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

Mandarin-speaking 6-year-olds can use preboundary pitch range expansion to disambiguate compounds from lists

Feng Xu[1], *Ping Tang*[2], *Katherine Demuth*[1], *Nan Xu Rattanasone*[1, 3]

[1]Centre for Language Sciences, Department of Linguistics, Macquarie University
[2]School of Foreign Studies, Nanjing University of Science and Technology
[3]Hearing Research Centre, Macquarie University
feng.xu2@hdr.mq.edu.au, ping.tang@njust.edu.cn,
katherine.demuth@mq.edu.au, nan.xu@mq.edu.au

Abstract

Pitch can be used for marking boundaries and chunking utterances into different units, e.g., compounds (N_1+N_2, e.g., *jellybeans*) and their corresponding list forms (N_1, N_2, e.g., *jelly, beans*). Unlike English, where 5-6-year-olds can use different pitch patterns to mark boundaries in an adult-like manner, Mandarin is a tonal language using pitch for both word meanings (lexical tone) and utterance meanings (e.g., preboundary pitch range expansion). While lexical tones are early acquired, it was unclear when Mandarin-speaking children can use pitch cues to disambiguate compounds and lists. A total of 41 adults and 29 6-year-olds participated in an elicited production task. The pitch range of N_1 was measured (highest minus lowest pitch (f0)). Our results showed that similar to adults, 6-year-olds produced a larger pitch range over N_1 in lists compared to compounds for rising and falling tones. However, no pitch range expansion was found for the high-level tone in children or adults. These patterns suggested that 6-year-olds are adult-like in producing preboundary pitch cues that disambiguate compounds and lists. These findings are discussed in terms of the modulation of pitch information at the word and phrase levels and the role of pitch as a cue in response to boundaries in Mandarin.

Index Terms: pitch, boundary, Mandarin, children

1. Introduction

Children need to acquire the ability to prosodically chunk an utterance into meaningful units, such as a compound referring to one object (N_1+N_2, e.g., *jellybeans*) or a list referring to two separate objects (N_1, N_2, e.g., *jelly, beans*). Compared to compounds where two nouns are combined into a single word unit, speakers usually insert a boundary between N_1 and N_2 in the list. For example, English-speaking 5-year-olds can distinguish compounds and lists using a range of prosodic cues, including pitch [1, 2]. However, in tonal languages like Mandarin, pitch is used primarily to contrast lexical meaning. Children must therefore learn to use pitch for both word-level and utterance-level meanings in response to boundaries, which might pose learning challenges. This study therefore aims to examine Mandarin-speaking children's ability to modify pitch in response to boundary insertion.

English is a stress-initial language where the first syllable is typically realized with a higher pitch compared to the following syllables [3]. For example, the initial syllable in the compound *jellybeans* carries a higher pitch compared to all following syllables, whereas in the list form *jelly, beans*, the initial syllables in *jelly* and *beans* both receive a higher pitch.

Unlike English, Mandarin is a tonal language with four lexical tones mainly differing in pitch contours: a high-level tone (T1), a rising tone (T2), a low-dipping tone (T3), and a falling tone (T4), as shown in Figure 1 [4]. The pitch realization for the prosodic phrasing is superimposed over tonal realizations [4, 5]. In Mandarin, the pitch range of lexical tones in words preceding a boundary is expanded, especially for simple contours such as T2 and T4 [6, 7, 8, 9, 10]. Therefore, preboundary pitch range expansion in Mandarin could be used as a cue for disambiguating between compounds and lists. For example, N1 in the list "*xiong(T2), mao(T1)*" (bear, cat) is preboundary and will be more fully realized – similar to its canonical T2 form (see Figure 1). However, in the compound "*xiong(T2)-mao(T1)*" (panda), there should not be a boundary following N1 and the pitch range of T2 should be reduced/less fully realized compared to N1 in the list form[1].

