

# Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan

Puisan Wong<sup>a)</sup>

Department of Otolaryngology—Head and Neck Surgery, Wexner Medical Center, The Ohio State University, 915 Olentangy River Road, Columbus, Ohio 43212

(Received 3 March 2012; revised 30 October 2012; accepted 6 November 2012)

This study used the same methodology in Wong [J. Speech Lang. Hear. Res. **55**, 1423–1437 (2012b)] to examine the perceived accuracy of monosyllabic Mandarin tones produced by 4- and 5-year-old Mandarin-speaking children growing up in Taiwan and combined the findings with those of 3-year-olds reported in Wong [J. Speech Lang. Hear. Res. **55**, 1423–1437 (2012b)] to track the development of monosyllabic tone production in preschool children. Tone productions of adults and children were collected in a picture naming task and low-pass filtered to remove lexical information and reserve tone information. Five native-speakers categorized the target tones in the filtered productions. Children's tone accuracy was compared to adults' to determine mastery and developmental changes. The results showed that preschool children in Taiwan have not fully mastered the production of monosyllabic Mandarin tones. None of the tones produced by the children in the three age groups reached adult-like accuracy. Little developmental change was found in children's tone accuracy during the preschool years. A similar order of accuracy of the tones was observed across the three age groups and the order appeared to follow the order of articulatory complexity in producing the tones. The findings suggest a protracted course of development in children's acquisition of Mandarin tones and that tone development may be constrained by physiological factors. © 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4768883]

PACS number(s): 43.70.Ep, 43.70.Kv [BRM]

Pages: 434–443

## I. INTRODUCTION

Large discrepancies have been reported on children's age of acquisition of lexical tones (hereafter tones), a linguistic feature used in a majority of the languages in the world for lexical contrast (Fromkin, 1978; Yip, 2002). Most studies suggested very early mastery of tones and reported that children produced tones correctly in various contexts at around 2 years of age well before the mastery of consonants and vowels (e.g., Tse, 1978, for Cantonese tones; Tuaycharoen, 1977, for Thai tones; and Clumeck, 1980; Hua and Dodd, 2000; Hua, 2002; Li and Thompson, 1977, for Mandarin tones). A few recent studies, however, suggested a much more protracted course of tone acquisition and reported that 3-year-old children growing up in Taiwan and the U.S. learning Mandarin as a first language had not produced the four Mandarin tones in monosyllabic words with adult-like accuracy (Wong *et al.*, 2005; Wong, 2012a,b). Thus, the age of acquisition of tones and the course of tone development remain unclear. This study used the same methods adopted in Wong *et al.* (2005) and Wong (2012b) to examine the perceived accuracy of the monosyllabic tones produced by 4- and 5-year-old children and combined the findings with those in Wong (2012b) with 3-year-olds to track the tone development in preschool children learning Mandarin as their first language in Taiwan.

Like other tone languages, Mandarin uses pitch or fundamental frequency (F0) to convey semantic contrasts. There

are four phonemically distinctive tones in fully stressed syllables in Mandarin. Tone 1 (T1) has a high and level F0 contour, Tone 2 (T2) has a rising contour, and Tone 4 (T4) has a falling F0 contour. The F0 contour of T3 varies depending on context. It is typically a dipping (falling-rising) tone in isolated monosyllables and a low falling tone in nonfinal position (Chao, 1968; Xu, 1997), though some Mandarin speakers in Taiwan produce T3 as a low falling tone in both contexts (Sander, 2008). At utterance final position, T3 is either a dipping or a low falling tone (Duanmu, 2007). When it is produced in connected speech, T3 changes into T2 when it is followed by another T3.

Physiologically, the production of Mandarin tones is primarily achieved by the contraction and relaxation of the laryngeal muscles to adjust the tension and mass per unit of the vocal folds, which in turn change the rate of vibration of the vocal folds and result in a change of F0 (see Wong, 2012a, for a review). Several EMG studies that examined specifically the neuromuscular activities in producing Mandarin tones showed that the cricothyroid (CT) muscle is the primary muscle for F0 rising while the sternohyoid (SH) muscle is for F0 lowering (Hallé, 1994; Sagart *et al.*, 1986).

Different degrees of articulatory control are required for producing the four Mandarin tones. The falling T4 contour appears to require the least laryngeal control for speakers with high F0 because the production of T4 involves mostly an increase in activities of the CT muscle at the beginning of the tone to reach a high F0 followed by a relaxation of the CT muscle to create the F0 fall (Hallé, 1994; Sagart *et al.*, 1986). For speakers with low pitch ranges, the SH muscle is activated at the end of T4 to bring the F0 to a low pitch. The

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: pswresearch@gmail.com

production of T1 involves more active control of the laryngeal muscles than the production of T4. At tonal onset, there is an increase in the activity of the CT muscle to reach a high F0. Then a moderately high level of activity of the CT muscle is sustained to maintain the high F0 throughout the syllable. The production of T2 and T3 involves the coordination of the CT and SH muscles. For T2, the SH muscle is activated to produce the initial fall of the tone to reach a low F0. Then the activity of the CT muscle increases to create a sharp rise in the F0. The production of T3 involves a high activity of the SH muscle at tonal onset to create a sharp F0 fall to reach an F0 even lower than in T2. Then there is a decrease of activity in the SH muscle coupled with an increase of activity of the CT muscle to produce the final rising in the F0 contour. T3 appears to be articulatorily the most complex for children also because it requires the most involvement and control of the SH muscle and previous acoustic data suggested that children have more difficulties producing low F0 targets that involve the use of the more frequently used CT muscle (Wong, 2012a). Thus, it appears that the articulatory control for children to produce the four tones in the order from the least to the most complex is T4, T1, T2, and T3.

Conflicting results have been found in the timeline for children to master the four Mandarin tones. Some studies suggested that the four tones are acquired quickly and early before the age of two. Hua and Dodd (2000), which examined the phonological development in 129 children between 1;6 and 4;6, reported only two tone errors in all the speech samples collected in a picture naming task and a picture description task. The findings of Hua (2002), which reported longitudinal data of four children between one and 2 years of age, supported that the four tones were stabilized before 2 years of age. Chao (1973) examined the phonological acquisition in a girl at 2;4 and reported that the child made few errors with the four tones at the outset of the study.

Two studies reported slightly later acquisition of Mandarin tones. Li and Thompson (1977) conducted a 7-month longitudinal study to follow the tone productions of 17 children from the age of 1;6 to 3;0 and found that children did not produce the four tones accurately until they produced utterances longer than 2–3 words (Li and Thompson, 1977). Clumeck (1980) conducted a longitudinal study to follow the development of Mandarin tones in two Mandarin-learning children and reported that T2 and T3 were not fully mastered by the children at the end of the study when the children were at 2;10 and 3;5 of age (Clumeck, 1980).

