ISCA Archive
http://www.isca-speech.org/archive

4th International Symposium on
Tonal Aspects of Languages (TAL-2014)
Nijmegen, The Netherlands
May 13-16 2014

Prosodic focus marking in child and adult Mandarin Chinese

Anqi Yang [1], *Aoju Chen*[1, 2]

[1] Utrecht Institute of Linguistics, Utrecht University, the Netherlands
[2] Max Planck Institute for Psycholinguistics, the Netherlands

a.yang1@uu.nl, aoju.chen@uu.nl

Abstract

This study investigates how Mandarin Chinese speaking children and adults use prosody to mark focus in spontaneous speech. SVO sentences were elicited from 4- and 8-year-olds and adults in a game setting. Sentence-medial verbs were acoustically analysed for both duration and pitch range in different focus conditions. We have found that like the adults, the 8-year-olds used both duration and pitch range to distinguish focus from non-focus. The 4-year-olds used only duration to distinguish focus from non-focus, unlike the adults and 8-year-olds. None of the three groups of speakers distinguished contrastive focus from non-contrastive focus using pitch range or duration. Regarding the distinction between narrow focus from broad focus, the 4- and 8-year-olds used both pitch range and duration for this purpose, while the adults used only duration.

Index Terms: focus, tone, Mandarin Chinese; L1 acquisition

1. Introduction

Focus is an information structural category referring to the new information in a sentence to the receiver [1, 2]. This study involves three kinds of focus, i.e. (non-contrastive) narrow focus, contrastive (narrow) focus, and broad focus. The first two differ from each other in contrastivity, while they differ from broad focus in the size of the focal constituent (e.g. a specific word vs. a whole sentence). Previous studies on prosodic focus marking in child language have shown that children learn to use the phonological and phonetic parameters to mark focus in a gradual process. For example, English-speaking children can use accentuation to highlight contrastive focus by the age of three, and from three to six this use of accentuation is further consolidated [3, 4]. Dutch-speaking children can use accentuation to mark focus at the age of four or five but become adult-like in choice of accent type only at the age of seven or eight [5, 6]. Further, they cannot vary the phonetic realisation of a pitch accent in terms of pitch range for focus-marking purposes until the age of seven or eight. The use of duration for this purpose is still not acquired at the age of seven or eight [7].

Compared to our knowledge of children speaking West Germanic languages, much less is known on prosodic focus marking in children acquiring Mandarin Chinese. [8] was a first study investigating how Mandarin-speaking children use prosody to mark focus. It was found that in spontaneous speech the 4-year-olds used duration to distinguish focus from both post- and pre-focus, but used pitch range to distinguish focus from only pre-focus. Further, they used neither pitch range nor duration to distinguish contrastive narrow focus from non-contrastive narrow focus. In addition, they used pitch range but not duration to distinguish narrow focus (both contrastive and non-contrastive narrow focus) from broad focus. The 8-year-olds differed from the 4-year-olds only in that they used both pitch range and duration to distinguish focus from both post- and pre-focus.

To study development in children's prosodic focus marking, it is essential to compare children's production to adults' production. In [8], the children's prosodic focus-marking in spontaneous speech was not discussed in comparison to adults' prosodic focus. Extensive studies have been done on adult Mandarin using read speech. It has been found that a focused constituent has a longer duration, a wider pitch range and/or a higher pitch level than the same constituent in the broad focus condition [e.g. 9, 10, 11]. Furthermore, the post-focus part of the sentence is usually compressed in pitch (i.e. spoken with a lower pitch level or a smaller pitch range) and duration, while the pre-focus part undergoes little change in pitch or duration [e.g. 10, 12, 13, 14,]. As for the difference between contrastive and non-contrastive focus, no systematic durational or pitch range difference had been found between them in all tonal contexts or sentence positions [11, 15]. However, it is not self-evident that adults' use of prosody in focus marking in read speech is a suitable model to compare children's prosodic focus marking in spontaneous speech to, because substantial differences between read speech and spontaneous speech in the use of prosody have also been reported for languages like German [16].

