



# The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children

First Language  
2018, Vol. 38(1) 26–46  
© The Author(s) 2017



Reprints and permissions:  
[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)  
DOI: 10.1177/0142723717733920  
[journals.sagepub.com/home/fla](http://journals.sagepub.com/home/fla)



**Anqi Yang**

Tianjin University, China

**Aoju Chen**

Utrecht University, The Netherlands

## Abstract

This study investigates how children acquire prosodic focus-marking in Mandarin Chinese. Using a picture-matching game, we elicited spontaneous production of sentences in various focus conditions from children aged four to eleven. We found that Mandarin Chinese-speaking children use some pitch-related cues in some tones and duration in all tones in an adult-like way to distinguish focus from non-focus at the age of four to five. Their use of pitch-related cues is not yet fully adult-like in certain tones at the age of eleven. Further, they are adult-like in the use of duration in distinguishing narrow focus from broad focus at four or five but in not using pitch-related cues for this purpose at seven or eight. The later acquisition of pitch-related cues may be related to the use of pitch for lexical purposes, and the differences in the use of pitch in different tones can be explained by differences in how easy it is to vary pitch-related parameters without changing tonal identity.

## Keywords

Acquisition, development, focus, Mandarin Chinese, prosody

## Introduction

Speakers use a range of linguistic means to distinguish information new to listeners from information already known to listeners. Listeners rely on the linguistic expressions of changes in information structure to efficiently process information. Prosody is used for marking information structure in many languages (e.g. Gussenhoven, 2007; Vallduví &

---

### Corresponding author:

Aoju Chen, Utrecht University, Trans 10, 3512 JK, Utrecht, The Netherlands.  
Email: [aoju.chen@uu.nl](mailto:aoju.chen@uu.nl)

Engdahl, 1996). Generally, speakers realise a word with more prosodic prominence when conveying new and/or contrastive information (also known as ‘focus’) than otherwise. However, there are notable differences between languages in the exact realisation of prosodic prominence. Some languages (e.g. Swedish, German, English, Dutch) can vary prosodic prominence via both phonological means, i.e. making coarse-grained variation in prosodic parameters (e.g. accenting or not accenting a word), and phonetic means, i.e. making fine-grained changes in prosodic parameters within a phonological category (e.g. changes in pitch span of a sentence-level accent or a lexical tone). But other languages can only use phonetic means (e.g. Mandarin Chinese, Cantonese). Languages can also differ in the consistency of the form–function mapping between prosody and focus. For example, in Dutch, the subject-noun of a sentence is nearly always accented regardless of whether it is focal or not (A. Chen, 2009); the object-noun is usually unaccented when non-focal in read speech but can be either unaccented or accented in spontaneous speech (A. Chen, 2011b). The relation between accentuation and focus is thus not consistent in Dutch. It changes with the position of the constituent in a sentence and the modality of speech, and is thus probabilistic by nature. In contrast, in Central Swedish, a word is realised with a prominence-lending high tone only when it is focal (e.g. Bruce, 1998, 2007). Additionally, the specific prosodic parameters involved in focus-marking can be simultaneously used for lexical purposes. For example, pitch is used to distinguish words in tone and pitch-accent languages (e.g. Mandarin Chinese, Cantonese, Central Swedish).

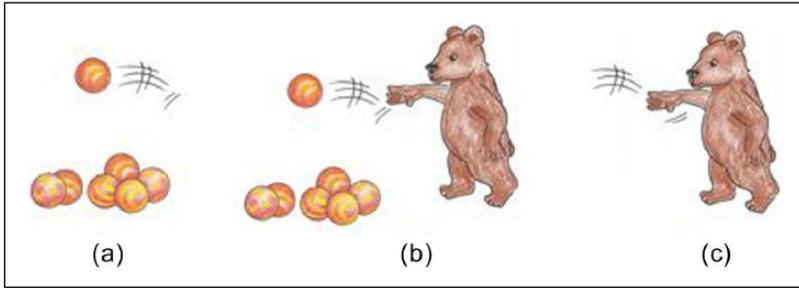
The question that arises in the context of language development is how the above-mentioned differences between languages influence the acquisition of prosodic focus-marking in different languages. Recent years have seen a significant increase in the number of studies examining the use of prosody in focus-marking in children acquiring a Germanic language. These studies have shown that children can already use accentuation to mark focus at the age of three to four but not necessarily in an adult-like way, and become adult-like in both the placement of accent and the type of accent (i.e. the shape of the pitch pattern) at about the age of eight (e.g. Hornby & Hass, 1970; MacWhinney & Bates, 1978; Wells, Peppé, & Goulondris, 2004, on English; Müller, Höhle, Schmitz, & Weissenborn, 2005, on German; A. Chen, 2011a, 2011b, on Dutch). On the other hand, the use of phonetic means as an alternative to phonological means, especially the use of duration, is not adult-like even at the age of eight (A. Chen, 2009, 2015, on Dutch). However, this developmental path to adult-like prosodic focus-marking may not be generalisable to children acquiring a language that differs from West Germanic languages. In the present study, we have investigated children’s use of prosody in focus-marking in Mandarin Chinese (hereafter Mandarin), in order to shed light on the acquisition of prosodic focus-marking from the perspective of a tone language.

Mandarin presents itself as an interesting case for the study of the acquisition of prosodic focus-marking for three reasons. First, as mentioned above, Mandarin only uses phonetic means to realise focus, different from West Germanic languages which primarily use phonological means and only use phonetic means when phonological means do not suffice. Yang and A. Chen (under revision) examined the prosodic realisation of three types of focus: (non-contrastive) narrow focus (i.e. focus on one content word in a sentence), contrastive (narrow) focus (i.e. contrast conveyed by one word in a sentence), and

whole-sentence focus or broad focus, in spontaneous speech. They found that narrow focus was realised with a longer duration and a larger pitch span than non-focus, as found in read speech (Y. Chen & Braun, 2006; Xu, 1999). Narrow focus was also realised with a longer duration and a slightly wider pitch span than broad focus, but was not distinguished from contrastive focus, different from findings on read speech (e.g. Y. Chen & Braun, 2006; Shih, 1988; Xu, 1999). Will the extensive use of phonetic means in prosodic focus-marking influence the rate of acquisition in Mandarin-speaking children, compared to children acquiring West Germanic languages? There are two plausible but opposing answers to this question. On the one hand, Mandarin-speaking children may learn prosodic focus-marking at an earlier age than children acquiring a West Germanic language, because of extensive exposure to the use of phonetic means for focus-marking in the input. On the other hand, they may learn prosodic focus-marking at a similar age to children acquiring a West Germanic language, because children, regardless of their native language, may need a similar amount of time to establish the form–function mapping between prosodic variation and focus in the input and develop sufficient control of prosodic parameters in production.

Furthermore, pitch is not only varied to mark focus but also to distinguish words in Mandarin. More specifically, Mandarin has four lexical tones, i.e. a high level tone (Tone 1), a rising tone (Tone 2), a low tone (Tone 3) and a falling tone (Tone 4), which are primarily identified by pitch movements (e.g. Chao, 1965; Lin, 2007). For example, the syllable ‘ma’ can mean ‘mother’ in Tone 1, ‘hemp’ in Tone 2, ‘horse’ in Tone 3 and ‘to scold’ in Tone 4. This raises the question of whether lexical uses of pitch will affect the order in which the use of pitch and duration is acquired for focus-marking. On the one hand, speakers need to maintain the shape of the pitch contour in a word for the sake of the identity of the lexical tone, thus leaving limited acoustic space for pitch variation for focus-marking purposes. This suggests that speakers need to execute precise control of pitch in order to vary pitch for focus-marking purposes and leave the identity of the lexical tones intact at the same time. Research has shown that young children have difficulty in pitch control over a stretch of speech as long as or longer than a word (A. Chen, 2009). We may thus predict that Mandarin-speaking children take a longer time to become adult-like in the use of pitch than in the use of duration for focus-marking purposes. On the other hand, Mandarin-speaking children produce lexical tones with considerable accuracy as early as three years of age (e.g. Wong, 2012; Wong, Schwartz, & Jenkins, 2005; Zhu, 2002). They may thus have developed considerable sensitivity to pitch variation and are skilful with pitch control in production, which might in turn facilitate their acquisition of the use of pitch for focus-marking purposes. Along this line of reasoning, we may predict that Mandarin-speaking children master the use of pitch earlier than the use of duration for focus-marking purposes. These conflicting predictions need to be tested experimentally.