Recent studies by Wong and colleagues reported the most lengthy course of Mandarin tone acquisition. Based on perceptual judgments of adult native speakers, Wong *et al.* (2005) and Wong (2012b) showed that 3-year-old Mandarin-speaking children growing up in the U.S. and in Taiwan exhibited similar developmental patterns and did not produce the four tones in familiar monosyllabic words with adult-like accuracy. The acoustic analyses in Wong (2012a) suggested that the order of accuracy of the four tones from the highest to the lowest for 3-year-old children was T4, T1, T2, and T3, following the order of articulatory complexity in producing the tones.

The large discrepancy in the age of acquisition of Mandarin tones in the earlier studies vs the later studies conducted by Wong and colleagues may be accounted for by methodological differences. All earlier studies did not control for lexical expectation in determining children's tone accuracy. Only one adult, who was usually the author, rated children's tone accuracy in naturally produced speech (i.e., unfiltered speech with segmental information available) with lexical and contextual supports. In addition, none of the studies included an adult control/reference group and the criteria for determining tone mastery was not specified in most earlier studies. Only two studies stated the criteria for determining tone stabilization. Hua and Dodd (2000) defined age of stabilization of a tone as the youngest age at which 90% of the children in that age group attained 66.7% accuracy of the tone and Hua (2002) defined age of stabilization as the age at which a child reached 66.7% accuracy in his/her spontaneous productions and continued to attain the same or higher accuracy rate in subsequent sessions. Moreover, most earlier studies did not control for production contexts and determined tone accuracy by collapsing children's tone productions in contexts varying from monosyllabic to multisyllabic productions, from isolated productions to connected speech, and from spontaneous productions to imitations (e.g., Hua and Dodd, 2000; Hua, 2002; Li and Thompson, 1977).

The recent studies by Wong and colleagues employed much tighter experimental control and differed from the earlier studies on Mandarin tone acquisition in several aspects. First, to ensure children's target productions, only tones that were elicited in a picture naming task were used. Second, to control for production contexts, only monosyllabic tones spontaneously produced in isolation were analyzed. Third, to determine mastery, an adult control group was included to compare with children's tone productions. Fourth, perceptual judgments were made by multiple adults who were blind to the experiment. Fifth, to prevent lexical biases and expectation, the judges were asked to determine children's and adults' tones in low-pass filtered speech in which lexical information was eliminated while tone information was preserved.

To obtain a fuller picture of children's acquisition of monosyllabic Mandarin tones, this study employed the more tightly controlled experimental design in Wong *et al.* (2005) and Wong (2012b) and expanded their work on 3-year-old children to examine monosyllabic tone productions by 4- and 5-year old children growing up in Taiwan. By combining the findings in Wong (2012b) and those in the current studies, the investigation sought to answer the following questions: (1) Do preschool children growing up in Taiwan produce monosyllabic tones with adult-like accuracy? (2) What is the order of acquisition/accuracy in the four tones produced by preschool children in Taiwan? (3) How does Taiwan children's tone accuracy improve during preschool years?

## II. METHOD

Tone productions collected from children and adults in a picture naming task were low-pass filtered to eliminate lexical information. Five judges were recruited in Taiwan to make judgments on the tones in the filtered stimuli.

TABLE I. Demographic information and language background of 3-year-old children.<sup>a</sup>

ID#	Age	Attending preschool	Mother education level	Father education level	Mandarin language percentile score	% of time listening to Mandarin	% of time listening to Taiwanese	% of time listening to English	% of time speaking Mandarin	% of time speaking Taiwanese	% of time speaking English
TC301	2;10	Yes	Master's	Ph.D.	39 <sup>d</sup>	90	5	5	100	0	
TC305	3;3	Yes	College	College	59	83	2	15 <sup>b</sup>	95	0	5
TC308	3;4	Yes	Junior high school	Senior high school	65	80	20	5	90	10	0
TC310	3;3	Yes	Junior college	Junior college	63	80	20		95	5	
TC312	3;1	Yes	Junior college	Junior college	37	85	5	10 <sup>c</sup>	95	0	5
TC313	3;5	Yes	College	College	82	90	10		95	5	
TC315	2;10	Yes	Ph.D.	Ph.D.	64	97	1.5	1.5	99	0.5	0.5
TC318	2;10	No	Junior college	Junior college	39	90	10		95	5	
TC319	2;11	No	College	Ph.D.	30	90	5	5	90	5	5
TC320	3;6	Yes	College	College	78	80	20		95	5	
TC321	3;0	No	College	College	93	90	10		95	5	

<sup>a</sup>Source: From Wong (2012b), Table 1. Reprinted with permission.

<sup>b</sup>With an English-speaking domestic helper at home.

<sup>c</sup>From movies and video tapes.

<sup>d</sup>Percentile rank of the total language score in the Language Disordered Scale of Preschoolers (Lin and Lin, 1994).

## A. Speakers

### 1. Child speakers

Thirty-three children growing up in Chiayi, Taiwan participated in the study. Eleven of them were 3-year-olds, (6M, 5F, mean age = 3;1, age range = 2;10 to 3;6), 11 were 4-year-olds (6F 5M, mean age = 4;1, range = 3;7–4;6) and 11 were 5-year-olds (7F 4M, mean age = 5;0, range = 4;7–5;4). Table I, which is a reprint of Table I in Wong (2012b), shows the demographic information of the 3-year-olds. Demographic information of the 4- and 5-year-old children is

presented in Table II. All children were administered a hearing screening, and the Language Disorder Scale of Preschoolers (學前兒童語言障礙評量表) (Lin and Lin, 1994), a Mandarin speech and language test normed in Taiwan. Additionally, a language sample was collected using conversations and storytelling. Parents provided developmental and language background of the children.

Because Taiwanese (Southern Min) is prevalent in Taiwan, to ensure that the children were learning Mandarin as their first language, all children had to meet the following inclusion criteria: (1) children attended Mandarin-speaking

TABLE II. Demographic information and language background of the 4- and 5-year-old children.

