


ARTICLE

The acquisition of phonological alternations: The case of the Mandarin tone sandhi process

Ping Tang^{1*} , Ivan Yuen², Nan Xu Rattanasone², Liqun Gao^{3*} and Katherine Demuth⁴

¹Nanjing University of Science and Technology and Macquarie University, ²Macquarie University, ³Beijing Language and Culture University and ⁴Macquarie University

*Corresponding author. Email: ping.tang@njust.edu.cn; gaolq@blcu.edu.cn

(Received 07 May 2018; revised 19 July 2019; accepted 29 July 2019)

Abstract

Phonological processes can pose a learning challenge for children, where the surface form for an underlying contrast may vary as a function of the phonological environment. Mandarin tone sandhi is a complex phonological process that requires knowledge about both the tonal and the prosodic context in which it applies. The present study explored the *productive knowledge* of tone sandhi processes by 108 3- to 5-year-old Mandarin-speaking children and 33 adults. Participants were asked to produce novel tone sandhi compounds in different tonal contexts and prosodic structures. Acoustic analysis showed that 3-year-olds have abstracted the tone sandhi process and can productively apply it to novel *disyllabic* words across tonal contexts. However, even 5-year-olds still differed from adults in applying tone sandhi in response to the *trisyllabic* prosodic structure. The results are discussed in terms of the factors that influence how tone sandhi processes, and phonological alternations more generally, are acquired.

Keywords: Mandarin Chinese; phonological acquisition; phonological alternation; phonological representations; tone sandhi

During the process of acquiring language, children need to develop an understanding of how words are phonologically represented, and then store and retrieve these words from their mental lexicon during both perception and production. This is complicated by the presence of phonological processes that occur when words or morphemes come together, which can lead to phonological alternations. In English, for instance, certain consonants can change place of articulation as a function of the following context, for example, *ten* can be realized as [tem] in the phase of *ten pounds*, as a result of *assimilation*. In this case, children need to understand that [tem] is a variant of the word *ten*, and correctly use this variant in appropriate contexts (i.e., before another word with an initial bilabial stop).

While there is a large body of literature investigating how children learn various types of phonological processes, these investigations have tended to focus on processes involving vowels and consonants, mostly in Germanic and Romance

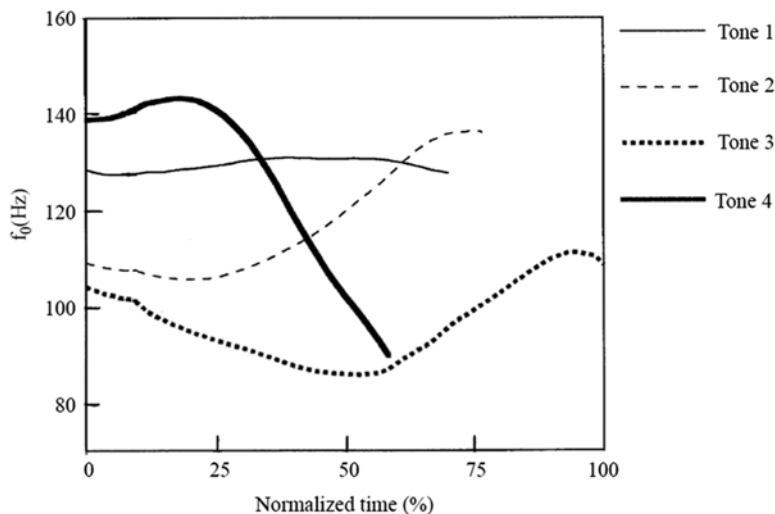


Figure 1. Mandarin Chinese lexical tone pitch contours, from Xu (1997, p. 67).

languages (e.g., Albright & Hayes, 2011; Kazazis, 1969; Kerkhoff, 2007; Skoruppa, Mani, & Peperkamp, 2013; van de Vijver & Baer-Henney, 2013; Zamuner, Kerkhoff, & Fikkert, 2006). However, 60%–70% of human languages also involve tonal contrasts (Yip, 2002). This raises questions about how children learning tonal languages acquire phonological alternations that involve tone changes, or tonal processes.

One of the most well-studied tonal processes is the tone sandhi phenomenon, leading to tonal changes in specific tonal and prosodic contexts. Mandarin Chinese has four lexical tones as well as various tone sandhi processes. Among those most well studied is the tone 3 sandhi process, typically modifying the realization of the third tone (T3) when followed by another T3 within the same prosodic domain (Yip, 2002, p. 180). Thus, tone sandhi is a complex phonological process conditioned by both the tonal context and the prosodic structure in which it appears (Shih, 1997). The aim of the present study was thus to explore when and how children develop *productive* knowledge of the tone sandhi process, with correct implementation across tonal contexts and prosodic structures. The findings could have implications not only for the learning of tonal alternations but also for learning phonological alternations more generally.

Lexical tones

Mandarin Chinese has four lexical tones contrasting in pitch contour, that is, Level tone 1 (T1), rising tone 2 (T2), dipping tone 3 (T3), and falling tone 4 (T4; see Figure 1). Word meaning differs when carrying different tones (e.g., /ma1/ *mother*, /ma2/ *hemp*, /ma3/ *horse*, and /ma4/ *scold*). These four tones, with associated word meanings, appear early in acquisition, during the single-word stage of development, although confusion between T2 (the rising tone) and T3 (the dipping tone) continues into the 2/3-word stage of development (Li & Thompson, 1977). By the age of 3,

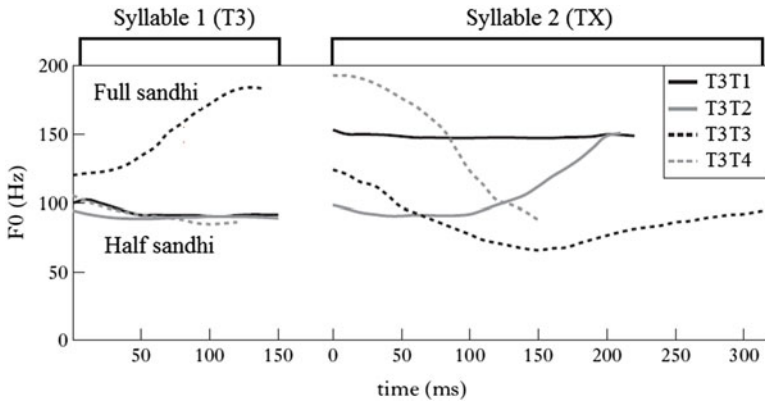


Figure 2. Mandarin Chinese tone sandhi pitch contours in disyllabic words (adapted from Zhang & Lai, 2010, p. 163).

when children begin to combine words into longer sentences, all lexical tones have generally been acquired (Li & Thompson, 1977).

Tone sandhi application as a function of tonal context in disyllabic words

Among the four lexical tones, T3 is unique in that it has highly context-conditioned phonetic realizations. T3 can only be realized in its citation/canonical form (a dipping tone) in isolation or in word-final position. In the initial position of a disyllabic word, however, its realization is governed by the tone sandhi process, which alters the surface tone of T3 as a function of the following tonal context (Chen, 2000, p. 20). For example, in a T3T3 disyllabic word, the first T3 changes to T2, with a rising pitch; this phonological process is referred to as the *full sandhi* process. In contrast, T3 is realized as a low-falling tone before T1/2/4, and this phonological process is referred to as the *half sandhi* process. Both are exemplified in (1) and Figure 2. The tone sandhi process therefore results in multiple allophonic variants of T3, including a dipping tone, a rising tone, and a low-falling tone. Full mastery of tone sandhi processes thus requires children to correctly use these variants in appropriate tonal contexts.

(1) *Mandarin tone sandhi processes*

Full sandhi process: T3 → T2 (rising tone) / _ T3

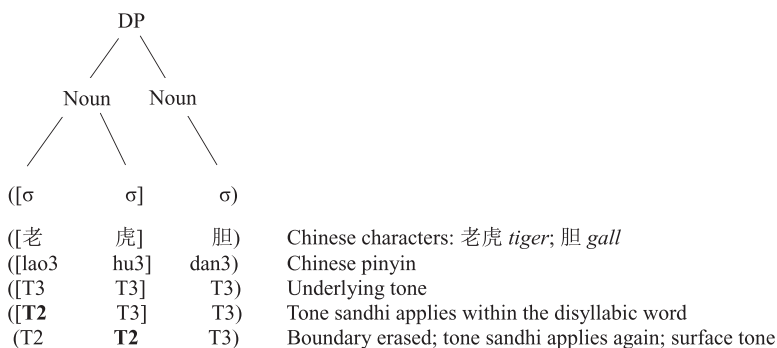
Half sandhi process: T3 → low-falling tone / _ T1/2/4

Tone sandhi application as a function of prosodic structure in trisyllabic words

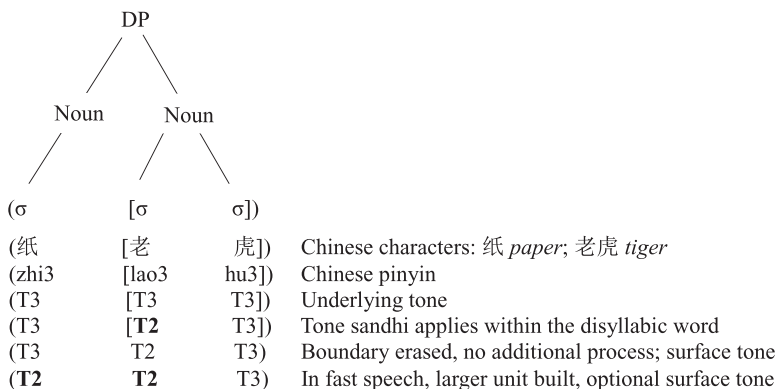
The application of tone sandhi is more complex for trisyllabic words. In a word containing T3T3T3 as underlying tones, the full tone sandhi process may apply recursively, depending on the internal prosodic structure (Shih, 1997).

Consider the two trisyllabic prosodic words: ([lao3 hu3] dan3) *tiger's gall* and (zhi3 [lao3 hu3]) *paper tiger*. The former involves a left-branching structure ([T3T3] T3) and the latter involves a right-branching structure (T3 [T3T3]), where “left and right” refer to the position of the disyllabic prosodic unit (foot) where tone sandhi first applies within the larger trisyllabic prosodic word. The tone sandhi process thus applies first within the disyllabic unit [$\sigma\sigma$], changing its tones from T3T3 to T2T3. With the innermost bracket erased, the tone sandhi process then applies again, within the higher prosodic domain of the entire trisyllabic prosodic word. If there are still two adjacent T3T3 syllables (as in the left-branching case in [2]), the tone sandhi process applies again. As a result of this recursive tone sandhi process, the left-branching structure will generate **T2T2T3** as the surface output form (where the bold font indicates the result of tone sandhi application). In contrast, the right-branching structure in (3) surfaces with T3**T2T3** output (Shih, 1997).