Relatedly, although they can produce the four tones with considerable accuracy by the age of three, Mandarin-speaking children do not acquire the four tones at the same rate. Specifically, Tone 1 and Tone 4 have generally been found to be acquired earlier than Tone 2 and Tone 3 by young children (e.g. Clumeck, 1977; Li & Thompson, 1977; Wong, 2012; Zhu, 2002). There is no consensus yet on why one tone is acquired later than another.



**Figure 1.** Sample pictures for a trial eliciting narrow focus on the sentence-initial word. Picture (a) was the experimenter's picture with the subject missing; picture (b) was the child's picture containing all the information; picture (c) contained the information missing from picture (a).

Researchers have attributed the relatively later acquisition of certain tones to production-related difficulties (e.g. the complexity of articulatory control), perception-related challenges (e.g. the difficulty of tonal distinction in perception), or the frequency of the tones in the input received by children in infancy and toddlerhood (Wong et al., 2005; Zhu, 2002). Whatever the reasons may be, the findings imply that the four tones are not equally easy to acquire. This raises the question of whether Mandarin-speaking children's use of pitch for focus-marking purposes varies between tones, especially at younger ages. Given the relatively later acquisition of Tones 2 and 3, we may predict later adult-like use of pitch for focus-marking in Tones 2 and 3 than in Tones 1 and 4.

To address the three above-mentioned issues and resolve conflicting predictions regarding the effects of only using phonetic means to mark focus and lexical use of pitch, we have investigated how Mandarin-speaking four- to eleven-year-olds use pitch and duration for focus-marking purposes in spontaneous speech. Specifically, we have examined how they use pitch and duration to distinguish: (1) Narrow focus from non-focus, i.e. pre-focus and post-focus (Effect of focus); (2) Narrow focus from broad focus (Effect of focal constituent size); and (3) Narrow focus from contrastive focus (Effect of contrastivity).

## Method

### *The picture-matching game*

We adapted the picture-matching game used in A. Chen (2011b) to elicit spontaneous production. In this game, the child was supposed 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 lay 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 three pieces of information. On each trial, the experimenter showed one of her pictures (e.g. sample picture (a) in Figure 1) to the child, described the picture

and asked a question about it or made a remark about the missing information (in the contrastive focus condition). The child then took a look at the corresponding picture (e.g. sample picture (b) in Figure 1) in his/her pile and responded to the experimenter's question or remark. The experimenter then looked for the right picture (e.g. sample picture (c) in Figure 1) in the messy pile and matched it with her own picture to form a pair. As rules of the game, the child was asked to answer the experimenter's questions in full sentences and not to reveal his/her pictures to the experimenter. Prior to the picture-matching game, a picture-naming task was conducted to ensure that the child would use the intended words to refer to the entities in the pictures. This procedure also rendered all the entities in the pictures referentially accessible.

### Research design

One-hundred-and-sixty question-answer dialogues were embedded in the picture-matching game to elicit 160 SVO sentences in five focus conditions: narrow focus on the subject-noun in sentence-initial position (NF-i), responding to who-questions; narrow focus on the object-noun in sentence-final position (NF-f), responding to what-questions; narrow focus on the verb in sentence-medial position (NF-m), responding to what-does-X-do-to-Y questions; contrastive focus in sentence-medial position (CF-m), correcting the experimenter's remark about the action; broad focus over the whole sentence (BF), responding to what-happens questions, as illustrated in (a). Including narrow focus in three sentence-positions made it possible to study the effect of narrow focus on the sentence-medial verbs (NF-m) compared to the same verbs in pre-focus (or NF-f) and post-focus (or NF-i). Comparing the verbs in NF-m, CF-m and BF allowed us to study the prosodic difference between different focus types. Examples of question-answer dialogues in the five focus condition are given in (1) below, in which the digits represent tones in Mandarin, and the referents are referred to with the definite article in the English glossary because they have been introduced in the picture-naming task.

(1) Examples of question-answer dialogues between the experimenter (E) and participant (P):

- (NF-i)** E: 看! 球。球在空中。看起来有小动物扔球。谁扔球?  
Look! The ball. The ball is in the air. It looks like someone throws the ball. Who throws the ball?
- P: [小熊] 扔 球。  
[xiao3 xiong2] reng1 qiu2.  
[The little bear] throws the ball.
- (NF-f)** E: 看! 小猫, 小猫的胳膊伸出去了。看起来小猫扔东西。小猫扔什么?  
Look! The little cat. The little cat stretches out its arm. It looks like the little cat throws something. What does the little cat throw?
- P: 小猫 扔 [笔]。  
xiao3 mao1 reng1 [bi3].  
The little cat throws [the pen].

- (NF-m)** E: 看！小兔，还有书。看起来小兔要弄书。小兔怎么弄书？  
Look! The little rabbit, and the book. It looks like the little rabbit does something to the book. What does the little rabbit do to the book?
- P: 小兔 [扔] 书。  
xiao3 tu4 [reng1] shu1.  
The little rabbit [throws] the book.
- (CF-m)** E: 看！小熊，还有菜。看起来小熊要弄菜。我猜：小熊[埋]菜。  
Look! The little bear, and the vegetables. It looks like the little bear will do something to the vegetables. I will make a guess: The little bear [buries] the vegetables.
- P: 小熊 [扔] 菜。  
xiao3 xiong2 [reng1] cai4.  
The little bear [throws] the vegetables.
- (BF)** E: 看！我的图片是模糊的。什么都看不清。你的图片上讲了什么？  
Look! My picture is very blurry. I cannot see anything clearly. What happens in your picture?
- P: [小狗 扔 菜]。  
[xiao3 gou3 reng1 cai4].  
[The little dog throws the vegetables].

The *svo* sentences were unique combinations of four disyllabic subject-noun phrases starting with the word *xiao3* ‘little’ (one noun phrase per lexical tone regarding the second word in each phrase), eight monosyllabic verbs (one verb per lexical tone per group), and eight monosyllabic object-nouns (one noun per lexical tone per group), as shown in Table 1. Each group-1 verb was combined once with each group-1 object-noun and each group-2 verb was combined once with each group-2 object-noun, leading to 160 verb phrases (4 tones in verbs  $\times$  4 tones in objects  $\times$  2 groups of verbs and object-nouns  $\times$  5 focus conditions = 160 VPs). We used two groups of verbs so that we had two realisations of each tone in the verbs and could avoid data loss regarding the tones in the case that a child mispronounced one verb. As we could not find four nouns representing four tones among the words reported to be acquired by Mandarin-speaking four-year-olds (Liu, Shu, & Li, 2007) that could form semantically appropriate VPs with all the eight target verbs, we paired each group of verbs with their ‘own’ object-nouns, thus having two groups of object-nouns. The subject-nouns were then approximately evenly distributed over the verb phrases, forming 160 *svo* sentences. This procedure made sure that in each focus condition, each tone in the verbs was combined with each tone in the preceding subject-noun and with each tone in the following object-noun.

The 160 sentences were then split evenly into two lists (List 1 and List 2) of 80 sentences. Each list contained all verb-object tonal combinations but not all verb-object word combinations. The trials on each list were randomised in a way that trials from the same focus condition did not appear after each other and the focused constituent of a trial was not mentioned in its preceding trial. Approximately half of the participants in each age group produced the sentences in List 1 and the other half produced the sentences in List 2.

