CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Acquisition of weak syllables in tonal languages: acoustic evidence from neutral tone in Mandarin Chinese

Ping TANG[1]*, Ivan YUEN[1], Nan XU RATTANASONE[1], Liqun GAO[2]*, and Katherine DEMUTH[1]

[1]Department of Linguistics, ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney and [2]School of Communication Science, Beijing Language and Culture University, Beijing
*Corresponding authors: Ping Tang, Department of Linguistics, ARC Centre of Excellence in Cognition and its Disorders, 16 University Avenue, Australian Hearing Hub, Balaclava Road, North Ryde, NSW 2109 Australia E-mail: ping.tang1@students.mq.edu.au. Liqun Gao, School of Communication Science, Beijing Language and Culture University, 15 Xueyuan Road, Haidain District, 100083, Beijing, China. E-mail: gaolq@blcu.edu.cn

Abstract

Weak syllables in Germanic and Romance languages have been reported to be challenging for young children, with syllable omission and/or incomplete reduction persisting till age five. In Mandarin Chinese, neutral tone (T0) involves a weak syllable with varied pitch realizations across (preceding) tonal contexts and short duration. The present study examined how and when T0 was acquired by 108 Beijing Mandarin-speaking children (3–5 years) relative to 33 adult controls. Lexicalized (familiar) and non-lexicalized (unfamiliar) T0 words were elicited in different preceding tonal contexts. Unlike previous reports, the present study revealed that children as young as three years have already developed a phonological category for T0, exhibiting contextually conditioned tonal realizations of T0 for both familiar and unfamiliar items. However, mastery of adult-like pitch and duration implementation of T0 is a protracted process not completed until age five. The implications for the acquisition of weak syllables more generally are discussed.

Introduction

During early phonological acquisition, young children demonstrate challenges in acquiring weak syllables (Gerken, 1994). For example, it has been shown that English-learning children at age five still omit weak syllables that appear pretonically, before a stressed syllable, e.g., *banana* produced as 'nana' (e.g., Ingram, 1974; Haelsig & Madison, 1986; Kehoe, Stoel-Gammon, & Buder, 1995; Demuth, 1996). Even when such syllables are produced, they can persist in having longer (unreduced) vowel durations (e.g., Yuen, Demuth, & Johnson, 2011). Similar phenomena have

also been observed in other languages, including Dutch (Fikkert, 1993), French (Demuth & Johnson, 2003), German, and Spanish (Kehoe & Lleó, 2003; Lleó, 2006).

However, most of these studies have focused on Germanic and Romance languages, which use vowels and consonants to contrast word meanings, and involve some type of stress or phrasal lengthening that provides the context for syllable omission/reduction. As most of the languages around the world are tonal, using lexical tone in addition to consonants and vowels to contrast meanings (Yip, 2002), but often having a more limited role for 'stress', this raises the question of how 'weak' syllables are acquired in such languages.

In Mandarin Chinese, for instance, in addition to the four lexical tones which occur on stressed/full syllables, there is also a toneless category, i.e., neutral tone, which occurs only on weak (short) syllables. Thus, neutral tone syllables share the acoustic attribute of being a short syllable, similar to weak syllables in Germanic and Romance languages (Fry, 1955). In addition, as a toneless category, neutral tone exhibits contextually conditioned acoustic realizations after different lexical tones. However, little is known about whether children learning Mandarin show difficulty in acquiring the acoustic aspects of neutral tone. Therefore, the goal of the present study was to explore when neutral tone is acquired by Mandarin-speaking children, crucial for providing a comprehensive account of these children's phonological acquisition above the level of the segment.

Mandarin Chinese has four lexical tones that contrast in pitch contour, i.e., Level tone 1 (T1), Rising tone 2 (T2), Dipping tone 3 (T3), and Falling tone 4 (T4; see Figure 1). Word meaning varies as a function of lexical tone (e.g., /ma1/ 'mother', /ma2/ 'hemp', /ma3/ 'horse', and /ma4/ 'scold'). These four tones appear early in acquisition, i.e., during the single-word stage of development (Li & Thompson, 1977), although confusion between T2 (the rising tone) and T3 (the dipping tone) continues into the 2/3-word stage of development. By the age of three, when children begin to combine words into longer sentences, all lexical tones are reportedly acquired (Li & Thompson, 1977).

The toneless category, neutral tone, is also called the 'fifth tone' or T0. Neutral tone is only carried by weak (short) syllables, appearing in the final position of a word. Neutral tone syllables can be classified into three semantic types: (morphological) suffix, reduplicative, and lexeme types, as exemplified in Table 1 (Li & Thompson, 1977; Hua & Dodd, 2000). Of these various types, some neutral tone syllables belong to part of a disyllabic lexicalized/holistic word, which are not productive, such as the /**tsi0**/ in /tʰu4 **tsi0**/ 'rabbit'. Other neutral tone syllables, such as the possessive particle /**tɤ0**/, can be combined with any noun, to form non-lexicalized/new neutral tone words (e.g., /tʂu1 **tɤ0**/ 'pig's'; Table 1). The noun suffix /tsi0/ also has a full tone counterpart /tsi3/ 'child' bearing a tone 3; this is true for many neutral tone syllables. While these neutral tones were historically derived from their full tone counterparts, most of them are phonologically and semantically distinct from the full tone counterparts today (Shen, 1992).

Unlike lexical tones, a feature of neutral tone is that it is phonologically 'under-specified', i.e., it does not have its own fixed tone. As a toneless category, its pitch implementation varies as a function of the preceding tone. Adults realize neutral tone as a fall in pitch after T1, T2, and T4, but as a rise or a level tone after T3[1] (see Figure 2; Tang, 2014). Neutral tone syllables are also reduced compared to

---

[1]The mechanism that drives this contextual variation is debatable. See Yip (1989) for a tonal spreading (phonological) account and Chen and Xu (2006) for a mid-target (phonetic) account. Yet these accounts agreed on the phonetic observations that neutral tone varies its tonal contour as a function of preceding
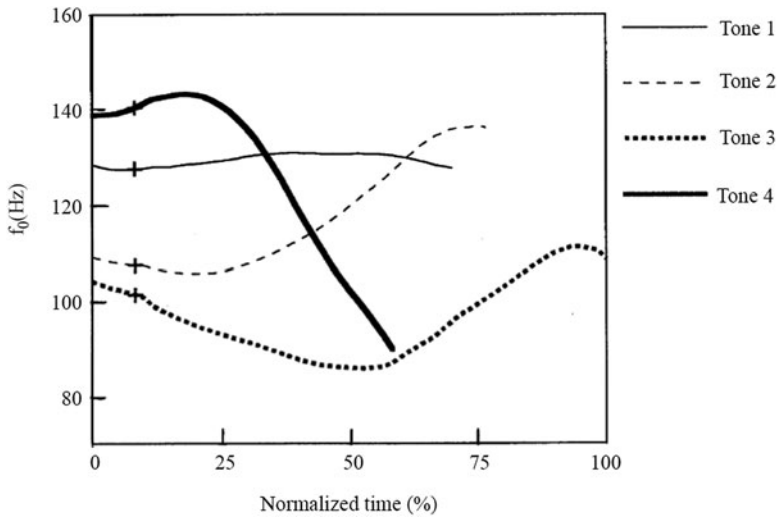
Figure 1. Mandarin Chinese lexical tone pitch contours, sourced from Xu (1997).

Table 1. Neutral Tone Types and Lexical Tone Counterparts

Type	Example	Productivity	Lexical tone counterpart
Suffix	Possessive particle /tɤ0/, e.g., /tʂu1 tɤ0/ 'pig's'	Productive	–
	Classifier particle /kɤ0/, e.g., /san1 kɤ0/ 'three items'	Productive	/kɤ4/
	Noun suffix /tsi0/, e.g., /tʰu4 **tsi0**/ 'rabbit'	Lexicalized	/tsi3/
Reduplicative	/ti4 **ti0**/ 'older brother'	Lexicalized	/ti4/
Lexeme	/tʰou2 **fa0**/ 'hair'	Lexicalized	/fa1/

lexical tone syllables, although the magnitude of shortening again varies as a function of the preceding tone. The overall mean duration of a neutral tone syllable is longer following a T3 syllable (about 70% of the preceding T3 syllable duration) than following T1/2/4 syllables (about 50% to 60% of the preceding T1/2/4 syllables; Cao, 1992; Tang, 2014). In other words, the preceding tonal context influences the tonal contour and duration of neutral tone. Learning to implement the pitch and durational features of this phonologically under-specified tonal category may therefore present a challenge for young children, as they must also learn to correctly modify its realization according to the tonal context.