(2) A left-branching trisyllabic prosodic word ([T3T3] T3).



(3) A right-branching trisyllabic prosodic word (T3 [T3T3]).



However, in the case of the structure in (3), it has also been suggested that, especially in fast speech, the tone sandhi process applies recursively from left to right within the entire trisyllabic word, regardless of the internal prosodic structure.

This then results in identical **T2T2T3** surface forms for both the left- and right-branching prosodic words in (2) and (3), respectively (Chen, 2000, p. 379). Thus, for the right-branching items such as (zhi3 [lao3 hu3]) *paper-tiger* there may be some variable surface realizations. However, this argument was only based on the author's informal observation (Chen, 2000): there has yet to be any acoustic investigation of this phenomenon to verify a possible speaking rate explanation for such variable surface realizations.

Previous studies of tone sandhi acquisition

Studies have often reported the acquisition of phonological processes/alternations being challenging (e.g., Albright & Hayes, 2011; Kazazis, 1969; Kerkhoff, 2007; van de Vijver & Baer-Henney, 2013; Zamuner, Kerkhoff, & Fikkert, 2006), where learners need to compare morphologically related forms, choosing a basic or an underlying form/representation, and then abstracting a grammatical structure that can generate the various surface forms (Albright & Hayes, 2011). This could also be the case for Mandarin-speaking children when learning the tone sandhi process, as the various acoustic realizations across tonal contexts (i.e., full and half sandhi contexts) may complicate learning how and where it applies.

Although Mandarin tone sandhi processes have been well studied in adults, they have received less attention with respect to when and how they are acquired by children. A few studies have examined children's production of lexicalized disyllabic tone sandhi words such as "shou3 zhi3" *finger*, revealing the early emergence of appropriate tone realization on these words before age 3 (e.g., Hua & Dodd, 2000; Li & Thompson, 1977; Xu Rattanasone, Tang, Yuen, Gao, & Demuth, 2018). Evidence from a perception study also suggests that children at age 3 are sensitive to tone sandhi mispronunciations on lexicalized tone sandhi words (Wewalaarachchi & Singh, 2016). However, "wug" tests with *novel* words (Berko, 1958) investigating the productive mastery of phonological processes often revealed errors even when children can correctly produce alternations on *real* words (van de Vijver & Baer-Henney, 2013; Zamuner et al., 2006). This raises the question of whether 3-year-olds can generalize the tone sandhi process, productively applying it to novel items they have never heard before.

Note that application of tone sandhi in the **T3T3T3** trisyllabic prosodic words requires children to build either a left-branching or a right-branching prosodic word structure. Given that right-branching **T3T3T3** prosodic words can be realized with two possible surface forms (**T3T2T3** ~ **T2T2T3**; Chen, 2000, p. 379), the *variational learning model* would predict that variability in the input may cause children to assume a different grammar from that of the adult, leading to later acquisition (Yang, 2002). Therefore, we might expect that adultlike tonal representations for the right-branching structure will take longer to acquire compared to the left-branching forms, where no such variability is found.

To our knowledge, there is only one study eliciting 3- to 5-year-olds' trisyllabic tone sandhi productions. In her thesis, Wang (2011; Chap. 6) examined children and adult controls' productions of **T3T3T3 noun + noun** compounds with two structures, that is, ([T3T3] T3) vs. (T3 [T3T3]), such as ([shui3 guo3] niao3)

fruit-bird and (shui3 [lao3 hu3]) *water-tiger*. Based on the author's perceptual transcriptions, it was observed that children at 3 years have adultlike tone sandhi productions for both structures, producing T2T2T3 for ([T3T3] T3) prosodic word structures, and T3T2T3 for (T3 [T3T3]) prosodic word structures. However, the prosodic structure of the stimuli used in her study was confounded with lexical familiarity, where the child could easily parse a T3T3T3 as a familiar monosyllabic word plus familiar disyllabic word, leading to high accuracy in tone sandhi production. Thus, the interaction between children's tone sandhi application and prosodic structure has yet to be fully tested.

In summary, it is unclear when children have abstracted the tone sandhi process to be able to productively apply it to novel compounds in response to different tonal contexts and prosodic structures. This issue was addressed in the present study by conducting a series of "wug" tests. We first elicited novel disyllabic tone sandhi compound words across tonal contexts, testing children's productive knowledge of the tone sandhi process in these prosodically simple contexts. We then elicited novel trisyllabic T3T3T3 tone sandhi compound words in two different prosodic structures (left-branching vs. right-branching prosodic words), testing children's productive knowledge of tone sandhi application as a function of prosodic structure. In this case, all trisyllabic compounds were composed of three monosyllabic words to eliminate any potential bias of lexicalized/known disyllabic word structures.

We addressed two questions. Can 3- to 5-year-olds productively apply the tone sandhi process to novel disyllabic compounds? If so, can they also extend this to novel trisyllabic compounds, showing sensitivity to prosodic context? Two main hypotheses were formulated.

Hypothesis 1 (H1): Disyllabic words

As phonological processes can be challenging to acquire, we predicted that (a) the younger children (3-year-olds) might not *productively* use the tone sandhi process with *novel* disyllabic compounds, and (b) children's ability to apply the tone sandhi process to novel disyllabic compounds would become acoustically more adultlike with age.

Hypothesis 2 (H2): Trisyllabic words

As the trisyllabic T3T3T3 word involves *recursive* sandhi application in the left-branching structure, but *single* or variably *recursive* sandhi application in the right-branching structure, it might be more challenging to learn the application of tone sandhi processes on the right-branching items. We therefore predicted that (a) the younger children (3-year-olds) might not correctly apply the tone sandhi process to novel trisyllabic T3T3T3 compounds, especially for the more variably realized right-branching prosodic words (i.e., T2T2T3 or T3T2T3), and (b) children's ability to apply the tone sandhi process to novel trisyllabic compounds would become acoustically more adultlike with age.

Table 1. Number of participants in each age group

Age group	Male	Female	Total
3 years (3;1-3;12, $M = 3;8$, $SD = 3$ months)	13	31	44
4 years (4;1-4;12, $M = 4;5$, $SD = 3$ months)	20	22	42
5 years (5;1-6;2, $M = 5;7$, $SD = 5$ months)	9	13	22
Adults (19-25 years, $M = 20$, $SD = 2$ years)	9	24	33

Method

Participants

One hundred and eight children aged 3, 4, and 5 years were recruited and tested in the affiliated kindergarten of Beijing Language and Culture University (see Table 1). All 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. There were fewer 5-year-olds than the other two age groups because many had already graduated from the kindergarten at the time of testing. In addition, 33 adult university students 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

A total of 32 words were selected as stimuli, including 8 monosyllabic TX (T1-T4) words, 16 disyllabic T3TX novel compounds, and 8 novel trisyllabic T3T3TX compounds. All were picturable and associated with a line drawing. The 8 monosyllabic words and 16 disyllabic compounds were used to address H1 and test children's productive use of the full and half tone sandhi process in various tonal contexts. The 8 trisyllabic T3T3TX compounds, including 6 T3T3T1/2/4 controls and 2 T3T3T3 targets with left-branching and right-branching prosodic structures, were used to address H2 and investigate children's ability to apply the tone sandhi process to novel compounds with different prosodic structures.

Eight monosyllabic stimuli were used to compose the disyllabic compounds. The 4 first syllable words were all T3 animal names, and the second syllable words all referred to inanimate objects, with tones varying from T1 to T4 (see Table 2). All words fell within the top 50% of the most frequent monosyllabic words in language input to children below 3 years, according to the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018) from the CHILDES database (MacWhinney, 2000).

Sixteen T3TX disyllabic compounds were then generated using these known monosyllabic words, resulting in 4 disyllabic full sandhi (T3T3) compounds and 12 disyllabic half sandhi (T3T1/2/4) compounds (see Table 3).

Eight trisyllabic compounds were also constructed, 4 with left-branching structure ([T3T3] TX) and 4 with right-branching structure (T3 [T3TX]; see Table 4). In the left-branching structures, the same disyllabic item (/tʰsɿ maʃ/ *purple-horse*) was combined with each of the monosyllabic words (TX) from the

Table 2. *Monosyllabic stimuli list*

Group	Target word	Tones	Meaning	Phonetic transcription
a	马	T3	<i>horse</i>	/ma3/
	鸟	T3	<i>bird</i>	/niau3/
	狗	T3	<i>dog</i>	/kou3/
	鼠	T3	<i>mouse</i>	/ʃu3/
b	书	T1	<i>book</i>	/ʃu1/
	球	T2	<i>ball</i>	/tʃiu2/
	鼓	T3	<i>drum</i>	/ku3/
	画	T4	<i>picture</i>	/hua4/

Table 3. *Disyllabic stimuli list*

Context	Compounds	Underlying tones	Surface/target tones	Meaning	Phonemic transcription
Full sandhi	马鼓	T3T3	T2 T3	<i>horse-drum</i>	/ma3 ku3/
	鸟鼓	T3T3	T2 T3	<i>bird-drum</i>	/niau3 ku3/
	狗鼓	T3T3	T2 T3	<i>dog-drum</i>	/kou3 ku3/
	鼠鼓	T3T3	T2 T3	<i>mouse-drum</i>	/ʃu3 ku3/
Half sandhi	马书	T3T1	Low-falling T1	<i>horse-book</i>	/ma3 ʃu1/
	马球	T3T2	Low-falling T2	<i>horse-ball</i>	/ma3 tʃiu2/
	马画	T3T4	Low-falling T4	<i>horse-painting</i>	/ma3 hua4/
	鸟书	T3T1	Low-falling T1	<i>bird-book</i>	/niau3 ʃu1/
	鸟球	T3T2	Low-falling T2	<i>bird-ball</i>	/niau3 tʃiu2/
	鸟画	T3T4	Low-falling T4	<i>bird-painting</i>	/niau3 hua4/
	狗书	T3T1	Low-falling T1	<i>dog-book</i>	/kou3 ʃu1/
	狗球	T3T2	Low-falling T2	<i>dog-ball</i>	/kou3 tʃiu2/
	狗画	T3T4	Low-falling T4	<i>dog-painting</i>	/kou3 hua4/
	鼠书	T3T1	Low-falling T1	<i>mouse-book</i>	/ʃu3 ʃu1/
	鼠球	T3T2	Low-falling T2	<i>mouse-ball</i>	/ʃu3 tʃiu2/
	鼠画	T3T4	Low-falling T4	<i>mouse-painting</i>	/ʃu3 hua4/

monosyllabic word set (e.g., [zi3 ma3] gu3; *purple-horse drum*; Table 2). Similarly, in the right-branching structures, the leftmost monosyllabic word was kept constant as /tsi3/ *purple* to combine with a disyllabic unit composed of 2 monosyllabic words in a T3TX sequence (e.g., zi3 [ma3 gu3]) *purple horse-drum*.¹ The left-branching structure contained **Noun + Noun** items (“purple-horse” + “drum”),