**Table 1.** Words that occurred in the sentences.

	Tone 1	Tone 2	Tone 3	Tone 4
Subjects	小猫 xiao3 mao1 'little cat'	小熊 xiao3 xiong2 'little bear'	小狗 xiao3 gou3 'little dog'	小兔 xiao3 tu4 'little rabbit'
Group-1 verbs	扔 reng1 'throw'	埋 mai2 'bury'	剪 jian3 'cut'	运 yun4 'transport'
Group-2 verbs	浇 jiao1 'water'	闻 wen2 'smell'	舔 tian3 'lick'	卖 mai4 'sell'
Group-1 objects	书 shu1 'book'	球 qiu2 'ball'	笔 bi3 'pen'	菜 cai4 'vegetable'
Group-2 objects	花 hua1 'flower'	梨 li2 'pear'	草 cao3 'grass'	树 shu4 'tree'

### *Participants and procedure*

Three groups of children, i.e. 12 four- to five-year-olds (average age: 5;2; range: 4;6–5;10), 10 seven- to eight-year-olds (average age: 7;10; range: 7;2–8;3) and 12 ten- to eleven-year-olds (average age: 10;9; range: 10;1–11;5) participated in the experiment. They were recruited from Beijing 21st Century kindergarten and primary school, and they all spoke Mandarin as their native language without any detectable regional accent. Twelve adult native speakers of Mandarin (average age: 19 years; range: 18–20 years; six females and six males) were tested as a control group. They were undergraduates from Beijing Forestry University at the time of testing, and all spoke Mandarin without any detectable regional accent. Three female native speakers of Mandarin administered the experiment after having received intensive training on how to conduct it. The children were tested individually in quiet rooms in their kindergarten or school. Two experimental sessions – one eliciting 40 sentences with the group-1 words and the other eliciting 40 sentences with the group-2 words – were held on two different days for each four- to five-year-old, and on the same day for the older children but with a break in between. The children were also allowed to take a break during an experimental session whenever necessary. Each experimental session lasted about 35 minutes and was audio-recorded at a sampling rate of 44.1 kHz with 16 bits using a ZOOM H1 recorder. The experimental sessions with the four- to five-year-olds were also video-recorded for training purposes. The adults were tested following the same experimental procedure, but were informed that the game was of a simple nature because it would also be played with children.

### *Annotation*

The audio recording from each participant was first orthographically annotated in Praat (Boersma & Weenink, 2013). Then usable sentences were selected for phonetic annotation (59% for the four- to five-year-olds, 82% for the seven- to eight-year-olds,

and 90% for the ten- to eleven-year-olds), and unusable ones were excluded from further analysis.<sup>1</sup> A target sentence was considered unusable only in one of the following cases: (1) the participant produced the target sentence before the experimenter asked the question, (2) the experimenter asked a different question than the intended question in that trial, (3) the experimenter did not provide an adequate description of the picture before she asked a question, (4) the sentence was produced at a noisy moment of the kindergarten or school, (5) the sentence was produced with word insertion, deletion, or replacement, (6) the sentence was produced with self-repair or clearly perceivable hesitation, or (7) the sentence was produced with perceivable marked intonation, such as chanting-, singing- and howling-like intonation, or with laughter. As we were interested in the use of pitch span, defined as the difference between the maximum pitch (pitch-max) and minimum pitch (pitch-min) of the target words, and duration in focus-marking, we then annotated the target words in the usable sentences for word boundaries, following standard procedures (Machač & Skarnitzl, 2009), and pitch-max and pitch-min, taking the tonal targets of the lexical tones into account (Xu, 1997; Xu & Wang, 2001).

## Analysis and results

As mentioned earlier, the target words were focal in the NF-m, BF and CF-m conditions, post-focal in the NF-i condition, and pre-focal in the NF-f condition. We analysed the effect of narrow focus (i.e. NF-m vs NF-i/NF-f) and focus types differing in size (i.e. NF-m vs BF) or contrastivity (i.e. NF-m vs CF-m) on the use of each duration- or pitch-related prosodic cue. The data of the adult control group (Yang & A. Chen, under revision) were included in the statistical analysis for comparison reasons.

Mixed-effects modelling in the program R with the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) was used to assess the data. The dependent variables were duration, pitch span, pitch-max and pitch-min of the target words. The random factors were *SPEAKER* (i.e. the participants) and *SENTENCE* (i.e. the target sentences). The fixed factors were *AGE*, *FOCUS* and *TONE*. *AGE* referred to the four age groups and thus had four levels (i.e. three groups of children and one group of adults. The adults was set as the reference category<sup>2</sup>). *FOCUS* referred to the focus conditions. For each analysis, two conditions were compared to each other in order to address a specific question. *FOCUS* thus always had two levels. *TONE* referred to the tones of the target words, and had four levels (i.e. Tone 1, 2, 3 and 4).

To find out whether a particular prosodic cue was used to distinguish two focus conditions, models were built using the aforementioned factors. Starting from an ‘empty’ model (hereafter Model 0) containing only the random factors, we added the main effects of the fixed factors, the two-way interactions between each two fixed factors, and the three-way interaction between all of them to the model in a stepwise fashion, building seven additional models (Table 2). The ANOVA function in R was used to compare models in order to derive the model with the best fit.

For each analysis, we report on the best-fit model according to the model comparisons, and statistically significant main effects or interactions according to the summary of the best-fit model. To control the false discovery rate (i.e. the proportion of significant results that are actually false positives) in multiple mixed-effects models on the data obtained from the same participants, we adopted the Benjamini–Hochberg procedure using a strict false discovery rate value 0.1, instead of the recommend 0.25 (Benjamini

**Table 2.** Model build-up procedure.

Model	Factor added
Model 0	
Model 1	FOCUS
Model 2	AGE
Model 3	TONE
Model 4	FOCUS : AGE
Model 5	FOCUS : TONE
Model 6	AGE : TONE
Model 7	FOCUS : AGE : TONE

& Hochberg, 1995; Simes, 1986), by means of the Excel spreadsheet developed by McDonald (2016). This procedure revealed that all effects with a  $p$ -value smaller than 0.05 and one effect with a  $p$ -value equal to 0.05 remained to be significant and that five effects with a  $p$ -value slightly larger than 0.05 (0.052 ~ 0.058) turned out to be significant as well. We decided to focus on the main effects and interactions emerging from the best-fit models with a  $p$ -value smaller than 0.05.

As the model summary of the best-fit model does not straightforwardly show the difference between two focus conditions in the use of a particular prosodic cue in each age group, we did additional mixed-effects modelling on each age group to obtain a clearer picture on interactions involving the factor AGE. If the best-fit model contains the three-way interaction of FOCUS, AGE and TONE, we discuss how the speakers in each age group distinguished the two focus conditions in each tonal category by examining the interaction of FOCUS and TONE in each age group. If the best-fit model contains the two-way interaction of FOCUS and AGE, we discuss how the speakers in each age group distinguished the two focus conditions by examining the main effect of FOCUS in each age group. We do not discuss the main effects of the factors when the interactions involving these factors are significant, two-way interactions when three-way interactions involving the same factors are significant, and interactions that do not involve the factors FOCUS and AGE.

### *Effect of narrow focus (Narrow focus vs non-focus)*

#### *Narrow focus vs pre-focus (NF-m vs NF-f)*

**WORD DURATION.** The summary of the best-fit model (Model 4) for the analysis on word duration showed that the interaction of FOCUS and AGE was significant. Subsequent analysis on each age group showed that the word duration was significantly longer in narrow focus than in pre-focus in all the four age groups: the four- to five-year-olds (69 ms longer,  $SE = 0.009$ ,  $df = 50.730$ ,  $t = -7.752$ ,  $p < 0.001$ ,  $r^2 = 0.535$ ,  $\Omega^2_0 = 0.527$ ), the seven- to eight-year-olds (61 ms longer,  $SE = 0.008$ ,  $df = 62.270$ ,  $t = -7.621$ ,  $p < 0.001$ ,  $r^2 = 0.651$ ,  $\Omega^2_0 = 0.640$ ), the ten- to eleven-year-olds (33 ms longer,  $SE = 0.006$ ,  $df = 63.230$ ,  $t = -5.350$ ,  $p < 0.001$ ,  $r^2 = 0.579$ ,  $\Omega^2_0 = 0.567$ ), and the adults (19 ms longer,  $SE = 0.005$ ,  $df = 60.820$ ,  $t = -3.995$ ,  $p < 0.001$ ,  $r^2 = 0.625$ ,  $\Omega^2_0 = 0.613$ ) (Figure 2). However, the children varied word duration to a much larger extent than the adults. Among the children, the four- to five-year-olds and seven- to eight-year-olds varied word duration to a larger degree than the ten- to eleven-year-olds.