However, relative to lexical tone studies, neutral tone has received much less attention with respect to when it is acquired. According to the few references on this

tonal contexts. In the present study, the focus was to examine when children can acquire neutral tone, despite its varied realizations.
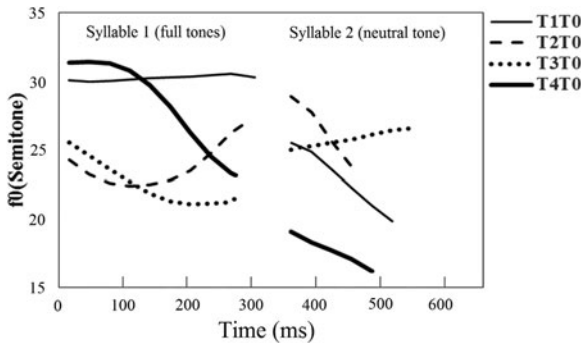
Figure 2. Pitch contour of Mandarin neutral tones (T0) in different tonal contexts (after T1–4), sourced from Tang (2014). Note that the duration of T0 is different when following different tones, i.e., longer (around 70% of the preceding syllable) after T3 than T1/2/4 (about 50% to 60% of the preceding syllable).

topic, neutral tone seems to be acquired much later than lexical tones, not being mastered until around the age of 4;6. For example, Hua and Dodd (2000) reported that no three-year-olds correctly produced all neutral tone words, and only 36% of four-year-olds could do so, suggesting that the acquisition of neutral tone is a protracted process for Mandarin-speaking children. Three types of errors were previously observed in children's neutral tone productions: (1) substituting the neutral tone with the lexical tone counterpart; (2) lengthening the neutral tone syllable; or (3) omitting the neutral tone syllable (for children under age 3) (Hua & Dodd, 2000). These results add to the expectation that neutral tone might be challenging to acquire.

However, in analysing children's neutral tone productions, previous studies were only based on subjective perceptual/auditory judgements, i.e., the authors transcribed children's productions and judged their neutral tone error patterns based on their own perceptual observations. So far there has been no ACOUSTIC investigation of children's neutral tone productions; it is therefore still unclear how and when children develop adult-like acoustic realizations of the neutral tone CATEGORY. That is, do children show contextually conditioned pitch and duration variations for neutral tone, like adults? Can children reduce the neutral tone duration to the same degree as adults? These issues were addressed in the present study by conducting an acoustic analysis of children's neutral tone productions across different tonal contexts, comparing their productions to that of adult controls.

Moreover, previous studies only examined children's neutral tone productions in lexicalized items, i.e., noun suffixes, reduplicatives, etc., where the neutral tone syllable was part of a lexicalized word. However, it might be the case that a child who has successfully produced a lexicalized neutral tone word, such as /ma1 ma0/ 'mother', may know nothing about the neutral tone category, but may be simply repeating the disyllabic word they know from their language input. Thus, previous investigations have mainly tapped into children's word knowledge (i.e., vocabulary) rather than examining their productive knowledge in generalizing the neutral tone category to new words, as in a 'wug' task (e.g., Berko, 1958). In the present study, we therefore wanted to know when children develop a phonological representation for the neutral tone that can be productively generalized to form new words. For

example, the neutral tone possessive particle /tɤ0/ can be productively composed with any noun to form a new possessive word.

To gain a better understanding of children's acquisition of the neutral tone category, the current study therefore examined the pitch and duration realization of three- to five-year-olds' neutral tone productions. This age range was selected since previous studies had reported that neutral tone syllable omission is not a problem for children above age three (Hua, 2002), and neutral tone is not fully acquired at 4;6 (Hua & Dodd, 2000). Two types of neutral tone words were adopted as stimuli, i.e., lexicalized/familiar and non-lexicalized/unfamiliar items. The non-lexicalized items were disyllabic words containing a monosyllabic noun and a monosyllabic possessive particle /tɤ0/, which does not have a lexical tone counterpart, e.g., /tʂu1 tɤ0/ 'pig's'. Exploring children's knowledge of these non-lexicalized items allows us to examine whether they have developed a robust neutral tone category generalizable to learning new words. Lexicalized neutral tone words (e.g., /ti4 **ti0**/ 'younger brother') were then included as known word control items. These words were all kinship terms which are familiar to young children and are reported to emerge early (Hua, 2002). Neutral tone syllables in these words also have a lexical tone counterpart that is identical to the first syllable of the disyllabic (reduplicated) word. The lexicalized neutral tones were therefore used to test whether children would substitute neutral tone with its lexical tone counterpart, as reported by Hua and Dodd (2000) and Hua (2002).

Based on previous studies (Li & Thompson, 1977; Hua & Dodd, 2000; Hua 2002), we assumed that the younger children would not have developed a robust neutral tone category, i.e., the contextually conditioned acoustic realization (pitch: falling after T1/2/4 and rising/level after T3; duration: shorter after T1/2/4 than T3) and reduced duration of neutral tone. Therefore, they would not be able to produce adult-like acoustic realizations of neutral tone for either the lexicalized or non-lexicalized items, at least not in the youngest group, i.e., three-year-olds. We made the following predictions:

Hypothesis 1 (H1): (a) young children (i.e., 3-year-olds) would face challenges in producing contextually conditioned neutral tone PITCH in an adult-like way for both lexicalized (reduplicative) and non-lexicalized (possessive) items; (b) the PITCH of children's neutral tone productions would become more adult-like with age.

Hypothesis 2 (H2): (a) young children (i.e., 3-year-olds) would face challenges in producing contextually conditioned neutral tone DURATION in an adult-like way, producing longer durations than adults for both the lexicalized and non-lexicalized syllables; and (b) the duration of children's neutral tone productions would become more adult-like with age.

Method

Participants

One hundred and eight children aged 3;1 to 6;2 were recruited and tested in the affiliated kindergarten of Beijing Language and Culture University. All children spoke Mandarin as their first language and were born and raised in Beijing. According to reports from the kindergarten, the recruited children did not have any speech, hearing, language, or intellectual difficulty. These children were divided into three age-groups (3-, 4-, and 5-year-olds) within a range of approximately one year (see

Table 2. Number of Participants in Each Age Group

Age group	Male	Female	Total
3 yrs. (3;1–3;12, mean: 3;8, SD: 3 months)	13	31	44
4 yrs. (4;1–4;12, mean: 4;5, SD: 3 months)	20	22	42
5 yrs. (5;1–6;2, mean: 5;7, SD: 5 months)	9	13	22
Adults (19–25 yrs., mean: 20, SD: 2 years)	9	24	33

Table 2). Since most five-year-olds had graduated at the time of testing, there were fewer five-year-olds than the other two age groups. Two young six-year-old children (6;1 and 6;2) were also included in the group of five-year-olds. In addition, 33 adult university students (mean: 20 yrs.; range: 19–25 yrs.) were recruited as controls. The adults were all native monolingual speakers of Mandarin and were born and raised in Beijing as well. The study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Panel.

Stimuli

The investigation of neutral tone was part of a larger study investigating Mandarin-speaking children's acquisition of tones in context. To investigate the acquisition of neutral tone, eight disyllabic words were selected as stimuli. These consisted of four lexicalized kinship reduplicative words and four non-lexicalized possessive words. Within each disyllabic word, the first syllable was T1, T2, T3, or T4 and the second syllable was always T0 (see Table 3).

Note that, in each of the four disyllabic LEXICALIZED reduplicative stimuli, both syllables shared the same segmental information, but differed in tone. Each of the four NON-LEXICALIZED disyllabic possessive stimuli consisted of a monosyllabic animal's name plus the possessive particle /tɤ0/. All disyllabic reduplicative words and the first syllable of the possessive words fell within the top 20% of the most frequent disyllabic and monosyllabic words in the language input to children below three years, according to the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018) from the CHILDES database (MacWhinney, 2000).

Procedure

Children were tested in a quiet room at their kindergarten, and the adults were tested in a quiet room at their university. Only one participant was tested in each session. During test sessions, participants wore a head-mounted cardioid-directional condenser microphone (AKG-C520) which was connected to a solid-state recorder (Marantz PMD661MKII).