ID#	Age	Gender	Mother education level	Father education level	Mandarin language score <sup>a</sup>	% of time listening to Mandarin	% of time listening to Taiwanese	% of time listening to English	% of time speaking Mandarin	% of Time speaking Taiwanese	% of Time speaking English
TC411	3;7	M	University	University	18	90	10		98	2	
TC426	3;7	F	University	University	53	90	9	1	90	9	1
TC410	3;8	M	Junior college	Junior college	67	80	20		95	5	
TC421	3;10	F	University	University	24	80	20		90	10	
TC413	4;0	F	Senior high school	University	53	85	15	5	100		
TC418	4;0	M	Senior high school	Senior high school	63	80	20		90	10	
TC416	4;4	F	Senior high school	Senior high school	31	70	20	10	90	10	
TC423	4;4	M	Senior high school	Junior college	46	85	10	5	80	10	10
TC401	4;6	M	Ph.D.	Ph.D.	56	90	10		99	1	
TC417	4;6	F	Senior high school	Senior high school	35	80	20		90		
TC522	4;6	F	Junior college	University	51	90		10	90		10
TC517	4;7	M	Senior high school	Senior high school	47	85	10	5	90	5	5
TC526	4;7	F	University	University	58	85	10	5	90	5	5
TC546	4;9	M	Junior college	Junior college	56	80	15	5	90	5	5
TC521	4;11	F	University	Junior college	56	90	10		90	10	
TC530	4;11	F	Junior college	Junior college	43	85	10	5	85	10	5
TC536	5;0	M	University	Master's	76	95	5		95	5	
TC545	5;1	F	Junior college	Senior high school	51	95	5		98	2	
TC518	5;3	F	Junior college	University	76	95	3	2	96	2	2
TC533	5;3	F	Junior college	University	58	90	7	3	97	1	2
TC516	5;4	M	Junior college	Master's	48	85	5	10	90	5	5
TC520	5;4	F	University	Junior college	43	90	8	2	94	3	3

<sup>a</sup>Percentile rank of the total language score in the Language Disordered Scale of Preschoolers (Lin and Lin, 1994).

schools, (2) Mandarin was the dominant language in the home, (3) children had limited exposure to Taiwanese and spoke Mandarin over 80% of the time, (4) children attained a percentile rank higher than the 16th percentile (i.e., one standard deviation below the mean) in the Language Disorder Scale of Preschoolers, (5) parents reported no concerns and abnormality in the children's developmental and language history, (6) no limitation or atypicality was observed in the language sample collected, and (7) children passed the hearing screening (see Tables I and II for details).

## 2. Adult speakers

The tones produced by the four mothers reported in Wong *et al.* (2005) and Wong (2012b) were used as the referents.

## B. Stimuli for tone production

The same 24 pictures adopted in Wong *et al.* (2005) and Wong (2012b) depicting 24 monosyllabic words familiar to young children were used. Twelve of the words were singletons without minimal pairs while the other 12 words were in minimal pairs contrasting the four tones. The Appendix shows the word list, which is an adaptation of the Appendix in Wong *et al.* (2005).

## C. Tone collection procedures

Children's tone productions were collected in the child's school or in a quiet room in National Chung Cheng University in Taiwan by the same experimenter who collected the tones in Wong *et al.* (2005) and Wong (2012b). The pictures were presented to the children one by one and simple questions such as “這是什麼 [What is this]” or “他在幹嗎 [What is he doing]?” were used to elicit the productions. The children labeled each picture two times. If the children's productions were unintelligible, too soft, or being interrupted by noise or another voice, the child was asked to repeat his/her production. Children's productions were recorded on a digital recorder using a dynamic microphone. After the picture labeling task, the standardized Mandarin test and the hearing screening were administered, and then a language sample of the child was collected. Adult productions were collected by asking the mothers to label the pictures two times to the child.

## D. Judges

The five Taiwan judges, who rated the tones produced by 3-year-old Taiwan children, reported in Wong (2012b) also judged the tones in this study. Additionally, these judges also rejudged the tones produced by 3-year-old children growing up in the U.S. reported in Wong *et al.* (2005). Wong (2012b) performed in-depth group and pairwise analyses on the interrater reliability of these five Taiwan judges and the ten judges recruited in the U.S. employed in Wong *et al.* (2005). The results showed high interrater reliability between the two groups of judges and between each judge and any of the other 14 judges after corrected for chance. Exclusion of any judge had little effect on the interjudge reliability or degree of agreement of the judges. These findings led the

author to conclude that all judges showed high degree of agreement and interjudge reliability in their tone categorizations and future studies may use significantly fewer judges [see Tables 3 and 4 in Wong (2012b) for details].

The five Taiwan judges (3F, 2M) had a mean age of 20 years (range = 18;3–23;9). All of them were students at National Chung Cheng University in Taiwan and had resided in Taiwan since birth. They all used Mandarin at home and Mandarin was their strongest and dominant language. All judges passed a hearing screening and reported no speech or hearing difficulties or other concerns in a language background questionnaire.

## E. Stimuli for tone judgment

### 1. Training stimuli

The training stimuli were the same as those in Wong *et al.* (2005) and Wong (2012b). Forty-eight (4 tones  $\times$  12 monosyllables) naturally produced unfiltered monosyllabic words were used in two training blocks to familiarize the judges with the procedures for tone judgment. Half of the stimuli were morphemes in Mandarin (e.g., kuan3) while the other half was nonsense syllables with legitimate segmental constructions in Mandarin (e.g., gua2). Each of the two training blocks had 12 morphemes (4 tones  $\times$  3 morphemes) and 12 nonsense words (4 tones  $\times$  3 nonsense monosyllables).

### 2. Experimental stimuli

The experimental stimuli for the judges included 734 usable monosyllabic tones produced by the 33 three- to five-year-old Taiwan children. Usable productions were defined as the spontaneously produced monosyllabic target words produced in isolation by children. Productions of non-target words, productions interrupted by other voices or noises and productions that were too soft or not produced in isolation were all excluded. Only the first two usable productions for each word were included. The word “mao4,” which was excluded in Wong *et al.* (2005) and Wong (2012b) because it was not produced in isolation by the children, was also excluded in this study. Three-year-old children produced 175<sup>1</sup> usable productions (47, 43, 35, and 50 for the four tones, respectively) (Wong, 2012b). Four-year-old children produced a total of 258 usable tokens (62, 75, 60, and 61 productions for the four tones, respectively) and the 5-year-olds produced 301 tokens (86, 83, 67, and 65 productions for the four tones, respectively). In addition, the judges also listened to the 92 tone productions by the four mothers (24, 24, 24, and 20 for T1, T2, T3, and T4, respectively), reported in Wong (2005), and another 348 tones for another study.

Following the procedures in Wong *et al.* (2005) and Wong (2012b), all the experimental stimuli were low-pass filtered such that the pitch information was retained while the lexical information was eliminated. Adult productions were low-pass filtered at 400 Hz while child productions were low-pass filtered at 500 Hz. All productions were normalized to the same root mean square value for intensity and were blocked by speakers.