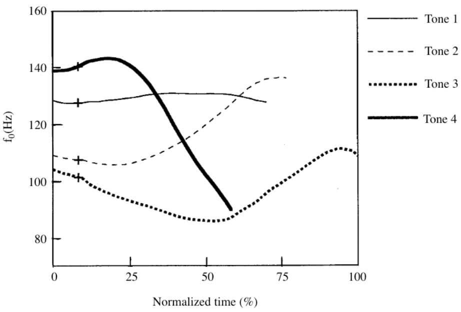

Figure 1: *Lexical tones in isolation in Mandarin. From [4].*

Unlike English-speaking children, Mandarin-speaking children must learn to modify pitch associated with lexical and phrasal meanings. While lexical tones are acquired at an early age, children's tonal realizations are not fully adult-like, even by the age of 5 years [13, 14, 15, 16]. Therefore, learning to map pitch to phrasal meanings in addition to lexical tone may be a challenge for Mandarin-speaking children and may

[1] Mandarin has a tone sandhi rule regarding T3, a complex contour. In a compound containing T3+T3 the first noun will be changed to T2 and in T3+T1/T2/T4 contexts it will change to a low-falling tone. How tone sandhi is implemented in compounds vs. lists by Mandarin-speaking children have been previously investigated, see in [11, 12].

10.21437/SpeechProsody.2024-9

therefore be delayed compared to children learning non-tonal language languages such as English.

This study therefore aims to answer the following two research questions: a) Can Mandarin-speaking 6-year-olds apply preboundary pitch expansion? b) If yes, are their productions adult-like? We hypothesize that Mandarin-speaking 6-year-olds cannot apply preboundary pitch range expansion as their ability to map pitch to both word-level and utterance-level meanings is still developing, and their productions may not be adult-like.

2. Methods

2.1. Participants

There were 43 adults and 29 typical-developing 6-year-olds (mean age: 6;1; range: 6;0-6;9) participating in this study. All adults were locals from Beijing. Child participants were recruited from kindergartens in Hebei, where Standard Chinese is used in teaching. No adults or children report any speech, hearing, or cognitive disorders. This study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Committee. All adult participants provided consent to participate in the study while the principals of kindergartens consented for children to participate in the study.

2.2. Materials

A total of sixteen picturable Noun-Noun items (N_1-N_2) with T1, T2, and T4 were selected to form compounds (N_1N_2) and their related list forms (N_1, N_2), including four items used in the practice trials and twelve items used in the target trials[2]. Thus, a total of 12 target pairs were generated. We adopted the paradigm used in [2] to ensure that children understood that there are two different structures. Compounds and their related list forms were embedded in a carrier sentence "*Zhe4-li3 you3…*" (Here are …) following a fruit as a filler (N_3), constructing a two-item list in the compound condition (N_1N_2, N_3) and a three-item list in the list condition (N_1, N_2, N_3). All items were presented on the screen using colored clipart pictures, as illustrated in Figure 2.

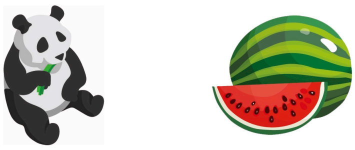

Figure 2: *Example of the compound condition.*

2.3. Procedure

During the testing phase, either a set of two pictures to elicit compounds or a set of three objects for the list forms was shown on the screen, and participants were asked: "What are here?" They were instructed to answer the question by completing the carrier sentence using the names of the objects from left to right. For example, "*Zhe4-li3 you4 xiong2-mao1 he2 xi1-gua1*" (Here are panda and watermelon) was elicited in the compound condition, as illustrated in Figure 2. No conjunctions were allowed between N_1 and N_2. The compound condition and the list condition were presented in a pseudo-randomized sequence, in which pictures of compounds would never occur right after their corresponding list forms and vice versa. No more than three compound conditions or list conditions would be shown successively to avoid any carryover effects and learning effects. According to this, two test versions were generated and counterbalanced across participants. Participants' speech was audio-recorded using a Marantz PM661 solid-state recorder and an AKG C520L head-worn microphone at a sampling rate of 44100 Hz. Each participant produced 24 target sentences and two adults were excluded due to the recording problems, hence resulting in 984 sentences from 41 adults and 696 sentences from 29 children for the pitch data extraction.

[2] T3 was excluded as it will induce tone sandhi and changes in tonal category. As this study investigates preboundary pitch range expansion, T3 was not included.