The current study thus has two goals: (1) to find out how Chinese adults use prosody to mark focus in spontaneous speech, and whether the previous findings based on read speech can hold for spontaneous speech; (2) to find out how children differ from adults in prosodic focus marking in spontaneous speech by taking the tonal targets of lexical tones in account in the prosodic analysis. These goals are addressed in three specific questions on prosodic focus marking: (1) How Mandarin-speaking children and adults use prosody to distinguish focus from non-focus (i.e., post- and pre-focus), (2) how they distinguish contrastive narrow focus from non-contrastive narrow focus. (3) and how they distinguish focus in different constituent-sizes (narrow focus vs. broad focus).

2. Method

The method was identical to [8] regarding target sentences and the elicitation method. Details on these aspects of the method are given here for the sake of completeness.

2.1. Target sentences

The target sentences were 80 four-syllable SVO sentences, deriving from four subject nouns, four verbs, and four object nouns. These words were selected from the words that are acquired by Mandarin-speaking children by the age of three or four [17]. All four lexical tones occurred in the verbs and object-nouns (Table 1). Each verb was combined each object noun, leading to 16 VPs. Each VP occurred in five focus

conditions, leading to 80 VPs: (1) Narrow focus on the subject in sentence-initial position (NF-i); (2) Narrow focus on the verb in sentence-medial position (NF-m); (3) Narrow focus on the object in sentence-final position (NF-f); (4) Contrastive focus on the verb in sentence-medial position (CF-m); (5) Broad focus over a whole sentence (BF). Four disyllabic subject nouns (xiao3-mao1, 'little cat', xiao3-xiong2, 'little bear', xiao3-gou3, 'little dog' and xiao3-tu4 'little rabbit') were evenly distributed over these 80 VPs, forming 80 sentences (4 tones in the verbs x 4 tones in the object x 5 focus conditions). These sentences were split evenly into two lists (List 1 & 2) of 40 sentences such that each list contained all target words and eight of the sixteen tonal combinations concerning the verbs and object nouns. Half of the participants produced the sentences on List 1 and the other half produced the sentences on List 2.

Verbs		Object nouns	
T1 rēng (throw)	T2 mái (bury)	T1 shū (book)	T2 qíu (ball)
T3 jiǎn (cut)	T4 yùn (transport)	T3 bǐ (pen)	T4 cài (vegetable)

Table 1 : Verbs and object nouns

2.2. Speech elicitation

To elicit the target sentences, question-answer dialogues illustrated in the following examples (1) to (5) were embedded in a picture-matching game adapted from [5].

(1) Exp: Look! A book, and the book is in the air. It looks like someone throws the book. Who throws the book?
Child: [The rabbit] throws the book. (NF-i)

(2) Exp: Look! A rabbit, and there is also a book. It looks like the rabbit does something to the book. What does the rabbit do to the book?
Child: The rabbit [throws] the book. (NF-m)

(3) Exp: Look! A rabbit, and its arm is stretched out. It looks like the rabbit throws something. What does the rabbit throw?
Child: The rabbit throws [the book]. (NF-f)

(4) Exp: Look! A rabbit, and a book. It looks like the rabbit will do something to the book. I will make a guess: The rabbit cuts the book.
Child: The rabbit [throws] the book. (CF-m)

(5) Exp: Look! This picture is very blurring. I cannot see anything clearly. What happens in the picture?
Child: [The rabbit throws the book]. (BF)

In the game, the child was asked to assist the experimenter to match pictures that could go together. Three piles of pictures were used. The experimenter and the child each held a pile of pictures. The third pile was laid around on the table in a seemingly 'messy' fashion. The experimenter's pictures showed incomplete stories in which there was always something missing, namely, the subject, the action, the object or all the three pieces of information. The child's pictures always contained all the three pieces of information. In every trial, the experimenter showed a picture of hers to the child, described the picture and asked a question about it, as illustrated in the examples (1) to (5). The child took a look at the corresponding picture in his pile and answered or corrected the experimenter (in the CF condition). According to the information provided by the child, the experimenter looked for the right picture in the messy pile and matched it with her own picture. Crucially, as rules of the game, the child was asked to answer the experimenter's question in full sentences and not to reveal his pictures to the experimenter.