## F. Procedures for tone judgment

Each judge attended a total of ten sessions in five consecutive days to categorize the filtered tones. Each session was 30 to 45 min long and the two sessions on each day was separated by at least 1 h. On the first day, the judges filled out a language background questionnaire, received a hearing screening and listened to a block of training stimuli. The training stimuli were presented using the same computer program that presented the experimental stimuli (see below). The judges identified the tones of the stimuli by clicking the corresponding button on the computer screen. All judges identified the tones in the training block with high accuracy (mean accuracy = 95%, range = 92%–100%). Then the judges listened to and categorized 151 filtered tones in nine blocks. In the following sessions, the judges categorized 5–7 blocks of stimuli with a total of 166–196 productions. The two training blocks were randomly selected and presented to the judges on every other session to monitor the level of attention in the judges. The judges attained 85% to 94% accuracy in the training blocks in all sessions. In the seventh session, the judges recategorized the tones they had categorized in a previous session to establish intrajudge reliability.

All tone judgment sessions were conducted in a quiet room in National Chung Cheng University. The judges listened to the filtered stimuli through a pair of headphones. A customized computer program (Tagliaferri, 2005) was used to run the experiment. After the judge clicked the start button on the computer screen, the program randomly selected a block of filtered stimuli and presented the stimuli one at a time in a random order. Because listening to filtered tones without lexical support could be more effortful particularly when the production did not resemble any of the four tones, the judges were allowed to listen to the stimuli as many times as needed before making their decision by clicking one of the four buttons that indicated. Tone 1, Tone 2, Tone 3, and Tone 4 on the screen. After the judge had made a decision, the next trial started. After finishing a block of stimuli, the judge could take a break or click the start button to move on to the next block of stimuli. The responses of the judges were automatically recorded.

## III. DATA ANALYSIS

To answer the research questions set forth in the introduction, first, interjudge and intrajudge reliability of the judges was examined. Then, the accuracy rates of the children's and adults' tones were compared to determine whether the children produced the monosyllabic tones with adult-like accuracy. After that, the accuracy rates of children's four tones were compared to determine the order of mastery of the four tones by the children. Finally, correlation analyses were performed to examine the developmental trends in children's tone accuracy.

### A. Interjudge reliability

The five judges were the same judges employed in Wong (2012b). Fleiss's kappas, which adjusted the overall agreement for chance agreement, were used to measure the inter-rater reliability of the five judges. A kappa coefficient above

0.80, between 0.60 and 0.80, between 0.40 and 0.60, between 0.20 and 0.40, and between 0.00 and 0.20, represents, respectively, almost perfect, substantial, moderate, fair, and slight agreement. A kappa value of 0.00 indicates agreement at chance level. The kappa coefficients of the five judges on their judgments of the adults' productions only and child- and adult-productions combined were 0.84 and 0.60, respectively, indicating almost perfect and substantial agreement in the judges, respectively.

### B. Intrajudge reliability

Cohen's kappa, a more conservative measure that controls for the likelihood of chance agreement, was used to determine intrajudge reliability. Like the criteria for Fleiss's kappa, a kappa coefficient between 0–0.2, 0.20–0.40, 0.40–0.60, 0.60–0.80, and 0.80–1 represents slight, fair, moderate, substantial, and almost perfect agreement, respectively. The five judges obtained Cohen's kappa coefficients of 0.63, 0.75, 0.82, 0.78, and 0.64, respectively, for the agreement of their first and second judgment of the 166 tones they rerated in the seventh session, indicating substantial to almost perfect intrajudge agreement on tone categorization.

### C. Perceived accuracy of children's productions

Perceived accuracy of the productions was defined as the judges' correct identification of the target tones. Table III(a), which was adapted from Table 5 in Wong (2012b),<sup>1</sup> shows the response patterns of the judges on the tones produced by the adults and the 3-year-old children growing up in Taiwan. As shown in the shaded cells in the left panel, adults' target tones were perceived by the judges with ceiling accuracy (93%, 96%, 82%, and 94% for the four tones, respectively). The relatively lower perceived accuracy of T3 has been reported in previous studies on adults' tones

TABLE III. Confusion matrices of the tones produced by the Taiwan children and adults.<sup>a</sup>

		(a) Tones Produced by 3-year-old Children and adults							
		Adult (% Identification)				3-year-olds (% Identification)			
Target Tone		T1	T2	T3	T4	T1	T2	T3	T4
T1		93	6	0	1	71	20	6	3
T2		0	96	4	0	10	53	34	3
T3		0	18	82	0	9	26	52	14
T4		2	0	4	94	16	1	16	67

		(b) Tones Produced by 4- and 5-year-old Children							
		4-year-olds (% Identification)				5-year-olds (% Identification)			
Target Tone		T1	T2	T3	T4	T1	T2	T3	T4
T1		56	35	6	3	63	28	7	3
T2		8	57	29	6	5	43	50	1
T3		12	33	43	11	5	29	53	14
T4		6	0	15	78	7	0	13	80

<sup>a</sup>Table III(a) was adapted from Table 5 in Wong (2012b). Five additional productions by 3-year-olds were added to those reported in Wong (2012b) (see footnote 1 for details).

(Gandour, 1978; Shen and Lin, 1991; Whalen and Xu, 1992). Children's tones, on the other hand, were perceived by the judges with much lower accuracy. Three-year-old children produced the four tones with 71%, 53%, 52%, and 67% accuracy, respectively [Table III(a), right panel]. Table III(b) shows that 4-year-old children produced the four tones with 56%, 57%, 43%, and 78% accuracy, respectively, and 5-year-old children produced the four tones with 63%, 43%, 53%, and 80% accuracy, respectively. Apparently, 4- and 5-year-old children's tone accuracy was more comparable to that of the 3-year-old children [Table III(a), right panel] but lower than those of adults [Tables III(a) and III(b)]. Results of Chi-square tests presented in Table IV(a) confirmed that none of the four tones produced by the three groups of children (i.e., 3-, 4-, and 5-year-olds) were adult-like (all  $p$ -values < 0.01).

The errors of 3-, 4-, and 5-year-old children's tone productions showed similar confusion patterns in the judges. T1 and T3 errors in all three groups of children were mostly perceived as T2 by the judges, while children's incorrect T2 and T4 productions were mostly perceived as T3 (Table III). Comparing to the adults, children's productions showed more diverse confusion patterns. Adults' T1 was never perceived as T3 and their T2 and T3 productions were never perceived as T1 or T4. However, all these confusion patterns were observed in children's productions, suggesting some qualitative differences in children's and adults' tone productions.

#### D. Order of accuracy of the four tones in children

To determine which tones were better mastered by the children, Chi-square tests were used to compare the accuracy

rates of the four tones produced by the children. The results in Table V showed that the order of accuracy of the four tones from the highest to the lowest accuracy for 3-, 4-, and 5-year-old was  $T4 = T1 > T2 = T3$ ,  $T4 > T1 = T2 > T3$ , and  $T4 > T1 > T2 = T3$ , respectively.

