Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions

Puisan Wong

Department of Otolaryngology—Head and Neck Surgery, College of Medicine, The Ohio State University

915 Olentangy River Road, Columbus, OH 43212, United States

Abstract

This study aimed to provide insights into children's development of lexical tone production by combining both perceptual and acoustic analyses. Duration and fundamental frequency analyses were performed on the monosyllabic Mandarin lexical tones produced by the 13 three-year-old children and four female adults reported in Wong, Schwartz & Jenkins (2005). Seven acoustic parameters that are strongly associated with the tonal judgments of 10 Mandarin-speaking judges were identified. Qualitative differences of the seven parameters in adult correct, child correct and child incorrect tone productions were compared and interpreted with reference to the perception data. The results confirmed that three-year-old children do not produce adult-like tones in isolated monosyllabic words. Even children's tones that are correctly categorized by adult listeners are phonetically different than adults' tones. The four tones from the most to the least adult-like are Tone 4 (Falling), Tone 1 (High Level), Tone 2 (Rising) and Tone 3 (Falling-Rising), perhaps corresponding to the complexity of speech motor control for producing these tones. Children demonstrate more difficulties producing low fundamental frequencies than high fundamental frequencies. The findings support the position that tone acquisition is a protracted process which may be affected by production complexities.

Lexical tone (hereafter, tone), the use of pitch to differentiate lexical and grammatical meanings, is an essential and important component of a majority of the world's languages. Yet, it remains unclear when and how lexical tones are mastered, what contributes to the order of acquisition of tones and how children's tone productions are different than those of adults. To gain better insights into children's acquisition of Mandarin lexical tones, this study examined the acoustic characteristics of three-year-old children's correct and incorrect Mandarin lexical tone productions in isolated monosyllabic words.

Mandarin is the tone language spoken by the greatest number of speakers. Each syllable in Mandarin is a morpheme and carries one of the four full tones or a neutral tone. In monosyllabic words in isolation, Tone 1 (T1), Tone 2 (T2), Tone 3 (T3) and Tone 4 (T4) have a high level, low rising, falling-rising (dipping), and high falling pitch or fundamental frequency (F0) contour, respectively. The same syllable produced with the four distinctive F0 contours has different meanings in Mandarin. For example, the syllable /ma/ means "mother", "hemp", "horse" and "scold" when produced with the four tones, respectively. Figure 1 shows the mean F0 contours of the four tones produced in an isolated syllable /ma/ by eight male adults reported in Xu (1997, Figure 2). Among the four tones, T3 undergoes the most contextual variations. When produced in isolation, it is a dipping tone. In non-final positions, it becomes a rising tone (T2) when preceding another T3, but a low falling tone when preceding any other tones. In final positions in connected speech, it is produced either as a low level tone or a dipping tone (Duanmu, 2007). The neutral tone occurs in short unstressed syllables in 5-7% of the words in Mandarin (Xu & Wang, 2009) and does not occur in monosyllabic words; it was, therefore, not examined in this study.

The primary and sufficient perceptual cue for Mandarin tones is the F0 contour (Fu & Zeng, 2000; Luo & Fu, 2004; Massaro, Cohen, & Tseng, 1985). When F0 information is available, all other acoustic cues such as amplitude (Gårding, Kratochvil, Svantesson, & Zhang, 1986; Howie, 1976; Whalen & Xu, 1992), vowel duration (Fu & Zeng, 2000; Ho, 1976; Shen, 1990) or vocal quality such as creaky voice (Gårding et al., 1986) are negligible, though they have been found to covary with the F0 contours and serve as cues in tone perception in the absence of F0 cues (Fu & Zeng, 2000; Whalen & Xu, 1992).

Tones are produced primarily by regulating the frequency of vibration of the vocal folds (i.e., fundamental frequency, F0) by the laryngeal muscles. The articulatory gestures for Mandarin tone production start at the beginning and terminate at the end of the syllable, regardless of the segment composition (type of consonant or vowels) or segmental structures of the syllable (Xu, 1998; Xu, 1999). According to Xu's Target Approximation Model of Tonal Contour Formation, the underlying pitch targets for the four Mandarin tones are high, rise, low and fall, respectively, suggesting that the main articulatory goals for the productions of the four tones are to achieve a high, rising, low, and falling F0 contour, respectively, before the end of the syllable (Xu, 1997; Xu, 1999; Xu & Wang, 2001).

Physiologically, the rate of vibration of the vocal folds is to a large extent controlled by laryngeal muscle activity. A handful of EMG (electromyographic) studies using hooked wire EMG electrodes inserted into the laryngeal muscles have investigated tonal production for Mandarin Chinese (Hallé, 1994; Sagart, Halle, Boysson-Bardies, & Arabia-Guidet, 1986) and Thai (Erickson, 1976; Erickson, 1993; Erickson, to appear). These studies have reported that an increase in CT activity occurs with high F0; relaxation of the CT muscle occurs with low F0, and frequently, SH also occurs with low F0, especially when F0 drops below a certain mid level

(Erickson, 1993; Hallé, 1994). Contraction of the CT muscle results in an increase in the length and tension of the vocal folds, and consequently an increase in the rate of vibration of the folds to produce a high F0. Relaxation of the CT muscle results in more slack, shorter vocal folds and consequently lower F0. Compared to the CT muscle, the SH muscle has a less direct connection to vocal folds, and is more complex, but it seems to work in certain cases to lower the larynx and, the by-product of this is to increase the thickness of the vocal folds and reduce the tension (see Erickson, 1993; Hallé, 1994; Honda, 1995; Honda, Hirai, Masaki, & Shimada, 1999 for details).

According to the studies about Mandarin tones, the production of high level F0 for Mandarin T1 involves increased CT activity. T2 which is characterized by an initial fall followed by a sharp rise, first shows increased SH activity and then decreased SH activity along with increased CT activity. When T3 is produced in connected speech (i.e., a low falling F0), only SH activity is seen. No data is available for T3 produced in isolation, but we assume that we would see a similar pattern as seen for T2—increased SH activity and then decreased SH activity along with increased CT activity. T4 is a falling tone, and is produced by first a peak in CT activity, followed by a relaxation of the CT muscle, and then, depending on the pitch range of the speaker, there may or may not be SH activity. For speakers with a high pitch range, the F0 fall of T4 is brought about simply by relaxing the CT muscle without any SH activity; however, for speakers with low pitch ranges, SH activity occurs to bring the F0 contour to a very low pitch level (Hallé, 1994; Sagart et al., 1986). The physiological mechanisms for producing the four tones are summarized in Table 3.