In order to familiarise the child with the game procedure, the experimenter started the game with five practice trials involving all five focus conditions. Prior to the practice session, the experimenter conducted a picture-naming task to make sure that the children would use the intended words to refer to the entities in the pictures.

2.3. Participants

Thirty-six children from three age groups (4-5 years, 7-8 years, 10-11 years, 12 per group) participated in the experiment. They were tested individually in a quiet room in their kindergartens or schools in Beijing. In addition, fifteen university students speaking Mandarin were tested as controls in Beijing, following the same procedure. The adults and children were both audio and video-recorded during the experiments. The current paper presents results from four 4-year-olds, four 8-year-olds and four adults.

2.4. Annotation and acoustic analysis

The audio recording from each speaker was orthographically annotated in Praat, and usable sentences were then carefully selected from the recordings. Responses deviating from the target sentences in choice of word or word order or produced with self-repairs and hesitations were considered unusable and excluded from further analysis. In total, 70 sentences from the 4-year-olds, 140 sentences from the 8-year-olds, and 142 sentences from the adults were included in the analysis.

The usable sentences were then acoustically annotated at the word level and at the pitch level. Landmarks indicating word-onsets and word-offsets and the locations of the maximum pitch and minimum pitch within each word were inserted in Praat textgrids for each sentence. The maximum and minimum pitch were labelled taking the tonal targets into consideration, following [18].

In this paper, we concentrate on the sentence-medial verbs. The verbs were on-focus in the NF-m condition, pre-focus in the NF-f condition and post-focus in the NF-i condition and were thus ideal for direct comparisons between focus and pre-/post-focus. The duration and pitch range (the difference between the maximum and the minimum pitch) of the verbs were calculated and analysed as dependent variables.

To address the first research question, namely, how focus differs from non-focus, we compared the prosody of the verbs in the NF-m condition (focused) with that in the NF-i (post-focus) and NF-f (pre-focus) conditions. To address the question on contrastivity, we compared the prosody of the verbs in the NF-m (non-contrastive narrow focus) condition with that in the CF-m condition. To address the question about size of focused constituent, we compared the prosody of the verbs in the NF-m and CF-m combined conditions with that of the BF condition.

3. Statistical analysis and results

Mixed-effect modeling was used to assess the effect of focus-related fixed factors (or independent variables) and the effect of interactions between the focus-related fixed factors and the other fixed factors on the dependent variables, i.e. duration and pitch range of the verbs. The focus-related fixed factors were FOCUS (focus vs. post- and pre-focus), CONTRASTIVITY (contrastive vs. non-contrastive narrow focus), and SIZE (narrow focus vs. broad focus). The other fixed factors were AGE (4-year-olds, 8-year-olds, and adults)

and TONE (four tones). The random factor was SPEAKER. In the analyses on the effect of the fixed factors, two models were built for each fixed factor, one with only the random factor, and one with both the random factor and the fixed factor. The two models were then compared to each other. A statistically significant difference between these two models indicated a main effect of the fixed factor. We then looked at the interaction between the focus-related fixed factor and the other fixed factors. In what follows, we will focus on the main effects of the fixed factors FOCUS, CONTRASTIVITY, SIZE and significant interactions between one of these factors and AGE.

3.1. The use of pitch range

3.1.1. Effect of focus

Regarding the comparison between focus and post-focus, the mixed-effect modelling showed that the main effect of FOCUS was significant ($p < .05$). So were the main effects of AGE ($p < .05$) and TONE ($p < .05$). There was a significant interaction between FOCUS and AGE ($p < .05$), but not between FOCUS and TONE ($p = .41$). We then examined the effect of FOCUS within each age group separately, and found that the main effect of focus was significant for the 8-year-olds ($p < .05$), and adults ($p < .05$), but was not significant for the 4-year-olds ($p = .74$). These results suggested that the 8-year-olds and adults used a wider pitch range in the focused verbs than in the post-focal verbs, but the 4-year-olds did not (Figure 1).