2.4. Data analysis

All recordings were annotated in Praat [17]. As the domain of lexical tones in Mandarin is mainly located in the rhyme part, we adopted the following criteria to define the onset and offset over the vowel. The onset was defined as the beginning of the second formant (F2) and the beginning of the periodic waveform. The end of high formants and the end of voicing in the waveform were regarded as the offset. Pitch was extracted by the Praat script [18] at 10 equidistant points within the annotated vowel part. To better match the scale of human perception, the original extracted F0 values in Hertz were converted to semitones using a reference frequency of 40 Hz. To examine whether there is preboundary pitch range expansion, only the pitch range of N_1 (by subtracting the lowest F0 from the highest F0) was compared in compounds and lists.

Statistical analysis was performed by linear mixed-effect models in R [19] using lmerTest package [20]. To select the model of best fit in each analysis, models were fitted with the maximal random structures first and were then simplified to avoid singularity or convergence problems. The final model was selected by a likelihood ratio test where a higher likelihood ratio represented a better model fit [21]. Post-hoc analysis was carried out using emmeans package [22] in R for further pairwise comparisons of multi-level factors with Tukey's HSD adjustment.

3. Results

Figure 3 shows the pitch range of N_1 by children and adults in both the compound condition and the list condition. To compare the difference in pitch range of tones across groups and conditions, a linear-mixed effect model (LMM) was fitted, with three fixed factors "Condition" (2 levels, Compound vs. List), "Tone" (3 levels, T1 vs. T2 vs. T4), and "Group" (2 levels, Children vs. Adults), and two random factors "Speaker" and "Item". The results (Table 1) showed a significant main effect of Condition and Tone and significant "Condition × Tone" interaction. No other significant main effects or interactions were detected.

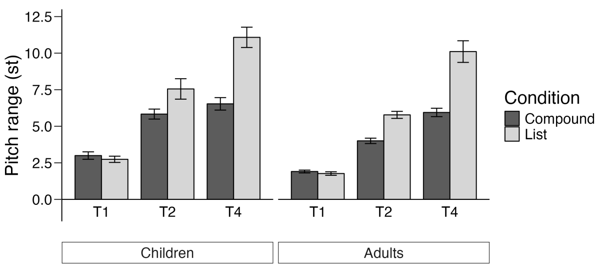

Figure 3: *Pitch range of Noun 1 (N_1) in the Compound and List condition with +/- 1SE in Children and Adult.*

Table 1: *Statistical results of LMM. F-values, degree of freedom (df) and p-values are provided, where *** indicates statistical significance at p < .001*

Fixed effects	df1	df2	F	*p*
Condition	1	15	91.98	< .001***
Tone	2	21	159.87	< .001***
Group	1	58	2.26	0.138
Condition × Tone	2	10	35.87	< .001***
Condition × Group	1	60	1.63	0.207
Tone × Group	2	61	1.72	0.188
Condition × Tone × Group	2	1228	2.21	0.110

Model: lmer(Pitch Range ~ Condition * Tone * Group + (Condition + Tone | Subject) + (Condition | Item))

A post-hoc analysis using Tukey's HSD was conducted to examine the Condition by Tone interaction (see Table 2 for results). The pairwise comparisons revealed that not all lexical tones showed distinct pitch range expansions in lists compared to compounds. The pitch range of T2 and T4 were more expanded in the list than the compound condition, but no pitch range difference was found for T1. This indicated that the two dynamic tones, T2 and T4, indeed showed preboundary pitch range expansion.

Table 2: *Post-hoc analysis of the pitch range of the Compound condition vs. the List condition in different lexical tones.*

Contrast	Tone	*β*	SE	df	*t*	*p*
Compound vs. List	T1	-0.06	0.26	12.0	-0.25	0.809
	T2	-1.79	0.28	11.5	-6.30	< .001***
	T4	-3.36	0.32	11.0	-10.38	< .001***

4. Discussion

This study examined whether Mandarin-speaking 6-year-olds produce preboundary pitch range expansion to disambiguate compounds and lists and, if so, whether their production is adult-like. We hypothesize that Mandarin-speaking 6-year-olds cannot apply preboundary pitch range expansion as their ability to map pitch to both word-level and utterance-level meanings is still developing, and their productions may also not be adult-like. Our findings did not support these hypotheses and instead showed that Mandarin-speaking 6-year-olds can produce preboundary pitch range expansion in dynamic tones (i.e., T2 and T4) to disambiguate compounds and lists. However, no pitch range expansion was found for the level tone (T1). These results are in line with the adult patterns, and indeed, we found no group differences in pitch range between children and adults. These results suggest that Mandarin-speaking children can implement correct preboundary pitch range expansion contingent on different tonal categories in an adult-like manner by age 6.

