllable utterances shows the interaction of focal accent and boundary tone. The duration of the final in N2, i.e., the last syllable in the target utterance, was always the longest compared to its preceding syllables despite focus positions. When N2 bore the focal accent, its final is longer than the N2s in other focus conditions. When A1 and N1 bore a focal accent, they were also the longest among other A1s and N1s, but not necessarily longer than other syllables in the same utterance. Focused V and A2 were neither the longest V and A2 among different focus conditions nor longer than their adjacent syllables. Furthermore, according to the duration ratio patterns, no post-focus duration shortening was found in all focus conditions.

Finally, as presented in Table 3, hesitation pauses were frequently produced by children. One or more hesitation pauses were found in 31.5% of the five-syllable utterances. The position of these pauses was mostly after N1 and V, that is, at word or prosodic word boundaries.

Table 3: *Hesitation pauses in the production of utterances with contrastive focus.*

Position (after)	**Total #**	**# of pauses in one utterance**	**Total #**
A1	1 0.9%	1	30 27.8%
N1	24 22.2%	2	3 2.8%
V	12 11.1%	3	2 0.9%
A2	1 0.9%		

4. Discussion and conclusions

Adult-like mastery of focus-to-prosody mapping requires the knowledge of the phonological representation of prosodic forms, the control over the phonetic implementation of prosodic properties, and the ability to use prosodic cues to appropriately convey and efficiently perceive focus information in communication. The development of focus-to-prosody mapping is closely related to cognitive domains, especially the development of executive function. For example, attention allocation and inhibition play an important role in quickly detecting the prosodic prominence, perceiving its meaning, and switching from previous assertions to upcoming ones [2,5].

With that in mind, the first part of this study investigated whether Mandarin-speaking children aged 4;6-6;0 were capable of using prosodic devices to encode contrastive focus so as to convey pragmatic meanings in communicative contexts. In this picture-guessing game, children were expected to give corrective answers with a contrastive focus on a certain syllable based on their interlocutors' questions and the visual prompt. Possible failures include 1) producing a broad focus sentence without highlighting the relevant alternatives, and 2) misplacing the focal accent which means emphasizing the information. Our results showed in the given interactive context, children had no difficulties in realizing the focus type. However, the phonological realization of focus positions was occasionally problematic. The interaction of focal accent and boundary tone may account for failing to accent and misaccent syllables at the sentence-final position. Errors happening within prosodic words indicated children's underdeveloped ability to integrate Mandarin word and phrase prosody.

The second part of this study examined how children implement the acoustic features in the contrastive focus utterance, especially on the focused and post-focus syllables. The acoustic analysis reflected children's awareness and ability to modulate the H and L target of the focused syllable so that the F0 range expands. Due to the lack of L target in some post-focus syllables, results regarding PFC need to be interpreted with caution. However, there was evidence of compression of F0 range and lowering of F0 register in the realization of focus prosody by children. Other prosodic modulation including peak delay and final lengthening can also be seen in the results. Additionally, the findings about hesitation pause in children's spontaneous production reveal the immaturity in language planning when dealing with complex tasks.

Our investigations into preschool children's prosodic acquisition are still in progress. Future work will not only incorporate more children from a larger age span but also explore the interplay between focus prosody and sentence types, speech acts, etc. More factors such as the length of the whole utterance and the variation of the tones being corrected in caretakers' production will be taken into account in analysis.

To conclude, this study investigates 4.5-6-year-old Mandarin-speaking children's spontaneous speech production of contrastive focus in terms of pragmatic use, phonological marking, and phonetic implementation. Results suggest the adult-like phonological and phonetic marking of contrastive focus is still developing by the age of 6. The findings contribute to the understanding of the development of prosody-information structure in preschool children.

5. 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).

6. References

[1] R. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 2008.

[2] M. Krifka, "Basic notions of information structure," *Acta Linguistica Hungarica*, vol. 55, no. 3–4, pp. 243–276, Dec. 2008.

[3] K. Lambrecht, *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press, Cambridge, 1994.

[4] R. L. Stalnaker, "Common ground." *Linguistics and Philosophy*, vol. 25, no. 5–6, pp. 701–721, Dec. 2002.

[5] M. Zimmermann and E. Onea, "Focus marking and focus interpretation," *Lingua*, vol. 121, no. 11, pp.1651–1670, Sep. 2011.

[6] M. A. K. Halliday, "Notes on transitivity and theme in English: Part 2," *Journal of Linguistics*, vol. 3, no. 2, pp. 199–244, Oct. 1967.

[7] Y. Chen, *The Phonetics and Phonology of Contrastive Focus in Standard Chinese*, Doctoral Dissertation, State University of New York at Stony Brook, 2003.

[8] C. Gussenhoven, *The Phonology and Tone of Intonation*. Cambridge: Cambridge University Press, Cambridge, 2004.

[9] J. Katz and E. Selkirk, "Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English," *Language*, vol. 87, no. 4, pp. 771–816, Dec. 2011.