**PITCH SPAN.** The summary of the best-fit model (Model 6) for the analysis on pitch span showed that the interaction of FOCUS and AGE was significant. Subsequent analysis on each age group showed that the pitch span was significantly wider in narrow focus than in pre-focus in the seven- to eight-year-olds (1.8 st wider,  $SE = 0.487$ ,  $df = 59.400$ ,  $t = -3.777$ ,  $p < 0.05$ ,  $r^2 = 0.757$ ,  $\Omega^2_0 = 0.745$ ) and ten- to eleven-year-olds (1.5 st wider,  $SE = 0.400$ ,  $df = 58.060$ ,  $t = -3.817$ ,  $p < 0.05$ ,  $r^2 = 0.737$ ,  $\Omega^2_0 = 0.729$ ), similar to the adults (0.7 st wider,  $SE = 0.295$ ,  $df = 60.650$ ,  $t = -2.517$ ,  $p < 0.05$ ,  $r^2 = 0.666$ ,  $\Omega^2_0 = 0.652$ ), but not in the four- to five-year-olds (Figure 3).

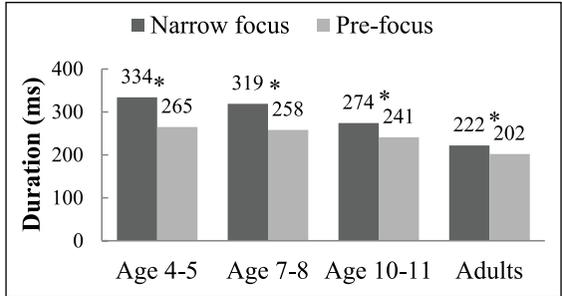


Figure 2. The use of duration in distinguishing narrow focus from pre-focus.

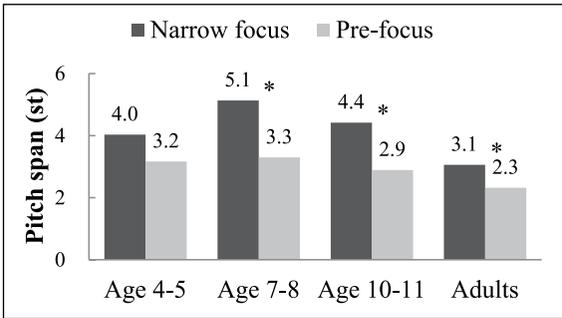
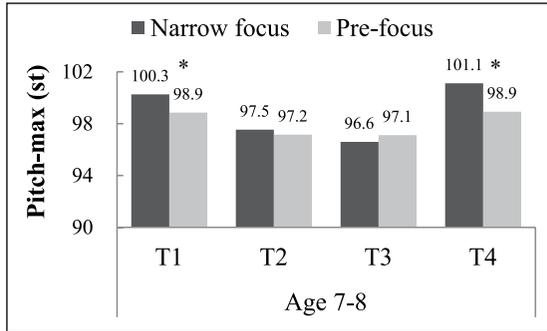


Figure 3. The use of pitch span in distinguishing narrow focus from pre-focus.

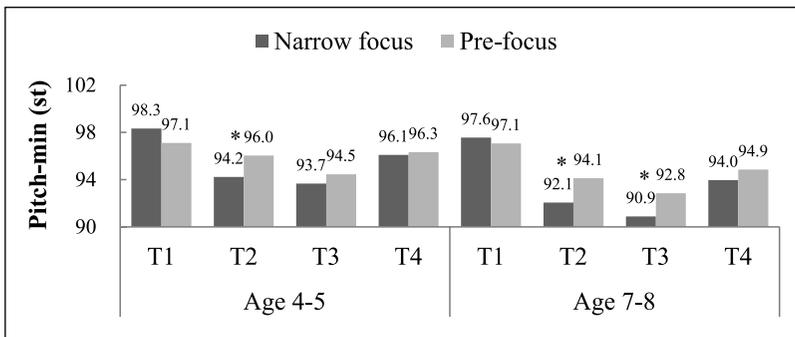
**PITCH-MAX.** The summary of the best-fit model (Model 7) for the analysis on pitch-max showed that the interaction of FOCUS, AGE and TONE was significant. Subsequent analysis revealed no significant main effect of FOCUS or significant interaction of FOCUS and TONE in the four- to five-year-olds, indicating that they did not use pitch-max to distinguish narrow focus from pre-focus, regardless of tone, similar to the adults. In the seven- to eight-year-olds, there was a significant interaction of FOCUS and TONE. They used a significantly higher pitch-max in narrow focus than in pre-focus for words in Tone 1 (1.4 st higher,  $SE = 0.569$ ,  $df = 59.390$ ,  $t = -2.463$ ,  $p < 0.05$ ,  $r^2 = 0.894$ ,  $\Omega^2_0 = 0.892$ ) and Tone 4 (2.2 st higher,  $SE = 0.573$ ,  $df = 61.040$ ,  $t = -3.821$ ,  $p < 0.001$ ,  $r^2 = 0.816$ ,  $\Omega^2_0 = 0.814$ ), but not for words in Tone 2 or Tone 3 (Figure 4). In the ten- to eleven-year-olds, there were significant main effects of FOCUS and TONE, but no significant interaction, suggesting that they used a higher

pitch-max in narrow focus than in pre-focus (0.7 st higher,  $SE = 0.246$ ,  $df = 62.710$ ,  $t = -2.653$ ,  $p < 0.05$ ,  $r^2 = 0.889$ ,  $\Omega_0^2 = 0.889$ ), regardless of tone.



**Figure 4.** The use of pitch-max in distinguishing narrow focus from pre-focus in different tones in the seven- to eight-year-olds.

**PITCH-MIN.** The summary of the best-fit model (Model 7) for the analysis on pitch-min showed that the interaction of FOCUS, AGE and TONE was significant. Subsequent analysis revealed a significant effect of FOCUS, but no significant interaction of FOCUS and TONE in the ten- to eleven-year-olds, indicating that they used a lower pitch-min in narrow focus than in pre-focus (0.9 st lower,  $SE = 0.344$ ,  $df = 57.750$ ,  $t = 2.622$ ,  $p < 0.05$ ,  $r^2 = 0.887$ ,  $\Omega_0^2 = 0.886$ ), regardless of tone, similar to the adults (0.5 st lower,  $SE = 0.222$ ,  $df = 61.450$ ,  $t = 2.148$ ,  $p < 0.05$ ,  $r^2 = 0.983$ ,  $\Omega_0^2 = 0.983$ ). In both the four- to five-year-olds and the seven- to eight-year-olds we found a significant interaction of FOCUS and TONE. The four- to five-year-olds used a lower pitch-min in narrow focus than in pre-focus for words in Tone 2 (1.8 st lower,  $SE = 0.691$ ,  $df = 34.490$ ,  $t = 2.627$ ,  $p < 0.05$ ,  $r^2 = 0.757$ ,  $\Omega_0^2 = 0.720$ ), but not for words in the other tones (Figure 5). The seven- to eight-year-olds used a lower pitch-min in narrow focus than in pre-focus for words in Tone 2 (2.0 st lower,  $SE = 0.630$ ,  $df = 54.090$ ,  $t = 3.266$ ,  $p < 0.05$ ,  $r^2 = 0.785$ ,  $\Omega_0^2 = 0.770$ ) and Tone 3 (1.9 st lower,  $SE = 0.836$ ,  $df = 75.500$ ,  $t = 2.341$ ,  $p < 0.05$ ,  $r^2 = 0.855$ ,  $\Omega_0^2 = 0.836$ ), but not for words in the other tones (Figure 5).