Large discrepancies have been found in the age of acquisition of Mandarin lexical tone production by children presumably due to limited number of studies and different methods

adopted (See Wong, Schwartz, & Jenkins, 2005 for a review). Most developmental studies on tone acquisition determined children's tone accuracy via the judgments of one rater listening to unprocessed natural speech. Two case studies (Chao, 1973/1951; Hua, 2002) that involved one and four children and one large study that involved 129 children growing up in Beijing (Hua & Dodd, 2000) reported that children as young as two years of age produced the four Mandarin tones correctly in various contexts including multisyllabic words in isolation and coarticulated tone productions in conversations and picture description. Another case study that involved three children showed variation in children's tone acquisition. Two of the children had not mastered the tones in isolated words and in utterance final position at 32 and 34 months (Clumeck, 1977a; Clumeck, 1977b). Another larger study involving 17 children did not report the age of acquisition of tones but stated that children mastered the four tones when they started to produce utterances longer than 2-3 words (Li & Thompson, 1977). Two, more recent studies employed multiple judges, used filtered stimuli to control for lexical expectation in the judges' tonal judgments, and compared child productions to adult productions. These studies found a much more protracted course of Mandarin tone acquisition. Children as old as three years of age did not produce the four tones in isolated monosyllabic words with adult-like accuracy (Wong et al., 2005); even children as old as five and six years of age did not produce adult-like tones in coarticulated disyllabic words (Wong, 2008; Wong & & Strange, submitted). The findings that children's accuracy rates of the same tones varied depending on the syllable position and were significantly lower when the F0 contours of the two coarticulated tones were more complex led Wong (2008) to conclude that children's tone accuracy was limited by speech motor constraints (Wong, 2008; Wong & Strange, submitted).

Several issues concerning children's acquisition of Mandarin tones remain unresolved. First, the age of acquisition of tones requires further quantitative specification. The studies that reported a lengthy process for tone acquisition used filtered stimuli for tone judgment to control for lexical biases (Wong et al., 2005; Wong, 2008; Wong & Strange, submitted). Though there was evidence that the judges were able to identify correct tone productions in filtered speech because adults' target tones were accurately identified with ceiling accuracy in filtered speech, it was less clear whether the tones that the judges categorized incorrectly in filtered speech were qualitatively different. Given that most studies reported early acquisition of Mandarin lexical tones, confident conclusions cannot be drawn from the results of these studies until further evidence is provided.

Second, it remains unclear how children's tone productions are different from those of adults. All developmental tone studies described above used perceptual judgments to determine tone accuracy. Thus, other than accuracy rates and substitution patterns, little other information was provided on children's tone production.

Third, the order of acquisition of the four tones in terms of which of children's tones approach the adult-form earlier is inconclusive. Some studies reported that T1 and T4 were acquired before T2 and T3 (Clumeck, 1980; Li & Thompson, 1977; Hua, 2002). Based on the findings with four children, Hua (2002) suggested that the production of T1 was stabilized before T4, while the order of acquisition of T2 and T3 was less clear. Clumeck (1980), however, reported that T2 was the first tone acquired by one of the three children. Wong, et al., (2005) reported that T3 was the most difficult for three year-old children and T1, T2 and T4 were produced with comparable accuracy rates.

Fourth, factors that affect the order of tone acquisition are unclear. Wong (2008) proposed that children's tone accuracy is related to F0 complexity and tone acquisition is limited by the maturation of speech motor control. If this hypothesis holds, one would predict that the order of accuracy of the four tones will follow the complexity of motor control for producing the four tones. Solid conclusions on the relation between speech motor control and tone accuracy cannot be drawn from studies that determined children's tone accuracy based on adults' perceptual categorization of children's tones.

One way to better address the above issues is to perform acoustic analysis on the tones produced by the adults and children in Wong et al., (2005) and compare the acoustic and perceptual findings. If the acoustic data support the reported perceptual judgment data, that is, if qualitative differences can be found in the acoustic measures of the tones that were judged as correct vs. incorrect, we can be more certain that children's tones that were incorrectly categorized by the judges in filtered speech were indeed qualitatively different and, therefore, we can be more confident to accept the findings in Wong et al., (2005) that three-year-old children are not able to produce adult-like Mandarin tones in monosyllabic words. In addition, acoustic analysis can provide more detailed information on children's articulatory gestures for tone productions, which will allow us to better characterize children's tone productions, examine the speech motor control in children's tone productions, and provide more insightful information on children's acquisition of tones. The more detailed acoustic data on children's tone production may also provide a more sensitive measure for children's tone and reveal more fine-grained differences on the order of acquisition of the four tones.

In all, the present study attempted to address the unresolved issues on three-year-old children's acquisition of monosyllabic Mandarin lexical tones stated above by performing

acoustic analysis on the monosyllabic tones produced by the children and adults in Wong et al., 2005 and comparing both the acoustic and previously reported perceptual findings in the study. The specific aims for the study are: (1) to characterize the correct and incorrect monosyllabic tones produced by children and provide acoustic evidence that three-year-old children's tones are qualitatively different than adults' tones, (2) to determine the order of acquisition, defined as degree of adultlikeness, of monosyllabic Mandarin tones by 3-year-old children, and (3) to investigate whether three-year-old children's tone accuracy is related to the complexity of speech motor control for producing the tones.

Method

This study performed acoustic analysis on the adult and child tone productions collected and reported in Wong et al., (2005). The following provides a brief description. More detailed information of the procedures for data collection and perceptual tone judgment can be found in the original study.

Tone Productions

Thirteen normally developing three-year-old (age range = 2;10 – 3;4) Mandarin-speaking children (6 girls and 7 boys) and four of their mothers participated in the study. All children were exposed to Mandarin at home and had limited exposure to English, other languages or other dialects (see Wong, et al., (2005) for their language scores and other details). All their parents, except one, attained college degrees or higher. The four mothers were native speakers of Mandarin. Mandarin was their strongest language and they spoke only Mandarin at home.

