3-year-olds Produce Pitch Contours Consistent with Mandarin Tone 3 Sandhi

Nan Xu Rattanasone[1], Ping Tang[1], Ivan Yuen[1], Liqun Gao[2], & Katherine Demuth[1]

[1]Department of Linguistics, Center for Language Sciences, Macquarie University, Australia
[2]Beijing Language and Culture University, China

nan.xu@mq.edu.au; ping.tang1@students.mq.edu.au; ivan.yuen@mq.edu.au; gaolq@blcu.edu.cn; Katherine.demuth@mq.edu.au

Abstract

Few studies have examined the acoustic properties of children's tonal productions. This study examined the productions of lexical tone and Tone 3 sandhi from 27 Mandarin-speaking 3-year-olds. Mandarin is a language with four lexical tones, T1 high level, T2 rising, T3 dipping and T4 falling, and both Full and Half Tone 3 sandhi processes. The results showed that 3-year-olds are producing all four lexical tones and tone sandhi, at least for lexicalized forms.

Index Terms: Lexical tone acquisition, tone sandhi, Mandarin

1. Introduction

Languages that have lexical tone make use of manipulations in pitch height and pitch contours to change the meanings of words. Whereas in English rising and falling pitch contours over a single word are typically associated with prosodic information such as focus, in lexical tone languages these changes can alter the meaning of the word. Mandarin is one of the best studied lexical tone languages, with the largest number of speakers. However, despite the pervasiveness of tone languages, relatively little is known about the acquisition of tones compared to other phonological contrasts (i.e., vowels and consonants). Even less is known about the acquisition of tone sandhi, where the surface tone changes depending on the tone context, e.g., Mandarin Tone 3 sandhi.

Mandarin has a four-tone system with one level and three contour tones; Tone 1 (T1) is High Level ('mā': *mother*), Tone 2 (T2) is Rising ('má': *hemp*), Tone 3 (T3) is Dipping ('mǎ': *horse*), and Tone 4 (T4) is Falling ('mà': *reprimand*). While all four tones appear in the productions of Mandarin-speaking children by the 1-word stage of development, confusion between Tones 2 and 3 (Rising and Dipping tones) continues into the 2/3-word stage of development, finally disappearing as longer sentences are produced[1]. Compared to other tone languages, acquisition of the Mandarin tonal system appears to be more protracted[2]. For example, for Cantonese, a language with a six-tone system, children have acquired all tones by the age of 2[3]. This includes tones with very similar pitch contours, e.g., three level tones (high, mid & low) and two rising tones (high & mid). The same early acquisition of the full tone inventory is also observed in Thai, a language with 5 tones, where all tones were present by the 2-word stage[4]. Thai also has three level tones (High, Mid, & Low). Thus, the similarities between pitch contours and larger tone inventory does not appear to delay tone acquisition.

The delayed mastery of the tonal system in Mandarin might be related to tone change processes which do not occur in Cantonese or Thai. Mandarin Tone 3 sandhi is a tone change process that occurs when the underlying tone changes to different surface forms in various tone contexts. When two T3 occur in succession (T3-T3), the first becomes a rising tone and is called the *full sandhi*, but when T3 is followed by any other tone, it falls in pitch; this is called the *half sandhi*. Previous studies have reported that Tone 3 sandhi begins to be acquired by the 2/3-word stage of development, with few errors[1]. It is after this point that the confusion between T2 and T3 is resolved. One reason for this might be that in the *full sandhi* form T3 has a rising contour like T2, and children need to first learn the sandhi rule before understanding the difference between lexical T2 and Tone 3 sandhi. In Bantu languages such as Sesotho, where lexical and grammatical tone interact, tone sandhi processes are acquired only by 3 years or later, as children learn more about the grammar of the language[5].

Acquisition studies typically report only briefly on Tone 3 sandhi, with results typically based on auditory impressions of the children's productions. One study reported that tone sandhi emerges early, at around 2 years[6], but studies have not examined the acoustic realization of children's tone sandhi productions.

To examine this issue, productions of lexical tone and contexts where tone sandhi should occur were collected from Mandarin monolingual 3-year-olds using an elicitation task. Since so little is known about the acoustic realization of children's use of tone sandhi, real words were used, testing lexicalized forms of Tone 3 sandhi.