[10] M. Zimmermann and C. Féry, "Introduction," in *Information Structure: Theoretical, Typological, and Experimental Perspectives*, M. Zimmermann and C. Féry, Eds., Oxford: Oxford University Press, 2010, pp. 1–11.

[11] G. Bruce, "Swedish word accents in sentence perspective," *Working papers/Lund University, Department of Linguistics and Phonetics*, no. 12, pp. 61–70, 1977.

[12] J. B. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, Doctoral Dissertation, Massachusetts Institute of Technology, 1980.

[13] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology*, vol. 3, pp. 255–309, May. 1986.

[14] A. Arvaniti, "The autosegmental-metrical model of intonational phonology," In S. Shattuck-Hufnagel and J. Barnes (Eds.), *Prosodic Theory and Practice*. Cambridge, MA: MIT Press, in press.

[15] M. Lin and Z. Li, "Focus and boundary in Chinese intonation," *Proceedings of ICPhS*, vol. 17, pp. 1246–1249, Aug. 2011.

[16] M. Lin, *Hanyu Yudiao Shiyan Yanjiu [Experimental Studies on Chinese Intonation]*, Beijing: China Social Sciences Press, 2012.

[17] Y. Jia, Z. Xiong, and A. Li, "Putonghua jiaodian zhongyin dui yuju yingao de zuoyong [Effect of focus stress on pitch patterns in Mandarin]," *Chinese Journal of Phonetics*, vol. 1, pp. 118–124, Apr. 2008.

[18] Y. Chen, "Durational adjustment under corrective focus in Standard Chinese," *Journal of Phonetics*, vol. 34, no. 2, pp. 176–201, Apr. 2006.

[19] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *Journal of Phonetics*, vol. 27, no. 1, pp. 55–105, Jan. 1999.

[20] K. Ito, "Gradual development of focus prosody and affect prosody comprehension," in *The Development of Prosody in First Language Acquisition*, P. Prieto and N. Esteve-Gibert, Eds., Amsterdam: John Benjamins, 2018, pp. 247–270.

[21] A. Chen, "Get the focus right across languages: Acquisition of prosodic focus-marking in production," in *The Development of Prosody in First Language Acquisition*, P. Prieto and N. Esteve-Gibert, Eds., Amsterdam: John Benjamins, 2018, pp. 295–314.

[22] Q. Zhang, K. Lyu, Z. Chen, and P. Tang, "Acquisition of prosodic focus marking by three-to six-year-old children learning Mandarin Chinese," *Proceedings of INTERSPEECH 2021*, pp. 3925–3928, Sep. 2021.

[23] P. Tang, I. Yuen, K. Demuth, and N. X. Rattanasone, "The acquisition of contrastive focus during online sentence-comprehension by children learning Mandarin Chinese," *Developmental Psychology*, vol. 59, no. 5, pp. 845, May 2023.

[24] H-C. Chen, K. Szendrői, S. Crain, and B. Höhle, "Understanding prosodic focus marking in Mandarin Chinese: Data from children and adults," *Journal of Psycholinguistic Research*, vol. 48, pp. 19–32, Feb. 2019.

[25] A. Yang and A. Chen, "Prosodic focus marking in child and adult Mandarin Chinese," *Proceedings of the 4th International Symposium on Tonal Aspects of Language*, pp. 54–58, May 2014.

[26] A. Yang, *The Acquisition of Prosodic Focus-Marking in Mandarin Chinese- and Seoul Korean-Speaking Children*. Utrecht University, Nov. 2017.

[27] A. Yang and A. Chen, "The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children," *First Language*, vol. 38, no. 1, pp.26–46, Feb. 2018.

[28] Z. Liu, *Baihan Shuangyu Ertong Jiaodian Yunlv Fazhan Yanjiu [Research on the development of Chinese Prosodic Focus in Bai-Han Bilingual Children]*. Beijing: China Social Sciences Press, 2022.

[29] J. Lu and P. Tang, "The Production of Contrastive Focus by Children Learning Mandarin Chinese," *Proceedings of 2022 13th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pp. 36–40, Dec. 2022.

[30] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer*, (version 6.1.54). Accessed: Oct. 9, 2021. [Online]. Available: http://www.praat.org/

[31] Z. Xiong, *xSegmenter: Yinduan Zidong Qiefen yu Biaozhu Gongju [xSegmenter: An Automatic Segmentation and Annotation Tool for Segments]*. Accessed: Nov. 16, 2021. [Online]. Available: https://mp.weixin.qq.com/s/INU_ph214iQ-HqQ2envn8g

[32] Z. Li, "Yunlv yanjiu he yunlv biaoyin [Prosodic studies and prosodic labelling]," *Yuyan Wenzi Yingyong [Applied Linguistics]*, vol. 25, no. 1, pp. 105–109, Feb. 1998.

[33] A. Li, "Chinese prosody and prosodic labeling of spontaneous speech," *Proceedings of Speech Prosody 2002*, pp. 39–46, April 2002.