Each participant was presented with 24 colored pictures representing 24 monosyllabic words (4 tones x 6 words). The 24 words were selected based on the results of a word familiarity test presented to 15 children not participating in the study (see Wong et al., (2005) for details).

Half of the 24 words formed six minimal pairs that contrasted the six tonal combinations of the four tones. The other half were singletons that did not form minimal pairs (See Appendix A in Wong et al., (2005) for the list of stimuli). The 13 children labeled each picture two times to the examiner. The four mothers labeled the pictures two times to their child after the child had finished all the testing activities. All productions were recorded on a DAT recorder via a condenser microphone with 16 bit rate and 22.1 kHz sampling rate. With all contexts included (i.e., in isolation, in disyllabic words, in longer utterances, and in duplicated syllables), the children produced 62, 54, 62, and 59 target syllables, for T1, T2, T3, and T4, respectively, showing comparable number of productions for the four tones. In order to control for coarticulatory effect, only target words produced spontaneously in isolation were included. If the first production could not be used due to noisy background, overlapping of voices or non-isolated productions, the second production was selected. None of the children produced the word "mao4" 'hat' in isolation; the word was, therefore, excluded. Thus, the final set of usable productions involved 23 target words (6 words each with T1, T2, and T3 and 5 words with T4), five minimal pairs (no T2-T4 contrast due to the loss of mao4), 92 adult productions (24 productions for T1, T2, and T3 and 20 productions for T4) and 198 child productions (55, 54, 47, and 42 productions for T1, T2, T3 and T4, respectively). Each adult produced 22-23 usable productions and each child produced 12-19 usable tokens.

Perceptual Judgment of Tones

The 290 adult and child productions were excised and saved as individual wave files. Ten Mandarin-speaking adults were recruited to judge the tones. All judges reported that Mandarin was their dominant language. All judges learned Mandarin at birth, except for one judge who learned Mandarin at 3 years of age (see Wong et al., 2005 for details). Seven of the judges

reportedly were also exposed to Taiwanese, Shanghainese, or Cantonese. To control for lexical biases on the judges' tone judgment, the adult and child productions were low-pass filtered at 400 Hz and 500 Hz, respectively, to eliminate most of the segmental information and retain the F0 information. All productions were normalized to the same root mean square value for intensity. The judges listened to the filtered productions and categorized the tones in a four-alternative forced choice task in a sound treated booth under headphones using a customized computer program (Tagliaferri, 2005) (See Wong et al., 2005 for details).

Acoustic Analyses

Acoustic analyses were performed on 289 tone productions by the adults and children. The word 'Tang1' produced by the child C06 was excluded because the duration of the production was too short to generate F0 information using the Praat script.

Each monosyllabic tone production was manually segmented into two portions: consonant (C) and rime (R) using a custom written script -- Prosody Pro v. 3.1 (Xu, 2005-2010)-- for Praat (Boersma & Weenink, 1992). Because acoustic-phonetic information about the consonants was largely eliminated in the filtered stimuli, segmentation was performed on the original unfiltered stimuli. The script showed simultaneously a waveform, a spectrogram, and a label window for the sound file on the screen. A phonetically trained Mandarin-speaking doctoral student marked and labeled the two segments. The onset of C started from the beginning of the initial consonant, which was defined as the onset of the release burst, the fricative noise, the nasal murmur and the onset of the first vocal pulse for the voiceless stops, fricatives and affricates, nasals, and approximants, respectively. In order to avoid F0 fluctuations due to the initial consonants (House & Fairbanks, 1953; Umeda, 1981) and provide sufficient window length to generate the initial and final F0 values for the rime, the onset of R (i.e., the offset of C)

was marked at the end of the fourth vocal cycle of the vowel, and the offset of R was marked at the beginning of the fourth observable vocal pulse from the end of the rime in the waveform. The segmentation measurements based on the unfiltered stimuli were then applied to the filtered counterpart. In cases when the segmentation based on the unfiltered stimuli did not mark the end of the fourth vocal pulse from the onset of the rime or the beginning of the fourth pulse from the end of the syllable in the filtered stimuli, the segmentation was adjusted.

F0 extraction and measurements were done on the filtered tokens using the same custom written Praat script (Xu, 2005-2010). Vocal pulse markings generated by Praat were inspected for any erroneous markings (missing pulses and doubling pulse markings) and corrected manually. All segmentation and vocal pulse markings were re-checked for accuracy and consistency by the author.

The script computed several measurements including maximum (max) F0, minimum (min) F0, and mean F0 for each of the 2 segments of each tone production. Each segment was also divided into 10 intervals equal in time. The mean F0 for each interval was calculated and a conservative smoothing algorithm was applied to remove local spikes in the F0 values (See description of the algorithm in Xu, 1999). All the F0 values were saved in a text file. In addition to the durations of C and R, the duration of each $1/10^{th}$ of the segment, and the duration between segment onset and maximum F0 and the duration from segment onset to minimum F0 were computed with an adapted version of ProsodyPro.

Because vocal pulses could not be reliably detected in most of the C segments, particularly in voiceless consonants, F0 data generated from C were eliminated from further analysis. Note that the duration measure of C was retained to compute the syllable duration, which corresponds to the duration for the articulators to produce the tones, according to the

Target Approximation Model of Tonal Contour Formation (Xu, 1997; Xu, 1999; Xu & Wang, 2001), mentioned above. Thus, except for syllable duration, all acoustic analyses were based exclusively on the F0 and duration values extracted from the rime portion of the tone productions. To facilitate comparisons and to convert the F0 scale to a psycho-acoustic pitch scale with equal perceptual intervals (Nolan, 2003), all F0 values were converted to semi-tones using 1 Hz as the reference frequency.