#### E. Developmental trend of tone accuracy

A weak correlation between age (in month) and tone accuracy was observed in preschool children. Pearson correlation coefficients between age and percent correct accuracy of the four tones were  $-0.04$ ,  $-0.15$ ,  $0.20$ , and  $0.28$  for the four tones respectively. None of the correlations reached significance ( $p$ -values ranged from 0.11 to 0.81). Figure 1 shows children's accuracy rates of the four tones as a function of age. As shown, the coefficients of determination ( $r^2$ ) were 0.00, 0.02, 0.04, and 0.08 for the four tones, respectively, indicating that age accounted for 0.18%, 2.19%, 3.88%, and 8.07% of the variance in children's accuracy of the four tones, respectively. Chi-square tests were used to compare the accuracy of the tones among the three age groups. The results presented in Table IV(b) showed that 3-year-olds produced T1 better than 4-year-olds. Four-year-old produced T2 better than 5-year-olds. Five-year-olds produced T3 better than 4-year-olds and 4- and 5-year-olds produced T4 better than 3-year-olds [Table IV(b)].

#### F. Number of harmonics in adults' and children's filtered tones

Given the surprising findings that preschool children in Taiwan make little improvement in their monosyllabic tone accuracy and have not produced adult-like Mandarin

TABLE IV. Comparisons of tone accuracy among different age groups.<sup>a</sup>

	(a) Adult Productions vs Child Productions											
	3-year-olds vs Adults				4-year-olds vs Adults				5-year-olds vs Adults			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
$\chi^2$	23.41	65.26	27.22	27.87	53.02	65.33	50.75	12.60	41.74	103.56	30.72	11.13
N	355	335	295	350	430	495	420	405	550	535	455	425
df	1	1	1	1	1	1	1	1	1	1	1	1
p-value	0.000** <sup>b</sup>	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.001**
Cramer's v	0.26	0.44	0.30	0.28	0.35	0.36	0.35	0.18	0.28	0.44	0.26	0.16
Better group	A	A	A	A	A	A	A	A	A	A	A	A

	(b) Child Productions											
	3-year-olds vs 4-year-olds				3-year-olds vs 5-year-olds				4-year-olds vs 5-year-olds			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
$\chi^2$	12.21	0.33	3.34	9.35	4.87	5.30	0.03	12.23	2.80	11.53	5.72	0.17
N	545	590	475	555	665	630	510	575	740	790	635	630
df	1	1	1	1	1	1	1	1	1	1	1	1
p-value	0.000**	0.566	0.068	0.002*	0.027	0.021	0.858	0.000**	0.094	0.001**	0.017*	0.681
Cramer's v	0.15	0.02	0.08	0.13	0.09	0.09	0.01	0.15	0.06	0.12	0.10	0.02
Better group	TWC3	n.s. <sup>c</sup>	n.s.	TWC4	n.s.	n.s.	n.s.	TWC5	n.s.	TWC4	TWC5	n.s.

<sup>a</sup>Data for 3-year-old productions were taken from Wong (2012b).

<sup>b</sup>Here \*\* indicates significance at 0.01 level and \* indicates significance at 0.05 level after Bonferroni correction for multiple comparisons. For the better group, "A," "TWC3," "TWC4," "TWC5," represent adults, 3-year-olds, 4-year-olds, and 5-year-olds, respectively.

<sup>c</sup>"n.s." indicates no significant difference.

TABLE V. Order of Accuracy of the Four Tones Produced by the Children.

Age Group	Tone	$\chi^2$	N	df	p	V	Order	Overall Pattern
TWC3 <sup>a</sup>	T1 vs T2	15.58	450	1	0.000** <sup>b</sup>	0.19	T1>T2 <sup>c</sup>	T4 = T1 > T2 = T3
	T1 vs T3	15.63	410	1	0.000**	0.20	T1>T3	
	T1 vs T4	1.03	485	1	0.311	0.05	T1 = T4	
	T2 vs T3	0.04	390	1	0.840	0.01	T2 = T3	
	T2 vs T4	9.18	465	1	0.002*	0.14	T4>T2	
TWC4	T3 vs T4	9.45	425	1	0.002*	0.15	T4>T3	T4 > T1 = T2 > T3
	T1 vs T2	0.07	685	1	0.796	0.01	T1 = T2	
	T1 vs T3	10.50	610	1	0.001**	0.13	T1>T3	
	T1 vs T4	33.54	615	1	0.000**	0.23	T4>T1	
	T2 vs T3	9.82	675	1	0.002*	0.12	T2>T3	
	T2 vs T4	39.14	680	1	0.000**	0.24	T4>T2	
TWC5	T3 vs T4	77.99	605	1	0.000**	0.36	T4>T3	T4 > T1 = T2 > T3
	T1 vs T2	31.21	845	1	0.000**	0.19	T1>T2	
	T1 vs T3	7.32	765	1	0.007*	0.10	T1>T3	
	T1 vs T4	25.84	755	1	0.000**	0.19	T4>T1	
	T2 vs T3	6.65	750	1	0.010	0.09	T3 = T2	
	T2 vs T4	99.63	740	1	0.000**	0.37	T4>T2	
	T3 vs T4	53.07	660	1	0.000**	0.28	T4>T3	T4 > T1 > T2 = T3

<sup>a</sup>Results were the same as in Wong (2012b) despite the fact that 5 more productions of the 3-year-olds were included in this study than in Wong (2012b) (see footnote 1 and Table 6 in Wong, 2012b).

<sup>b</sup>Here \*\* indicates significance at 0.01 level and \* indicates significance at 0.05 level after Bonferroni correction for multiple comparisons.

<sup>c</sup>Throughout ">" indicates "has a higher accuracy than." "V" presents Cramer's V.