The procedure used was an elicited production task. In eliciting the reduplicative kinship terms, four pictures were presented in a sequence to the participant on a computer screen. In each picture, a pair of family members depicted in line drawings were presented side-by-side, such as an older brother (illustrated by a tall boy) and a younger sister (illustrated by a shorter girl). To elicit the word /kɤ1 kɤ0/ 'older brother', for example, the Mandarin-speaking experimenter would point to the picture and ask the participant: "这个男生叫这个女孩妹妹，那么这个女孩叫这个

Table 3. Target Words Were of Two Types (Reduplicative and Possessive). Within Each Disyllabic Word, the Lexical Tone of the First Syllable Was Varied To Be Either T1, T2, T3, or T4 and the Tone of the Second Syllable Was Always the Neutral Tone (T0).

Target word	Word type	Lexicalization	Tones	Meaning	Phonetic transcription
哥哥	Reduplicative	Lexicalized	T1 + T0	Older brother	/kɤ1 **kɤ0**/
爷爷	Reduplicative	Lexicalized	T2 + T0	Grandpa	/je2 **je0**/
姐姐	Reduplicative	Lexicalized	T3 + T0	Older sister	/tɕie3 **tɕie0**/
弟弟	Reduplicative	Lexicalized	T4 + T0	Younger brother	/ti4 **ti0**/
猪的	Possessive	Non-lexicalized	T1 + T0	Pig's	/tʂu1 **tɤ0**/
牛的	Possessive	Non-lexicalized	T2 + T0	Cow's	/niu2 **tɤ0**/
狗的	Possessive	Non-lexicalized	T3 + T0	Dog's	/kou3 **tɤ0**/
鹿的	Possessive	Non-lexicalized	T4 + T0	Deer's	/lu4 **tɤ0**/

男生什么?" 'This boy calls this girl "younger sister", so what does the girl call the boy?' Once the participant produced the target word, e.g., /kɤ1 kɤ0/ 'older brother', the experimenter proceeded to the next picture to elicit the next target word. If the target word was not produced, the experimenter would rephrase and repeat the question. For example, if instead of saying the target word 'older brother' a child said 'boy', the experimenter would rephrase the question and ask them again, i.e., "他叫 她妹妹, 她叫他什么?" 'He calls her "younger sister", so what does she call him?' The target words were never used when the experimenter rephrased the question. All participants succeeded in producing the target word after one or two attempts. The tone of the first syllable varied across items, providing the tonal context for variable acoustic realization of the neutral toned syllable as well.

To elicit the possessive items, two animals depicted in line drawings were simultaneously presented side by side, e.g., a pig (/tʂu1/) and a cow (/niu2/). To elicit the target neutral tone /tɤ0/ in a disyllabic word, such as /tʂu1 **tɤ0**/ 'pig's' and /niu2 **tɤ0**/ 'cow's', the experimenter first introduced the pig and the cow, respectively – "这是一只猪, 这是一头牛" 'This is a pig; this is a cow' – and then pressed a button to play a pre-programmed animation (e.g., tail spinning) on one of the animals, (e.g., a pig). During the animation, the experimenter asked: "这个尾巴是谁 的?" 'Whose tail is it?' After the participant produced the target word (e.g., /tʂu1 tɤ0/ 'pig's'), the experimenter then pressed the button to trigger an animation of a spinning tail on the other animal, for example, the cow, and asked the same question to elicit another target item (e.g., /niu2 tɤ0/ 'cow's'). Note that, again, the first word/ syllable (in this case, the animal) varied in tone, so that the acoustic realization of the neutral toned syllable 'tail' would be different in each newly formed disyllabic word. The same procedure and protocol were followed for the rest of the possessive stimuli. The order of the target words and their corresponding pictures were counterbalanced across participants to avoid any order effect. If participants did not produce the target words, the experimenter would rephrase and repeat the question. For example, instead of saying the target word in isolation "猪的" 'pig's', some

children produced the target word in a clause, e.g., "是猪的" '**is** pig's'. In this case, the experiment would ask them to repeat without the verb "是" 'is'. All participants succeeded in producing the target words after one or two attempts. A total of eight test items were collected per participant, with four reduplicative words and four possessive stimuli.

Coding and measurements

The vowels were acoustically coded for both duration and pitch using Praat (Boersma & Weenink, 2016). To do this, the temporal landmarks (vowel onset and vowel offset) were identified for each syllable of the word based on the onset and offset of the second formant (F2) information in the spectrogram (see Figure 3). One of the reduplicative stimuli /je2 **je0**/ 'grandfather' contained the glide /j/, making it difficult to separate /j/ from /e/. Therefore, for this item, the glide was also included as part of the vowel, with the sonority trough in /j/ between the two vowels used to demarcate the first from the second syllable. This did not affect the overall durational measures because our analysis of neutral tone duration was normalized with reference to the first syllable. Since the neutral toned syllable of all remaining test words began with either a stop or affricate, such as /kɣ1 **kɣ0**/ 'older brother' or /tɕie3 **tɕie0**/ 'older sister', closure duration helped to demarcate the two syllables. This is illustrated in Figure 3. Duration and pitch contour (as measured by f0) were then extracted from each annotated syllable. F0 points were tracked within the annotated interval, using the default autocorrelation algorithm in Praat (Boersma, 1993); these were checked and manually revised to correct for any 'doubling' or 'halving' errors in pitch tracking. The revised pitch track was then interpolated and smoothed with a bandwidth of 20 Hz. Ten percent of the items were re-coded by a second trained native speaker of Mandarin. Inter-rater reliability on the duration of annotated vowels was good (r = 0.92).

Two types of measures were derived from the vowel portion of neutral tone syllables: F0 and normalized duration. The pitch contour of the neutral tone was based on 10 pitch points taken from each vowel. In order to minimize tonal coarticulation and pitch perturbation from neighbouring consonants, the initial and final 5% of the vowel proportion were excluded. In the analysis, pitch values were transformed to semitones from Hz values with 50 Hz as the reference to match human perception, using the following formula:

$$\text{Semitone} = 12^{*}\log_{2}\big(\text{target Hz}/50\big)$$

The neutral tone duration was the temporal difference between the vowel onset and offset of the neutral tone; the lexical tone duration was the temporal difference between the vowel onset and offset of the lexical tone. The normalized vowel duration was a ratio between the neutral tone duration in milliseconds (ms) and the lexical tone duration in ms, using the following formula:

$$\text{Normalized Neutral Tone Duration} = \frac{\text{Raw Neutral Tone Duration(ms)}}{\text{Raw Preceding Lexical Tone Duration(ms)}}$$

Thus, a ratio that was smaller than 1 indicated that the neutral tone duration was reduced relative to the lexical tone.
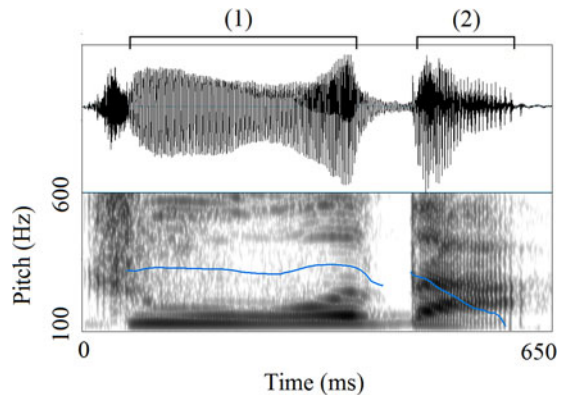
Figure 3. Waveform and spectrogram for the word /tʂu1 tɤ0/ 'pig's'. This token was produced by a four-year-old boy. (1) and (2) illustrate the vowel portion of syllables /tʂu1/ and /tɤ0/, respectively.

Statistical analysis

A total of 1104 tokens were included in the analysis, with 842 tokens from children and 262 from adults. An additional 24 tokens produced by children were excluded from the analysis due to poor acoustic quality, including environmental noise, whispered speech, or productions with prolonged creak.

The data were analysed using R (R Core Team, 2016). To quantify the pitch contour of children's and adults' neutral tone productions, a second-order orthogonal polynomial equation was used to model tone production, using the POLY function of R. The second-order polynomials were adopted since the most complex pitch contour of tones in our data has only a convex or concave contour shape. The poly equation generated three parameters to characterize each pitch contour, i.e., (1) the intercept, (2) the linear trend, and (3) the quadratic trend. With reference to Mirman (2014), the intercept captures the f0 contour's overall HEIGHT (a large intercept indicates a high pitch height and vice versa). The linear trend models the GENERAL DIRECTION OF THE CONTOUR (a positive linear trend indicates a rising pitch and vice versa). The curvature approximates the PITCH CONTOUR by estimating the shape deviated from a linear trend (a positive quadratic trend indicates a concave in the linear approximation of a pitch contour, whereas a negative quadratic trend indicates a convex in an f0 contour, and a small positive quadratic trend indicates a level f0 contour). These parameters were used to evaluate any group differences in the f0 contour.