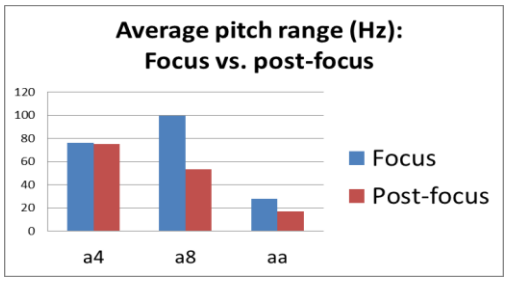

Figure 1: Pitch range: focus vs. post-focus

Regarding the comparison between focus and pre-focus, the mixed-effect modelling showed that there were significant main effects of FOCUS ($p < .05$) and AGE ($p < .05$), but was no significant main effect of TONE ($p = .09$). There was a significant interaction between FOCUS and AGE ($p < .05$), but not between FOCUS and TONE ($p = .11$). We then examined the effect of FOCUS within each age group separately, and found that the main effect of focus was significant for the 8-year-olds ($p < .05$), and adults ($p < .05$), but was not for the 4-year-olds ($p = .08$). Thus, the 8-year-olds and adults used a wider pitch range in the focused verbs than in the pre-focal verbs, but the 4-year-olds did not (Figure 2).

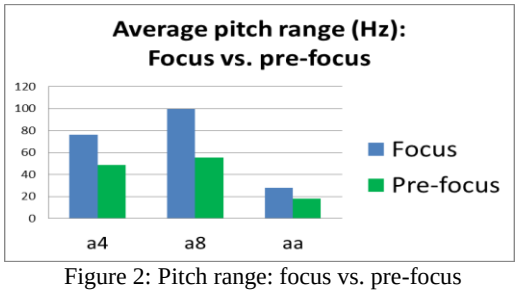

Figure 2: Pitch range: focus vs. pre-focus

3.1.2. Effect of contrastivity

Comparing contrastive narrow focus and non-contrastive narrow focus, the mixed effect modelling showed that there was no significant main effect of CONTRASTIVITY ($p = .47$), but were significant main effects of AGE ($p < .05$) and TONE ($p < .05$). There was no significant interaction between AGE and FOCUS ($p = .75$) or between TONE and FOCUS ($p = .45$). The results indicated that neither the adults nor the children used pitch range to differentiate contrastive narrow focus from non-contrastive narrow focus.

3.1.3. Effect of focal constituent size

To examine the effect of focal constituent size, we grouped the two narrow focus conditions, i.e., contrastive-narrow focus and non-contrastive narrow focus together, and compared them with broad focus. The mixed-effect modelling showed that the main effects of SIZE, AGE and TONE were all significant ($p < .05$). There was a significant interaction between SIZE and AGE ($p < .05$), but was not between SIZE and TONE ($p = .82$). We then examined the effect of SIZE within each age group separately, and found that the main effect of focus was significant in both the 4- and 8-year-olds ($p < .05$), but not in the adults ($p = .88$). Thus, the children used a wider pitch range in the narrow focus conditions than in the broad focus condition, but adults did not do so (Figure 3).

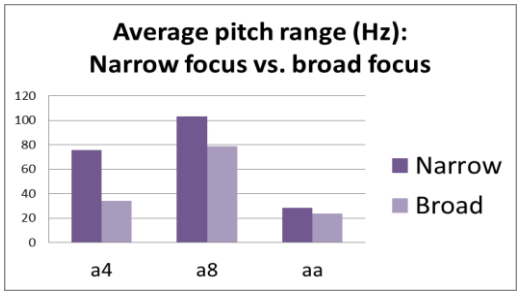

Figure 3: Pitch range: narrow vs. broad focus

3.2. The use of duration

3.2.1. Effect of focus

Regarding the comparison between focus and post-focus, the mixed-effect modelling showed that the main effects of FOCUS, AGE, and TONE were all significant ($p < .05$). The interaction between FOCUS and AGE ($p = .21$) and that between FOCUS and TONE ($p = .54$) were not significant. Similarly, comparing focus and pre-focus, we found that the main effects of FOCUS, AGE, and TONE were all significant ($p < .05$). The interaction between FOCUS and AGE ($p = .28$) and that between FOCUS and TONE ($p = .43$) were not significant. The results revealed that both the children and adults used a longer duration for the focused verbs than for the post- and pre-focal ones (Figure 4).