Duration and seven additional acoustic parameters were derived from the measurements obtained from the Praat script to test the acoustic characteristics of the tone productions. Table 1 presents the definitions and the purposes for the selection of the parameters. Syllable duration was selected to test whether children produced the four tones with adult-like durations. The other seven parameters were chosen for three reasons. First, they measured the pitch targets and the unique characteristics of the F0 contours for the tones. "Pitch Shift" measured the F0 range of the tone regardless of the direction of the F0 slope and can be used to index the degree of levelness of the produced tone. It was, therefore, used to measure and compare the flatness of the F0 contours in T1 productions. "Height of Mean F0" indexed the mean F0 level in the produced tone (Token Mean) relative to the mean F0 of the speaker (Speaker Mean). Thus, it was used to measure how well the speakers reached the high tonal target for T1 and the low overall F0 for T3. "Height of Min F0" indexed the height of minimum F0 in the produced tone (Token Min) relative to the Speaker Mean and was used to measure how well the speaker reached the high tonal target for T1 and the low tonal target for T3. "Directional Excursion" measured the degree of positive and negative F0 span and "Slope" measured how steep the positive and negative F0 slopes were. These two acoustic parameters were selected to test and compare how well the rising and falling tonal target were reached in T2 and T4. Previous studies reported that the

temporal location of the turning point, which corresponds to the minimum F0 in T2 and T3, in the F0 contour was associated with T2 and T3 perception (Shen & Lin, 1991). Thus, "Timing of Min F0" and "Timing of Max F0" were used to measure and compare how early the minimum F0 and how late the maximum F0 occurred in T2 syllables, and how late the minimum F0 and how early the maximum F0 occurred in T4 syllables.

Another reason for selecting the specific acoustic parameters was that, except for syllable duration, the extreme (highest and lowest) values of the acoustic parameters were expected to best associate with the characteristics of the F0 contour of one and only one of the four tones. This property would allow correlation analyses on the association between the acoustic parameters and the perception of the four tones (more details below). To illustrate, low values of "Pitch Shift", high positive values of "Height of Mean F0" or high positive values of "Height of F0" should predict T1 perception only. None of the other tones has as a small F0 range, high Token Mean, or high Token Min as T1. Large positive values of "Directional Excursion" and "Slope", small values of "Timing of Min F0", or large values of "Timing of Max F0" should associate with T2 perception only. Large negative values of "Mean F0 Height" or "Min F0 Height" should associate with T3 perception while large negative values of "Directional Excursion" or "Slope", large values of "Timing of Min F0", or small values of "Timing of Max F0" are associated with T4 perception only. Parameters whose extreme values were expected to associate with more than one tone were not selected. For example, final F0, which would be similarly high for both T1 and T2, or initial F0, which would be low for both T2 and T3 and high for T1 and T4, were not selected.

In addition only parameters that could be computed relatively reliably and accurately using a Praat script were selected. The reason was to reduce as much as possible the amount of

labor and time needed for performing acoustic analysis on tones such that the methods adopted in this study can be applied to measure a larger amount of tone productions in future studies.

Data Analysis and Results

Correlation analyses.

To determine which acoustic parameters explained the most variation in the judges' tone judgments, correlation analyses were performed. First, for each of the 289 tone productions, the values of the eight parameters listed in Table 1 were obtained. Then, the percent of judges who categorized the tone into each of the four tone categories was calculated. For example, the word Qiu2 'ball' produced by child C02 was categorized as T1, T2, T3 and T4 by six, one, three, and none of the judges, respectively. It was, therefore, recorded as 60%, 10%, 30% and 0% of the time being perceived as T1, T2, T3, T4, respectively. After that, because all distributions of the data significantly violated the normality assumption for parametric statistics, Spearman's rank order correlation was conducted on the 289 productions to determine the association between each of the eight acoustic parameters and the percent of judgments for each of the four tones.

As presented in Table 1, syllable duration did not associate well with the judges' perception of any of the four tones. The correlations between the other seven acoustic parameters and the tone decisions of the judges were highly predictable. Acoustic parameters that were unique for a certain tone yielded strong correlations with the judgments of that specific tone (highlighted in bold in Table 1), while having weak correlations with other tones (presented in gray). For example, as expected, "Pitch Shift" had a strong negative correlation with the judges' perception of T1 (r_s = -.712), indicating that the smaller the F0 fluctuation was in the production, the more T1 judgments occurred. As predicted, "Pitch Shift" did not correlate well with the other

three tones because large values of "Pitch Shift" predicted both T2 and T4 and the values of "Pitch Shift" for T3 were neither the highest nor the lowest among the four tones.

Overall, the correlation results showed that T1 judgments associated strongly with small F0 ranges, high mean F0 and high minimum F0. T3 perception, on the other hand, associated strongly with low mean F0 and low minimum F0. T2 judgments associated strongly with large positive F0 excursion, steeper positive slopes in the F0 contours, and reaching minimum F0 early and maximum F0 later in the rime. T4 perceptions, on the contrary, associated strongly with large negative F0 excursion, steeper negative slope in the F0 contour, and reaching maximum F0 earlier and minimum F0 later in the rime.

Accuracy groups

To examine the acoustic characteristics in children's correct vs. incorrect tone productions, all tone productions were categorized into three groups based on the results of the tone judgments by the 10 judges. Productions in which the target tones were correctly identified by 8 or more of the 10 judges were categorized as "correct" productions. 86 adult productions and 104 child productions fell into this category. Productions in which the tones were correctly identified by 0 to 4 of the judges were considered "incorrect" productions. Three adult productions and 50 child productions belonged to this category. The target tones of three adult productions and 43 child productions (4, 15, 12 and 12 productions for T1, T2, T3, and T4, respectively) were correctly identified by 5-7 judges; these productions and the three adult "incorrect" productions were excluded for the accuracy group comparisons below. Thus, the accuracy group comparisons below involved 86 adult correct (AC) productions (23, 24, 19 and 20 productions for T1, T2, T3, and T4, respectively), 104 child correct (CC) productions (39, 29,

11 and 25 productions for T1, T2, T3, and T4, respectively), and 50 child incorrect (CI) productions (11, 10, 24, 5 productions for T1, T2, T3 and T4, respectively).

Tone Contours of AC, CC and CI Productions

To visually compare the F0 contours of the produced tones and to note individual differences in the tones produced by the adults and children, time normalized F0 contours of the tones in the rime of the syllables were plotted and are presented in Figure 2. Each line in each panel presents the F0 contour of one production and each panel presents the F0 contours of the same tone in different words produced by the same speaker. The panels with no F0 contours indicated that none of the child's productions fell into that accuracy category. The mean F0 values for each 1/10 of the duration of the rime computed by the Praat script described above was plotted with equal intervals on the abscissa to facilitate comparisons of the F0 contours across productions and speakers.

