Prosodic focus-marking in Chinese four- and eight-year-olds

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 use prosody to distinguish focus from non-focus, and focus types differing in size of constituent and contrastivity. SVO sentences were elicited from four- and eight-year-olds in a game setting. Sentence-medial verbs were acoustically analysed for both duration and pitch range in different focus conditions. The children started to use duration to differentiate focus from non-focus at the age of four. But their use of pitch range varied with age and depended on non-focus conditions (pre- vs. post-focus) and the lexical tones of the verbs. Further, the children in both age groups used pitch range but not duration to differentiate narrow focus from broad focus, and they did not differentiate contrastive narrow focus from non-contrastive narrow focus using duration or pitch range. The results indicated that Chinese children acquire the prosodic means (duration and pitch range) of marking focus in stages, and their acquisition of these two means appear to be early, compared to children speaking an intonation language, for example, Dutch.

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

1. Introduction

The term 'focus' refers to an information structural category and is defined as the new information in a sentence to the receiver [e.g. 1, 2]. This study involves three types of focus, i.e. narrow focus, contrastive focus and broad focus. The former two differ from the latter in the size of the focus constituent, e.g. a lexical word (narrow focus, contrastive focus) vs. a whole sentence (broad focus). Narrow focus and contrastive focus differ in that the latter conveys an explicit contrast to alternatives in the context.

Prosodic focus-marking in adult Mandarin Chinese (hereafter Mandarin) has been extensively studied. It is generally agreed that a focused constituent has a longer duration, a higher pitch level and/or a wider pitch range than the same constituent in the broad focus condition [e.g. 3, 4, 5] 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. 4, 6, 7, 8]. However, the difference between narrow focus and contrastive focus is less conclusive. Some researchers have reported that contrastive focus induces a wider pitch range than narrow focus in sentence-initial position when the focused constituent has a certain tonal composition [5]. Yet [9] have found neither pitch range nor durational differences between narrow and contrastive focus.

In contrast, little is known on how Mandarin-speaking children use prosody to mark focus. Studies on other languages have revealed that children learn to use prosody to mark focus in their respective languages in stages [10]. 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 [11, 12]. 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 [10, 13]. 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 [14]. The use of duration for this purpose is still not acquired at the age of seven or eight [14].

The current study is a first study examining Mandarin-speaking children's use of pitch and duration in focus-marking. As Mandarin uses pitch not only to mark focus and express other sentence-level meanings but also to distinguishing lexical meanings, acquiring Mandarin entails that children have to learn both functions of pitch. The question that arises is whether Mandarin-speaking children follow a similar developmental trajectory to children speaking a non-tonal language in prosodic focus-marking. As a first step towards addressing this question, we have investigated (1) how Mandarin-speaking children use prosody to distinguish focus from non-focus, (2) how they distinguish focus in different constituent-sizes (narrow focus vs. broad focus), and (3) how they distinguish contrastive focus from non-contrastive focus.

2. Method

2.1 Target sentences

We aimed to elicit 160 SVO sentences from participants: (5 focus conditions x 4 tones in the verbs x 4 tones in the object-nouns x 2 types of verbs and object nouns). The five focus conditions were: (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 lexical tones were used in the verbs and object-nouns. Two types of verbs and corresponding object nouns were included (Table 1). Four subject nouns (cat, bear, dog, and rabbit) were evenly distributed over the sentences. Crucially, all words were selected from the words that are acquired by Mandarin-speaking children by the age of three or four [15]. The 160 sentences were split evenly into two lists (List 1 & 2) of 80 sentences such that each list contained all target words and all tonal combinations but not all word combinations of the verbs and objects. Half of the participants produced the sentences on List 1 and the other half produced the sentences on List 2.