Linear mixed-effect models were built to compare the three parameters of tone production from children and adults, using the LME4 package (Bates, Mächler, Bolker, & Walker, 2014). All random slopes were included in the model to make it maximally generalizable across the data (Barr, Levy, Scheepers, & Tily, 2013). There are different approaches to estimate the significance of fixed factors in linear mixed-effects models (Luke, 2017). In the present study, the significance of fixed effects was estimated using the ANOVA function in the R package LMERTEST, which provides Satterthwaite's approximation to derive degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2015). This function reported omnibus effects for multilevel factors or interactions using *F*-tests, rather than comparisons with the baseline level using *t*-/*z*-tests. The main/interaction effects reported in the results were averaged across all levels of the other effects, and the effect of multilevel factors

was an omnibus effect (see Peters, Hanssen, & Gussenhoven, 2014, and Tang, Xu Rattanasone, Yuen & Demuth, 2017b, for examples). Relative to other approaches adopted to generate *p* values for fixed effects, such as the likelihood ratio test (LRT) for model comparison, the Satterthwaite's method is a good alternative as it outperforms LRT especially in cases with unbalanced and/or small sample designs (Kuznetsova *et al.*, 2015). When a significant main effect of a multilevel factor or a significant interaction effect was observed, Bonferroni adjusted post-hoc comparisons were performed on the multilevel factor, as well as interactions, using the LSMEANS package (Lenth, 2016).

Results

Pitch

The pitch contours of both the children's and the adults' productions are illustrated in Figure 4. Visual inspection of Figure 4 shows that, for both lexicalized and non-lexicalized neutral tone words, all child groups and adults produced a falling pitch for neutral tones after T1/2/4 and a level/rising pitch after T3, consistent with the pitch variations of neutral tone reported in previous studies (Tang, 2014).

We predicted in H1a that children might face challenges in producing contextually conditioned neutral tone PITCH in an adult-like way for both lexicalized (reduplicative) and non-lexicalized (possessive) items, and in H2b that the pitch of children's neutral tone productions would become more adult-like with age. To test these hypotheses, a linear mixed regression model was constructed using the three pitch parameters, i.e., pitch height (intercept), slope (linear trend), and curvature (quadratic trend), across 10 time steps to explore the group difference of pitch. Three fixed factors: 'Group' (3-, 4-, 5-year-olds and adults), 'Type' (reduplicative and possessive), and 'Tonal Context' (T1, T2, T3, and T4), and a random factor: 'Participant' were included in the model.

As children tend to have higher pitch than adults, the pitch height (intercept) would be predictably higher in children than adults. Our results confirmed this: the main effect of 'Group' was significant on the pitch height ($F(3,141) = 48.59$, $p < .001$; Table 4). Post-hoc analysis revealed that three-year-olds exhibited the highest pitch height, followed by four- and five-year-olds, with adults showing the lowest pitch height ('Appendix 1').

In terms of the SHAPE of the pitch contours (pitch slope or curvature), an interaction of 'Group × Tonal Context' was expected, as we predicted that children would not produce adult-like pitch variation of neutral tone. In addition, developmental differences between the younger children and adults were anticipated, though perhaps not between the oldest children and adults in pairwise comparisons.

The results of the model are presented in Table 4, which shows that there was a significant three-way 'Group × Type × Tonal Context' interaction on the PITCH SLOPE, and a significant two-way 'Group × Type' interaction on the PITCH CURVATURE. Bonferroni adjusted post-hoc tests were then performed on the two interactions to further compare the group difference in terms of neutral tone pitch contours across conditions.

Post-hoc comparisons of the PITCH SLOPE for the three-way interaction revealed three observations. Relative to adults, (a) three-year-olds showed a more falling pitch contour for the reduplicative neutral tone after T4, and for the possessive neutral tone after T1;
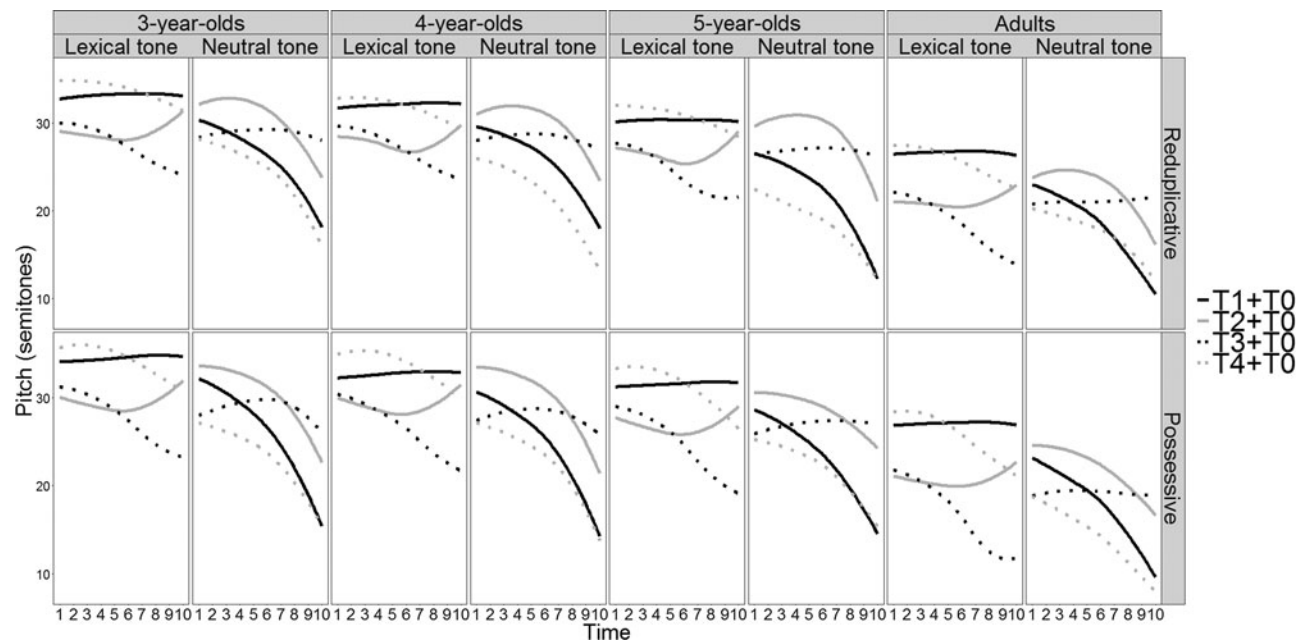
Figure 4. Pitch contours of lexical tone and neutral tone in two word types (reduplicatives on the top row, possessives on the bottom row) across tonal contexts (following T1/2/3/4 syllable) produced by children (3-, 4-, and 5-year-olds, from left to right) and adults. The duration on the *x*-axis is normalized.

Table 4. Results of Linear Mixed Regression Model with Second-order Polynomials on Pitch Points of Neutral Tone across Age Groups (3-, 4-, and 5-year-olds and Adults), Types (Reduplicatives and Possessives) and Tonal Contexts (T1–4). Items in bold indicate significant findings.

Trends	Factors	df 1	df 2	F	p
Intercept	**Group**	**3**	**141**	**48.59**	**<.001*****
	Type	1	141	1.31	0.254
	Tonal Context	**3**	**140**	**157.06**	**<.001*****
	Group × Type	**3**	**141**	**4.19**	**<.01****
	Group × Tonal Context	9	140	1.31	0.236
	Type × Tonal Context	**3**	**10283**	**9.54**	**<.01****
	Group × Type × Tonal Context	**9**	**10279**	**18.76**	**<.001*****
Linear trend	Group	3	141	2.32	0.078
	Type	**1**	**10214**	**73.71**	**<.001*****
	Tonal Context	**3**	**10213**	**1258.09**	**<.001*****
	Group × Type	**3**	**10214**	**8.36**	**<.01****
	Group × Tonal Context	9	10213	1.31	0.236
	Type × Tonal Context	**3**	**10214**	**9.99**	**<.01****
	Group × Type × Tonal Context	**9**	**10214**	**7.43**	**<.001*****
Quadratic trend	**Group**	**3**	**147**	**4.86**	**<.01****
	Type	1	10211	0.54	0.464
	Tonal Context	**3**	**10211**	**53.77**	**<.001*****
	Group × Type	**3**	**10211**	**4.43**	**<.01****
	Group × Tonal Context	9	10210	1.56	0.121
	Type × Tonal Context	**3**	**10211**	**11.7**	**<.001*****
	Group × Type × Tonal Context	9	10211	1.38	0.192

Notes. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Type * Context + (1 + Type + Context + Linear trend + Quadratic trend | Participant : Group).