Table 4. Trisyllabic stimuli list

Control/ target	Compound	Underlying tones/ structure	Surface/target tone	Meaning	Phonemic transcription
T3T3T1/2/4 (Control)	([紫马] 书)	(([T3T3] T1)	T2 T3T1	((purple horse book)	/tsi3 ma3 ʂu1/
	([紫马] 球)	(([T3T3] T2)	T2 T3T2	((purple horse) ball)	/tsi3 ma3 tɕiu2/
	([紫马] 画)	(([T3T3] T4)	T2 T3T4	((purple horse) painting)	/tsi3 ma3 hua4/
	(紫 [马书])	(T3 [T3T1])	T2 T3T1	(purple (horse book))	/tsi3 ma3 ʂu1/
	(紫 [马球])	(T3 [T3T2])	T2 T3T2	(purple (horse ball))	/tsi3 ma3 tɕiu2/
	(紫 [马画])	(T3 [T3T4])	T2 T3T4	(purple (horse painting))	/tsi3 ma3 hua4/
T3T3T3 (Target)	([紫马] 鼓)	(([T3T3] T3)	T2 T2T3	((purple horse) drum)	/tsi3 ma3 ku3/
	(紫 [马鼓])	(T3 [T3T3])	T3 T2T3~ T2 T2T3	(purple (horse drum))	/tsi3 ma3 ku3/

whereas the right-branching structure contained **Adjective** + **Noun** items (“purple” + “horse-drum”). In total, there were 6 T3T3T1/2/4 compounds and 2 T3T3T3 compounds. The 6 T3T3T1/2/4 compounds were used as controls as the full tone sandhi process is only applied on the first T3 syllable of these trisyllabic novel compounds (Shih, 1997). The 2 T3T3T3 novel compounds were thus the target items used to test participants’ sensitivity to applying tone sandhi as a function of different prosodic structures. According to the recursive application theory of Shih (1997), the expected surface tones of the 6 control items and the 2 target items are those exemplified in Table 4.

All trisyllabic novel compounds were picturable, with prosodic structure constructed via eliciting different novel compounds. For example, the left-branching target item ([zi3 ma3] gu3) *purple-horse drum* was depicted by a drum with a purple horse on it, while the right-branching target item (zi3 [ma3 gu3]) *purple horse-drum* was depicted by a horse-drum (a drum with a horse on it), which was then colored purple, resulting in the final representations depicted in Figure 3.

Procedure

Children were tested in a quiet room at the kindergarten; adults were tested in a quiet room at the university. During the test session, participants wore a head-mounted cardioid-directional condenser microphone (AKG-C520), which was connected to a solid-state recorder (Marantz PMD661MKII) in order to capture their word productions.

The experiment had three parts, eliciting monosyllabic, disyllabic, and trisyllabic items, respectively, each with an initial familiarization/practice phase. In Part 1,

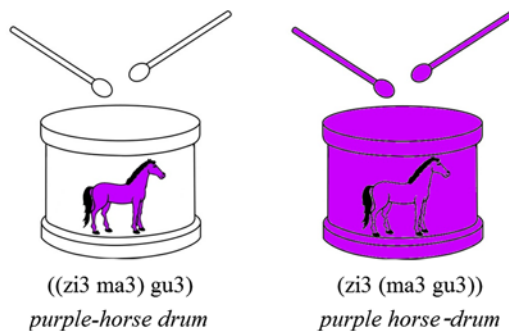


Figure 3. Two pictures used to elicit the left-branching trisyllabic item $[(zi3\ ma3)\ gu3]$ *purple-horse drum* and the right-branching trisyllabic item $(zi3\ [ma3\ gu3])$ *purple horse-drum*.

monosyllabic words were elicited in a picture-naming task. Prior to testing, there was a familiarization phase during which the Mandarin-speaking experimenter produced the names of all monosyllabic words in Table 2, including the 2 practice items: /mau1/ *cat* and /jan2/ *sheep*. The experimenter first presented a picture of an item on the computer screen and asked the participant to name the picture. If the participant failed to name it or produced it with a different word, the experimenter would correct it and ask the child to name it again. Once the participant was able to correctly produce the name of all the pictured words, the experimenter proceeded to the test phase. In the test phase, the experimenter showed pictures of all the objects one by one and asked the participant to name the pictures using the words produced during the familiarization phase. The order of items was randomized across participants. These data provided both a tonal and a lexicalized item control for assessing participants' knowledge of tone sandhi processes with novel word combinations in Parts 2 and 3.

In Part 2, the 16 disyllabic compounds were elicited in a novel word-compounding game. Four novel practice trials (with feedback, i.e., the correct target word) were first used to familiarize participants with the procedure for forming novel compounds: /mau1 ʃu1/ *cat-book*, /mau1 tɕ^{hiu}2/ *cat-ball*, /jan2 ku3/ *sheep-drum*, and /jan2 hua4/ *sheep-painting*. This was identical to the procedure then used for the test trials, where, for example, to elicit the disyllabic compound /ma3 ku3/ *horse-drum*, a picture of a horse and a picture of a drum were displayed on the left and right sides of a robot-like cartoon figure, respectively (see Step 1 in Figure 4). The experimenter then pressed a button to play a preprogrammed animation where the two pictures disappeared behind the robot and it jiggled to output a novel object: a horse-drum (see Step 2 in Figure 4). The experimenter then asked the participant to produce the name of the resulting item. The order of the test trials was randomized across participants.

In Part 3, the 8 trisyllabic compounds were elicited using a similar procedure. The elicitation methods for compounds with left- and right-branching structures were slightly different. To elicit the left-branching target item $[(zi3\ ma3)\ gu3]$ *purple-horse drum*, for instance, a purple-horse and a drum appeared on opposite sides of the robot (Step 1, Figure 5), and the experimenter asked the participant to

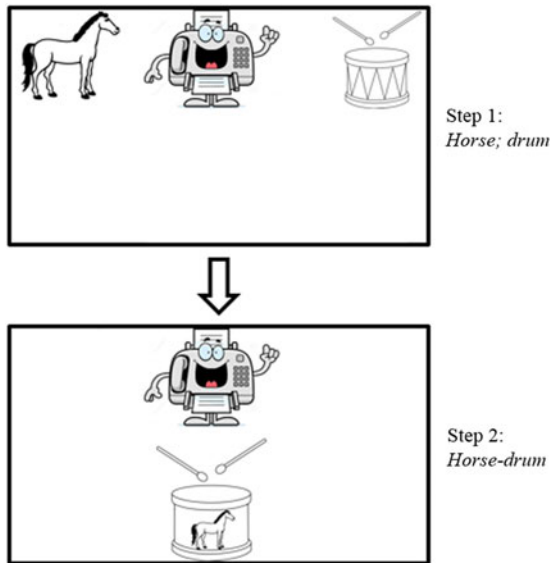


Figure 4. To elicit the compound “horse-drum,” a horse and a drum were first presented on either side of the robot. The horse and the drum were then combined into a new item “horse-drum” by the robot.

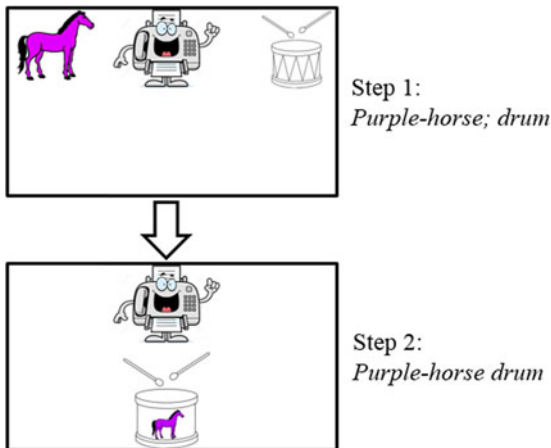


Figure 5. To elicit the compound “purple-horse drum,” a purple horse and a drum were first presented on opposite sides of the robot. The purple-horse and the drum were then combined into a new item “purple-horse drum” by the robot.

name both items. Next, the experimenter pressed a button to play the same animation of the robot (jiggling) to combine the purple-horse and the drum into a new object: a purple-horse drum (Step 2, Figure 5). The experimenter then asked the participant to name the novel compound.

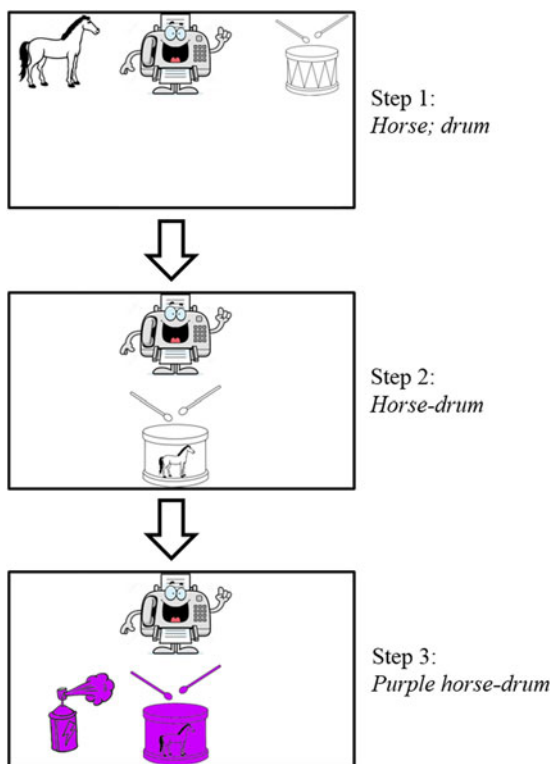


Figure 6. To elicit the compound “purple horse-drum,” a horse and a drum were first presented on opposite sides of the robot. The horse and the drum were then combined into a new item “horse-drum.” A can of spray-paint appeared and colored the horse-drum purple to create a new object “purple horse-drum.”