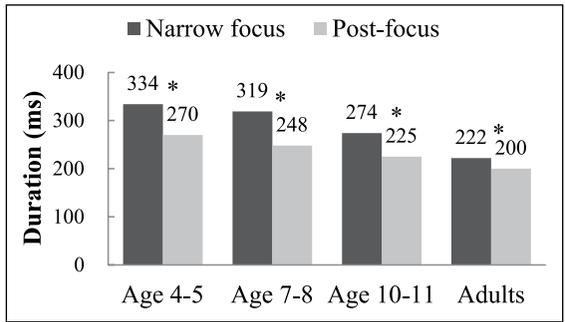


**Figure 5.** The use of pitch-min in distinguishing narrow focus from pre-focus in different tones in the four- to five-year-olds and seven- to eight-year-olds.

**INTERIM SUMMARY.** To distinguish narrow focus from pre-focus, the three groups of children all used a longer duration for the focused words than for the pre-focal ones, similar to the adults. Only the seven- to eight-year-olds and ten- to eleven-year-olds used a wider pitch span for the focused words than for the pre-focal ones, as the adults did. Different from the adults who expanded the pitch span mainly by lowering the pitch-min, the ten- to eleven-year-olds did that by both raising the pitch-max and lowering the pitch-min. The way in which the seven- to eight-year-olds expanded the pitch span of the focused words was conditioned by tone. They expanded the pitch span of the Tone 1 and Tone 4 words mainly by raising the pitch-max and expanded the pitch span of the Tone 2 and Tone 3 words mainly by lowering the pitch-min.

*Narrow focus vs post-focus (NF-m vs NF-i)*

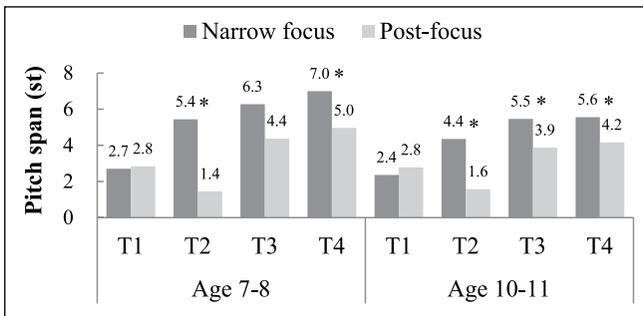
**WORD DURATION.** The summary of the best-fit model (Model 4) for the analysis on word duration showed that the interaction of FOCUS and AGE was significant. Subsequent analysis on each age group showed that the word duration was significantly longer in narrow focus than in post-focus in the four- to five-year-olds (64 ms longer,  $SE = 0.010$ ,  $df = 57.860$ ,  $t = -6.771$ ,  $p < 0.001$ ,  $r^2 = 0.601$ ,  $\Omega^2_0 = 0.591$ ), the seven- to eight-year-olds (71 ms longer,  $SE = 0.007$ ,  $df = 62.500$ ,  $t = -10.06$ ,  $p < 0.001$ ,  $r^2 = 0.676$ ,  $\Omega^2_0 = 0.669$ ), the ten- to eleven-year-olds (49 ms longer,  $SE = 0.007$ ,  $df = 63.020$ ,  $t = -7.242$ ,  $p < 0.001$ ,  $r^2 = 0.703$ ,  $\Omega^2_0 = 0.697$ ), and the adults (22 ms longer,  $SE = 0.005$ ,  $df = 61.600$ ,  $t = -4.706$ ,  $p < 0.001$ ,  $r^2 = 0.568$ ,  $\Omega^2_0 = 0.552$ ) (Figure 6). However, the children varied word duration to a much larger extent than the adults. Among the children, the four- to five-year-olds and seven- to eight-year-olds varied word duration to a larger degree than the ten- to eleven-year-olds.



**Figure 6.** The use of duration in distinguishing narrow focus from post-focus.

**PITCH SPAN.** The summary of the best-fit model (Model 7) for the analysis on pitch span showed that the interaction of FOCUS, AGE and TONE was significant. Subsequent analysis showed no significant main effect of FOCUS or significant interaction of FOCUS and TONE in the four- to five-year-olds, indicating they did not use pitch span to distinguish narrow focus from post-focus, different from the adults, who used a wider pitch

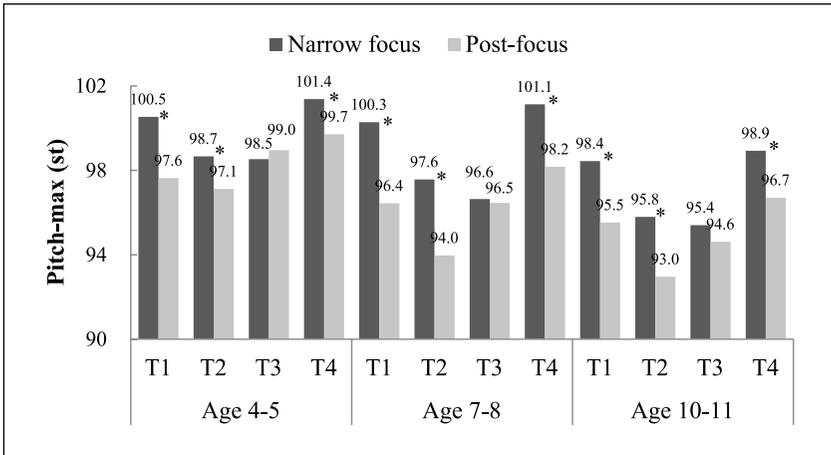
span in narrow focus than in post-focus regardless of tones (0.8 st wider,  $SE = 0.233$ ,  $df = 64.120$ ,  $t = -3.261$ ,  $p < 0.05$ ,  $r^2 = 0.687$ ,  $\Omega^2_0 = 0.677$ ). In both the seven- to eight-year-olds and ten- to eleven-year-olds, we found a significant interaction of FOCUS and TONE. The seven- to eight-year-olds used a significantly wider pitch span in narrow focus than in post-focus for words in Tone 2 (4.0 st wider,  $SE = 0.736$ ,  $df = 55.190$ ,  $t = -5.424$ ,  $p < 0.001$ ,  $r^2 = 0.741$ ,  $\Omega^2_0 = 0.734$ ) and Tone 4 (2.0 st wider,  $SE = 0.742$ ,  $df = 57.270$ ,  $t = -2.738$ ,  $p < 0.05$ ,  $r^2 = 0.842$ ,  $\Omega^2_0 = 0.833$ ). They also tended to use a wider pitch span in narrow focus than in post-focus for words in Tone 3, but this tendency only approached statistical significance (1.9 st wider,  $SE = 0.958$ ,  $df = 80.580$ ,  $t = -1.989$ ,  $p = 0.05$ ,  $r^2 = 0.578$ ,  $\Omega^2_0 = 0.548$ ). They did not vary pitch span for the same distinction for Tone 1 (Figure 7). The ten- to eleven-year-olds used a significantly wider pitch span in narrow focus than in post-focus for words in Tone 2 (2.8 st wider,  $SE = 0.698$ ,  $df = 56.890$ ,  $t = -3.980$ ,  $p < 0.001$ ,  $r^2 = 0.770$ ,  $\Omega^2_0 = 0.760$ ), Tone 3 (1.6 st wider,  $SE = 0.734$ ,  $df = 66.350$ ,  $t = -2.174$ ,  $p < 0.05$ ,  $r^2 = 0.726$ ,  $\Omega^2_0 = 0.679$ ) and Tone 4 (1.4 st wider,  $SE = 0.677$ ,  $df = 50.720$ ,  $t = -2.061$ ,  $p < 0.05$ ,  $r^2 = 0.648$ ,  $\Omega^2_0 = 0.621$ ), but not in Tone 1 (Figure 7).