(b) four-year-olds also showed a more falling pitch contour for the reduplicative neutral tone after T4, and for the possessive neutral tone after T2; (c) five-year-olds did not differ from adults in the neutral tone pitch contour for any condition (also reflected by small effect sizes between 5-year-olds and adults, i.e., Cohen's d < 0.1; see 'Appendix 1'). Consistent with these results, there were then also some pitch differences among the child groups: relative to five-year-olds, three-year-olds showed a more falling pitch contour for the possessive neutral tone after T1, and four-year-olds showed a more falling pitch contour for the possessive neutral tone after T2. However, there were no significant differences between the three- and four-year-olds. These results indicate that all child groups showed contextually conditioned pitch variation of neutral tone syllables across tonal contexts, i.e., falling

after T1/2/4 and rising/level after T3. While three- and four-year-olds exhibited a more falling pitch contour for neutral tones after T1/2/4, this did not distort their contextual pitch variation.

Post-hoc comparisons of the PITCH CURVATURE for the two-way interaction revealed two observations. Relative to adults, three- and four-year-olds produced a more curved pitch contour across all tonal contexts for the possessive neutral tone (see Figure 4). There were no other significant differences among groups (see 'Appendix 2'). These results suggest that the three- and four-year-olds exhibited a more falling pitch contour for neutral tone in the possessive than adults. In contrast, the five-year-olds did not differ from adults in pitch curvature for either the possessive or reduplicative neutral tone, and this is also supported by small effect sizes (Cohen's d: −0.12 for reduplicatives and −0.04 for possessives).

To further explore the developmental trend in the pitch contour of children's neutral tone, a linear regression model was constructed on children's productions with age coded as a continuous factor (in months). Three fixed factors 'Age' (from 38 to 74 months), 'Type' (reduplicative and possessive), and 'Tonal Context' (T1, T2, T3, and T4) and a random factor 'Participant' were included. The dependent variables were pitch slope and curvature since these two parameters reflected the SHAPE of pitch contours. We expected a main effect of 'Age' because we predicted that children's realization of the neutral tone contour would become more adult-like as they mature. The results are presented in Table 5, which shows that there was neither a main effect of 'Age' nor an interaction between 'Age' and other factors, and therefore a post-hoc test was not conducted. Thus, overall, the direction and curvature of the pitch contour for neutral tone did not exhibit developmental differences between the ages of three and five.

In summary, counter to the prediction in H1a, all child groups generally showed adult-like pitch variations across tonal contexts, namely, a falling pitch for neutral tones after T1/2/4 and a rising or level pitch for neutral tones after T3, for both lexicalized and non-lexicalized neutral tone types (see Figure 4). Our results are also partly consistent with H1b, showing that, although there was no general developmental trend in the neutral tone pitch contour as a function of children's age, the younger children (3- and 4-year-olds) showed a more falling pitch component in a few tonal contexts compared to adults. By five years, however, children produced neutral tone pitch contours that were very similar to those of adults across all tonal contexts.

Duration

The normalized durations of children's and adults' neutral tone productions are presented in Figure 5. Visual inspection shows that both children and adults produced longer duration for neutral tones following T3 compared to T1/2/4 across word types. This is consistent with the durational variations of neutral tone reported in previous studies (Cao, 1992; Tang, 2014).

We had predicted in H2a that three-year-olds might face challenges in producing contextually conditioned neutral tone duration in an adult-like way, producing longer duration than adults for both lexicalized and non-lexicalized neutral tones. We also predicted in H2b that the duration of children's neutral tone productions would become more adult-like with age. To test these hypotheses, a linear mixed-effects model was built for the normalized duration of neutral tone syllables with three fixed factors: 'Group' (3-, 4-, 5-year-olds, and adults), 'Type' (the reduplicative and the possessive words), preceding 'Tonal Context' (T1, T2, T3, and T4), and the random

Table 5. Results of Linear Mixed Regression Model with Second-order Polynomials on the Pitch Points in Children's Neutral Tone Productions Only, with Children's Age Coded as Continuous Factor (in Month). Three Fixed Factors Were Included: Children's Age (from 38 to 74 Months), Type (Reduplicatives and Possessives) and Tonal Context (T1–4). Items in Bold Indicate Significant Findings.

Trends	Factors	df 1	df 2	F	p
Linear trend	Age	1	108	3.10	0.081
	Type	1	8211	12.76	**<.001*****
	Tonal Context	3	8210	12.07	**<.001*****
	Age × Type	1	8213	1.17	0.35
	Age × Tonal Context	3	8210	0.94	0.422
	Type × Tonal Context	3	8212	3.64	**<.05***
	Age × Type × Tonal Context	3	8212	2.15	0.092
Quadratic trend	Age	1	112	1.84	0.178
	Type	**1**	**8206**	**5.02**	**<.05***
	Tonal Context	3	8205	0.52	0.672
	Age × Type	1	8206	1.75	0.166
	Age × Tonal Context	3	8205	1.16	0.323
	Type × Tonal Context	3	8205	1.65	0.177
	Age × Type × Tonal Context	3	8205	1.80	0.146

Notes. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Age * Type * Context + (1 + Linear trend + Quadratic trend | Participant).

factor 'Participant'. For H2a, the model anticipated a main effect of 'Group' since we predicted that children would not reduce the neutral tone duration to the same degree as adults; the model also expected an interaction of 'Group × Tonal Context' because children might not produce adult-like durational variation of neutral tone across all tonal contexts. For H2b, the model predicted significant pairwise comparisons between three-year-olds and adults, but perhaps not between five-year-olds and adults.

The results of the comparison are presented in Table 6, which shows as predicted a main effect of 'Group', but also a significant three-way interaction of 'Group × Type × Tonal Context'. A Bonferroni post-hoc test was then conducted on the three-way interaction to compare the durational difference of neutral tone between children and adults across word types and tonal contexts.

Results of the post-hoc test showed that children generally reduced the neutral tone duration to a similar degree as adults across tonal contexts and types, except for the following conditions: relative to adults, (a) three-year-olds produced longer durations for the reduplicative neutral tone after T3 and T4, and for the possessive neutral tone after T2; (b) four-year-olds exhibited longer duration only for the reduplicative neutral tone after T3; (c) five-year-olds did not differ from adults for the two neutral tone types (also reflected by small effect sizes: Cohen's d from –0.2 to 0.14; 'Appendix 3'). Moreover, the results also showed that, among the child groups, there was no difference between either the three- and four-year-olds or between the four-
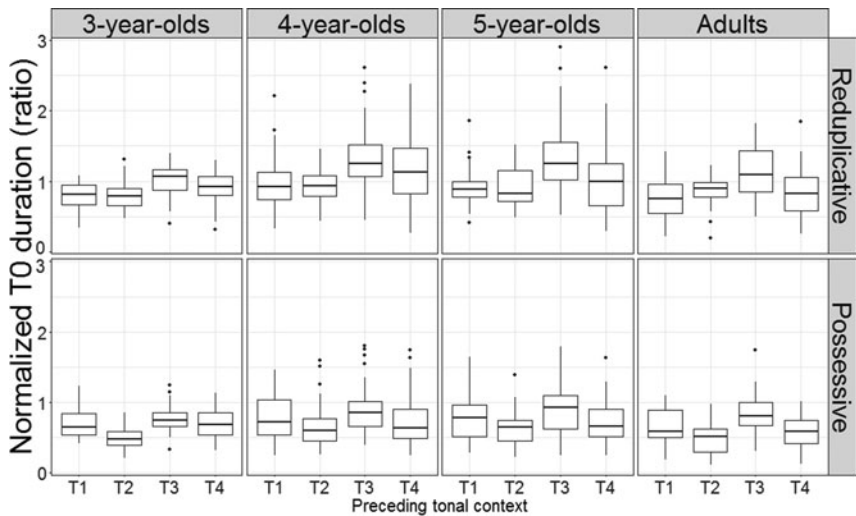
Figure 5. Normalized duration of neutral tone across word types (reduplicatives on the top row, possessives on the bottom row) and tonal contexts (following T1/2/3/4 syllable) produced by children (3-, 4-, and 5-year-olds, from left to right) and adults.

and five-year-olds. Yet relative to the five-year-olds, three-year-olds produced longer durations in the reduplicative words after T4, and in the possessive word after T2.[2]

To further explore the developmental trend in the duration of neutral tone among children, a linear regression model was constructed on children's neutral tone productions. Children's age (in months) was coded as a continuous factor, with the three fixed factors 'Age' (from 38 to 74 months), 'Type' (reduplicative and possessive), and 'Tonal Context' (T1, T2, T3, and T4) and a random factor 'Participant'. H2b would predict a main effect of 'Age'.