To elicit the *right-branching* item (zi3 [ma3 gu3]) *purple horse-drum*, a “horse-drum” was first generated, using identical steps to those for the left-branching items (Steps 1 and 2, Figure 6). The participant was asked to produce its name, as it also served as the input for the trisyllabic word. Step 3 was then added to render the novel object purple (Step 3, Figure 6) so as to elicit the novel compound “purple horse-drum.”

Compounds with left-branching and right-branching structures were elicited in two separate blocks to minimize the processing load for young children by not requiring them to switch between slightly different elicitation procedures. Both blocks began with a nonsandhi practice trial (e.g., [hong2 mao1] shu1]; *red-cat book* or [hong2 {mao1 shu1}]; *red cat-book*), followed by the 4 test trials. During practice, feedback was provided when the child used a nontarget word in forming the novel compounds. For instance, instead of using the target word “hong2” (red), some children produced “hong2 se4” (red-colored) for “(hong2 [mao1 shu1])” *red cat-book*, or “(hong2 se4 de0 [mao1 shu1])” *red-colour cat-book*. In this case, the experimenter would provide the target word “hong2” and prompt the child to use it to produce the novel compound again. This was done to ensure that the child

understood the task as no feedback was provided for the test items. The order of the two blocks and the order of testing trials within each block were randomized across participants.

Annotation and measurements

Participants' productions were acoustically coded in Praat (Boersma & Weenink, 2016). The vowel of each syllable in the compound was identified and annotated in terms of the onset and offset of the second formant (F2). Pitch tracks of annotated vowels were checked and manually revised to correct for any "doubling" or "halving" errors. The revised pitch track was then interpolated and smoothed with a bandwidth of 20 Hz. The f0 values were extracted at 10 equidistant points within each annotated vowel using the default autocorrelation algorithm in Praat (Boersma, 1993). The extracted f0 values were then converted from Hz to semitones with a reference of 50 Hz. Ten percent of the items were recoded by a second trained native speaker of Mandarin. Interrater reliability on the extracted pitch points was good ($r = .94$). To minimize possible interspeaker variation in pitch, the f0 values of each token were z-score normalized against the mean pitch across all tokens for each individual speaker using the following formula (Mean pitch_{Individual} and Standard deviation of pitch_{Individual} are the grand mean and standard deviations of all tokens per individual participant):

$$\text{Normalized pitch} = \frac{\text{Observed pitch} - \text{Meanpitch}_{\text{Individual}}}{\text{Standard deviation of pitch}_{\text{Individual}}}$$

Statistical analysis

A total of 4,444 tokens were included in the data analysis (1,377 tokens from 3-year-olds, 1,313 tokens from 4-year-olds, 698 tokens from 5-year-olds, and 1,056 tokens from adults). An additional 68 tokens (2% exclusion rate: 31 tokens from 3-year-olds, 31 tokens from 4-year-olds, and 6 tokens from 5-year-olds) 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 child and adult productions across conditions, a second-order orthogonal polynomial equation was fitted for each tone production, using the *poly* function of R. The second-order polynomials were adopted because the most complex pitch contour of tones in our data had only a convex or a concave contour shape. The polynomial function generated three parameters for each pitch contour: the intercept, the linear trend, and the quadratic trend. According to Mirman (2014), the three parameters, respectively, capture the *pitch onset* (as reflected in the intercept: the higher the intercept, the higher the pitch onset value), *direction* (slope as reflected in the linear trend: a positive value indicates a rising pitch and a negative value a falling pitch, with a large value representing steepness), and *curvature* (as reflected in the quadratic trend: a positive quadratic trend indicates a concave f0 contour, a negative quadratic trend indicates a convex contour; and a large

quadratic trend indicates a more curved f0 contour and vice versa). These parameters were used to evaluate any group differences in the overall f0 contour.

Linear mixed regression models were built to compare the tone productions between 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). The analysis of variance function, which provides Satterthwaite's approximation to degrees of freedom for estimating *p* values using *F* statistics in R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015), was used to test for statistical significance. The main and interaction effects reported in the results were averaged across all levels of the other effects (see Luke, 2017, for a detailed explanation and Peters, Hanssen, & Gussenhoven, 2014, for a practical example). When a significant main effect of a multilevel factor or a significant interaction effect was observed, Tukey–HSD post hoc comparisons were performed on the multilevel factor, as well as interactions, using lsmeans package (Lenth, 2016). The parameter estimates are provided in the results of post hoc comparisons rather than in the results of the main model.

Results

Tone sandhi in disyllabic words

In this section we address H1: (a) young children (i.e., 3-year-olds) might not productively apply the tone sandhi process to novel disyllabic compounds; and (b) children's ability to apply the tone sandhi process to novel disyllabic compounds will become more acoustically adultlike with age. We first examined children's lexical/citation tone productions to make sure that they had correct tonal representations for all four lexical tones, and then examined their disyllabic tone sandhi productions.

Lexical tones

Figure 7 illustrates the normalized lexical tone pitch contours for the three child groups compared to adults. All child groups showed very similar lexical tone pitch patterns, and were similar to the adults, with a level pitch for T1, a rising pitch for T2, a dipping pitch for T3, and a falling pitch for T4 (Figure 7). A linear mixed regression model was built using the three pitch parameters (intercept, slope, and curvature) across 10 time points to explore any group differences in tone contour. Two fixed factors, group (3-, 4-, 5-year-olds and adults) and tone (T1, T2, T3, and T4), and the random factor participant were entered into the model. As we expected all children to have acquired all four lexical tone categories, in line with previous studies (e.g., Hua & Dodd, 2000; Li & Thompson, 1977), we predicted that the model would predict no effect of group on the pitch shape, that is, the pitch slope (linear trend) and the pitch curvature (quadratic trend).

The results are presented in Appendix A, showing that there were significant main effects of group and tone and a significant interaction between group and tone on the pitch intercept (pitch onset). Note that the intercept parameter only captures pitch onset of a lexical tone contour and that raw pitch values were normalized

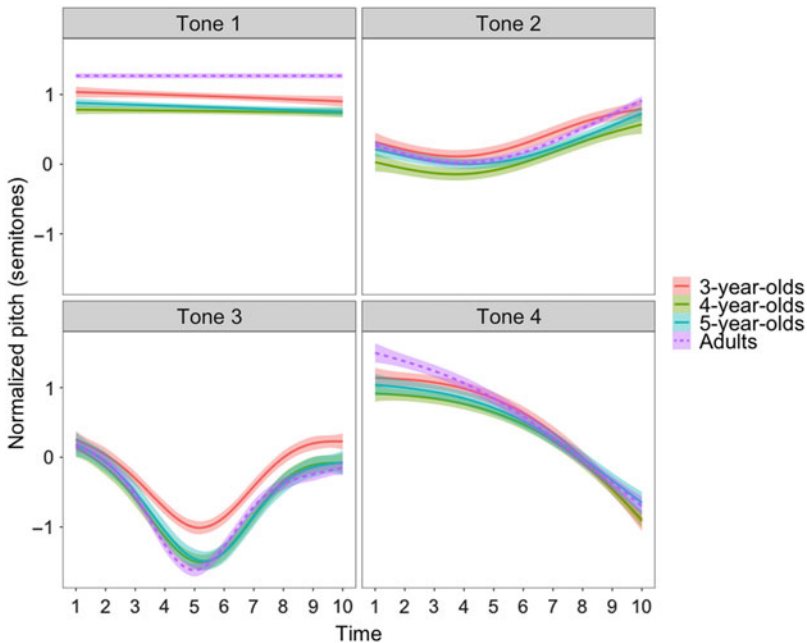


Figure 7. Pitch contour of lexical tone productions from children (3-, 4-, and 5-year-olds) and adults.

against the grand mean and standard deviation. Therefore, the normalized pitch onset can deviate from the grand mean. As pitch onsets of T1 and T4 are higher than those of T2 and T3, the Group \times Tone interaction indicates that adults had higher T1 and T4 onsets than children. As Mandarin Chinese is characterized by distinct tone contours, our study focused on pitch slope and curvatures as parameters to examine any group differences. Thus, although the pitch intercept parameter was included in further statistical modeling, results of this parameter will not be further discussed.

Regarding the pitch slope and curvature, however, the results showed that there was no main effect or interaction on these two parameters, indicating that the shape of children's lexical tone pitch contours was not significantly different from that of adults. Furthermore, a significant main effect of tone on the pitch shape parameters was also observed. The post hoc test on this main effect showed that the pitch shape was significantly different between any two lexical tones (Appendix B), suggesting that adults and children produced tonal contrasts in a consistent way. This corroborates previous findings showing that, by 3 years, children have established appropriate representations for Mandarin lexical tones (e.g., Hua & Dodd, 2000; Li & Thompson, 1977).

Disyllabic tone sandhi compounds

We now turn to the results of the disyllabic novel compounds. Figure 8 shows 3-, 4-, and 5-year-olds' and adults' tone sandhi productions across the two types of tonal

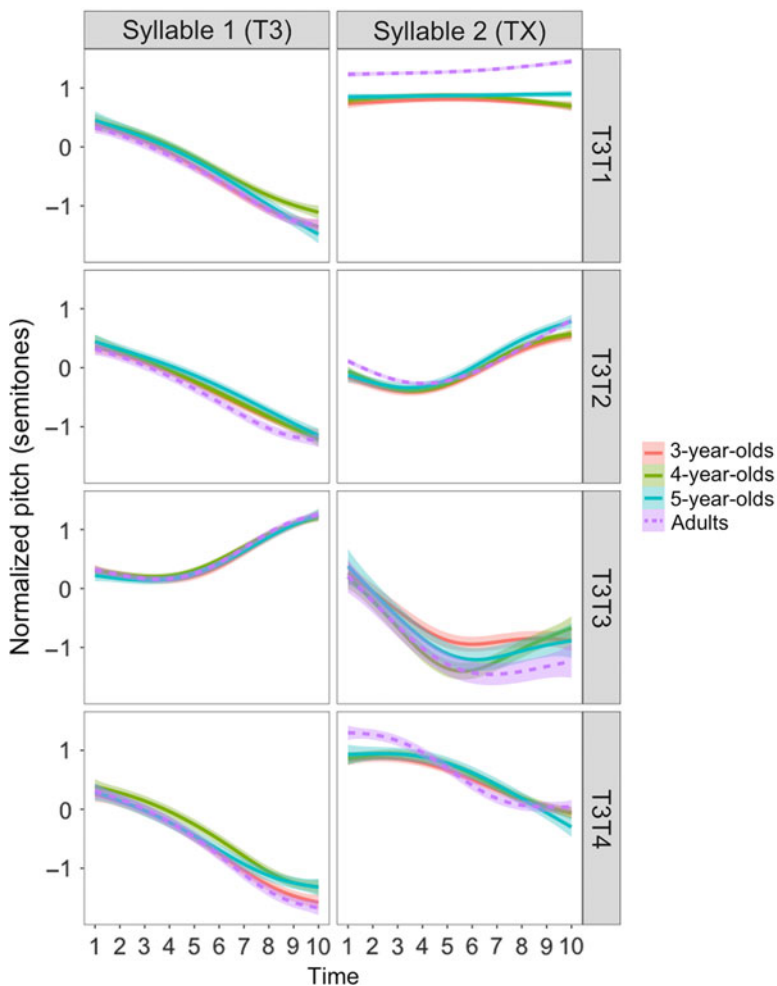


Figure 8. Pitch contours of T3TX productions from children (3-, 4-, and 5-year-olds) and adults, where T3T3 items result in full sandhi compounds and T3T1/2/4 items result in half sandhi compounds.

contexts: one triggering full sandhi (T3T3), and the others triggering half sandhi (T3T1/2/4). Visual inspection of Figure 8 suggests that children from all three age groups have correctly applied the tone sandhi process to novel disyllabic compounds across both types of tonal context. That is, they changed the tone of the first syllable (T3) to a rising tone (T2) in the full sandhi context, and changed it to a low-falling tone in the three half sandhi contexts.