**Figure 7.** The use of pitch span in distinguishing narrow focus from post-focus in different tones in the seven- to eight-year-olds and ten- to eleven-year-olds.

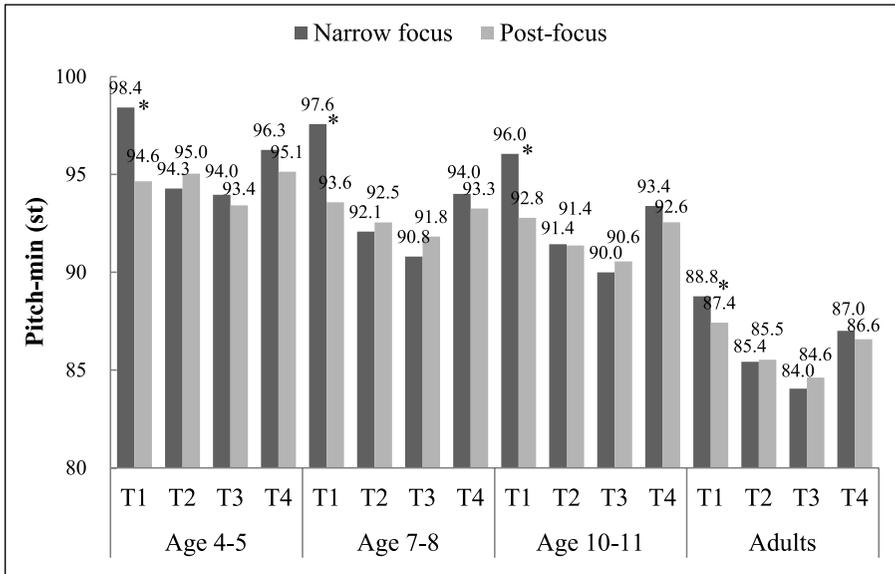
**PITCH-MAX.** The summary of the best-fit model (Model 7) for the analysis on pitch-max showed that the interaction of FOCUS, AGE and TONE was significant. Subsequent analysis showed a significant interaction of FOCUS and TONE in each group of children. The four- to five-year-olds used a significantly higher pitch-max in narrow focus than in post-focus for words in Tone 1 (2.9 st higher,  $SE = 0.688$ ,  $df = 63.280$ ,  $t = -4.224$ ,  $p < 0.001$ ,  $r^2 = 0.721$ ,  $\Omega^2_0 = 0.715$ ), Tone 2 (1.6 st higher,  $SE = 0.693$ ,  $df = 59.540$ ,  $t = -2.230$ ,  $p < 0.05$ ,  $r^2 = 0.646$ ,  $\Omega^2_0 = 0.598$ ) and Tone 4 (1.7 st higher,  $SE = 0.664$ ,  $df = 55.320$ ,  $t = -2.506$ ,  $p < 0.05$ ,  $r^2 = 0.545$ ,  $\Omega^2_0 = 0.476$ ), but not for words in Tone 3 (Figure 8). The seven- to eight-year-olds also used a significantly higher pitch-max in narrow focus than in post-focus for words in Tone 1 (3.9 st higher,  $SE = 0.740$ ,  $df = 58.260$ ,  $t = -5.196$ ,  $p < 0.001$ ,  $r^2 = 0.895$ ,  $\Omega^2_0 = 0.892$ ), Tone 2 (3.6 st higher,  $SE = 0.753$ ,  $df = 62.240$ ,  $t = -4.783$ ,  $p < 0.001$ ,  $r^2 = 0.848$ ,  $\Omega^2_0 = 0.843$ ) and Tone 4

(2.9 st higher,  $SE = 0.749$ ,  $df = 61.000$ ,  $t = -3.946$ ,  $p < 0.001$ ,  $r^2 = 0.864$ ,  $\Omega^2_0 = 0.860$ ), but not for words in Tone 3 (Figure 8). The ten- to eleven-year-olds were similar to the two other groups of children, also using a significantly higher pitch-max in narrow focus than in post-focus for words in Tone 1 (2.9 st higher,  $SE = 0.555$ ,  $df = 63.820$ ,  $t = -5.247$ ,  $p < 0.001$ ,  $r^2 = 0.883$ ,  $\Omega^2_0 = 0.882$ ), Tone 2 (2.8 st higher,  $SE = 0.564$ ,  $df = 67.380$ ,  $t = -5.029$ ,  $p < 0.001$ ,  $r^2 = 0.886$ ,  $\Omega^2_0 = 0.882$ ) and Tone 4 (2.2 st higher,  $SE = 0.546$ ,  $df = 60.060$ ,  $t = -4.061$ ,  $p < 0.001$ ,  $r^2 = 0.914$ ,  $\Omega^2_0 = 0.913$ ), but not for words in Tone 3. Differently, the adults used a significantly higher pitch-max in narrow focus than in post-focus, regardless of tones (1.1 st higher,  $SE = 0.209$ ,  $df = 61.740$ ,  $t = -5.214$ ,  $p < 0.001$ ,  $r^2 = 0.977$ ,  $\Omega^2_0 = 0.977$ ) (Figure 8). However, all the three groups of children realised narrow focus with a substantially larger increase in pitch-max ( $> 1.1$  st) than the adults when they did use pitch-max systematically.



**Figure 8.** The use of pitch-max in distinguishing narrow focus from post-focus in different tones in the three groups of children.

**PITCH-MIN.** The summary of the best-fit model (Model 7) for the analysis on pitch-min showed that the interaction of FOCUS, AGE and TONE was significant. Subsequent analysis showed a significant interaction of FOCUS and TONE in each age group. The four age groups all used a higher pitch-min in narrow focus than in post-focus for words in Tone 1, but the children produced a larger difference in pitch-min between the two focus conditions than the adults (3.8 st higher for the four- to five-year-olds,  $SE = 0.605$ ,  $df = 167.230$ ,  $t = -6.230$ ,  $p < 0.001$ ,  $r^2 = 0.721$ ,  $\Omega^2_0 = 0.715$ ; 4.0 st higher for the seven- to eight-year-olds,  $SE = 0.554$ ,  $df = 50.400$ ,  $t = -7.215$ ,  $p < 0.001$ ,  $r^2 = 0.895$ ,  $\Omega^2_0 = 0.892$ ; 3.2 st higher for the ten- to eleven-year-olds,  $SE = 0.610$ ,  $df = 52.580$ ,  $t = -5.352$ ,  $p < 0.001$ ,  $r^2 = 0.883$ ,  $\Omega^2_0 = 0.882$ ; 1.3st higher for the adults,  $SE = 0.423$ ,  $df = 55.200$ ,  $t = -3.187$ ,  $p < 0.05$ ,  $r^2 = 0.984$ ,  $\Omega^2_0 = 0.984$ ) but not for words in the other tones (Figure 9).



**Figure 9.** The use of pitch-min in distinguishing narrow focus from post-focus in different tones in each age group.

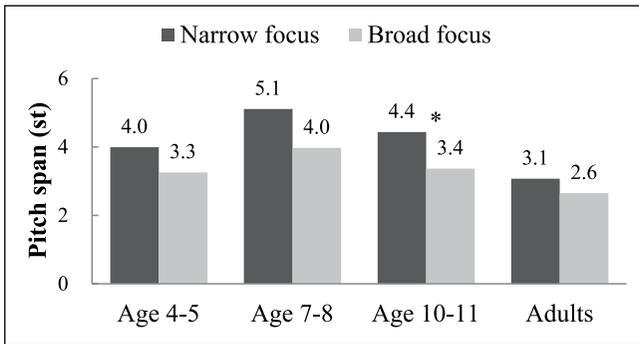
*INTERIM SUMMARY.* To distinguish narrow focus from post-focus, the three groups of children all used a longer duration in narrow focus than in post-focus, similar to the adults. However, different from the adults, who used a wider pitch span in narrow focus than in post-focus, the ten- to eleven-year-olds used pitch span in this manner for words in Tone 2, Tone 3 and Tone 4, and the seven- to eight-year-olds used pitch span in this manner for words in Tone 2 and Tone 4. The four- to five-year-olds, however, did not vary pitch span for this purpose. Further, different from the adults, who used a higher pitch-max in narrow focus than in post-focus, the three groups of children only used pitch-max in this manner for words in Tone 1, Tone 2 and Tone 4 but to a larger extent than the adults. In addition, the children all used a higher pitch-min in narrow focus than in post-focus for words in Tone 1, similar to the adults but again to a larger extent than the adults. Thus, in the two groups of older children, the pitch span difference between narrow focus and post-focus for words in Tone 2 and Tone 4 was caused by a difference between focus and post-focus in pitch-max, similar to the adults' production. Last, the three groups of children all used a higher pitch register (i.e. a higher pitch-max and a higher pitch-min) in narrow focus than in post-focus for words in Tone 1, similar to the adults.