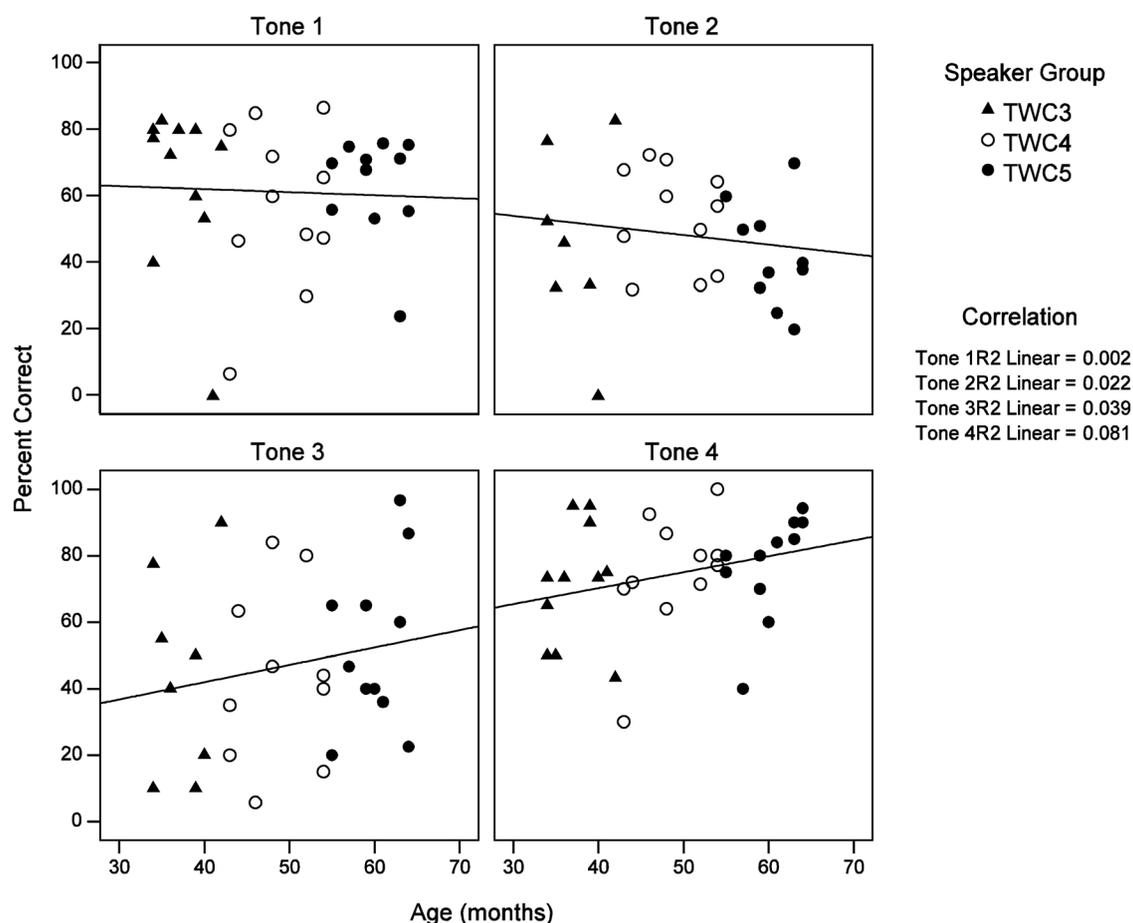


FIG. 1. Development of the accuracy of the four tones in 3- to 5-year-old children.

monosyllabic tones, we compared the number of harmonics in adults' and children's filtered productions to rule out the possibility that the findings were confounded by less pitch-related information in children's than in adults' filtered tones. The results showed the opposite effect. The number of harmonics in children's filtered productions (mean = 2.68, SD = 0.71, range = 1–6) was significantly more than in adults' filtered productions (mean = 2.39, SD = 0.63, range = 1–4), [ $U(N = 826) = 26916.50, z = -3.456, p = 0.001, r = 0.12$ ], indicating that the lower accuracy rates in children's productions were not due to less amount of pitch information in the filtered tones in children's productions than in adults' productions.

#### IV. DISCUSSION

The first research question addressed in this study was whether preschool children in Taiwan produce adult-like monosyllabic tones. Contrary to the findings in the earlier studies and the prevalent impression that children produced correct Mandarin tones in all contexts very early by the age of 2 years, the findings in this study showed that preschool children in Taiwan have not mastered the production of the four Mandarin tones in familiar monosyllabic words. Children as old as 5 years of age did not produce any of the four monosyllabic tones with adult-like accuracy (Tables III and IV).

The second question investigated children's order of acquisition/accuracy of the four tones. Previously, based on perceptual judgments, Wong *et al.* (2005) reported that the order of accuracy of the four tones produced by 3-year-old children growing up in the U.S. was  $T4 = T1 = T2 > T3$ , and Wong (2012b) reported that the order of perceived accuracy of the four tones produced by 3-year-old children growing up in Taiwan was  $T4 = T1 > T2 = T3$  (see also the top cluster in Table V). Wong (2012a) and an ongoing study performed acoustic analyses on tones produced by 3-year-old children growing up in the U.S. and Taiwan, respectively. By combining both acoustic and perceptual results, the two studies suggested that 3-year-old children growing up in the U.S. and in Taiwan followed the same order of acquisition of the four tones, namely,  $T4 > T1 > T2 > T3$ .

Acoustic analysis will be needed to gain more information on the phonetic characteristics of the tones produced by 4- and 5-year-old children and to fine tune the order of acquisition of the four tones in 4- and 5-year-old children. However, the perceptual findings in this study on the tones produced by 4- and 5-year-old children growing up in Taiwan suggested a similar order of accuracy reported in Wong (2012a). In both age groups, T4 was the best or one of the best tones produced by the children. The accuracy of T1 was not higher than T4, and not lower than T2 or T3. T2 accuracy was not better than T4 or T1 and not lower than T3. T3 was always the worst tone or one of the worst tones produced. Thus, it appeared that 3-, 4-, and 5-year-old children growing up in Taiwan and 3-year-old children growing up in the U.S. followed a similar order of acquisition of the tones (i.e., T4, T1, T2, and T3).

The third research question sought to investigate the developmental trends in the accuracy of monosyllabic tones in preschool children. Overall, little development was observed

in preschool children's accuracy of the four tones in monosyllabic words, as indicated by the weak correlations between age and the accuracy of the four tones (Fig. 1). T1 exhibited the least developmental change (Fig. 1 top left panel). Age accounted for less than 0.2% of the variance in the accuracy of T1 in preschool children. Though the Chi-square test showed that 3-year-old children produced T1 with higher accuracy than 4-year-old children (Table IV), no difference was found in the accuracy between 4- and 5-year-olds or between 3- and 5-year-olds (Table IV).

Similarly, little development was observed in T2 accuracy in Taiwan preschool children. Age accounted for only 2% of the variance in children's accuracy of T2 (Fig. 1). The Pearson correlation of  $-0.15$  and the finding that 4-year-old children produced T2 with higher accuracy rates than 5-year-old children (Table IV) seemed to suggest slight regression in T2 productions in 5-year-old children.

Among the four tones, T3 seemed to show slightly better development in children's productions. Age accounted for 4% of the changes in children's T3 accuracy (Fig. 1). Five-year-old children produced T3 with higher accuracy than 4-year-old children (Table IV). Our ongoing acoustic analyses of the tones produced by 3-year-old children growing up in Taiwan showed that most 3-year-old children in Taiwan (88%) produced T3 in isolation with a low falling F0 contour instead of a typical dipping F0 contour while more 4- and 5-year-old Taiwan children, 45% and 55%, respectively, produced a dipping contour for T3. Given that the dipping F0 contour of T3 and the rising contour of T2 have been found to be confusing due to the similarity in the shapes of the F0 contours of the two tones (Gandour, 1978; Shen and Lin, 1991; Whalen and Xu, 1992), and the perception of T2 and T3 is affected by the degree of initial fall and the location of the turning point in the F0 contour (Shen and Lin, 1991), it is speculated that if 5-year-olds regressed in their T2 productions while increasing their use of a dipping tone for T3, the regression might be a result of adjusting the phonetic categories for T2 and T3. Longitudinal studies and detailed acoustic analyses of children's tones will be required to provide more information in this respect.