We then conducted a linear mixed regression model to look for any group differences in sandhi production. The comparison focused on the first T3 syllable of the disyllabic compound, as this is where the tone sandhi change takes place. Two fixed factors, group (3-, 4-, 5-year-olds, and adults) and context (before T3 and before T1/2/4), with the random factor participant were included in the model.

According to H1(a), if young children did not productively apply the tone sandhi process, the model would predict a main effect of group and an interaction between group and context on the pitch shape (slope and curvature). According to H1(b), if children's ability to apply the tone sandhi process becomes more adultlike as they get older, a post hoc test on the main effect of group would reveal a smaller pitch difference between older children and adults compared to that between young children and adults.

The results are presented in Appendix C. Neither the main effect of group nor the interaction between group and context was significant for pitch shape (slope and curvature), suggesting that children's sandhi productions were not significantly different from adults for either full or half sandhi syllables. Moreover, the significant main effect of context suggests that, for both child and adult groups, the pitch contour of T3 syllables was different when produced before T3 and before T1/2/4 syllables. A post hoc test of context showed that, relative to the half sandhi context (before T1/2/4), the pitch contour of T3 syllables was more rising and curved: slope, $\beta = 11.6$, $SE = 0.16$, $t(22084) = 71.38$, $p < .001$; curvature, $\beta = 2.25$, $SE = 0.16$, $t(22082) = 13.81$, $p < .001$, when they were produced before another T3 syllable (i.e., the full sandhi context), consistent with the acoustic feature of a rising T2. Thus, children from all age groups were able to productively apply tone sandhi processes to novel disyllabic compounds in both full and half sandhi tonal contexts.

Tone sandhi in trisyllabic words

In this section we address H2: (a) 3-year-olds might not apply the full sandhi process to novel trisyllabic (T3T3T3) compounds, and (b) children's tone sandhi application to novel trisyllabic compounds would become acoustically more adultlike with age. We first examined children's and adults' productions of trisyllabic control items (underlying T3T3T1/2/4), where the tone sandhi process does not interact with the prosodic structure (Shih, 1997). We then examined their tone sandhi application on the target items (underlying T3T3T3) to test children's ability to build different prosodic structures to guide recursive tone sandhi application.

Control items: T3T3T1/2/4 compounds

Figure 9 shows children's and adults' productions of trisyllabic tone sandhi control words, ([T3T3] T1/2/4) versus (T3T3T1/2/4), where the tone sandhi process only applies to the first syllable of the T3T3 sequence. The surface tones should thus be T2T3T1/2/4 for both structures. Visual inspection of Figure 9 showed that all groups produced the correct surface tones across structures.

To examine any group differences in surface tone of the T3T3T1/2/4 control items, a linear mixed regression model was constructed using the three pitch contour parameters (intercept, slope, and curvature) of the first T3 syllable within the trisyllabic compound, as this is where the tone sandhi process takes place. Two fixed factors, group (3-, 4-, 5-year-olds, and adults) and structure (left-branching and right-branching), and a random factor, participant, were entered into the model. On the basis of previous observations that prosodic structures do not change the surface realization of T3T3T1/2/4 sequence in real words (Shih, 1997), we wanted

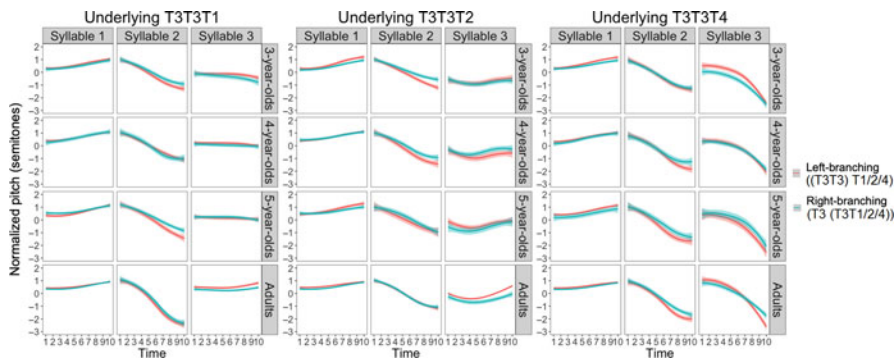


Figure 9. Pitch contours of child (3-, 4-, and 5-year-olds) and adult productions of trisyllabic tone sandhi controls: ([T3T3] T1/2/4) and (T3 [T3T1/2/4]).

to verify this for the novel items used in the current study and therefore included “structure” as a factor in the model. The results are presented in Appendix D, which show that there was no main effect of structure or group nor any interaction between structure and group on the pitch shape, indicating that the prosodic structure did not interact with their tone sandhi application on these control items, and that children’s acoustic realizations were not significantly different from those of adults.

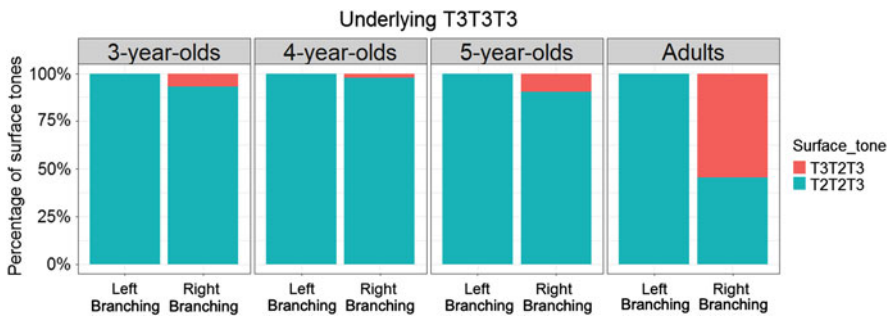
Target items: T3T3T3 compounds

For the target trisyllabic tone sandhi items, the prosodic structure was expected to guide sandhi application resulting in different surface tone realizations for the two different prosodic structures. Recall that, in the left-branching structure ([T3T3] T3), tone sandhi applies to the leftmost disyllabic unit first, and then again at the level of the trisyllabic prosodic word, resulting in a surface **T2T2T3** tonal realization. In the right-branching structure (T3 [T3T3]), tone sandhi applies to the rightmost disyllabic unit first, and this is then incorporated into the trisyllabic prosodic word, resulting in a **T3T2T3** surface tone realization. However, this right-branching structure has also been reported to “optionally” surface as **T2T2T3**, thought to occur in fast speech (Chen, 2000, p. 379).

Because the right-branching structure can exhibit two possible surface realizations of Syllable 1 (i.e., **T3T2T3** and **T2T2T3**), it was necessary to check whether this is the case. Standard deviations (*SD*) of the pitch slope for Syllable 1 from both the left- and right-branching structures were compared (see Table 5 for descriptive data). Adults’ *SD* for the right-branching structure was twice that of the left-branching structure (0.2 vs. 0.48), whereas children’s *SD* did not show this pattern. As suggested by Chen (2000), adults produced more variable surface outputs for the right-branching structure. Therefore, it is necessary to check whether the large variability fell into the two possible **T2T2T3** or **T3T2T3** forms as suggested in Chen (2000). Surface forms for the trisyllabic items were therefore first coded perceptually by a trained native Mandarin speaker (the first author). Acoustic comparisons were

Table 5. Descriptive data of the F0 slope (mean, SD) of Syllable 1 in children's and adults' T3T3 productions

Structure	Group	Slope (mean)	Slope (SD)
([T3T3] T3)	3-year-olds	0.37	0.28
	4-year-olds	0.34	0.24
	5-year-olds	0.33	0.26
	Adults	0.21	0.2
(T3 [T3T3])	3-year-olds	0.41	0.36
	4-year-olds	0.29	0.25
	5-year-olds	0.32	0.31
	Adults	-0.12	0.48

**Figure 10.** Proportion of participants who produced surface tones **T2T2T3** versus **T3T2T3** for ([zi3 ma3] *purple-horse drum* and [zi3 [ma3 gu3]] *purple horse-drum*, based on perceptual coding.

made between adults and children to explore any group differences in the phonetic implementation of identical surface tones.

Figure 10 shows the proportion of children and adults who produced the two surface tone patterns for two prosodic structures. Ten percent of the tokens were reassessed by another trained native speaker, with good interrater reliability ($r = .96$).