Overall, the F0 contours of adults' tone productions showed the expected F0 shapes for the four tones: high and level for T1, rising for T2, dipping for T3 and falling for T4. There was little variation in the F0 contour across the different words produced by the same adult speaker for T1 and T2 (Figure 2). The seemingly more varied F0 contours observed in T3 (e.g., deep U-shape F0 contours) and T4 (e.g., sharp F0 drops at the end of T4) were mostly due to the presence of glottal fry, which occurred when the speaker produced the tones with extremely low F0 causing the vocal folds to vibrate slowly with extreme jitter (unequal vocal periods).

Based on visual inspection of Figure 2, the F0 contours of the children's correct productions generally showed the typical shapes of each of the four tones. However, many of the children's correct T1 productions were not as flat as the adults' T1. Some of the productions drifted slightly downward (e.g., productions of C01 and C03); while some moved slightly

upwards (e.g., productions of C11). The rising slopes of many children's correct T2 productions were not as steep as adults' T2 (e.g., productions of C04, C10). Children as a group had only eleven correct T3 productions; these productions by and large had similar F0 shapes as those of adults. There were some variations in the F0 shapes of T4 across child speakers but a sharp falling F0 is noted in the second half of the rime in most productions. The precipitous falls for the production by C10 and one of the T4 productions by C13 were due to glottal fry.

The F0 contours of children's incorrect productions were much more varied. Visual inspection showed that some T1s were produced with a falling F0 (e.g., C10 and C12), some with a rising F0 (e.g., C05, C11), and some with a relatively low F0 (e.g., C06). Children's errors in T2 involved having flat F0 contours (e.g., C03, C10), falling F0 contours (e.g., C02), and going too low before turning (i.e., having a big dip similar to T3, e.g., C14). Children's T3 errors involved not reaching a low F0 before rising (e.g., C05, C13), having falling F0 contours (e.g., C01 and C02), and having level F0 contours (e.g., C10 and C04). Most of children's T4 errors had a falling F0. However, the slope was less steep than the correct productions. Individual error patterns were noted. Some children showed consistent errors in the F0 contours for the same tone (e.g., T2 and T3 of C02, T3 of C05), whereas some showed varied patterns across different words for the same tone (e.g., C08,C12). C02 seemed not to have very distinct tone categories as her F0 contours were similar for T2, T3 and T4.

Figure 3 shows the average F0 plots of the four tones of the three accuracy groups. Children had higher F0 contours than adults. Visual inspection showed that, as a group, children's correct T1 productions were not as flat as adults' T1. The larger F0 difference at the onset and smaller F0 difference at the offset between the mean f0 contours of children's and adults' correct T2 productions in the second panel suggested that children's correct T2

productions did not go as low at the beginning and did not go as high at the end of the rime as the adults' T2 productions. Statistical analyses revealed that the F0 differences were significant between AC and CC productions at the beginning of T2, whereas the F0 differences at the end of the rime did not reach statistical significance (see below). Children's correct T3 seemed to get to the minimum F0 later than adults' T3, though the difference was not statistically significant (see below). Children's correct T4 productions appeared not to be as low at the end of the syllable indicated by the larger F0 difference between the adults' and children's F0 contours, though the difference did not reach statistical significance (see below). Due to the large variations in children's incorrect pronunciations (Figure 2), their average F0 plots (Figure 3) were less representative of their productions.

Comparing AC, CC and CI

Syllable duration and the acoustic parameters that strongly correlated with each of the four tones (see Table 1) are listed in Table 2. Though syllable duration was not found to highly correlate with any of the four tones (Table 1), it was included to examine any duration differences in the tones among the three accuracy groups. Due to the violations of the assumptions of normality and homogeneity of variance for parametric statistics, a Mann-Whitney U test was used to compare the differences of the selected acoustic parameters among the three accuracy groups. Table 2 presents the results and Figure 4 presents the box and whisker plots of the distributions of the parameters in which statistical significance were found among the three accuracy groups.

Syllable durations were not significantly different among the three groups for the production of any of the four tones (Table 2). However, the three groups differed in the other acoustic parameters measured. As indicated in Table 2 and Figure 4, not all children's correct

tone productions were adult-like. Children's correct T1 productions were neither as level nor as high as adults' as indicated by the significantly larger pitch shifts, smaller differences between the Token Mean and Speaker Mean, and reduced differences between Token Min and Speaker Mean F0 (see AC vs. CC in Table 2). Children's correct T2 productions had adult-like slopes and timing in reaching the minimum F0 and maximum F0. However, their F0 ranges did not span as much as those of adults. Further analysis revealed that the minimum F0 in children's correct T2 productions were significantly higher than those of adults (z (N=53) = -2.126, p= .033, Mann Whitney-U test) while their maximum F0 were not significantly different than those of adults (z (N=53) = -1.001, p = .317, Mann Whitney-U test), suggesting that children produced correct T2 with higher onset F0 and reached an F0 comparable to adults towards the end of the F0 contour. The eleven CC productions of T3 were adult-like in terms of the degree to which a low F0 was reached. No statistical difference was found in the timing of reaching the minimum F0 in the rime between the adult and child correct productions, z (N=30) = -.1.399, p=.162, Mann Whitney-U Test. However, the mean F0 of the CC productions were higher than the mean F0 of adults (Figure 4) suggesting that children's correct T3 productions did not maintain an F0 as low as adults throughout the syllable. Children's correct T4 productions were adult-like in all the five acoustic parameters measured, with p-values ranged from 0.08 to 0.65, Mann Whitney-U test) (Table 2 and Figure 4). Results of Mann Whitney-U test also found non-significant differences in the final F0 between children's and adults' productions (z(N=45) = -1.783, p=.075).