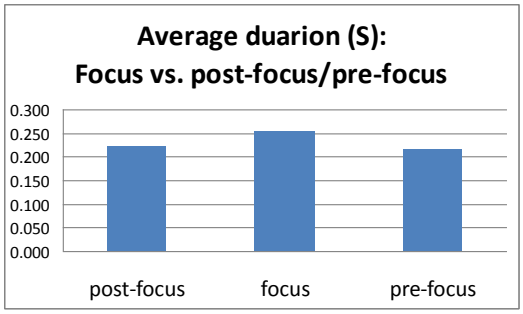

Figure 4: Duration: focus vs. post-/pre-focus

3.2.2. Effect of contrastivity

Comparing contrastive narrow focus and non-contrastive narrow focus, the mixed effect modelling showed that there was no significant main effect of CONTRASTIVITY ($p = .49$), but were significant main effects of AGE ($p < .05$) and TONE ($p < .05$). There was no significant interaction between AGE and FOCUS ($p = .62$) or between TONE and FOCUS ($p = .98$). The results indicated that the speakers regardless of age did not use duration to differentiate contrastive narrow focus from non-contrastive narrow focus.

3.2.3. Effect of focal constituent size

Regarding the comparison between narrow focus (both contrastive and non-contrastive) and broad focus, the mixed-effect modelling showed that the main effects of SIZE, AGE, and TONE were all significant ($p < .05$). The interactions between FOCUS and AGE ($p = .82$) and between FOCUS and TONE ($p = .06$) were not significant. Therefore, regardless of age, the speakers used a longer duration for the verbs in the narrow focus conditions than the ones in the broad focus condition (Figure 5).

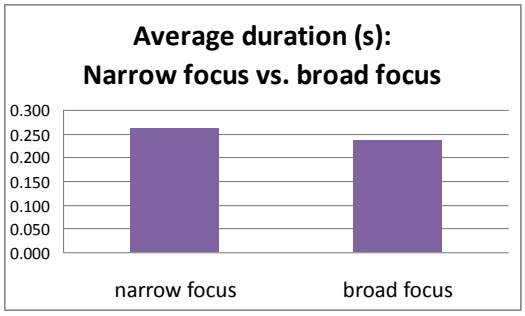

Figure 5: Duration: narrow vs. broad focus

4. Discussion and conclusions

4.1 spontaneous vs. read speech in adult Mandarin

In spontaneous speech we found that adults used duration but not pitch range to distinguish narrow focus from broad focus. However, previous studies [e.g. 10] based on read speech showed that narrow focus was distinguished from broad focus in both duration and pitch range. Moreover, we found that contrastive and non-contrastive narrow focus were not distinguished in pitch range or duration, similar to the findings on spontaneous speech in [15] but different from the findings on read speech in [11], according to which pitch played a role, though rather limited. Therefore, there are differences between read speech and spontaneous speech in adults' use of prosody to distinguish focus types in Mandarin Chinese.

4.2 Children vs. adults

The children could use both pitch range and duration to distinguish focus from post- and pre-focus at eight, similar to the adults. Further, similar to the adults, the children used neither pitch range nor duration to distinguish contrastive narrow focus from non-contrastive narrow focus already at four. However, the children used both pitch and duration to distinguish narrow focus (both contrastive and non-contrastive) from broad focus at both four and eight, different from the adults, who used duration only. It seems that the children 'overused' prosody to distinguish focus types differing in size.

4.3 Chinese- vs. Dutch-speaking children

Dutch-speaking children's prosodic focus marking has been relatively extensively studied at both the phonetic and phonological level [5, 6, 7, 19], using similar method to the current study. It thus makes sense to make a comparison between these two groups of children. Compared to the Dutch-speaking children who could not use pitch range or duration to distinguish focus from non-focus at four and could use pitch range at eight [7], the Chinese-speaking children could already use duration for this purpose at four and both duration and pitch range at eight. This suggests that the primary cues in the ambient language (i.e., phonetic cues in Chinese, and phonological cues like accent type and accent placement in Dutch [5, 6]), may be acquired earlier than the secondary cues (i.e. phonetic cues in Dutch). With regard to the order of acquiring the use of pitch range and duration, the Chinese-speaking children could use duration to distinguish focus from non-focus earlier than pitch range, whereas the Dutch-speaking children could use pitch range earlier than duration. The different order of acquisition may be related to the fact that in Chinese pitch is the main cue to lexical tone but in Dutch duration is the main cue to lexical stress [20]. The lexically highly relevant cue in each language may thus be acquired later for focus-marking purposes.