All children and adults produced **T2T2T3** as the surface tone for the left-branching item ([T3T3] T3). However, the children differed from adults in the choice of surface tone for the right-branching item (T3 [T3T3]), with most of the children producing **T2T2T3**. In contrast, 55% of adults (18 out of 33 participants) used the **T3T2T3** surface form, whereas the other 45% (15 out of 33 participants) used the **T2T2T3** surface form (recall that each participant produced only 1 item per prosodic structure). To investigate the proportion of surface tone patterns for each prosodic structure, Chi-square tests were then conducted. The results showed that, for the child groups, counts of **T2T2T3** surface productions did not differ between the two prosodic structures: 3-year-olds, $\chi^2(1) = 2.828$, $p = .093$; 4-year-olds, $\chi^2(1) = 0.329$, $p = .566$; 5-year-olds, $\chi^2(1) = 2.86$, $p = .091$. For adults, however,

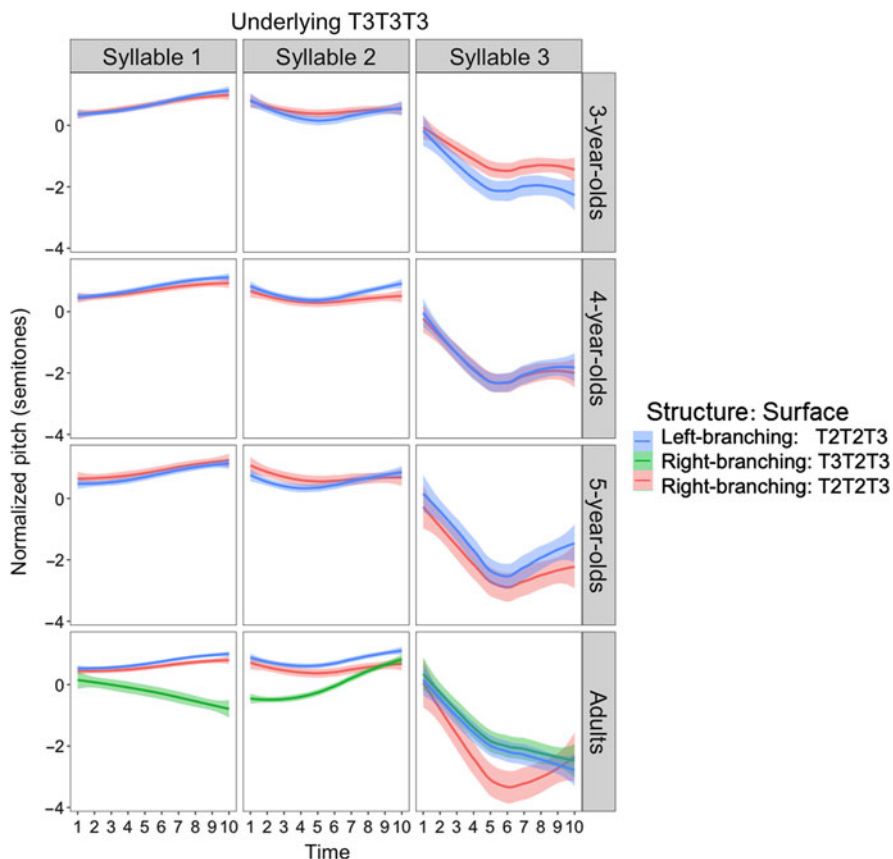


Figure 11. Pitch contours of child (3-, 4-, and 5-year-olds) and adult productions of target trisyllabic tone sandhi items ([zi3 ma3] gu3) *purple-horse drum* and (zi3 [ma3 gu3]) *purple horse-drum*.

there was a significant difference between the two conditions, $\chi^2(1) = 31.871$, $p < .001$. Thus, children used **T2T2T3** as the surface tone for *both* trisyllabic prosodic word structures, while adults used **T2T2T3** for the left-branching structure, but both **T3T2T3** and **T2T2T3** for the right-branching structure. In other words, children produced the same surface tones (**T2T2T3**) as adults in the left-branching structure, but differed from adults in sandhi application in the right-branching structure, with all but a few children (four 3-year-olds, one 4-year-old, and two 5-year-olds) producing **T2T2T3** in contrast to adults' variable production of both **T3T2T3** and **T2T2T3**.

Figure 11 shows the overall child and adult acoustic realizations of pitch for the **T3T3T3** items as a function of the two different prosodic structures. Note the two different realizations for the right-branching forms for the adults: both **T3T2T3** and **T2T2T3**. We then performed two acoustic analyses to examine whether child trisyllabic tone sandhi productions were adultlike and whether the adult **T2T2T3** output for the right-branching structure was a result of faster speaking rate, as proposed by Chen (2000).

To answer the first question, we acoustically compared children's productions with those from adults. As adults produced different surface tones for the two structures (i.e., **T2T2T3** for the left-branching structure and **T2T2T3/T3T2T3** for the right-branching structure), the two structures were analyzed separately. For the left-branching items, where both children and adults only used **T2T2T3** as the surface pattern, two linear mixed regression models were built for the first two syllables (where the tone sandhi process occurs), with group as a fixed factor (3-, 4-, 5-year-olds, and adults) and participant as a random factor. The results showed that there was no main effect of group on the pitch slope or curvature for the two syllables (Appendix E), indicating that children's left-branching surface productions were not different from those from adults. For the right-branching items, where children used **T2T2T3** as the surface pattern and adults used both **T2T2T3** and **T3T2T3** as surface tones, another two linear mixed regression models were built for the first two syllables with "Group_Surface tone" as a fixed factor (3-, 4-, 5-year-olds, adults_ **T2T2T3**, and adults_ **T3T2T3**), where adult productions were further categorized into two levels (i.e., adults_ **T2T2T3** and adults_ **T3T2T3**) based on the perceived surface tones, and participant as a random factor. The results showed a significant main effect of "Group_Surface tone" on the pitch slope of both syllables, Syllable 1: $F(4, 1187) = 13.42, p < .001$; Syllable 2: $F(4, 1187) = 112.53, p < .001$ (Appendix F). Post hoc tests showed that the first two syllables of children's and adults' **T2T2T3** productions were significantly different from those of adults' **T3T2T3** productions, while there was no difference between children's and adults' **T2T2T3** productions (Appendix G). This provided acoustic evidence that adults used two different surface patterns for the right-branching structure, and children's productions of right-branching items were not different from those of adults with the same surface pattern.

It has been proposed that faster speaking rate may optionally change the surface tone of the right-branching structure from **T3T2T3** to **T2T2T3** in adult speech, leading to variation in tone sandhi application for this structure (Chen, 2000, p. 379). To examine whether the adult **T2T2T3** productions for the right-branching structure were a result of fast speaking rate, we compared the normalized syllable duration (as a proxy of speaking rate) of adult right-branching surface outputs (**T3T2T3** vs. **T2T2T3**). Syllable duration was normalized to control for any individual differences in speaking rate. The mean syllable durations of all three syllables (i.e., duration of vowel portion of /tsi³ ma³ ku³/) were measured and normalized against the grand mean duration across all tokens produced by each individual speaker. A linear mixed-effects model was constructed on the normalized syllable duration for all right-branching adult prosodic words, with surface-realization as a fixed factor (**T3T2T3** vs. **T2T2T3**) and participant as a random factor. If there was a speaking rate difference, we predicted a main effect of surface-realization, where the normalized syllable duration would differ between the two output forms. However, the result did not support this prediction: the main effect of surface-realization was not significant: $F(1, 31) = 1.03, p = .31$; the average normalized syllable duration for the **T3T2T3** productions was 0.84 ($SD = 0.23$, range: 0.35 to 1.4), and for the **T2T2T3** productions was 0.87 ($SD = 0.31$, range: 0.26 to 1.4). Thus, the normalized syllable duration did not differ between the **T3T2T3** and **T2T2T3** forms for the right-branching prosodic words, suggesting that speaking rate was not the source of these variable adult surface forms.

Discussion

Phonological processes can often pose a learning challenge for children. In this study, we explored the acquisition of tone sandhi processes in Mandarin Chinese by 3- to 5-year-old children. We asked two questions. The first concerned children's ability to productively apply the tone sandhi process to novel disyllabic compounds in appropriate tonal contexts. The second concerned children's ability to construct different prosodic structures to guide single or recursive sandhi rule application in novel trisyllabic compounds. We addressed these two questions by conducting acoustic analyses to examine the fine-grained pitch contours of children's tone sandhi productions across tonal contexts and prosodic structures. To our knowledge, very few studies have carried out similar analyses of children's tone sandhi productions, either with novel compounds or using acoustic methods.

Regarding the productive knowledge of the tone sandhi process, we hypothesized in H1 that 3-year-olds would not be able to productively apply the tone sandhi process to novel disyllabic compounds, but that this would become more adultlike with age. However, our results provide acoustic evidence that children from all age groups (3, 4, and 5 years) were able to productively apply the (full and half) tone sandhi process to novel disyllabic compounds, changing underlying T3 to T2 before another T3, and to a low-falling tone before T1/2/4. This result therefore extends our understanding of children's tone sandhi knowledge from the early emergence of tone sandhi productions in lexicalized items (by 3 years; Hua & Dodd, 2000; Li & Thompson, 1977) to the productive application of tone sandhi processes in novel items. These findings are also consistent with the reported acquisition of certain grammatical tone sandhi processes in Bantu languages, also acquired by the age of 3 (Demuth, 1993).

The early acquisition of tone sandhi processes is somewhat surprising given that previous studies have suggested that phonological alternations can often pose a learning challenge for young children (Albright & Hayes, 2011). One potential factor contributing to this early acquisition of simple tone sandhi processes here might be the presence of these processes in the input infants hear. For instance, Tang, Xu Rattanasone, Yuen, and Demuth (2017) found that, despite the slow speaking rate in Mandarin infant-directed speech (IDS), which might induce more boundaries between syllables and block the application of tone sandhi processes, mothers consistently used tone sandhi in their speech to their 12-month-olds, thus providing Mandarin-learning infants with adequate evidence from which to learn these tone sandhi processes.

We then explored children's ability to apply the tone sandhi process to novel trisyllabic words, where learners need to use different prosodic structures to guide the application of tone sandhi. Recall that the expected surface tone patterns for the left-branching and right-branching structures were **T2T2T3** and **T3T2T3** ~ **T2T2T3**, respectively (Chen, 2000; Shih, 1997).

Results from the perceptual and acoustic analyses found that both adults and children produced the **T2T2T3** surface tones for the left-branching structure, and did not differ from one another in the phonetic implementation of these forms. This suggests that, by 3 years, children are already able to recursively apply the tone sandhi process to novel left-branching trisyllabic items. This extends previous findings from Wang (2011) that 3-year-olds can produce **T2T2T3** as the surface form

for left-branching compounds when these contain a familiar disyllabic lexicalized item (e.g., [ʃhui3 guo3] niao3] *fruit-bird*). This early ability to recursively apply the tone sandhi process in the left-branching structures can also be attributed to the language input, where Mandarin-speaking mothers consistently produce T2T2T3 for left-branching structures in IDS (Tang et al., 2017).

However, adults and children differed in their surface outputs for the right-branching trisyllabic words, where half of the adults produced the T3T2T3 surface form, but the other half, and most of the children, produced the recursive T2T2T3 surface form. This is *not* consistent with previous findings that 3-year-olds generally produce the T3T2T3 surface pattern for right-branching items such as (shui3 [lao3 hu3]) *water-tiger* (Wang, 2011), where the (lao3 hu3) *tiger* is a lexicalized disyllabic word. This raises many questions as to why the adults showed variability in their realization of the right-branching trisyllabic compounds in the present study,² and why children generally used this form.