2. Methods

2.1. Participants and design

A total of 27 3-year-olds (13 boys; 14 girls) participated in the study. The mean age of the children was 3;10 (range 3;6 – 3;11). They were recruited in Beijing from the preschool associated with the Beijing Language and Culture University. The study was conducted in accordance with the ethics protocol approved by Macquarie University's Human Ethics Panel. All child participants received fun stickers for their participation and the preschool received book donations for all children to use at the center.

A within-subjects design was used, and all children were tested on all target items.

2.2. Stimuli

A total of 34 high-frequency disyllabic words were used (Table 1). To elicit the lexical tones, 12 disyllabic words with T1, T2, and T4 as the first syllable and T1 to T4 as the second syllable were chosen. It was not possible to find enough words all beginning with T1 to avoid tone co-articulation effects, so a variety of tone contexts were used. It was also not possible to

avoid words ending in nasal /n/ and /ŋ/ codas, which can have the effect of lowering the pitch of the syllable.

To elicit Full sandhi, five disyllabic T3-T3 words were chosen. For Half sandhi five words, each with T3 as the first syllable and T1, T2, and T4 as the second syllable were chosen, yielding 20 stimulus words for tone sandhi contexts. An additional two practice items in the form of T3-T3 (a puppy and a pony) were used at the beginning of each session but were not analyzed.

Most syllables were CV in form, and where possible contained a stop or fricative/affricate onset to facilitate acoustic coding. However, a few contained a lateral or nasal onset, and some contained a nasal coda. Two versions of the test were created, each with a different randomization for the order of presentation of the words.

Table 1. *List of Disyllabic Stimuli Words*

	Tones	Pinyin	Meaning
Practice	T3T3	xiao-gou	puppy
	T3T3	xiao-ma	pony
Full	T3T3	lao-shu	mice
sandhi	T3T3	lao-hu	tiger
	T3T3	xiao-niao	chick (bird)
	T3T3	yu-san	umbrella
Half	T3T1	xiao-mao	kitten
sandhi	T3T1	jian-dao	scissors
	T3T1	kao-ya	Peking duck
	T3T1	yu-yi	raincoat
	T3T1	bing-gan	biscuit
	T3T2	er-huan	earing
	T3T2	kou-hong	lipstick
	T3T2	cai-hong	rainbow
	T3T2	cao-mei	strawberry
	T3T2	xiao-niu	calf
	T3T4	li-wu	present
	T3T4	shou-tao	gloves
	T3T4	tu-dou	potatoes
	T3T4	kong-que	peacock
	T3T4	tan-ke	tanker
Lexical	T1T1	xi-gua	watermelon
Tone	T1T2	ying-tao	cherries
	T1T3	ban-ma	zebra
	T1T4	ji-dan	egg
	T2T1	long-xia	lobster
	T2T2	liang-xie	sandals
	T2T3	ping-guo	apple
	T2T4	qin-cai	celery
	T4T1	li-zhi	lychee
	T4T2	qi-qiu	balloon
	T4T3	chi-bang	wings
	T4T4	da-xiang	elephant

2.3. Materials

A total of 34 non-proprietary photographic images representing each of the 32 test and 2 practice items were selected from google images. The images were presented one at a time using Microsoft PowerPoint 2013 delivered on an Apple iPad 2. The recordings were collected using a Zoom H2 digital voice recorder with lapel mic and the recordings were exported as PCM files.

2.4. Procedure

Testing was conducted in a quiet area in the preschool. Each child was greeted by the native Mandarin-speaking experimenter. The task was explained as a picture naming game where children named the pictures on the screen and receive stickers for playing the game. Two practice trials were given and for children who could not provide an answer after three prompts, the experimenter provided the answer, e.g., "puppy". The child was then asked to repeat the label before moving to the next item. The children were encouraged to provide answers independently during the practice trials. All children were able to perform the elicitation task, however, there were two items where a majority of the children could not identify, i.e., T3-T2 *rainbow* and T2-T1 *lobster*. For these items, the experimenter produced the items but the imitations from the children were not analyzed.

2.5. Data Analysis

The productions were acoustically coded in Praat[7] by a trained coder who is a native speaker of Mandarin. The tones were extracted from the vocalic portion of the syllable (and nasal if present). The vocalic portion was identified from the onset of higher formants to the cessation of higher formants in the first syllable and offset of voicing as indicated by the onset of the second syllable. In cases where the second syllable had a nasal onset, anti-resonance and simplification of the waveform was used to identify the onset of the second syllable. Using Praat[7], the average F0 for each syllable were extracted in 10 equal steps.