The results of the model are presented in Table 7, showing that there was a significant main effect of 'Age' on the duration of neutral tone. No interaction between 'Age' and other factors was found. To further explore the main effect of 'Age', a Pearson correlation test was used to examine the relationship between children's age and the duration of their neutral tone. The results revealed a significant but weak negative correlation between these two parameters ($r(840) = -0.134$, $p < .001$), suggesting that (normalized) neutral tone duration becomes shorter, and more adult-like, as children mature.

In summary, the results from the analysis of neutral tone duration indicate that all children generally showed adult-like durational variation of neutral tone syllables across tonal contexts, with a shorter neutral tone duration after T1/2/4 than T3 (see Figure 5). In addition, children also generally reduced the neutral tone syllable duration to the same degree as adults, though in a few tonal contexts the three- and four-year-olds

---

[2]One of the anonymous reviewers pointed out that the procedure to elicit the possessives might induce the use of contrastive focus, because we presented two objects (i.e., a cow and a pig) and then played an animation (tail rotation) on each animal to elicit the target words 'cow's' and 'pig's', respectively. This might lead to the use of contrastive focus induced by the order effect. However, acoustic analysis revealed that the order in which the items were presented and produced yielded the same durational ratio for the same words. Thus, there is no order effect: $p = .09$, suggesting no contrastive focus.

Table 6. Results of Linear Mixed-effect Model of the Normalized Duration of Neutral Tone across Age Groups (3-, 4-, and 5-year-olds and Adults), Types (Reduplicatives and Possessives) and Tonal Contexts (T1–4). Items in Bold Indicate Significant Findings.

Factors	df 1	df 2	F	P
Group	**3**	**139**	**6.03**	**<.01****
Type	**1**	**142**	**220.88**	**<.001*****
Tonal Context	**3**	**168**	**44.38**	**<.001*****
Group × Type	3	142	2.06	0.109
Group × Tonal Context	9	168	1.05	0.400
Type × Tonal Context	**3**	**549**	**15.22**	**<.001*****
Group × Type × Tonal Context	**9**	**547**	**2.28**	**<.05***

Notes. R code for this model: Normalized duration ~ Group * Type * Context + (1 + Type + Context | Participant : Group).

Table 7. Results of Linear Mixed Regression Model on the Duration in Children's Neutral Tone Productions Only, with Children's Age Coded as Continuous Factor (in Months). Three Fixed Factors Were Included: Children's Age (from 38 to 74 Months), Type (Reduplicatives and Possessives) and Tonal Context (T1–4). Items in Bold Indicate Significant Findings.

Factors	df 1	df 2	F	P
Age	**1**	**107**	**6.58**	**<.05***
Type	**1**	**110**	**5.28**	**<.05***
Tonal Context	3	133	2.02	0.114
Age × Type	1	110	0.00	0.988
Age × Tonal Context	3	133	1.01	0.389
Type × Tonal Context	3	423	2.52	0.057
Age × Type × Tonal Context	3	424	1.30	0.273

Notes. R code for this model: Normalized duration ~ Group * Type * Context + (1 + Type + Context | Participant : Group).

produced longer neutral tone durations than adults, e.g., the reduplicative neutral tone after T3/4 and the possessive neutral tone after T2. This suggests that these young children might be inconsistent in their realization of duration for neutral tone. By five years, however, children exhibited adult-like realizations of neutral tone duration in all tonal contexts. Thus, children's neutral tone productions generally become shorter as they get older.

Discussion

This study investigated the acoustic realization of neutral tone by three- to five-year-old Mandarin-learning children. The results showed that tree-year-olds have already developed the neutral tone category for both lexicalized/familiar and non-lexicalized/ unfamiliar neutral tone items, as reflected by their contextually conditioned tonal realizations. However, adult-like acoustic implementation of neutral tone was more protracted, with differences in pitch and duration between child and adult productions disappearing by age five.

These results thus provide partial support for previous studies (e.g., Li & Thompson, 1977; Hua & Dodd, 2000) that reported that children at 4;6 still did not acquire neutral tone, tending to replace the pitch of neutral tone with a full lexical tone and/or lengthened the duration of neutral tones, showing a limited understanding of the neutral tone category. However, our results showed that three- and four-year-olds have already developed the neutral tone category, albeit with occasional use of more falling pitch contours and longer durations.

A possible reason for the different results between the present and previous studies might be related to the task difference. The present study used a new word- formation task (for possessives) and a picture-naming task (for reduplicatives), whereas previous studies only employed a picture-naming task for known words (Li & Thompson, 1977; Hua & Dodd, 2000). The new word-forming task requires children to generate a new disyllabic word by combining a lexical tone with a neutral tone. This taps into children's productive knowledge of neutral tone as a phonological category. In contrast, the picture-naming task adopted in previous studies only taps children's word knowledge (i.e., vocabulary). Therefore, the word-formation task used in the present study is more challenging than the picture-naming task adopted in previous studies, for both the familiar words (kinship reduplicatives) and the unfamiliar (possessive) words. In addition, in the present study, the picture-naming task required children to use neutral tone in communicative situations, i.e., producing a kinship term to indicate the relationship between two relatives, i.e., grandma vs. grandpa. In previous studies, however, the picture-naming task only required children to name a neutral tone object, i.e., "xing1 xing0" 'star'. Given the more complex tasks (both the word-formation task and the picture-naming task) used in the present study, one might have expected a higher proportion of errors. However, our results suggest that this was not the case. The different results between the present and previous studies must therefore be driven by other factors.

One possibility for the different results might be in the coding methods used. The present study used acoustic analysis to investigate the fine-grained pitch and durational realization of children's neutral tone productions, whereas previous studies used a subjective auditory transcription method, where the accuracy of children's neutral tone productions was determined by a single transcriber, with no reported inter-transcriber reliability for tones (though Hua & Dodd, 2000, and Hua, 2002, report inter-transcriber reliability for consonants and vowels). As children tend to speak more slowly than adults, neutral tone in children's productions will be longer than that in adult productions, and this could have biased the transcriber's judgement in previous studies, leading them to misinterpret neutral tone in slower speech as a full lexical tone. Indeed, our data suggest that the mean raw duration of neutral tone productions is 0.21s for three- and four-year-olds, 0.18s for five-year-olds, and 0.15s for adults. It has also been shown that phonetic expectation can bias perceptual transcription (Oller & Eilers, 1975). It is precisely for this reason that we used ratios in the current study (syllable 2/syllable 1) to compare child and adult productions. This showed that children's syllable ratios (comparing the duration of the first and second (T0) syllable) for disyllables, were adult-like, with only occasional lengthening of the neutral tone syllable by three- and four-year-olds. In a future study it would be interesting to compare perceptual and acoustic analysis/ coding of children's neutral tone production to determine the extent to which the two would yield similar results.

Another possibility that might explain the different results between the current study and previous studies might be related to the stimulus difference. The current study examined children's neutral tone productions in two different word types, i.e., reduplicatives and possessives; whereas Li and Thompson (1977) and Hua and Dodd (2000) also examined neutral tone in noun suffixes and lexemes. It might be the case that these different types of neutral tone pose different challenges for young children. This deserves further investigation in future studies.