Children's incorrect tone productions were mostly different from children's correct productions (Table 2). Children's incorrect T1 productions had significantly larger F0 fluctuations and lower minimum F0 than CC productions (Tables 2, Figure 4). The F0 contours of children's incorrect T2 productions ranged from having significantly smaller positive

excursions than those of CC productions to having negative excursions (Table 2, Figure 4). The F0 slopes of their productions ranged from reduced steepness to negatively sloped (Figure 4). Children's incorrect T2 productions also tended to reach the minimum F0 significantly later and the maximum F0 much earlier in the rime than CC productions (Tables 2, Figure 4). The minimum F0 of children's incorrect T3 productions was significantly higher than CC productions (Table 2, Figure 4). Children's incorrect T4 productions reached maximum F0 at around the same time in the syllable as in CC productions (Table 2). However, their F0 contours were significantly less negatively excursed, had much reduced negative slopes and reached the minimum F0 earlier than in CC productions (Table 2 and Figure 4).

Discussion

Acquisition of tones

Results of the current study provide additional evidence that tones can be accurately and reliably categorized in filtered stimuli with degraded segmental information and, thus, categorization of tones in filtered stimuli is a sensitive and reliable method for judging tone accuracy. The judges' tone categorizations on filtered speech corresponded strongly with the distinctive F0 characteristics of the tones. As shown in Table 1, acoustic parameters that were distinctive for a specific tone yielded strong correlations with the identification of that tone, while non-distinctive acoustic parameters or acoustic characteristics that were shared by more than one tone yielded weak correlations with the judges' tone categorization. T1 perception was highly correlated with high and level F0 contours. Higher mean F0, higher minimum F0, and narrower F0 ranges all yielded more T1 categorizations from the judges. T2 judgments were highly associated with steep positive slopes. Large positive F0 excursions, steeper positive F0 slopes, and early minimum F0 and late maximum F0 all gave rise to more T2 categorizations. T3

categorizations were strongly associated with achieving a low tonal target (low minimum F0) in the syllable and having median tonal onsets and offsets (overall, lower mean F0). T4 perception was highly correlated with steep falling F0 contours. Large negative F0 excursions, steeper negative F0 slopes, and reaching maximum F0 early and minimum F0 later in the rime yielded more T4 categorizations.

Second, the tones that were categorized as correct vs. incorrect based on the judgments on filtered stimuli corresponded well with the differences observed in the F0 plots (Figure 2) and had qualitatively different acoustic characteristics that deviated from the pitch target for that tone (Table 2). To illustrate, adults' tone productions had the typical F0 shapes (Figure 2) and were identified with close to ceiling accuracy in filtered speech (96%, 96%, 83% and 98%, see Wong et al., 2005). Children's tones that were correctly identified by the judges had F0 contours and acoustic characteristics that resembled adults' tones (Figures 2, 3, & 4, Table 2), whereas children's tones that were identified as incorrect had different F0 contours (Figure 2) and had significantly different acoustic characteristics than adult and child correct productions (Figures 2, 3, 4, Table 2). These results not only confirmed that the judges in Wong et al., (2005) were able to accurately identify the correctly produced tones, they also indicated that the judges' categorization of incorrect tones were valid and acoustically based.

Results of the current study also support the finding that the acquisition of Mandarin lexical tone is a protracted process and is not completed as early as most previous studies have suggested. The significant differences found in the distinctive acoustic parameters for the four tones in children's correct vs. incorrect productions support the claim of Wong et al. (2005) that three-year-old children learning Mandarin as a first language have not yet mastered the production of the four Mandarin lexical tones in monosyllabic words.

How children's tones differ from adults

Based on the acoustic parameters measured, even children's tones correctly categorized by adult native speakers were not the same as those of adults. Table 3 presents the important acoustic discriminants, the physiological mechanisms for producing the tones and the major findings of the study. The F0 contours of children's correct T1 productions were neither as level nor as high as adults' T1, as indicated by the significantly larger pitch shift and lower minimum F0 and mean F0 (Tables 2 and 3). Children's correct T2 productions had adult-like F0 slopes and attained relatively similar timing in reaching the minimum and maximum F0. Yet, their F0 ranges were much more reduced, and they started the tone at a much higher F0 than adults. Based on the eleven productions, children's correct T3 productions reached a minimum F0 as low as adults, but the mean F0 of their T3 productions was higher than adults' mean F0 for T3, indicating that the F0 contours of their T3 were not maintained at frequencies as low as adults' T3. Children's correct T4 productions were the most adult-like. The syllable durations, F0 excursions, degrees of F0 slopes, and relative timing for maximum and minimum F0 were all comparable to adults' (Tables 2 and 3, Figure 4).

Most of the acoustic parameters in children's incorrect productions were significantly different than children's correct productions. Children's incorrect T1 productions were characterized by even more fluctuated F0 and much lower F0 than children's correct productions as indicated by significantly larger F0 ranges and much reduced differences between the minimum F0 in the token and the mean F0 of the speaker (Figure 4, Tables 2 and 3). Children's incorrect T2 productions were significantly different from children's correct productions in the four acoustic parameters tested. Their F0 slopes and F0 excursion sizes tended to be much reduced or have a reversed (falling) direction. They reached the minimum F0 much later and the

maximum F0 much earlier in the syllable (Tables 2 & 3, Figure 4). Children's incorrect T3 productions had more varied F0 contours (Figure 2) and did not reach a low F0 target . Children made fewer T4 errors (only 5 productions fell in the child incorrect category, though 12 productions were judged incorrectly by 3 to 5 of the 10 judges and were excluded from analysis). The F0 shapes of their incorrect T4 were less varied and they mostly involved reduced falling slopes, more limited F0 excursion sizes, and reached minimum F0 earlier than children's correct productions (Tables 2 and 3, Figure 4).