3. Results

To determine that 3-year-olds are producing lexical tones, analysis of the F0 changes over time were conducted. If children are specifying the four tones, there should be significant linear (rising or falling) or quadratic (dipping) trends emerging over time for all four tones. For Tone 3 sandhi, F0 changes were analyzed separately for Full and Half sandhi contexts. Analysis was conducted over the first syllable only where tone sandhi was expected to manifest. If children are producing Tone 3 sandhi then F0 changes should be different across these sandhi contexts with linear trends (rising vs. falling) for Full and Half sandhi.

3.1. Lexical Tone

A linear mixed effects regression model (LMEM) was conducted with F0 as the dependent variable. Polynomial equations (up to quadratic) were fitted for the 10 Time points with Tone type (T1 to T4) as the other fixed factor. Random intercepts for each child and item were fitted for the effect of tone (F0 ~ poly(Time, 2)*Tone + (1 | Child) + (1 | Word)). See results on Table 2 and Figure 1.

The results show a main effect of Tone type. Mean F0 for T2 (t = -5.991, p < .01) and T3 (t = -13.164, p < .01) are significantly lower than T1 (the referent category). The interactions show a significant linear trend for T2 (t = 2.832, p = .005) and a positive effect on the intercept, suggesting a linear increase over time (rising contour). There were significant linear (t = -5.717, p < .01) and quadratic (t = 7.013, p < .01) trends for T3. The negative effect on the intercept for the linear

trend and positive effect for the quadratic trend suggest that T3 decreased before rising over time (dipping contour). The significant linear trend for T4 (t = -7.274, p < .01), and the negative effect on the intercept suggest a linear decrease over time (falling contour). These F0 changes over time (rising for T2, dipping for T3 and falling for T4) are consistent with the expected tone contours for each tone.

Table 2: *Results for F0 of lexical tone across 10 time points* (Kenward-Roger approximations were made for degrees of freedom; **p < .01)

	Estimate	Std. Error	t value	p (KR)	
(Intercept)	286.736	9.211	31.130	0.000	
Main Effects					
poly1	24.047	146.413	0.164	0.870	
poly2	63.949	122.340	0.523	0.601	
T2	-43.331	7.232	-5.991	0.000	**
T3	-95.168	7.229	-13.164	0.000	**
T4	-0.774	7.229	-0.107	0.915	
Interactions					
poly1:T2	448.293	158.306	2.832	0.005	**
poly2:T2	299.838	158.280	1.894	0.058	
poly1:T3	-902.687	157.886	-5.717	0.000	**
poly2:T3	1107.326	157.886	7.013	0.000	**
poly1:T4	-1148.486	157.886	-7.274	0.000	**
poly2:T4	115.326	157.886	0.730	0.465	

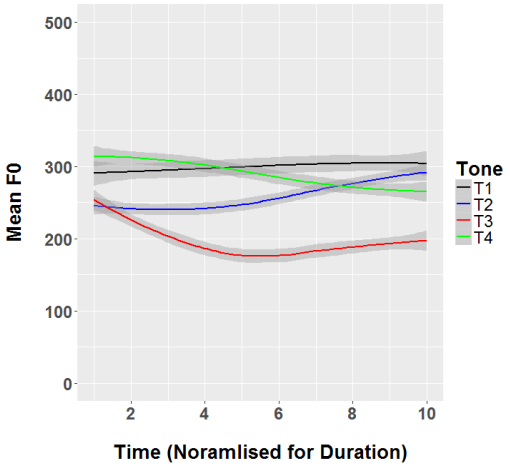

Figure 1. *Mean F0 at 10 time points for the four lexical tones*

3.2. Tone 3 Sandhi

To examine whether 3-year-olds are distinguishing between Full and Half sandhi, two LMEMs estimating polynomial trends over time were conducted separately for Full and Half sandhi contexts. The same model and factors were used as above after removing Lexical Tone as a factor. See results on Table 3 and Figure 2.

The results show significant linear and quadratic trends for both Full and Half sandhi contexts. For Full sandhi, the positive effect on the intercept for both the linear (t = 7.227, p < .01) and quadratic (t = 3.803, p = .001) trends suggest a steep rise, then slowing over time. For Half sandhi, the negative effect on the intercept for the linear trend (t = -12.729, p < .01) and positive effect for the quadratic trend (t = 5.913, p < .01) suggest a falling then slight rising (dipping) contour over time. While the rising contour is consistent with Full sandhi, the dipping contour is not consistent with Half sandhi which should have a falling contour.