Adults: Speaking rate

One possible explanation for adults' surface variation in the right-branching trisyllabic compounds was that the T2T2T3 surface pattern results from a fast speaking rate, as suggested by Chen (2000, p. 379). To test this possibility, we compared the (normalized) durational difference between adults' T3T2T3 and T2T2T3 outputs in the right-branching productions. There was no durational difference between the two, indicating that adults' surface variation was not driven by speaking rate, at least in the present study. Moreover, Tang et al. (2017) observed no surface tonal variation for right-branching trisyllabic words across different registers, even though the speaking rate differed as a function of IDS (hyperarticulated with a slow speaking rate), adult-directed speech (normal speaking rate), and Lombard speech (used in noise, also with a slow speaking rate); all (female) speakers consistently produced T3T2T3 for the trisyllabic right-branching word (xiao3 [ma3 yi3]) *small ant* across all registers (including at least 10 tokens for each register, for a total of 442 tokens produced across registers). That is, there was no variation in these participants' production of tone sandhi in this right-branching context. This suggests again that speaking rate may not be involved in adults' variable realization of tone sandhi on right-branching compounds, but this variation may be driven by other factors. Note, however, that in the Tang et al. (2017) study, like the Wang (2011) study, these trisyllabic compounds were composed of a monosyllable followed by a lexicalized disyllabic word.

We therefore considered three other factors that might help explain the adult variation in surface tones for the right-branching structure found in our results: (a) the frequency distribution of the left-branching versus right-branching structures in Mandarin; (b) the low lexical familiarity of the novel right-branching forms used in the present study; and (c) the monomorphemic status of Adj+N units in Mandarin.

Frequency distribution of left-branching versus right-branching compounds in Mandarin

One factor possibly accounting for adults' surface variation in the right-branching structure might be related to the frequency distribution between the left-branching

and right-branching structures in Mandarin Chinese. Duanmu (2012) conducted a corpus study on the word length in Mandarin Chinese, using the Lancaster Corpus of Mandarin Chinese (1 million words of written Mandarin Chinese; McEnery & Xiao, 2004), observing that 95% of trisyllabic noun compounds had left-branching structure (disyllable + monosyllable) with only 5% having right-branching structure (monosyllable + disyllable). Thus, perhaps adults' optional T2T2T3 surface pattern for the right-branching trisyllabic items is simply a default to the more frequent surface tonal pattern found on trisyllabic words.

Low lexical familiarity of novel trisyllabic compounds

Apart from the asymmetric frequency distribution between the two structures, the second factor leading to some adults' variable tonal realization for the right-branching structure might be related to the low lexical familiarity of the novel compounds used in the present study. In this study, both trisyllabic words were novel items, that is, ([zi3 ma3] gu3) *purple-horse drum* and (zi3 [ma3 gu3]) *purple horse-drum*, where the disyllabic components were also novel items, that is, “zi3 ma3” *purple-horse* and “ma3 gu3” *horse-drum*. Therefore, adults were not able to rely on their preexisting lexical knowledge to build prosodic structures for these items, and this might thus have led some adults to use the more frequent surface tonal pattern, that is, T2T2T3, for both types of items.

Monomorphemic status of Mandarin Adj+N compounds

In the left-branching structure “([zi3 ma3] gu3)” *purple-horse drum*, the adjective “zi3” *purple* was first combined with a noun “ma3” *horse* to form a disyllabic noun, then incorporated as part of a trisyllabic compound with the structure of ([Adj+N] N). This is in accord with the strong preference in Chinese for the monomorphemic status of Adj+N sequence (Xu, 2018). However, this is not the case for the right-branching structure “(zi3 [ma3 gu3])” *purple horse-drum*, whereby the adjective was combined with a novel disyllabic N+N compound as part of a noun phrase, with the structure of (Adj [N+N]). Therefore, the monomorphemic status of Adj+N sequence might potentially bias some adults to use the preferred left-branching structure to produce the right-branching items, thus leading to the T2T2T3 production, especially when the disyllabic component is not a lexicalized item.

Taken together, our results suggest that the low overall frequency of right-branching structures in Mandarin, along with the low lexical familiarity of the novel items used in the present study and the monomorphemic status of the Adj+N sequence, may all have led some adults to produce both structures with the preferred T2T2T3 pattern.

Children's T2T2T3 productions for the right-branching structure

We now turn to the children's results and discuss why they used only the T2T2T3 surface pattern for the right-branching item. As mentioned above, Tang *et al.* (2017) showed that, in children's early language input, Mandarin-speaking mothers

consistently produced T2T2T3 versus T3T2T3 for left-branching versus right-branching items, suggesting that children hear abundant evidence for the canonical surface realizations for these respective trisyllabic words in the input. Children's T2T2T3 productions then were not driven by variable tonal realizations of right-branching forms in their environment. Thus, it appears that, as for adults, children's T2T2T3 productions may be related to the higher frequency of left-branching forms in the overall input they hear, the low lexical familiarity of the novel disyllabic component, and/or the monomorphemic bias of the Adj+N unit.

Previous acquisition studies suggest that, in early language development, phonological structures that are more frequent in the language input are usually acquired and produced first (Levelt, Schiller, & Levelt, 2000; Stites, Demuth, & Kirk, 2004; Zamuner, Gerken, & Hammond, 2004), and that words with non-predominant phonological structures may sometimes undergo reorganization to adhere to the predominant phonological patterns in the input (Demuth, 1995, 1996). In English, for instance, the predominant metrical structure of words is trochaic (strong + weak; cf. Kager, 1989; Selkirk, 1984), so young children often truncate words with a weak-strong-weak prosodic pattern, such as "banana," to "Nana" (strong + weak; Gerken, 1994), or even shift stress to the initial syllable of a weak-strong word (e.g., *guiTAR* to *GUITar*; cf. Demuth, 1995). This might also be the case for tone sandhi in Mandarin Chinese. If the left-branching structure is predominant in children's language input, children might prefer this pattern and use its surface tones for both left- and right-branching items. We therefore calculated the frequency distribution of left-branching versus right-branching trisyllabic words using data from caregiver child-directed speech from the Chang Corpus (Chang, 1998) and the Tong Corpus (Deng & Yip, 2018; cf. CHILDES database, MacWhinney, 2000). A total of 2,237 trisyllabic words were found, of which 67% had left-branching structure and 33% had right-branching structure. Again, these trisyllabic words all consisted of a monosyllabic and a disyllabic (lexicalized) word. Thus, it is possible that this predominant prosodic structure in children's language input biased children's parsing of the novel trisyllabic word, leading to the T2T2T3 surface output.

However, the structural frequency distribution itself cannot not fully account for children's T2T2T3 productions for the right-branching item, as the results from Wang (2011) suggest that children at 3 years consistently produced T3T2T3 as the surface output for right-branching compounds such as (*shui3 [lao3 hu3]*) *water-tiger*. Therefore, we proposed that, similar to adults, the novel disyllabic component in our trisyllabic compounds and the monomorphemic status of the Adj + N sequence might help account for children's T2T2T3 realizations. The present study used three independent monosyllabic words to generate novel trisyllabic compounds (e.g., [*zi3 {ma3 gu3}*] *purple horse-drum*) in which the disyllabic component "*ma3 gu3*" *horse-drum* does not coincide with a lexicalized word, and children might have used the preferred/more frequent surface patterns, that is, T2T2T3. Moreover, as mentioned earlier, the monomorphemic status of the Adj+N sequence may have biased children toward the structure of "purple-horse," especially for these unfamiliar novel items. In contrast, the right-branching items such as (*shui3 [lao3 hu3]*) *water-tiger* used in Wang (2011) consisted of two familiar lexicalized items, where children then produced the expected T3T2T3 surface pattern.

Our results therefore indicate that Mandarin Chinese-speaking children (and some adults) rely on lexical knowledge, morphological structure, and structural frequency to guide their application of the tone sandhi rule. When the word is a familiar lexicalized item, and its syntactic structure aligns with the prosodic structure, 3-year-olds apply the tone sandhi process with the expected surface pattern (see Wang, 2011). However, when unfamiliar novel words are involved and the internal structure differs from the most common morphological and prosodic structure, children might prefer to use the most frequent structure. This suggests that the lexical knowledge may facilitate children's learning of both prosodic structure and phonological processes. Given that there was only one T3T3T3 test item for each structure in the current study, it would be useful in future studies to include more test items to probe these issues further, exploring within- and between-speaker variable application of tone sandhi as a function of lexical, morphological, and prosodic structure, for both children and adults.

Conclusion

This study examined the acquisition of a complex phonological process in Mandarin Chinese, the tone sandhi process. The results showed that 3-year-olds were able to acquire the tone sandhi process and productively apply it to novel disyllabic compounds in different tonal contexts. However, even 5-year-olds still differed from adults in applying tone sandhi processes in novel trisyllabic compounds, where the lexical knowledge and structural frequency in the input appear to influence children's application of the tone sandhi rule. This raises many questions about when and how these tonal alternations, and their variable realizations, are learned, and the factors that influence the establishment of adultlike tonal representations more generally.

Acknowledgments. We thank the Child Language Lab, the Phonetics Lab, and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback, and support. We thank Xin Cheng for helping with the reliability check. We also acknowledge the help from the Affiliated Kindergarten of Beijing Language and Culture University with data collection. This research was supported, in part, by a Macquarie University iMQRES scholarship to the first author, the Fundamental Research Funds for the Central Universities No. 30919011252, and the following grants: ARC CE110001021 and ARC FL130100014. The equipment was funded by MQSIS 9201501719.

NOTES

1. One of the anonymous reviewers points out that there is only one test item in the critical initial position, that is, purple. However, it is difficult to find multiple T3 test items for this position that 3-year-olds are familiar with. Adjectives more frequently appear in this position compared to nouns, especially for trisyllabic words. "Purple" is a T3 adjective, and this is highly frequent in children's language input, based on the Mandarin corpora (the Chang Corpus; Chang, 1998, and the Tong Corpus; Deng & Yip, 2018) from the CHILDES database. Other highly frequent T3 adjectives such as "xiao3" small would induce a relative contrast effect, requiring two items to be presented on the screen, complicating the display/procedure.
2. Note that the surface variation in adults' right-branching productions was also unlikely to be driven by the experimental order in which trials were presented, as the presentation order for the left-branching and right-branching forms was counterbalanced across participants, and there was no correlation between order of presentation and single versus recursive tone sandhi application.