The order of acquisition of tones

With the combination of the perceptual findings in Wong et al., (2005) and the acoustic results in the current study, more fine-tuned conclusions can be made about the order of acquisition (i.e., order from being the most to the least adult-like) of Mandarin tones by three-year-old children. Though Wong et al. (2005), reported comparable perceptual accuracy rates for children's T1 (78%), T2 (70%) and T4 (76%), results of the acoustic analysis suggested that T4 was the most adult -like because children's correct T4 productions were not significantly different than adults' T4 in any of the five acoustic parameters measured (Tables 2 & 3). Also, few of the children's T4 fell into the child incorrect category defined by this study and most of the children's T4 errors had F0 shapes resembling a falling F0 contour (Figure 2), though the F0 slopes were much shallower and the F0 excursion size was reduced. Children's T1 productions were likely to be the second most adult-like. They were not as good as the T4 productions because even children's correct T1 productions were significantly different than those of adults in terms of pitch shift, mean F0, and minimum (Table 2, Figure 4). Nevertheless, children's T1 appeared to be more adult-like than T2 because they yielded higher perceptual accuracy from adult listeners (78% vs. 70%), there were more T1 (39 productions, 72%) than T2 (29

productions, 54%) tokens that were categorized as correct productions, and incorrect T1 productions were not significantly different than children's correct productions in terms of minimum F0, whereas children's T2 incorrect productions were different than children's correct productions in all the acoustic parameters measured (Table 2). Children's T3 are the least adult-like because they were perceived with the lowest accuracy rates (44% correct out of all the 470 judgments made by the 10 judges, see Wong et al., (2005)) and only eleven of the 47 child T3 productions (23%) were judged correctly by eight or more judges (Table 2). Though the minimum F0 of the eleven correct productions was adult-like, the mean F0 was significantly higher than that of adults (Figure 4). For the children's incorrect productions of T3, the shapes of the F0 contours varied drastically (Figure 2) and the minimum F0 was higher than that of children's correct T3 (Tables 2 & 3). Taken together, the order of acquisition of the four monosyllabic tones in 3-year-old children from the most to the least adult-like are: T4, T1, T2 and T3.

Factors that affect tone acquisition

Various reasons can be proposed for the protracted long course of Mandarin tone acquisition and the observed order of acquisition of Mandarin tones in 3-year-old children. One plausible reason is perceptual difficulties: children's inability to produce the tones accurately may be attributed to their difficulties in perceiving the F0 differences in the tones. Wong et al., (2005) also collected data on the perception of monosyllabic lexical tones in these 13 children. The children perceived the four tones with accuracy rates of 90%, 87%, 69% and 90% in trials that involved tone minimal pairs, suggesting comparable perceptual accuracy for T1, T2, and T4 and lower accuracy for T3, a pattern that corresponds to that of children's overall tone

production accuracy determined by the 10 judges (79%, 70%, 44%, and 76% for the four tones, respectively).

The findings of the current study suggest support for Wong's (2008) claim that children's ability to accurately produce tones may be related to the complexity of articulatory control. That is, ease of motor control might play a role in the order in which children acquire the four tones: T4, T1, T2, T3. T4 might be the easiest to produce because, especially for speakers with a high pitch range, (e.g., Hallé, 1994), it can be produced by tensing the CT muscle, and then relaxing it. We suggest that T1 might require more control than T4 because it would involve sustained CT activity to maintain the high F0 throughout the tone. T2 might be more articulatorily complex than T1 because it involves suppression of CT and activation of SH for the initial F0 drop, followed by activation of CT (Hallé, 1994; Sagart et al., 1986). It also has been reported by a number of studies (e.g., Loeb & Allen, 1993; Ohala & Ewan, 1973; Snow, 1998; Sundberg, 1979; and Xu & Sun, 2002) that producing a rising pitch tends to be more effortful than a falling pitch. Finally, T3 might be the most articulatorily complex because it requires coordinated control of two muscles – the CT and the SH (e.g., Hallé, 1994). First, there is relaxation of the CT muscle together with SH activity at the beginning of the tone to get to the very low initial F0, which is then followed by relaxation of the SH and CT activity to achieve the final high F0. The notion that more complex intonation contours are difficult children to produce has also been reported by Koike and Asp (1981). Whether, in addition, pitch lowering muscles (i.e., SH) develop later than pitch raising muscles (i.e., CT) is a topic for further investigation.

To summarize, results of the acoustic analysis in this study support the report of protracted developmental course of Mandarin tone acquisition reported in Wong (2005) and Wong (2008) and confirm that children at three years of age do not produce the tones in isolated

monosyllabic words like adults. The combination of perception and acoustic analyses show that even children's tones that are correctly categorized by adult judges are not fully adult-like. The order of acquisition of the four tones from the most to the least adult-like is T4, T1, T2 and T3. The ease of motor control for producing the four tones may be a contributing factor to the order of tone acquisition by children. This needs to be explored in the future.

Author Note

Puisan Wong, Eye and Ear Institute, Department of Otolaryngology—Head and Neck Surgery, College of Medicine, The Ohio State University.

Correspondence should be addressed to Puisan Wong at the Eye and Ear Institute, Department of Otolaryngology—Head and Neck Surgery, College of Medicine, The Ohio State University, 915 Olentangy River Road, Columbus, Ohio 43212. Email: pswResearch@gmail.com.

The author thanks Yi Xu for his input and help in performing acoustic analyses, and his feedback on an earlier version of this manuscript. Thanks also go to Xiliang Liu for his technical support, Winifred Strange for the helpful discussions, the editor, Kenneth de Jong, and four anonymous reviewers for their insightful input, Jing Yang for her assistance in segmenting the stimuli and checking the fundamental frequencies, and Kimbrough Oller and Susan Nittrouer for editing an earlier version of the manuscript.

Table 1. Description of the acoustic parameters and their correlations with tone categorization

Acoustic Parameters	Definition	Purpose	Correlations between Acoustic Parameters and Tone Judgments (Spearman's rho, r_s)			
			T1	T2	T3	T4
Syllable duration (s)	Duration of C + R	To test tone duration differences	-.084	-.019	.211	-.087
Pitch shift (St)	Max f0 - min f0 in R	To test the flatness of the f0 in T1	**-.712**	.076	.265	.422
Height of mean f0 (St)	Mean f0 in R (Token Mean)- mean f0 across all the productions by the speaker (Speaker Mean)	To test the tone targets: "High" in T1 and "Low" in T3	**.531**	-.135	**-.688**	.034
Height of min f0 (St)	Min f0 in R (Token Min) - mean f0 across all the productions by the speaker (Speaker Mean)	To test the tone targets: "High" in T1 and "Low" in T3	**.709**	.051	**-.500**	-.418
Directional excursion (St)	Max f0 - min f0 in R, positive if min f0 precedes max f0, negative if max f0 precedes min f0	To test the amount of f0 displacement in T2 and T4	-.086	**.601**	.183	**-.716**
Slope (St/s)	Directional excursion/duration between max f0 and min f0 in R	To test the steepness of the slopes of f0 in T2 and T4	-.056	**.629**	.136	**-.726**
Timing of min f0	Duration between R onset and min f0 / Duration of R	To test the timing of min f0 in T2 and T4 by means of its proportional duration in the rime	-.054	**-.533**	.041	**.654**
Timing of max f0	Duration between R onset and Max f0 / Duration of R	To test the timing of max f0 in T2 and T4 by means of its proportional duration in the rime	-.129	**.576**	.087	**-.542**

"C" represents "consonant". "R" represents "rime". "St" represents" semi-tone". "s" represents "second". "/" represents division. Correlations of interest are in bold.