However, despite the early emergence of neutral tone representations, our study found that adult-like acoustic implementation of neutral tone was not fully achieved until five years. Relative to adults, the three- and four-year-olds occasionally showed a more falling pitch for neutral tones. These might be related to the acoustic features of children's early language input, where the pitch contour of Mandarin lexical tone and neutral tone is exaggerated, leading to tone hyperarticulation, i.e., more falling tone and rising pitch contours (e.g., Tang *et al.*, 2017a, 2017b). Alternatively, these more falling pitch contours might reflect a tendency for young children to replace neutral tone with a lexical falling tone T4. This could be examined in future studies. Our results also found that three- and four-year-olds did not shorten the duration of their neutral tone productions to the same degree as adults, especially for the lexicalized reduplicatives (children lengthened reduplicatives in more tonal contexts than possessives). Perhaps this difference is due to the fact that the lexicalized words are learnt at a younger age, with a lengthened duration, and children at three or four years have not yet updated the acoustic realization of these early-learnt forms. The more adult-like implementation of neutral tone in non-lexicalized items, in contrast, indicates that children HAVE developed a robust category for neutral tone and can generalize this productively to new words.

Our results therefore reveal a slightly different pattern of weak syllable acquisition from the acquisition of pretonic weak syllables in English. For instance, the predominant error pattern in English-learning children's weak syllable productions is syllable omission (Haelsig & Madison, 1986), and this phenomenon interacts with the stress pattern of words. For example, children were more likely to omit the weak syllable of a weak–strong word like *giraffe* than the weak syllable of a strong–weak word like *tiger* (Gerken, 1994; Demuth, 1996). In Mandarin, however, weak syllables are manifested in a toneless category with short duration and contextually conditioned tonal realization. The stress pattern of neutral tone words is always strong-weak, i.e., full tone + neutral tone, and therefore Mandarin-learning children do not typically omit neutral tone syllables. However, our results suggested that three- and four-year-olds still produced lengthened neutral tone productions. This is similar to findings in English where children sometimes produced weak syllables with longer (unreduced) vowel durations (e.g., Yuen *et al.*, 2011). Taken together, these studies suggest that mastering adult-like acoustic realizations of weak syllables is a protracted process, sometimes deleting and sometimes lengthening, depending on the specific linguistic contexts of the language.

Finally, there are some limitations in the present study. First, as this study was part of a larger study on the acquisition of tones in context, only a few tokens of each neutral tone context were tested. In addition, the number of participants was unbalanced across groups (fewer 5-year-olds), and this might have resulted in insufficient power for the group comparison. Future study of neutral tone could include more items and a more balanced number of participants across age groups to confirm the reliability and generalizability of the current results. It would also be interesting, in future

acoustic studies, to test children younger than three years, to investigate when and how the neutral tone category begins to be acquired.

Conclusion

The goal of this study was to conduct an acoustic analysis of how and when Mandarin-speaking children begin to produce adult-like pitch and durational cues for the short (weak) neutral tone. This is all the more challenging as both cues vary depending on the tone of the preceding syllable, leading to previous claims that neutral tone acquisition is a protracted process. However, our results show that children have extracted the phonological category of neutral tone from its varied surface forms by age three, though they continue to refine their acoustic implementation of neutral tone in terms of pitch contour and duration, becoming more adult-like by the age of five. This result is consistent with findings from other languages, suggesting that the mastery of adult-like weak syllable implementation is protracted. The acoustic analysis used here provides a framework for exploring these issues in a more nuanced manner, providing insight into the development of not only Mandarin tonal contrasts, but other (weak syllable) phonological processes as well.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*(3), 255–78.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: linear mixed-effects models using Eigen and S4. *R package version*, *1*(7), 1–23. Retrieved from <https://CRAN.R-project.org/package=lme4>.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2/3), 150–77.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, *17*(1193), 97–110.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0. 19. Retrieved from <www.praat.org>.

Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics*, *20*(3), 49–50.

Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese-speaking children's play narratives. *Narrative Inquiry*, *8*(1), 77–111.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech–evidence from f_0 patterns of neutral tone in Standard Chinese. *Phonetica*, *63*(1), 47–75.

Demuth, K. (1996). The prosodic structure of early words. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition* (pp. 171–84). Mahwah, NJ: Lawrence Erlbaum Associates.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics / Revue canadienne de linguistique*, *48*(3/4), 211–41.

Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: the Tong corpus. *Journal of Chinese Linguistics*, *46*(1), 69–92.

Fikkert, P. (1993). The acquisition of Dutch stress. In M. Verrips & F. Wijnen (Eds.), *Amsterdam Series in Child Language Development*, Vol. 1 (pp. 21–35). Amsterdam: Universiteit van Amsterdam, Instituut voor Algemene Taalwetenschap.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27(4), 765–68.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, 21(3), 565–84.

Haelsig, P. C., & Madison, C. L. (1986). A study of phonological processes exhibited by 3-, 4-, and 5-year-old children. *Language, Speech, and Hearing Services in Schools*, 17(2), 107–14.

Hua, Z. (2002). *Phonological development in specific contexts: studies of Chinese-speaking children*. Clevedon: Multilingual Matters.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3–42.

Ingram, D. (1974). Phonological rules in young children. *Journal of Child Language*, 1(1), 49–64.

Kehoe, M., & Lleó, C. (2003). A phonological analysis of schwa in German first language acquisition. *Canadian Journal of Linguistics / Revue canadienne de linguistique*, 48(3/4), 289–327.

Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech, Language, and Hearing Research*, 38(2), 338–50.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, 2(0). Retrieved from <https://CRAN.R-project.org/package=lmerTest>.

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, 69(1), 1–33.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185–99.

Lleó, C. (2006). The acquisition of prosodic word structures in Spanish by monolingual and Spanish–German bilingual children. *Language and Speech*, 49(2), 205–29.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–502.

MacWhinney, B. (2000). *The CHILDES Project: tools for analyzing talk*, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman & Hall/CRC.

Oller, K., & Eilers, R. E. (1975). Phonetic expectation and transcription validity. *Phonetica*, 31(3/4), 288–304.

Peters, J., Hanssen, J., & Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics*, 46, 185–209.

R Core Team (2016). R: A language and environment for statistical computing [Computer program]. Version 3.3.1. Retrieved from < https://www.R-project.org/> (last accessed 21 June 2016).

Shen, X. S. (1992). Mandarin neutral tone revisited. *Acta Linguistica Hafniensia*, 24(1), 131–52.

Tang, P. (2014). A study of prosodic errors of Chinese neutral tone by advanced Japanese students (Chinese version). *TCSOL Studies*, 56(4), 39–47.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017a). Phonetic enhancement of Mandarin vowels and tones: infant-directed speech and Lombard speech. *Journal of the Acoustical Society of America*, 142(2), 493–503.

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017b). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *Journal of the Acoustical Society of America*, 142(5), 2823–35.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61–83.

Yip, M. (1989). Contour tones. *Phonology*, 6(1), 149–74.

Yip, M. (2002). *Tone*. Cambridge University Press.

Yuen, I., Demuth, K., & Johnson, M. (2011). Prosodic structure in child speech planning and production. In *Proceedings from the 17th International Congress of Phonetic Sciences* (pp. 2248–51).

Appendix 1 Bonferroni adjusted post-hoc test of the LINEAR TREND (SLOPE) of the f0 contour of neutral tone syllables between children and adults across word types (reduplicatives and possessives) and tonal contexts (neutral tone syllables after T1, T2, T3, and T4 syllables). Items in bold indicate significant findings

Group difference	Type	Tonal context	β	SE	df	t	p	Cohen's d
3 yrs. to adults	Reduplicative	T1	0.89	1.06	651	0.83	1.000	0.05
		T2	−0.88	1.07	659	−0.82	1.000	−0.05
		T3	−0.75	1.06	651	−0.70	1.000	−0.04
		T4	**−3.71**	**1.07**	**659**	**−3.48**	**<.01****	**−0.19**
	Possessive	**T1**	**−2.98**	**1.07**	**659**	**−2.79**	**<.05***	**−0.15**
		T2	−2.50	1.07	659	−2.34	0.117	−0.13
		T3	−1.03	1.07	659	−0.96	1.000	−0.05
		T4	−0.45	1.07	659	−0.42	1.000	−0.02
4 yrs. to adults	Reduplicative	T1	1.33	1.08	651	1.24	1.000	0.07
		T2	0.18	1.08	651	0.16	1.000	0.01
		T3	−1.55	1.08	651	−1.44	0.902	−0.08
		T4	**−4.65**	**1.08**	**669**	**−4.29**	**<.001*****	**−0.23**
	Possessive	T1	−2.49	1.08	659	−2.31	0.128	−0.13
		T2	**−3.04**	**1.09**	**688**	**−2.78**	**<.05***	**−0.15**
		T3	−1.19	1.08	659	−1.10	1.000	−0.06
		T4	1.33	1.08	651	1.24	0.990	−0.08
5 yrs. to adults	Reduplicative	T1	−1.10	1.27	651	−0.86	1.000	−0.05
		T2	−0.31	1.27	651	−0.24	1.000	−0.01
		T3	−0.65	1.27	651	−0.51	1.000	−0.03
		T4	−2.04	1.30	700	−1.57	0.702	−0.08