Table 3. *Results for F0 of full and half sandhi contexts across 10 time points* (**p < .01)

	Estimate	Std. Error	t value	p (KR)
Full Sandhi				
(Intercept)	282.895	10.316	27.423	0.000
poly1	529.127	73.217	7.227	0.000**
poly2	160.272	42.148	3.803	0.001**
Half Sandhi				
(Intercept)	236.665	6.221	38.041	0.000
poly1	-627.542	49.301	-12.729	0.000**
poly2	244.237	41.305	5.913	0.000**

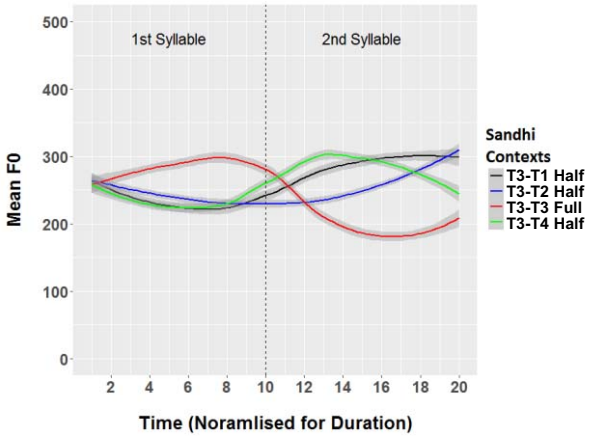

Figure 2. *Mean F0 at 20 time points for disyllabic word contexts eliciting full and half sandhi*

4. Discussion

The aim of this study was to examine the acoustic realizations for lexical tones and Tone 3 sandhi contexts in the productions of 3-year-olds. First, the analysis of lexical tone production suggested that children are producing contours consistent with that of the four lexical tones, i.e., level for T1, rising for T2, dipping for T3 and falling for T4. Secondly, the analysis of Tone 3 sandhi contexts suggested that they are also producing sandhi forms in these high frequency known words: the *full sandhi* had a rising contour while the *half sandhi* had a dipping contour. The *full sandhi* is consistent with the expected contour, however the results for *half sandhi* suggested that 3-year-olds might be producing the underlying lexical T3. Overall, the results suggest that, in addition to having a well specified lexical tone space, 3-year-olds are also producing *Full sandhi*.

These results raise several questions for future research, such as how adult-like the children's productions are, not only in terms of the contours but also on other measures. For example, T3 is typically associated with creaky voice in adult productions but it is unclear if children also use this cue. Adults also use a variety of acoustic cues to signal lexical tones, including differences in pitch onset, offset, range and turning points. Knowing whether children are using similar cues can better inform our understanding of what children are encoding when learning the tone space. There is also considerable individual variation (Figure 3), showing that some children have good tonal representations and tone sandhi productions

(child 11 & 12) while others show poor tone separation and sandhi production (child 3, 16 & 22).

5. Conclusions

Mandarin-speaking 3-year-olds can produce Tone 3 sandhi in *full sandhi* contexts on familiar/known lexicalized items. Future studies should examine children's ability to apply tone change processes to novel words.

6. Acknowledgements

This work is funded by the Macquarie University Research and Development Grant [1547620] awarded to the first author and the ARC Laureate Fellowship [FL130100014] awarded to the last author. We thank Dr. Ivan Yuen, members of the Child Language Lab, and the two anonymous reviewers for their helpful comments.

7. References

[1] Li, C. N., & Thompson, S. A., "The acquisition of tone in Mandarin-speaking children", Journal of Child Language, 4(02): 185–199, 1977.

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

[3] So, L. K. H., & Dodd, B. J., "The acquisition of phonology by Cantonese-speaking children", Journal of Child Language, 22: 473–495, 1995.

[4] Tuaycharon, P., "The phonetic and phonological development of a Thai baby: From early communicative interaction to speech", Unpublished PhD Thesis, University of London, 1977.

[5] Demuth, K., "Issues in the acquisition of the Sesotho tonal system". Journal of Child Language, 20: 275-301, 1993.