Verb – type1	Verb – type 2	Noun - type1	Noun - type 2
T1 rēng (throw)	T1 jiāo (water)	T1 shū (book)	T1 huā (flower)
T2 mái (bury)	T2 wén (smell)	T2 qíu (ball)	T2 lí (pear)
T3 jiǎn (cut)	T3 tiǎn (lick)	T3 bǐ (pen)	T3 cǎo (grass)
T4 yùn (transport)	T4 mài (sell)	T4 cài (vegetable)	T4 shù (tree)

Table 1: Two types of verbs and object nouns

2.2   Speech elicitation

To elicit the target sentences, question-answer dialogues between the experimenter and the child as illustrated in examples (1) to (5) were embedded in a picture-matching game adapted from [10].

(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's task was to help the experimenter to put pictures in matched pairs. Three piles of pictures were used. The experimenter and the child each held a pile of pictures; the third pile laid around on the table in a seemingly 'messy' fashion. The experimenter's pictures always missed some information, e.g. 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 (1) to (5). The child took a look at the corresponding picture in his pile and answered the question or made a correction (in the CF condition). The experimenter could then look for the right picture in the messy pile and matched it with her own picture to form a pair. 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 familarise 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 (four-five yrs, even-eight yrs, ten-eleven yrs, twelve 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, following the same procedure. Considering children's limited concentration capacity, the 80 sentences on each list were elicited in two sessions of $20-35$ minutes on two different days. The adults and children were both audio and video-recorded during the experiments. The current paper presents results from four four-year-olds and four eight-year-olds.

2.4   Annotation and acoustic analysis

The audio recording from each child was orthographically annotated at three levels using Praat: trial, question from the experimenter, and response from the child. 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, 432 sentences from the eight children 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. It is worth noting that Mandarin is a tone language, and each tone has a particular target to reach. According to [16], the pitch contour approximates to or reaches at the target towards the end of a syllable. In this study the tonal targets were taken into account. For Tone 2 (rising tone) and Tone 4 (falling tone), it was presumed that their pitch contour approach to or reach at the high/low target respectively at the syllable offset. To be more specific, the maximum pitch of Tone 2 was always labeled and measured on the right side of its minimum pitch, even though sometimes there was an even higher pitch occurring on the left side due to the influence of the preceding tone. Similarly, the minimum pitch of Tone 4 was obtained on the right side of its maximum pitch. For Tone 1 (flat tone), its maximum and minimum pitch were obtained regardless of their relative order of occurrence. The pitch contour of Tone 3 varied most, and three patterns were observed in the data, namely, fall-rise, rise and fall. When Tone 3 was realised as a fall-rise, it was assumed to have two targets to approach, first the low target and then the high target. In this case, the maximum pitch was obtained on the right side of the minimum pitch. When Tone 3 was realised as a fall, it was assumed to have a low target to approach, and its minimum pitch was obtained at the syllable offset. When Tone 3 was realised as a rise, it was assumed to have a high target to approach, and its maximum pitch was obtained at the syllable offset.

In this paper, we concentrated 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 pitch 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 in child Mandarin, 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 about size of focused constituent, we compared the prosody of the verbs in the NF-m and CF-m combined narrow focus condition with that of the BF condition. 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 (contrastive narrow focus) condition.

3.   Statistical analysis and results

Mixed-effect modeling was used to assess the effect of fixed factors and the effect of interactions between the fixed factors and the other fixed factors on the dependent variables, i.e.

duration and pitch range of the verbs. There are two kinds of fixed factors: those related to focus directly and the others. The focus-related fixed factors were FOCUS (focus vs. non-focus), SIZE (narrow focus vs. broad focus), and CONTRASTIVITY (contrastive focus vs. non-contrastive focus). The other fixed factors were AGE (four-year-olds and eight-year-olds), TONE OF VERB (four tones for verbs), and TYPE OF VERB (type1 and type2). 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 focus-related factor. We then looked at the interaction between the focus-related fixed factor and the other fixed factors.