	Possessive	T1	−0.65	1.27	651	−0.51	1.000	−0.03
		T2	1.69	1.28	674	1.31	1.000	0.07
		T3	1.29	1.27	651	1.01	1.000	0.06
		T4	1.19	1.31	729	0.91	1.000	0.05
3 yrs. to 4 yrs.	Reduplicative	T1	−0.44	1.00	651	−0.44	1.000	−0.02
		T2	−1.06	1.00	660	−1.06	1.000	−0.06
		T3	0.80	1.00	651	0.80	1.000	0.04
		T4	0.94	1.01	681	0.93	1.000	0.05
	Possessive	T1	−0.49	1.01	670	−0.49	1.000	−0.03
		T2	0.54	1.02	704	0.53	1.000	0.03
		T3	0.16	1.01	670	0.16	1.000	0.01
		T4	1.06	1.01	681	1.05	1.000	0.06
3 yrs. to 5 yrs.	Reduplicative	T1	1.99	1.21	651	1.64	0.604	0.09
		T2	−0.57	1.21	657	−0.47	1.000	−0.03
		T3	−0.10	1.21	651	−0.08	1.000	0.00
		T4	−1.67	1.24	712	−1.35	1.000	−0.07
	Possessive	**T1**	**−4.19**	**1.22**	**683**	**−3.43**	**0.01**	**−0.19**
		T2	−2.33	1.21	657	−1.92	0.330	−0.11

(*Continued*)

Appendix 1 (*Continued.*)

Group difference	Type	Tonal context	β	SE	df	t	p	Cohen's d
		T3	−2.32	1.21	657	−1.91	0.337	−0.11
		T4	−1.64	1.25	744	−1.31	1.000	−0.07
4 yrs. to 5 yrs.	Reduplicative	T1	2.43	1.22	651	2.00	0.279	0.11
		T2	0.49	1.22	651	0.40	1.000	0.02
		T3	−0.90	1.22	651	−0.74	1.000	−0.04
		T4	−2.61	1.25	719	−2.09	0.223	−0.11
	Possessive	T1	−1.84	1.22	657	−1.51	0.794	−0.08
		T2	**−4.73**	**1.24**	**706**	**−3.80**	**0.001*****	**−0.20**
		T3	−2.48	1.22	657	−2.03	0.257	−0.11
		T4	−2.69	1.27	751	−2.13	0.201	−0.11

Appendix 2 Bonferroni adjusted post-hoc test of the QUADRATIC TREND (CURVATURE) of the pitch contour of neutral tone between groups across word types (the reduplicative and the possessive). Items in bold indicate significant findings

Group difference	Type	β	SE	df	t	p	Cohen's d
3 yrs. to adults	Reduplicative	−0.76	0.54	329	−1.40	0.983	−0.11
	Possessive	**−1.88**	**0.54**	**331**	**−3.46**	**<.01****	**−0.27**
4yrs. to adults	Reduplicative	−0.89	0.55	330	−1.62	0.638	−0.13
	Possessive	**−1.94**	**0.55**	**336**	**−3.52**	**<.01****	**−0.27**
5yrs. to adults	Reduplicative	−0.98	0.65	333	−1.51	0.799	−0.12
	Possessive	−0.36	0.65	339	−0.55	1.000	−0.04
3yrs. to 4yrs.	Reduplicative	0.13	0.51	332	0.26	1.000	0.02
	Possessive	0.06	0.51	342	0.12	1.000	0.01
3yrs. to 5yrs.	Reduplicative	0.22	0.62	335	0.36	1.000	0.03
	Possessive	−1.52	0.62	344	−2.44	0.091	−0.19
4yrs. to 5yrs.	Reduplicative	0.09	0.62	335	0.15	1.000	0.01
	Possessive	−1.58	0.63	347	−2.51	0.075	−0.19

Appendix 3 Bonferroni adjusted post-hoc test of the DURATION of neutral tone syllables between children and adults across word types (the reduplicative and the possessive) and tonal contexts (neutral tone after T1, T2, T3, and T4). Items in bold indicate significant findings

Group difference	Type	Tonal context	β	SE	df	t	p	Cohen's d
3 yrs. to adults	Reduplicative	T1	0.18	0.07	200	2.51	0.078	0.25
		T2	0.13	0.06	223	2.06	0.243	0.20
		T3	**0.31**	**0.09**	**176**	**3.54**	**<.01****	**0.38**
		T4	**0.28**	**0.09**	**174**	**3.18**	**<.05***	**0.34**
	Possessive	T1	0.10	0.07	203	1.52	0.781	0.15
		T2	**0.20**	**0.06**	**251**	**3.39**	**<.01****	**0.30**
		T3	0.12	0.08	186	1.56	0.723	0.16
		T4	0.02	0.07	183	0.21	1.000	0.02
4 yrs. to adults	Reduplicative	T1	0.11	0.07	200	1.52	0.782	0.15
		T2	0.13	0.06	222	1.96	0.304	0.19
		T3	**0.33**	**0.09**	**176**	**3.69**	**<.01****	**0.39**
		T4	0.10	0.09	176	1.09	1.000	0.12
	Possessive	T1	0.09	0.07	202	1.29	1.000	0.13
		T2	0.15	0.06	257	2.51	0.076	0.22
		T3	0.13	0.08	186	1.59	0.678	0.17
		T4	0.04	0.08	185	0.54	1.000	0.06
5 yrs. to adults	Reduplicative	T1	0.00	0.09	200	0.04	1.000	0.00
		T2	0.06	0.08	222	0.74	1.000	0.07
		T3	0.14	0.10	176	1.31	1.000	0.14
		T4	0.07	0.11	180	0.66	1.000	−0.07

	Possessive	T1	−0.05	0.08	201	−0.63	1.000	−0.06
		T2	0.00	0.07	254	0.01	1.000	0.00
		T3	0.09	0.09	184	0.93	1.000	0.10
		T4	−0.18	0.09	194	−1.93	0.330	−0.20
3 yrs. to 4 yrs.	Reduplicative	T1	0.07	0.07	200	1.04	1.000	0.10
		T2	0.01	0.06	224	0.09	1.000	0.01
		T3	−0.02	0.08	176	−0.21	1.000	−0.02
		T4	0.18	0.08	178	2.20	0.176	0.23
	Possessive	T1	0.01	0.06	204	0.23	1.000	0.02
		T2	0.05	0.06	260	0.86	1.000	0.08
		T3	0.00	0.07	188	−0.05	1.000	−0.01
		T4	−0.03	0.07	187	−0.36	1.000	−0.04
3 yrs. to 5 yrs.	Reduplicative	T1	0.18	0.08	200	2.25	0.155	0.22
		T2	0.07	0.07	223	1.04	1.000	0.10
		T3	0.17	0.10	176	1.74	0.506	0.19
		T4	**0.35**	**0.10**	**182**	**3.45**	**<.01****	**0.36**
	Possessive	T1	0.15	0.08	202	2.01	0.277	0.20
		T2	**0.20**	**0.07**	**256**	**2.97**	**<.05***	**0.26**

(*Continued*)

Appendix 3 (*Continued.*)

Group difference	Type	Tonal context	β	SE	df	t	p	Cohen's d
		T3	0.04	0.09	185	0.40	1.000	0.04
		T4	0.19	0.09	197	2.20	0.172	0.22
4 yrs. to 5 yrs.	Reduplicative	T1	0.11	0.08	200	1.38	1.000	0.14
		T2	0.07	0.07	222	0.96	1.000	0.09
		T3	0.19	0.10	176	1.89	0.362	0.20
		T4	0.17	0.10	183	1.63	0.624	0.17
	Possessive	T1	0.14	0.08	202	1.80	0.438	0.18
		T2	0.15	0.07	261	2.21	0.167	0.19
		T3	0.04	0.09	185	0.44	1.000	0.05
		T4	0.22	0.09	198	2.47	0.086	0.25