5. Acknowledgements

This study is supported by a VIDI grant (276-89-001) from the NWO (Netherlands Organisation for Scientific Research) to the second author. We would like to express our special gratitude to Min Zhu, Jun Bian, Yian Liang, and Shushuang Yu from Beijing 21st Century International Kindergarten and School, the children and their parents for their full cooperation. We would also like to thank Mei Ou, Mengting Huang, Yun Li, and Xingzhi Yao from Beijing Forestry University for administering the tests with the adults. We thank Paula Cox for drawing the pictures, Frank Bijlsma, Sjef Pieters, and Alex Manus for the technical support, and Mattis van den Bergh for statistical support. Last, we thank Xiaoli Dong, Mengru Han, René Kager, Zenghui Liu, Anna Sara Romøren and Wim Zonneveld for their input..

6. References

[1] Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the representations of discourse referents.* Cambridge: Cambridge University Press.

[2] Gundel, J. K. (1999). "On different kinds of focus". In P. Bosch, & R. van de Sandt (eds.) *Focus: linguistic, cognitive, and computational*. Cambridge: Cambridge University Press.

[3] Hornby, P. A. & Hass, W. A. (1970). Use of contrastive stress by preschool children. *J. of Speech and Hearing Research*, 13, 359-399.

[4] MacWhinney, B. & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross- cultural developmental study. *Journal of verbal learning and verbal behavior*, 17, 539-558.
[5] Chen, A. (2011a). Tuning information packaging: intonational realization of topic and focus in child Dutch. *J. Child Lang*, *38*, 1055-1083.
[6] Chen, A. (2011b). The developmental path to phonological focus-marking in Dutch. In *Prosodic Categories: Production, Perception and Comprehension* (pp. 93-109). Springer Netherlands.
[7] Chen A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Vigário, S. Frota and M. J. Freitas (eds.), *Phonetics and Phonology: Interactions and interrelations* (pp. 91-106). Amsterdam: Benjamins.
[8] Yang, A. and Chen, A. (2014). Prosodic focus marking in child Mandarin Chinese. *Speech Prosody*. Dublin, 20-23 May, 2014.
[9] Shih, C. (1988). Tone and intonation in mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-109.
[10] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0contours. *Journal of Phonetics*, 27(1), 55-105.
[11] Chen, Y., & Braun, B. (2006). Prosodic realization of information structure categories in standard Chinese. Paper presented at the *Proceedings of Speech Prosody*, Dresden, Germany.
[12] Shih, C. (2000). A declination model of mandarin Chinese. *Intonation: Analysis, Modelling and Technology*, 243-268.
[13] Chen, Y. (2010). Post-focus F0 compression—Now you see it, now you don't. *Journal of Phonetics*, 38(4), 517-525. doi:10.1016/j.wocn.2010.06.004
[14] Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. Paper presented at the *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong. pp. 152-155.
[15] Greif, M. (2010). Contrastive focus in mandarin Chinese. Paper presented at the *Proc. Speech Prosody*, Chicago, UAS.
[16] de Ruiter, L (2010). *Studies on intonation and information structure in child and adult German*. Doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen.
[17] Li, P., Zhao, X., Liu, Y. and Levine, S. *Chinese Single-character Word Database*. [http://www.personal.psu.edu/pul8/psylin_norm/psychnorms.html]
[18] Xu, Y. and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319-337.
[19] Romøren, A. S. H. and Chen, A. (2014). Accentuation, pitch and pausing as cues to focus in child Dutch. In *Proceedings of the 38th Boston University Conference on Language Development* (BUCLD). Cascadilla Press.
[20] Heuven V.J.J.P. van & Sluijter A.M.C. (1996), Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America* 100: 2471-2485.