3.1   Focus vs. non-focus

3.1.1   Duration

Regarding the comparison between focus (verbs in NF-m) vs. post-focus (verbs in NF-i), the mixed-effect modelling showed that the main effect of FOCUS was significant ($p < 0.05$). As can be seen in Figure 1, the focused verbs in the NF-m condition were longer than the same verbs in the NF-i (post-focus) condition. There was also significant main effects of TONE OF VERB ($p < 0.05$) and TYEP OF VERB ($p = 0.01$), but no significant main effect of AGE ($p = 0.4$). No significant interaction was found between FOCUS and AGE ($p = 0.80$) or between FOCUS and TONE OF VERB ($p = 0.33$), but was found between FOCUS and TYPE OF VERB ($p < 0.05$). We then used subsets of data to look at the effect of FOCUS within the type 1 verbs and the type 2 verbs separately, and found that the main effect of focus was significant for both the type 1 verbs ($p < 0.05$) and the type 2 verbs ($p < 0.05$). This suggested that the interaction was caused by a gradient difference between type 1 verbs and type 2 verbs. The durational difference between focus and post-focus was larger in the type 2 verbs (0.08s) than in the type 1 verbs (0.04s).

Regarding the comparison between focus and pre-focus (verbs in NF-f), the mixed-effect modelling showed that the main effect of FOCUS was significant ($p < 0.05$). There was also significant main effects of TONE OF VERB ($p = 0.01$) and TYPE OF VERB ($p < 0.05$), but no significant main effect of AGE ($p = 0.44$). No significant interaction was found between FOCUS and AGE ($p = 0.82$), between FOCUS and TONE OF VERB ($p = 0.08$), or between FOCUS and TYPE OF VERB ($p = 0.14$).

The above results indicated that the children realized the focused verbs with a longer duration than the post-focal and pre-focal verbs, regardless the tones or the types of the verbs.
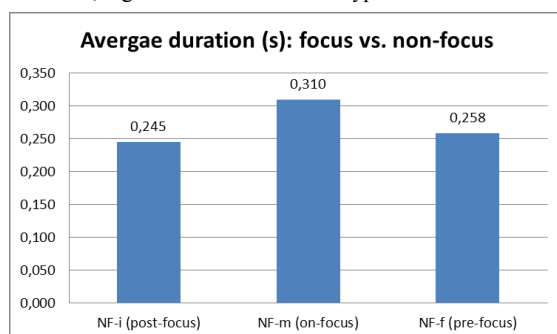

Figure 1: Verb duration in NF-i, NF-m, and NF-f

3.1.2   Pitch range

Regarding the comparison between focus and post-focus, the mixed-effect modelling showed that the main effect of FOCUS was significant ($p < 0.05$). The focused verbs in the NF-m condition had a wider pitch range (84Hz) than the same verbs in the NF-i (post-focus) condition (59Hz). The main effects of TONE OF VERB ($p < 0.05$) was also significant, but the main effect of AGE ($p = 0.53$) and TYPE OF VERB ($p = 0.1$) was not significant. A significant interaction was found between FOCUS and AGE ($p < 0.05$), and between FOCUS and TONE OF VERB ($p < 0.05$), but not between FOCUS and TYPE OF VERB ($p = 0.75$). We looked at the effect of FOCUS within each age, and found that the main effect of FOCUS was significant for the eight-year-olds ($p < 0.05$), but was not significant for the four-year-olds ($p = 0.72$) (Figure 2), so the eight-year-olds used pitch range to differentiate focus from post-focus, but the four-year-olds didn't. We then looked at the effect of FOCUS within each tone, and found that the main effect of FOCUS was significant for Tone 2 ($p < 0.05$) and Tone 4 ($p < 0.05$), but was not significant for Tone 1 ($p = 0.6$) or Tone 3 ($p = 0.28$).