Among the four tones T4 exhibited the most developmental changes. Age accounted for 8% of the variance in the accuracy of T4 (Fig. 1). Four- and five-year-old children produced T4 with relatively high accuracy at 78% and 80%, respectively, significantly better than 3-year-old children (Tables III and IV).

In light of the findings in the literature that adult native listeners confused T2 and T3, it seems likely that the low accuracy rates in children's tone productions were a reflection of adults' perceptual difficulties in identifying the tones rather than tone production errors made by children. However, the judges' high accuracy in determining adults' tone productions [Table III(a), left panel] indicated that adult listeners were able to identify monosyllabic tones in filtered speech with ceiling scores. This suggests that factors other than adults' tone perception difficulties have contributed to the low tone accuracy in children.

Factors that possibly have contributed to the lengthy course of Mandarin tones acquisition include perceptual

difficulty of tones in children. Children's low accuracy rates may indicate that they need extended time and substantial input to develop the phonemic categories of the four tones. Among the four tones, the phonemic and phonetic categories of T3 likely require the maximum amount of time to establish because the surface form of T3 varies in different contexts and the F0 contour of T3 is the most complex and is similar in shape to the F0 contour of T2.

However, the findings in this and recent studies appeared to suggest that perceptual difficulty may not be the primary factor for the lengthy course of acquisition of Mandarin tones in children. First, young children did not seem to have much difficulties identifying the four Mandarin tones. Wong *et al.* (2005) reported that 3-year-old Mandarin-speaking children growing up in the U.S. identified the four monosyllabic Mandarin tones in minimal pairs with relatively high accuracy rates at 90%, 87%, 69%, and 90%, respectively (Wong *et al.*, 2005). Yet, their monosyllabic tone production accuracy was significantly lower than their tone perception accuracy (Wong *et al.*, 2005). Second, young children's tone accuracy did not seem to improve with an increase in input. Three-year-old children growing up in Taiwan, where substantial exposure to Mandarin was available in the environment, did not produce the four tones better than 3-year-old children growing up in the U.S., where there was limited Mandarin exposure in the ambient environment (Wong *et al.*, 2005; Wong, 2012a,b). Findings in the present study showed that despite the fact that 4- and 5-year-old children growing up in Taiwan had more exposure to Mandarin tones than 3-year-olds, there was little improvement in tone accuracy with age (Fig. 1, Table IV). Moreover, tones that are expected to have equal perceptual difficulty were produced by children with varying accuracy. Xu (1997) showed that the same tone in the first syllable in disyllabic words have similar F0 contours regardless of the tones in the second syllable (i.e., small anticipatory tonal coarticulation), suggesting little difference in the perceptual difficulties for the same tones in the first syllable in different tonal contexts. Yet, Wong (2008) reported that children's production accuracy of the same tone in the first syllable varied and was lower when the offset F0 of the first tone was very different from the onset F0 of the second tone, requiring a large F0 transition from the first tone to the second tone at the beginning of the second syllable (Wong, 2008).

The physiological account proposed by Wong (2008), Wong (2012a), and Wong (2012b) seems to be a feasible explanation of the protracted course of development of tones. With reference to the findings in the EMG studies on adults' Mandarin tone productions (Hallé, 1994; Sagart *et al.*, 1986), Wong (2012a) postulated that the order of accuracy of children's four tones (i.e., T4, T1, T2, and T3, from the highest to the lowest accuracy) followed the order of articulatory complexity in producing the tones and suggested that tone acquisition is related to maturation of speech motor control. This view was supported by the finding that children have more difficulties producing disyllabic tone combinations that have more complex F0 contours than tone combinations with less complex F0 contours (Wong, 2008). For example, though T2–T3 (T2 followed by T3) and T3–T2 were both composed

of T2 and T3, the F0 contour for T2–T3 has a more complex F0 contour because a large F0 transition is needed to go from the high F0 offset of the rising tone in the first syllable to the low F0 onset in the low dipping tone in the second syllable. On the other hand, T3–T2 has a simpler F0 contour because less F0 transition is needed to go from the low F0 offset in the first syllable to the low F0 onset for the rising tone in the second syllable. Wong (2008) reported that adults produced the two tone combinations with 100% perceived accuracy while children produced T2–T3 with much lower accuracy (69% accuracy) than T3–T2 (81% accuracy).

If tone production is limited by physiological development, it is not surprising to find universal patterns in the development of tones in age-matched children acquiring Mandarin as a first-language in different linguistic environments. Given that motor skills take time to develop, young children have less mature speech motor control than adults (Smith, 2006; Smith *et al.*, 2006; Smith and Zelaznik, 2004), and the larynx of young children is not yet fully developed (Crelin, 1987; Hirano *et al.*, 1981; Kent and Vorperian, 1995), it is also not surprising to find comparable production accuracy in 3- to 5-year-old children. Future studies that systematically examine the relation between F0 complexity and tone accuracy and studies that directly measure children's articulatory control for producing tones will provide more insights in the relation between speech motor control and tone development.

The bilingual environment in Taiwan might be another factor that have contributed to the low accuracy in Taiwan children's tone productions. Wong (2012b) showed that though Mandarin-learning 3-year-old children growing up in the U.S. and in Taiwan showed similar error patterns and order of accuracy of the four tones, 3-year-old children in Taiwan made significantly more errors in T2 and T4 than the 3-year-old children in the U.S. Future studies that examine Mandarin tones produced by children in monolingual environments and other linguistic environments will help delineate the effect of ambient language on tone acquisition.

In all, the findings of the current study suggest a much more protracted development of Mandarin tones. Three- to five-year old children growing up in Taiwan have not yet produced any of the four Mandarin tones with adult-like accuracy. Little developmental change has been observed in the perceived accuracy the tones produced by preschool children. A common pattern was found in the order of accuracy of the four tones produced by 3- to 5-year-old children and the order appears to follow the order of articulatory complexity for producing the four tones.

## ACKNOWLEDGMENTS

This research was supported by funding from the East Asia and Pacific Summer Institute (EAPSI) of the National Science Foundation (NSF, OISE-0611641). The author thanks 蔡素娟 (Jane Tsay) for her assistance in subject recruitment and providing lab space for data collection, 莊淑蘭 (Sophia Chuang), 羅美如, the teachers and staff at Green World Children School and Chung Cheng Nursery School for their help in recruiting children, and the participants and their parents for their participation.