### *Effect of size of focal constituent (Narrow focus vs broad focus (NF-m vs BF))*

*WORD DURATION.* The summary of the best-fit model (Model 2) for the analysis on word duration showed that the main effects of FOCUS and AGE were significant, but the

interaction of FOCUS and AGE was not significant. Thus the speakers used a significantly longer word duration in narrow focus than in broad focus, regardless of age (20 ms longer,  $SE = 0.006$ ,  $df = 63.750$ ,  $t = -3.540$ ,  $p < 0.001$ ,  $r^2 = 0.683$ ,  $\Omega^2_0 = 0.682$ ).

**PITCH SPAN.** The summary of the best-fit model (Model 6) for the analysis on pitch span showed that the interaction of FOCUS and AGE was significant. Subsequent analysis on each age group showed that the pitch span was significantly wider in narrow focus than in broad focus in the ten- to eleven-year-olds (1.0 st wider,  $SE = 0.424$ ,  $df = 56.480$ ,  $t = -2.523$ ,  $p < 0.05$ ,  $r^2 = 0.749$ ,  $\Omega^2_0 = 0.740$ ), but not in the other two groups of children or in the adults (Figure 10).



**Figure 10.** The use of pitch span in distinguishing narrow focus from broad focus.

**PITCH-MAX.** The summary of the best-fit model (Model 6) for the analysis on pitch-max showed that the interaction of FOCUS and AGE was significant. Subsequent analysis on each age group showed that the pitch-max was significantly higher in narrow focus than in broad focus in the four- to five-year-olds (1.3 st higher,  $SE = 0.435$ ,  $df = 57.500$ ,  $t = -2.936$ ,  $p < 0.05$ ,  $r^2 = 0.616$ ,  $\Omega^2_0 = 0.588$ ), but not in the other two groups of older children or in the adults.

**PITCH-MIN.** The summary of the best-fit model (Model 6) for the analysis on pitch-min showed that the main effect of FOCUS was not significant, and it was not involved in any interaction. Thus, the speakers did not use pitch-min to distinguish narrow focus from broad focus, regardless of age.

**INTERIM SUMMARY.** The three groups of children all used a longer duration in narrow focus than in broad focus, similar to the adults. The seven- to eight-year-olds did not vary the pitch-related cues for this distinction, similar to the adults. The ten- to eleven-year-olds used a wider pitch span and the four- to five-year-olds used a higher pitch-max for the words in narrow focus than in broad focus, different from the adults.

### *Effect of contrastivity (Narrow focus vs contrastive focus (NF-m vs CF-m))*

**WORD DURATION, PITCH SPAN AND PITCH-MAX.** In the analyses on word duration, pitch span and pitch-max, the effect of FOCUS was not significant, and it was not involved

in any interaction. Thus, the children did not use these cues to distinguish narrow focus from contrastive focus, similar to the adults.

*PITCH-MIN.* The summary of the best-fit model (Model 6) for the analysis on pitch-min showed that the interaction of FOCUS and AGE was significant. Subsequent analysis on each age group showed that the pitch-min was significantly lower in contrastive focus than in narrow focus in the four- to five-year-olds (1.4 st lower,  $SE = 0.664$ ,  $df = 46.220$ ,  $t = -2.075$ ,  $p < 0.05$ ,  $r^2 = 0.706$ ,  $\Omega^2_0 = 0.667$ ), but not in the other two groups of children or in the adults.

*INTERIM SUMMARY.* The seven- to eight-year-olds and ten- to eleven-year-olds did not use any of the four phonetic cues to distinguish narrow focus from contrastive focus, similar to the adults. The four- to five-year-olds used a lower pitch-min for contrastive focus than for narrow focus, different from the adults.

## Discussion

We have examined how Mandarin-speaking four- to five-year-olds, seven- to eight-year-olds, and ten- to eleven-year-olds use duration and pitch to distinguish narrow focus from non-focus and two other types of focus, compared to adults in spontaneous speech.

With regard to the prosodic realisation of narrow focus, we have observed adult-like use of duration in distinguishing narrow focus from both pre-focus and post-focus at the age of four to five. However, children make no systematic use of pitch span for the same purposes at this age, unlike adults. Nevertheless, they can vary pitch-max and pitch-min to distinguish focus from post-focus and vary pitch-min to distinguish focus from pre-focus in an adult-like way in certain tonal categories. By the age of seven to eight, they can vary pitch span in an adult-like way to distinguish focus from pre-focus regardless of tone, and to distinguish focus from post-focus in Tone 2 and Tone 4. By the age of ten to eleven, they become adult-like in the use of pitch span in distinguishing focus from post-focus in all but Tone 1. In addition, children's use of pitch-max and pitch-min is still not fully adult-like at the age of seven to eleven, thus unadult-like in how they realise the difference in pitch span between different focus conditions. Additionally, it is worth noting that children vary the duration- and pitch-related cues to a greater degree than adults do. Similarly, English-speaking children were found to vary pitch span to a greater degree than adults in producing the falling accent H\*L (Astruc, Payne, Post, Vanrell, & Prieto, 2013). The more intensive use of duration and pitch in children might be explained by the 'Effort Code', according to which more articulatory effort leads to a larger pitch span (Gussenhoven, 2004) and consequently a longer duration to implement the change in pitch span (Xu & Wang, 2001). The children in our study may be more engaged in the game than adults, and thus take more effort in speech production, especially when telling something new.

As for the distinction of focus types differing in focal constituent size, we have found that four- to five-year-olds can already vary duration in an adult-like way in distinguishing narrow focus from broad focus. However, they also use a higher pitch-max for narrow focus than for broad focus, different from adults. By the age of seven to eight, children no longer distinguish these two types of focus using the pitch-related cues, similar to adults. Unexpectedly, at the age of ten to eleven, children use a wider pitch span for

narrow focus than for broad focus, different from adults. Ten- to eleven-year-olds' use of pitch span in distinguishing these two conditions is similar to adults' use of prosody for the same purpose in read speech (e.g. Xu, 1999), suggesting a more careful manner of speaking in this age group than in the other age groups.

Regarding the distinction of focus types differing in contrastivity, adults do not use any of the four prosodic cues to distinguish narrow focus from contrastive focus. At the age of four to five, children do not use duration to make the distinction but use a lower pitch-min in contrastive focus than in narrow focus, different from adults. At the age of seven to eight and onwards, they do not prosodically distinguish narrow focus from contrastive focus, like adults. A similar observation that young children use prosody to mark contrast while adults do not has been reported in German (Grünloh, Lieven, & Tomasello, 2015). We conjecture that young children may find it exciting to correct an adult with whom they are already familiar, and thus mark contrast with more prosodic prominence.