Comparing focus with pre-focus, the mixed-effect modelling showed that the main effect of FOCUS was significant ($p < 0.05$). The focused verbs in the NF-m condition had a wider pitch range (84Hz) than the same verbs in the NF-f (pre-focus) condition (57Hz). The main effect of TONE OF VERB ($p < 0.05$) was also significant, but the main effects of AGE ($p = 0.17$) and TYPE OF VERB ($p = 0.43$) were not significant. A significant interaction was found between FOCUS and TONE OF VERB ($p < 0.05$), but not between FOCUS and AGE ($p = 0.22$) or between FOCUS and TYPE OF VERB ($p = 0.43$). We looked at the effect of FOCUS within each tone using subsets of data. It was found that the main effect of FOCUS was significant for Tone 2 ($p < 0.05$) and Tone 4 ($p < 0.05$), but was not significant for Tone 1 ($p = 0.5$) or Tone 3 ($p = 0.35$).
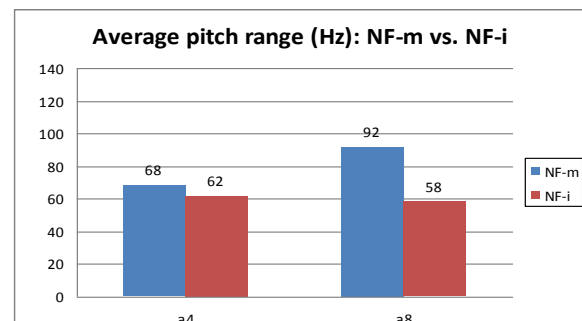

Figure 2: Pitch range in NF-m and NF-i for each age

The above results indicated that both the four- and eight-year-olds could use pitch range to differentiate focus from pre-focus, but only the eight-year-olds could use pitch range to differentiate focus from post-focus. In addition, without looking into each age group, we found that the children as a whole group used a wider pitch range to differentiate focus from post-focus and from pre-focus when the verbs were in Tone 2 and Tone 4, but not in Tone 1 and Tone 3. The four-year-olds' not using pitch range to distinguish focus from post-focus might be caused by their failure to use pitch range in tone 1- and tone-3-verbs.

3.2   Narrow focus vs. broad focus

To examine the realisation of narrow focus compared to that of broad focus, we grouped NF-m and CF-m together as a

combined narrow focus condition (hereafter, the "NF-m&CF-m" condition) with a small focal size, and compared it with the BF condition with a larger focal size. Mixed-effect modeling was adopted and the focus-related fixed factor was SIZE.

3.2.1   Duration

Comparing narrow focus with broad focus, the mixed-effect modelling showed that the main effect of SIZE was not significant ($p = 0.22$). In other words, children did not use duration to differentiate narrow focus from broad focus.

3.2.2   Pitch range

Regarding the comparison between narrow focus and broad focus, the mixed effect modelling showed that the main effect of SIZE was significant ($p < 0.05$). Figure 3 showed that the pitch range of the verbs in the NF-m&CF-m condition was larger than that in the BF condition. The main effect of TONE OF VERB ($p < 0.05$) was also significant, but the main effects of AGE ($p = 0.05$) and TYPE OF VERB ($p = 0.21$) were not significant. No significant interaction was found between FOCUS and AGE ($p = 0.35$), between FOCUS and TONE OF VERB ($p = 0.15$), or between FOCUS and TYPE OF VERB ($p = 0.05$). The results revealed that the four- and eight-year-olds used pitch range to differentiate narrow focus from broad focus, regardless of tones and types of verbs.
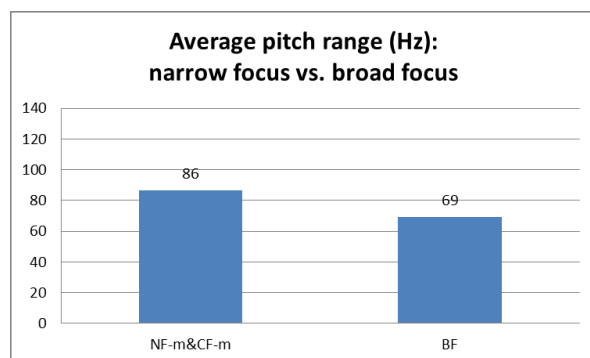

Figure 3: Pitch range of the verbs in NF-m&CF-m and BF

To sum up, children used pitch range but not duration to distinguish narrow focus from broad focus.

3.3   Contrastive focus vs. non-contrastive focus

Regarding the comparison between contrastive and non-contrastive focus, the mixed-effect modeling showed that the main effects of CONTRASTIVITY were not significant for duration ($p = 0.69$) or for pitch range ($p = 0.37$), indicating that the children did not use duration or pitch range to distinguish contrastive focus (CF-m) from non-contrastive (NF-m) focus.