Our results have several theoretical implications. First, the pitch cue appears to be used in response to boundaries rather than marking boundaries. This is because not all lexical tones were expanded in pitch range, only T2 and T4. This is consistent with previous studies that have shown pitch range variation in T2 and T4, while no obvious pitch range variation has been found for T1 [6, 7, 8, 9, 10]. Thus, pitch is not a robust cue to mark boundaries but rather occurs in response to boundaries, leading to different performances of pitch range variation contingent on tonal categories. Future studies should examine whether pitch height may differ across compounds and lists for the level tone (T1).

Second, the ability to manipulate pitch in response to boundaries for utterance-level meanings appears to have been acquired already by 6 years. This is consistent with previous studies on non-tonal language English, where children demonstrated the ability to manipulate pitch patterns to distinguish compounds vs. lists in response to boundaries by age 6 [2, 23]. Our results further extended these findings by showing that children speaking a lexical tone language Mandarin, have also acquired a similar ability to manipulate pitch contours for lexical tones in response to boundaries by age 6. This suggests that the reported delay in adult-like production of lexical tones does not appear to impede children's ability to acquire prosodic cues, in this case, preboundary pitch range expansion. While achieving adult-like lexical tone production might be a protracted process [13, 14, 15, 16], children can produce distinct categories much earlier. Together with our results, these findings suggest that having acquired tonal categories early might be enough for children to map pitch to word-level and utterance-level meanings. Achieving adult-like lexical tone productions therefore does not pose a challenge to the mapping of pitch at different prosodic levels. Our results further suggest that in a paradigm with limited conditions (compounds vs. lists) and using familiar words, prosodic productions by 6-year-olds are adult-like.

Finally, while this study has shed light on the acquisition of pre-boundary pitch range expansion by Mandarin-speaking children in response to boundary marking, it remains unclear when Mandarin-speaking children acquire boundary marking. Indeed, studies with non-tonal languages suggest that boundary marking is also acquired by this age. For example, English-speaking 5-year-olds can use longer word durations and insert a silent pause to indicate the presence of a boundary [2]. Our results suggest that Mandarin-speaking 6-year-olds must already have acquired the knowledge of where boundaries should occur in order to manipulate preboundary pitch contours. This raises questions about whether Mandarin-speaking children's production of boundary cues, such as duration, e.g., pre-boundary lengthening and pause, are also adult-like by age 6. These will be investigated in our future research to provide a more complete understanding of the development of boundary marking by Mandarin-speaking children and provide the basis for comparison with clinical populations such as children with cochlear implants.

5. Conclusion

The current study investigated whether Mandarin-speaking 6-year-olds can produce pre-boundary pitch range expansion correctly and whether their performance is adult-like. Results showed that they successfully applied pre-boundary pitch range expansion contingent on dynamic lexical tones in an adult-like manner. Given the different tonal contours in Mandarin, pre-boundary pitch range expansion would not be a robust cue to marking boundary. Further steps are required to examine the usage of prosodic cues for boundary marking by Mandarin-speaking children.

6. Acknowledgements

We thank the Centre for Language Sciences, the Department of Linguistics, and the Hearing Research Centre at Macquarie University for their support. We also acknowledge the assistance from teachers at Tuofu, Doudou Wu, and Aishang kindergartens with data collection. This research was supported, in part, by a Macquarie University iMQRES scholarship to the first author and the Hearing Innovation Grant to the last three authors. The equipment was supported by the School of Foreign Studies at Nanjing University of Science and Technology.

7. References

[1] X. Zhang, “A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese,” PhD Dissertation, University of Michigan, 2012.

[2] I. Yuen, N. X. Rattanasone, E. Schmidt, G. Macdonald, R. Holt, and K. Demuth, “Five-year-olds produce prosodic cues to distinguish compounds from lists in Australian English,” *Journal of Child Language*, vol. 48, no. 1, pp. 110–128, May 2020, doi: 10.1017/s0305000920000227.

[3] A. Cutler and D. M. Carter, “The predominance of strong initial syllables in the English vocabulary,” *Computer Speech & Language*, vol. 2, no. 3–4, pp. 133–142, Sep. 1987, doi: 10.1016/0885-2308(87)90004-0.