## APPENDIX

Target words for the picture naming task for adults and children.<sup>a</sup>

	Non Minimal Pairs				Minimal Pairs			
	Word	Tone	Pinyin	Translation	Word	Tone	Pinyin	Translation
1	哭	1	ku	cry	汤 <sup>b</sup> [湯] <sup>c</sup>	1	tang	soup
2	花	1	hua	flower	糖	2	tang	candy
3	灯[燈]	1	deng	lamp	三	1	san	three
4	门[門]	2	men	door	伞[傘]	3	san	umbrella
5	牛	2	niu	cow	书[書]	1	shu	book
6	球	2	qiu	ball	树[樹]	4	shu	tree
7	马[馬]	3	ma	horse	鱼[魚]	2	yu	fish
8	手	3	shou	hand	雨	3	yu	rain
9	笔[筆]	3	bi	pen	毛	2	mao	hair
10	二	4	er	two	帽	4	mao	hat
11	四	4	si	four	脚[腳]	3	jiao	foot
12	抱	4	bao	hold	叫	4	jiao	shout

<sup>a</sup>Source: adapted from the Appendix in [Wong et al. \(2005\)](#). Reprinted with permission from P. Wong, R. Schwartz, and J. Jenkins, "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children," *J. Speech, Lang., Hear. Res.* 48, 1079 (2005). © Copyright 2005. American Speech-Language-Hearing Association. All rights reserved.

<sup>b</sup>Simplified characters.

<sup>c</sup>Traditional characters, Pinyin is the official romanization system of Chinese.

<sup>1</sup>[Wong \(2012b\)](#) included children's first and second attempts of the target words only. This study selected the first two usable tokens from children's productions and included children's third or fourth attempt if the first two attempts were not usable. Thus, there were five more 3-year-old productions in this study than in [Wong \(2012b\)](#).

Chao, Y. R. (1973). "The Cantian idiolect: An analysis of the Chinese spoken by a twenty-eight-month-old child," in *Studies of Child Language Development*, edited by C. A. Ferguson and D. I. Slobin (Holt, Rinehart and Winston, New York), pp. 13–33.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese* (University of California Press, Berkeley, CA), Chap. 1, pp. 1–18.

Clumeck, H. (1980). "The acquisition of tone," in *Child phonology: Production*, edited by G. H. Yeni-Komshian, J. F. Kavanaugh, and C. A. Ferguson (Academic Press, New York), Vol. 1, pp. 257–275.

Crelin, E. S. (1987). *The Human Vocal Tract: Anatomy, Function, Development, and Evolution* (Vantage Press, New York), pp. 265.

Duanmu, S. (2007). *The Phonology of Standard Chinese*, 2nd ed. (Oxford University Press, New York), Chaps. 10 and 13, pp. 225–254 and 298–312.

Fromkin, V. A. (Ed.). (1978). *Tone: A Linguistic Survey*, 1st ed. (Academic Press, London), pp. 292.

Gandour, J. T. (1978). "The perception of tone," in *Tone: Linguistic Survey*, edited by V. A. Fromkin (Academic Press, New York), pp. 71–76.

Hallé, P. A. (1994). "Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese," *Lang. Speech* 37, 103–123.

Hirano, M., Kurita, S., and Nakashima, T. (1981). "The structure of the vocal folds," in *Vocal Fold Physiology*, 1st ed., edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo, Japan), pp. 33–41.

Hua, Z. (2002). *Phonological Development in Specific Context: Studies of Chinese-Speaking Children* (Multilingual Matters Limited, Clevedon, UK), Chaps. 3–4, pp. 45–105.

Hua, Z., and Dodd, B. (2000). "The phonological acquisition of Putonghua (modern standard Chinese)," *J. Child Lang.* 27(1), 3–42.

Kent, R. D., and Vorperian, H. K. (1995). "Development of the craniofacial-oral-laryngeal anatomy: A review," *J. Med. Speech Lang. Pathol.* 3, 145–190.

Li, C. N., and Thompson, S. A. (1977). "The acquisition of tone in mandarin-speaking children," *J. Child Lang.* 4(2), 185–199.

Lin, B., and Lin, N. (1994). *Language Disorder Scale of Preschoolers* (學前兒童語言障礙評量表) (Department of Special Education, National Taiwan Normal University, Taipei, Taiwan), pp. 36.

Sagart, L., Halle, P., Boysson-Bardies, B. d., and Arabia-Guidet, C. (1986). "Tone production in modern standard Chinese: An electromyographic investigation," *Cah. Linguistique Asie Orientale* 15(2), 205–221.

Sander, R. (2008). "Tonetic sound change in Taiwan mandarin: The case of tone 2 and tone 3 citation contours," in *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, Vol. 1, 87–107.

Shen, X. S., and Lin, M. (1991). "A perceptual study of Mandarin tones 2 and 3," *Lang. Speech* 34(2), 145–156.

Smith, A. (2006). "Speech motor development: Integrating muscles, movements, and linguistic units," *J. Commun. Disord.* 39(5), 331–349.

Smith, A., and Zelaznik, H. N. (2004). "Development of functional synergies for speech motor coordination in childhood and adolescence," *Dev. Psychobiol.* 45(1), 22–33.

Smith, B., McGregor, K. K., and DeMille, D. (2006). "Phonological development in lexically precocious 2-year-olds," *Appl. Psycholinguistics* 27(3), 355–375.

Tagliaferri, B. (2005). "Paradigm," [www.perceptionresearchsystems.com](http://www.perceptionresearchsystems.com) (Last viewed January 19, 2005)

Tse, J. K. (1978). "Tone acquisition in Cantonese: A longitudinal case study," *J. Child Lang.* 5(2), 191.

Tuaycharoen, P. (1977). "The phonetic and phonological development of a Thai baby: From early communicative interaction to speech," Ph.D. dissertation, University of London, UK.

Whalen, D. H., and Xu, Y. (1992). "Information for Mandarin tones in the amplitude contour and in brief segments," *Phonetica* 49(1), 25–47.

Wong, P. (2008). "Development of lexical tone production in disyllabic words by 2- to 6-year-old Mandarin-speaking children," Ph.D. dissertation, Graduate Center of the City University of New York.

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

Wong, P. (2012b). "Monosyllabic Mandarin tone productions by three-year-olds growing up in Taiwan and the U.S.: Interjudge reliability and perceptual results," *J. Speech Lang. Hear. Res.* 55, 1423–1437.

Wong, P., Schwartz, R. G., and Jenkins, J. J. (2005). "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children," *J. Speech Lang. Hear. Res.* 48(5), 1065–1079.

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

Yip, M. (2002). *Tone*, 1st ed. (Cambridge University Press, London, UK), Chap. 1, pp. 1–16.