[6] Hua, Z., & Dodd, B., "The phonological acquisition of Putonghua (Modern Standard Chinese)", Journal of Child Language, 27: 3–42, 2000.

[7] Boersma, P., & Weenink, D., "Praat: doing phonetics by computer [Computer program]", Version 6.0.18, retrieved 23 May 2016 from http://www.praat.org/.

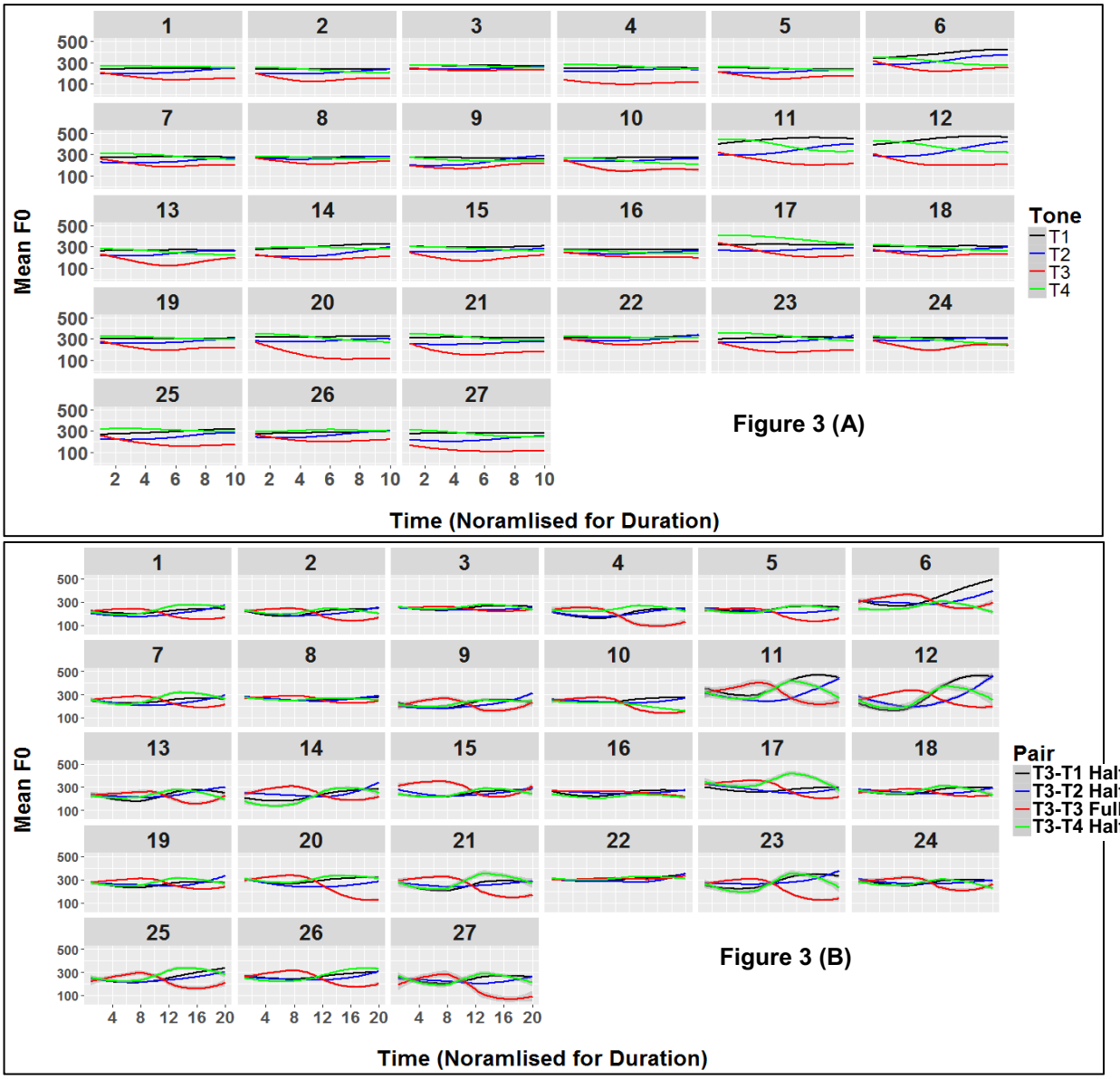

Figure 3. (A) *Mean F0 at 10 time points of lexical tones for each child*, (B) *Mean F0 at 20 time points for disyllabic words for sandhi contexts for each child*