4.   Discussion and Conclusions

This study aimed at finding out how Mandarin-speaking children use pitch range and duration to mark focus in comparison with non-focus, how they encode focus with different constituent size, and how they differentiate contrastive focus from non-contrastive focus. With regard to focus, the children from both age groups produced the focused words with a longer duration than the non-focused ones. Further, both the four- and eight-year-olds used a wider pitch range for the focused verbs than for the pre-focal ones, but

only the eight-year-olds used a wider pitch range for the focused verbs than for the post-focal ones. In addition, the children as a whole group used pitch range to differentiate focus from post-focus and pre-focus for the Tone 2 and Tone 4 verbs, but not for the Tone 1 and Tone 3 verbs. With regard to the size of the focal constituent, the children used pitch range but not duration to differentiate narrow focus from broad focus. With regard to contrastivity, children did not differentiate contrastive narrow focus from non-contrastive narrow focus using duration or pitch, similar to the findings on adult Mandarin [9].

The results had four implications. First, in previous studies on prosodic focus marking, the non-focus condition varies from study to study. Our results show that the defintion of the non-focus condition can influence the results on the use of prosody in distinguishing focus from non-focus in children. Second, to differentiate focus from non-focus, the children used duration regardless of lexical tone but used pitch range only in Tone 2- and Tone 4-verbs, while to differentiate narrow focus from broad focus they used only pitch range. As such selective uses of duration and pitch range have not been observed in adult Mandarin, these results may suggest that the children have not consolidated the use of pitch or duration. Third, as has been mentioned, to become adult like, Mandarin-speaking children have to acquire both the lexical and post-lexical functions of pitch. Previous studies on the acquisition of Mandarin tones showed that the production of Tone 4 is most adult-like in the production of Chinese 3-year-olds, followed by Tone 1, Tone 2, and Tone 3 [17]. However, in terms of focus-marking, we found that the children used pitch range to mark focus only when the verbs were in Tone 2 and Tone 4 but not in Tone 1 or Tone 3. These indicated that the acquisition of pitch range in focus marking may not be related to the order of tonal acquisition. However, To explicate children's use of pitch, not only pitch range but also the maximum and minimum pitch should be analyzed. Last, corss-linguistically, comparing to Dutch-speaking children, who have not acquired the use of duration in focus-marking at the age of seven or eight [14], the Chinese children acquired the use of duration for focus-marking quite early. Besides, the use of pitch range was in place in the Chinese four-year-olds, though not necessarily in all conditions. This suggested an earlier acquisition of pitch range as well in the Chinese children.

5.   Acknowledgements

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] Shih, C. (1988). Tone and intonation in mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-109.

[4] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0contours. *Journal of Phonetics*, 27(1), 55-105.

[5] Chen, Y., & Braun, B. (2006). Prosodic realization of information structure categories in standard Chinese. Paper presented at the *Proceedings of Speech Prosody*, Dresden, Germany.

[6] Shih, C. (2000). A declination model of mandarin Chinese. *Intonation: Analysis, Modelling and Technology*, 243-268.

[7] 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

[8] 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.

[9] Greif, M. (2010). Contrastive focus in mandarin Chinese. Paper presented at the *Proc. Speech Prosody*, Chicago, UAS.

[10] Chen, A. (2011a). Tuning information packaging: intonational realization of topic and focus in child Dutch. *J. Child Lang*, *38*, 1055-1083.

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

[12] MacWhinney, B. and 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.

[13] Chen, A. (2011b). The developmental path to phonological focus-marking in Dutch. In *Prosodic Categories: Production, Perception and Comprehension* (pp. 93-109). Springer Netherlands.

[14] 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.

[15] Li, P., Zhao, X., Liu, Y. and Levine, S. *Chinese Single-character Word Database*. [http://www.personal.psu.edu/pul8/psylin_norm/psychnorms.html]

[16] Xu, Y. and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33: 319-337.

[17] Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, *40*(1), 141-151