[4] Y. Xu, “Contextual tonal variations in Mandarin,” *Journal of Phonetics*, vol. 25, no. 1, pp. 61–83, Jan. 1997, doi: 10.1006/jpho.1996.0034.

[5] X. S. Shen, “Tonal coarticulation in Mandarin,” *Journal of Phonetics*, vol. 18, no. 2, pp. 281–295, Apr. 1990, doi: 10.1016/s0095-4470(19)30394-8.

[6] B. Wang, Y. Xu, and Q. Ding, “Interactive prosodic marking of focus, boundary and newness in Mandarin,” *Phonetica*, vol. 75, no. 1, pp. 24–56, Jun. 2017, doi: 10.1159/000453082.

[7] F. Shi and P. Wang, “Boundary tone and focus tone,” *Journal of Chinese Linguistics*, vol. 42, no. 1, pp. 93–108, Jan. 2014.

[8] M. Swerts, “Prosodic features at discourse boundaries of different strength,” *Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 514–521, Jan. 1997, doi: 10.1121/1.418114.

[9] M. Lin, Ed., *Break with filled pause in standard Chinese and F0 range of tones: a F0 normalization in utterances*. Proceedings of the 14th International Congress of Phonetic Sciences, 1999.

[10] Y. Sun and C. Shih, “Boundary-conditioned anticipatory tonal coarticulation in Standard Mandarin,” *Journal of Phonetics*, vol. 84, p. 101018, Jan. 2021, doi: 10.1016/j.wocn.2020.101018.

[11] P. Tang, N. X. Rattanasone, I. Yuen, L. Gao, and K. Demuth, “The development of abstract representations of tone sandhi.,” *Developmental Psychology*, vol. 55, no. 10, pp. 2114–2122, Oct. 2019, doi: 10.1037/dev0000781.

[12] P. Tang, I. Yuen, N. X. Rattanasone, L. Gao, and K. Demuth, “The acquisition of phonological alternations: The case of the Mandarin tone sandhi process,” *Applied Psycholinguistics*, vol. 40, no. 6, pp. 1495–1526, Sep. 2019, doi: 10.1017/s0142716419000353.

[13] N. X. Rattanasone, P. Tang, I. Yuen, L. Gao, and K. Demuth, “Five-Year-olds’ acoustic realization of Mandarin tone Sandhi and lexical tones in context are not yet fully Adult-Like,” *Frontiers in Psychology*, vol. 9, May 2018, doi: 10.3389/fpsyg.2018.00817.

[14] P. Wong, R. G. Schwartz, and J. J. Jenkins, “Perception and production of lexical tones by 3-Year-Old, Mandarin-Speaking children,” *Journal of Speech Language and Hearing Research*, vol. 48, no. 5, pp. 1065–1079, Oct. 2005, doi: 10.1044/1092-4388(2005/074.

[15] P. Wong, “Acoustic characteristics of three-year-olds’ correct and incorrect monosyllabic Mandarin lexical tone productions,” *Journal of Phonetics*, vol. 40, no. 1, pp. 141–151, Jan. 2012, doi: 10.1016/j.wocn.2011.10.005.

[16] R. Choo, “The Acquisition of segments and tones in Mandarin: an observational and Experimental study,” PhD Dissertation, University of York, 2019.

[17] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer [Computer program].” Nov. 30, 2023. [Online]. Available: https://ci.nii.ac.jp/naid/20001461274

[18] Z. Xiong, *A tutorial of speech corpus Construction and analysis*, 1st ed. Xi’an Jiaotong University Press, 2017.

[19] “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, 2023. [Online]. Available: https://www.R-project.org

[20] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “LMERTest Package: Tests in Linear Mixed Effects models,” *Journal of Statistical Software*, vol. 82, no. 13, Jan. 2017, doi: 10.18637/jss.v082.i13.

[21] J. C. Pinheiro and D. Bates, *Mixed-Effects models in S and S-PLUS*. Springer Science & Business Media, 2009.

[22] R. V. Lenth (2023). emmeans: Estimated Marginal Means, aka Least-Squares Mean. R package version 1.8.9, <https://CRAN.R-project.org/package=emmeans>.

[23] M. Yoshida, “Children’s use of prosodic information to understand and produce phrasal distinctions in American English,” PhD Dissertation, University of Texas at Dallas, 2005.