Furthermore, we have seen evidence that Mandarin-speaking children's use of pitch for focus-marking purposes varies between tones at younger ages. First, as mentioned earlier, to distinguish narrow focus from pre-focus, seven- to eight-year-olds expand the pitch span in Tone 1 and Tone 4 by raising the pitch-max, and expand the pitch span in Tone 2 and Tone 3 by lowering the pitch-min in narrow focus. Given that Tone 1 and Tone 4 start with a high tonal onset, while Tone 2 and Tone 3 start with a relatively low tonal onset (e.g. Xu, 1997), seven- to eight-year-olds might have chosen to expand the pitch span by raising the high tonal onset in Tone 1 and Tone 4, which led to an even higher pitch-max in Tone 1 and Tone 4, and lowering the low tonal onset in Tone 2 and Tone 3, which led to an even lower pitch-min in Tone 2 and Tone 3. This approach is not adult-like but may still be effective in achieving more prosodic prominence in focus. It has been shown that listeners associate a higher pitch of a high tone and a lower pitch of a low tone with more prominence in Dutch (Gussenhoven & Rietveld, 2000). Further, in distinguishing focus from post-focus, ten- to eleven-year-olds' use of pitch span in Tone 1 and their use of pitch span in terms of variation in pitch-max in Tone 3 are not adult-like. The later acquisition of the use of pitch span for Tone 1 and Tone 3 than for Tone 2 and Tone 4 post-focally may be related to tonal properties. Specifically, Tone 2 and Tone 4 are characterised by a sharp rise and a steep fall respectively, whereas Tone 1 is a flat tone and Tone 3 is usually realised as a low tone with a slightly falling pitch contour. Adult speakers vary the pitch span in Tone 1 and Tone 3 to a lesser degree than in Tone 2 and Tone 4 in focus-marking, probably to keep the tonal identities of Tone 1 and Tone 3 intact. Thus, the perceivable changes in pitch span in post-focal regions may also be relatively small in Tone 1 and Tone 3 in the input available to children. Accordingly, children may need more time to grasp these small changes from input in Tone 1 and Tone 3 than in Tone 2 and Tone 4.

Comparing Mandarin-speaking children with children acquiring a West Germanic language, we have found that Mandarin-speaking children begin to use phonetic means at an earlier age and become adult-like in the use of certain phonetic means at an earlier age, in line with our prediction based on the more extensive exposure to phonetic focus-marking in Mandarin. Specifically, Mandarin-speaking children can already vary duration to distinguish narrow focus from non-focus as well as from broad focus in an adult-like way by the age of four to five. In contrast, English-speaking children

do not vary duration to distinguish contrastive (i.e. 'Given-Shift' and 'New') from non-contrastive (i.e. 'Given-NonShift') words when they are both accented with a falling pitch accent at the age of three to four (Wonnacott & Watson, 2008), different from adults (Watson, Arnold, & Tanenhaus, 2005, as cited by Wonnacott & Watson, 2008). Dutch-speaking children do not vary duration to distinguish sentence-initial focus from non-focus when a falling accent is used in both cases at the age of seven to eight, different from adults (A. Chen, 2009). Furthermore, English- and German-speaking children vary pitch-related cues in an adult-like way to mark focus at the age of four to five when focus is contrastive (Müller et al., 2005; Wonnacott & Watson, 2008), while Dutch-speaking children of this age range do not show adult-like use of pitch in distinguishing focus from non-focus when focus is not contrastive (A. Chen, 2009). These findings imply that children's phonetic use of pitch in West Germanic languages may be related to the expression of contrast (A. Chen, 2015). However, Mandarin-speaking children at a similar age can vary pitch-max and pitch-min of certain lexical tones for focus-marking purposes even when the focused words do not carry contrast.

We have also found that the use of pitch for lexical purposes in Mandarin influences the order in which Mandarin-speaking children learn to use pitch and duration in focus-marking. We have predicted that the lexical use of pitch can either speed up the acquisition of pitch for focus-marking purposes (because children need to pay close attention to pitch from an early age in word learning) or slow it down (because there is limited acoustic space left for focus-related manipulation in pitch), compared to the acquisition of duration. Our results provide evidence for the latter prediction. Mandarin-speaking children master the use of duration for focus-marking purposes by the age of four to five, but their use of pitch for the same purposes still develops after the age of eleven. In contrast, children acquiring a West Germanic language use pitch-related cues earlier than duration in phonetic focus-marking.

### **Acknowledgements**

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 Hua Shu from Beijing Normal University and Mei Ou from Beijing Forestry University for their support. We thank Paula Cox for drawing the pictures, Frank Bijlsma, Sjef Pieters and Alex Manus for technical support, and Mattis van den Bergh and Huub 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 in the course of this study.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a VIDI grant awarded to the second author by the Netherlands Organisation for Scientific Research (grant number 276-89-001).

## Notes

1. The unusable sentences were evenly distributed over different focus conditions, even in the case of the four- to five-year olds (i.e. 7% in NF-m; 8% in NF-f; 7% in NF-i; 11% in BF; 8% in CF-m).
2. We previously did analyses for each age group separately, and found that the main pattern did not deviate from what we observe in the current analyses including all age groups. We thus keep the adults in the models as the reference category for the current analyses.

## References

- Astruc, L., Payne, E., Post, B., Vanrell, M. M., & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech, 56*, 229–253.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*, 289–300.
- Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3) [Computer software]. Available from <http://www.praat.org/>
- Bruce, G. (1998). *Allmän och svensk prosodi* [General and Swedish prosody]. Lund: Institutionen för Lingvistik, Lunds Universitet.
- Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In T. Riad, & C. Gussenhoven (Eds.), *Tones and tunes: Typological and comparative studies in word and sentence prosody* (Vol. 1, pp. 113–146). Berlin, Germany: Mouton de Gruyter.
- Chao, Y. R. (1965). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Vigário, S. Frota, & M. J. Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations* (pp. 91–106). Amsterdam, The Netherlands: John Benjamins.
- Chen, A. (2011a). The developmental path to phonological focus-marking in Dutch. In S. Frota, E. Gorka, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 93–109). Dordrecht, The Netherlands: Springer.
- Chen, A. (2011b). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language, 38*, 1055–1083.
- Chen, A. (2015). Children's use of intonation in reference and the role of input. In L. Serratrice, & S. Allen (Eds.), *The Acquisition of reference* (pp. 83–104). Amsterdam, The Netherlands: John Benjamins.
- Chen, Y., & Braun, B. (2006). Prosodic realization in information structure categories in standard Chinese. In R. Hoffmann, & H. Mixdorff (Eds.), *Speech prosody 3*. Dresden, Germany: TUD Press.
- Clumbeck, H. V. (1977). *Studies in the acquisition of Mandarin phonology* (Unpublished doctoral dissertation). University of California, Berkeley.
- Grünloh, T., Lieven, E., & Tomasello, M. (2015). Young children's intonational marking of new, given and contrastive referents. *Language Learning and Development, 11*, 95–127.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.
- Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 83–100). Heidelberg, Germany: Springer.
- Gussenhoven, C., & Rietveld, T. (2000). The behavior of H\* and L\* under variations in pitch range in Dutch rising contours. *Language and Speech, 43*, 183–203.

- Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech, Language, and Hearing Research*, 13, 395–399.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4, 185–199.
- Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge, UK: Cambridge University Press.
- Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, 39, 192–198.
- Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Praha, Czech Republic: Epocha Publishing House.
- 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.
- McDonald, J. (2016). *A spreadsheet to do the Benjamin-Hochberg procedure on up to 1000 P values*. Retrieved from <http://www.biostathandbook.com/multiplecomparisons.html>
- Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2005). Focus-to-stress alignment in 4- to 5-year-old German-learning children. *Proceedings of Generative Approaches to Language Acquisition*, 7, 393–407.
- Shih, C. (1988). Tone and intonation in mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83–109.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73, 751–754.
- Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34, 459–520.
- Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2005, March). *Not just given and new: The effects of discourse and task based constraints on acoustic prominence*. Poster presented at the 2005 CUNY Human Sentence Processing Conference, Tucson, AZ.
- Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31, 749–778.
- Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, 40, 141–151.
- Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, 48, 1065–1079.
- Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition*, 107, 1093–1101.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61–83.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27, 55–105.
- Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319–337.
- Yang, A., & Chen, A. (under revision). Prosodic realisation of focus in semi-spontaneous speech in Mandarin Chinese.
- Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon, UK: Multilingual Matters.