References

- Albright, A., & Hayes, B. (2011). Learning and learnability in phonology. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of phonological theory* (2nd ed., pp. 661–669). Malden, MA: Wiley-Blackwell.
- 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*, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. [Computer software]. R package version 1.7. <https://CRAN.R-project.org/package=lme4>
- Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*, 97–111.
- Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. [Computer program]. Version 6.19. Retrieved from <http://www.praat.org>
- Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese-speaking children's play narratives. *Narrative Inquiry*, *8*, 77–111.
- Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge: Cambridge University Press.
- Demuth, K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language*, *20*, 275–301.
- Demuth, K. (1995). Markedness and the development of prosodic structure. In J. N. Beckman (Ed.), *Proceedings of the North East Linguistic Society* (Vol. 25, pp. 13–25). Amherst, MA: University of Massachusetts, Graduate Linguistic Student Association.
- 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. 171–184). Mahwah, NJ: Erlbaum.
- Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: The Tong corpus. *Journal of Chinese Linguistics*, *46*, 69–92.
- Duanmu, S. (2012). Word-length preferences in Chinese: A corpus study. *Journal of East Asian Linguistics*, *21*, 89–114.
- Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, *21*, 565–584.
- Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, *27*, 3–42.
- Kager, R. (1989). *A metrical theory of stressing and destressing in English and Dutch*. Dordrecht: Foris Publications.
- Kazasis, K. (1969). Possible evidence for (near-) underlying forms in the speech of a child. *Proceedings of the Chicago Linguistic Society* (Vol. 5, pp. 382–388). Chicago: Chicago Linguistic Society.
- Kerkhoff, A. O. (2007). *Acquisition of morpho-phonology: The Dutch voicing alternation* (Unpublished doctoral dissertation, LOT).
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package "lmerTest." [Computer software]. R package version 2.0. <http://CRAN.R-project.org/package=lmerTest>
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*, 1–33.
- Levelt, C. C., Schiller, N. O., & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition*, *8*, 237–264.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*, 185–199.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behaviour Research Methods*, *49*, 1494–1502.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.
- McEnery, A., & Xiao, Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. *Religion*, *17*, 3–4.
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman & Hall/CRC.
- 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. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

- Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, **20**, 81–124.
- Stites, J., Demuth, K., & Kirk, C. (2004). Markedness vs. frequency effects in coda acquisition. In *Proceedings of the 28th annual Boston University conference on language development* (Vol. 2, pp. 565–576). Somerville, MA: Cascadilla Press
- Skoruppa, K., Mani, N., & Peperkamp, S. (2013). Toddlers' processing of phonological alternations: Early compensation for assimilation in English and French. *Child development*, **84**, 313–330.
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *Journal of the Acoustical Society of America*, **142**, 2823–2835.
- van de Vijver, R., & Baer-Henney, D. (2013). On the development of the productivity of plural suffixes in German. *Boston University Law Review*, **37**, 444–455.
- Wang, C. Y. (2011). *Children's acquisition of tone 3 sandhi in Mandarin* (Unpublished doctoral dissertation Michigan State University).
- Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, **7**, 1–14.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, **25**, 61–83.
- Xu, Z. (2018). The word status of Chinese adjective-noun combinations. *Linguistics*, **56**, 207–256.
- Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology*, **9**, 817.
- Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language*, **31**, 515–536.
- Zamuner, T. S., Kerkhoff, A., & Fikkert, J. P. M. (2006). Acquisition of voicing neutralization and alternations in Dutch. In D. Bammann, T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Boston University Conference on Language Development* (pp. 701–712). Somerville, MA: Cascadilla Press.
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, **27**, 153–201.

Appendix A

Table A.1. Results of linear mixed regression model with second-order polynomials on pitch points of lexical tones across age groups and tones

Pitch parameter	Factor	df 1	df 2	F	p
Intercept	Group	3	10610	11.04	<.001***
	Tone	3	10610	227.03	<.001***
	Group × Tone	9	10610	12.57	<.001***
Slope	Group	3	204	0.53	.659
	Tone	3	10502	352.98	<.001***
	Group × Tone	9	10505	1.64	.100
Curvature	Group	3	175	1.65	.180
	Tone	3	10492	417.61	<.001***
	Group × Tone	9	10494	0.85	.573

Note: Items in bold indicate significant findings. R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Tone + (1 + Linear trend + Quadratic trend | Participant). **p* < .05. ***p* < .01. ****p* < .001.

Appendix B

Table B.1. Results of pairwise comparison on the pitch shape (slope and curvature) for lexical tone contrasts

Tonal contrast	Pitch parameter	β	SE	df	t	p
T1–T2	Slope	-0.95	0.11	10504	-8.50	<.001***
	Curvature	-0.65	0.11	10500	-5.79	<.001***
T1–T3	Slope	-0.14	0.09	10509	-1.62	.367
	Curvature	-1.99	0.09	10504	-22.89	<.001***
T1–T4	Slope	2.37	0.11	10509	21.24	<.001***
	Curvature	0.72	0.11	10505	6.48	<.001***
T2–T3	Slope	0.81	0.09	10510	9.35	<.001***
	Curvature	-1.34	0.09	10505	-15.50	<.001***
T2–T4	Slope	3.32	0.11	10500	29.82	<.001***
	Curvature	1.37	0.11	10497	12.31	<.001***
T3–T4	Slope	2.51	0.09	10496	29.11	<.001***
	Curvature	2.71	0.09	10494	31.42	<.001***

Note: Items in bold indicate significant findings. * $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix C

Table C.1. Results of linear mixed regression model with second-order polynomials on pitch points of the tone sandhi syllable (the first syllable: T3) in disyllabic T3TX words across age groups and contexts

Pitch parameter	Factor	df 1	df 2	F	p
Intercept	Group	3	142	1.21	.303
	Context	1	22084	5871.83	<.001***
	Group × Context	3	22083	6.42	<.001***
Slope	Group	3	146	0.11	.981
	Context	1	22084	5095.09	<.001***
	Group × Context	3	22084	2.12	.092
Curvature	Group	3	249	0.63	.591
	Context	1	22082	19.62	<.001***
	Group × Context	3	22082	0.61	.633

Note: R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Context + (1 + Linear trend + Quadratic trend | Participant). Items in bold indicate significant findings. * $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix D

Table D.1. Results of the linear mixed regression model with second-order polynomials on pitch points of the tone sandhi syllable (the first syllable: T3) in trisyllabic T3T3T1/2/4 words across age groups and structures

Pitch parameter	Factor	<i>df</i> 1	<i>df</i> 2	<i>F</i>	<i>p</i>
Intercept	Group	3	137	84.61	<.001***
	Structure	1	7600	92.81	<.001***
	Group × Structure	3	7601	9.22	<.001***
Slope	Group	3	135	1.70	.171
	Structure	1	7639	0.97	.323
	Group × Structure	3	7641	2.32	.072
Curvature	Group	3	284	1.37	.251
	Structure	1	7593	0.11	.740
	Group × Structure	3	7592	0.44	.732

Note: R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group * Structure + (1 + Linear trend + Quadratic trend | Participant). Items in bold indicate significant findings. **p* < .05. ***p* < .01. ****p* < .001.

Appendix E

Table E.1. Results of the linear mixed regression model with second-order polynomials on pitch points of the tone sandhi syllables (the first two syllables: T3T3) in left-branching ([T3T3] T3) words across groups (3-, 4-, and 5-year-olds and adults)

Syllable	Pitch parameter	<i>df</i> 1	<i>df</i> 2	<i>F</i>	<i>p</i>
Syllable 1	Intercept	3	124	0.37	.773
	Slope	3	1144	1.08	.358
	Curvature	3	1144	1.72	.161
Syllable 2	Intercept	3	124	2.03	.113
	Slope	3	1144	1.24	.294
	Curvature	3	1144	2.33	.073

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

Appendix F

Table F.1. Results of the linear mixed regression model with second-order polynomials on pitch points of the tone sandhi syllables (the first two syllables: T3T3) in right-branching (T3 [T3T3]) words across groups with different surface tones (3-, 4-, and 5-year-olds, adults_ **T2T2T3**, and adults_ **T3T2T3**)

Syllable	Pitch parameter	df 1	df 2	F	p
Syllable 1	Intercept	4	128	15.53	<.001***
	Slope	4	1187	13.42	<.001***
	Curvature	4	1187	1.06	.374
Syllable 2	Intercept	4	128	3.23	<.05*
	Slope	4	1187	112.53	<.001***
	Curvature	4	1187	0.20	.937

Note: R code for this model: Pitch ~ (Linear trend + Quadratic trend) * Group + (1 + Linear trend + Quadratic trend | Participant). Items in bold indicate significant findings. * $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix G

Table G.1. Results of pairwise comparison on the pitch slope for right-branching (T3 [T3T3]) productions from children and adults (with T3**T2T3** and **T2T2T3** perceived surface tones)

Group difference	Syllable	β	SE	df	t	p
3 years–4 years	Syllable 1	0.13	0.06	1187	2.09	.225
	Syllable 2	-0.09	0.07	1187	-1.35	.663
3 years–5 years	Syllable 1	0.01	0.07	1187	0.13	1.000
	Syllable 2	0.12	0.08	1187	1.49	.568
3 years–adults_ T2T2T3	Syllable 1	0.26	0.08	1187	1.17	.121
	Syllable 2	-0.21	0.09	1187	-2.39	.119
3 years–adults_ T3T2T3	Syllable 1	1.63	0.08	1187	21.17	<.001***
	Syllable 2	-1.59	0.08	1187	-19.11	<.001***
4 years–5 years	Syllable 1	-0.12	0.07	1187	-1.61	.494
	Syllable 2	0.21	0.08	1187	2.60	.070
4 years–adults_ T2T2T3	Syllable 1	0.13	0.08	1187	1.61	.490
	Syllable 2	-0.12	0.09	1187	-1.38	.638
4 years–adults_ T3T2T3	Syllable 1	1.51	0.08	1187	19.44	<.001***
	Syllable 2	-1.50	0.08	1187	-17.97	<.001***
5 years–adults_ T2T2T3	Syllable 1	0.25	0.09	1187	1.73	.042
	Syllable 2	-0.33	0.10	1187	-1.33	.092

(Continued)

Table G.1. (Continued)

Group difference	Syllable	β	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
5 years –adults_T3T2T3	Syllable 1	1.63	0.09	1187	18.60	<.001***
	Syllable 2	-1.71	0.09	1187	-18.13	<.001***
Adults_T2T2T3–adults_T3T2T3	Syllable 1	1.37	0.10	1187	14.44	<.001***
	Syllable 2	-1.38	0.10	1187	-13.44	<.001***

Notes: Items in bold indicate significant findings. * $p < .05$. ** $p < .01$. *** $p < .001$.