Table 2. Results of pairwise comparisons on the acoustic parameters using Mann-Whitney U test

		Adult Correct vs. Child Correct (AC vs. CC)					Child Correct vs. Child Incorrect (CC vs. CI)				
Tone	Acoustic Parameter	U	N	z	p	r	U	N	z	p	r
T1	Syllable duration	381	62	-0.98	0.325	0.12	179	50	-0.83	0.406	0.12
	Pitch shift	206	62	-3.53	**0.000****	**0.45**	62	50	-3.57	**0.000****	**0.50**
	Height of mean f0	172	62	-4.03	**0.000****	**0.51**	185	50	-0.69	0.490	0.10
	Height of min f0	119	62	-4.80	**0.000****	**0.61**	117	50	-2.28	**0.022***	**0.32**
T2	Syllable duration	243	53	-1.88	0.061	0.26	143	39	-0.06	0.949	0.01
	Directional excursion	217	53	-2.34	**0.019***	**0.32**	38	39	-3.44	**0.001****	**0.55**
	Slope	322	53	-0.47	0.642	0.06	31	39	-3.67	**0.000****	**0.59**
	Timing of min f0	314	53	-0.61	0.542	0.08	13	39	-4.28	**0.000****	**0.69**
	Timing of max f0	294	53	-1.01	0.314	0.14	45	39	-3.31	**0.001****	**0.53**
T3	Syllable duration	65	30	-1.70	0.089	**0.31**	83	35	-1.74	0.082	0.29
	Height of mean f0	30	30	-3.21	**0.001****	**0.59**	88	35	-1.56	0.118	0.26
	Height of min f0	84	30	-0.88	0.378	0.16	68	35	-2.27	**0.023***	**0.38**
T4	Syllable duration	218	45	-0.73	0.465	0.11	57	30	-0.31	0.760	0.06
	Directional excursion	230	45	-0.46	0.648	0.07	8	30	-3.03	**0.002****	**0.55**
	Slope	227	45	-0.53	0.599	0.08	22	30	-2.25	**0.024***	**0.41**
	Timing of min f0	182	45	-1.78	0.075	0.27	22	30	-2.61	**0.009***	**0.48**
	Timing of max f0	183	45	-1.59	0.112	0.24	43	30	-1.13	0.258	0.21

* indicates statistical significance at .05 level and ** indicates statistical significance at .01 level after Bonferroni correction for multiple comparisons. "r" represents the effect sizes. Medium effect sizes are in bold and large effect sizes are underlined and in bold

Table 3. Description of the tones and summary of major findings

Order[1]	Tone	Important Acoustic Discriminant	Physiological Mechanisms for Producing the Tone	Perceptual Accuracy		Major Findings on the Selected Parameters	
				Adult Tone	Child Tone	Child Correct Tone	Child Incorrect Tone
1	T4	Falling F0: Steep negative slope	Decrease F0 by relaxation of CT. For speakers with low pitch range, SH is activated to get to the final low F0.	98%	76%	Same as AC	Shallower slopes, reduced excursion span, reach minimum F0 earlier than CC, but few CI productions
2	T1	High and level F0	Increase and maintain CT activity	96%	79%	Not as high nor as level as AC	Not as level nor as high as CC
3	T2	Rising F0: Steep positive slope	Activate SH (and likely other strap muscles) for the initial F0 fall followed by suppressing SH and activating CT	96%	70%	Narrower excursion span and higher onset F0 than AC	Vary from reduced positive excursions and slopes to negative excursions and slopes, reaching minimum F0 later and maximum F0 earlier
4	T3	Low F0: Low minimum and mean F0	High activity of SH (and likely other strap muscles) to get to the low F0, then decrease activity of SH and increase activity of CT	83%	44%	Few correct productions, higher mean F0 than AC	Higher minimum F0 than CC, many CI productions with very different F0 shapes

Order of acquisition of the four tones from the most adult-like to the least adult like. “CT” represents the cricothyroid muscle, “SH” stands for the sternohyoid muscle, “AC” represents adult correct production, “CC” stands for child correct productions, “CI” stands for child incorrect productions.

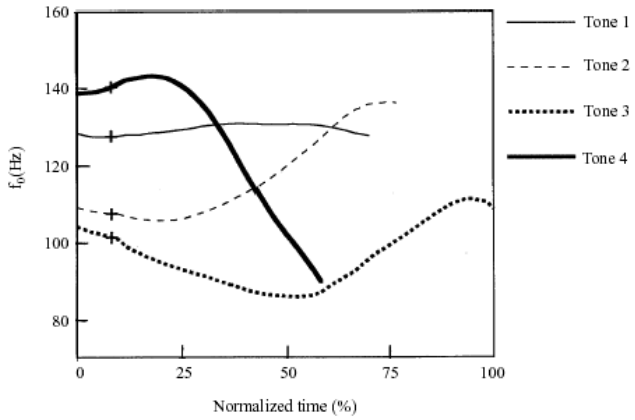

Figure 1. Mean F0 Contours of the Four Mandarin Tones Produced in Isolated Syllables by Adult. Mean F0 contours of 48 tokens of the four Mandarin tones produced in the syllable /ma/ in isolation by eight male adults. Time is normalized with all tones plotted with their average duration proportional to the average duration of Tone 3.

Reprinted from Journal of Phonetics, 25, Y. Xu, Contextual Tonal Variations in Mandarin, 61-83, Copyright (2011), with permission from Elsevier.

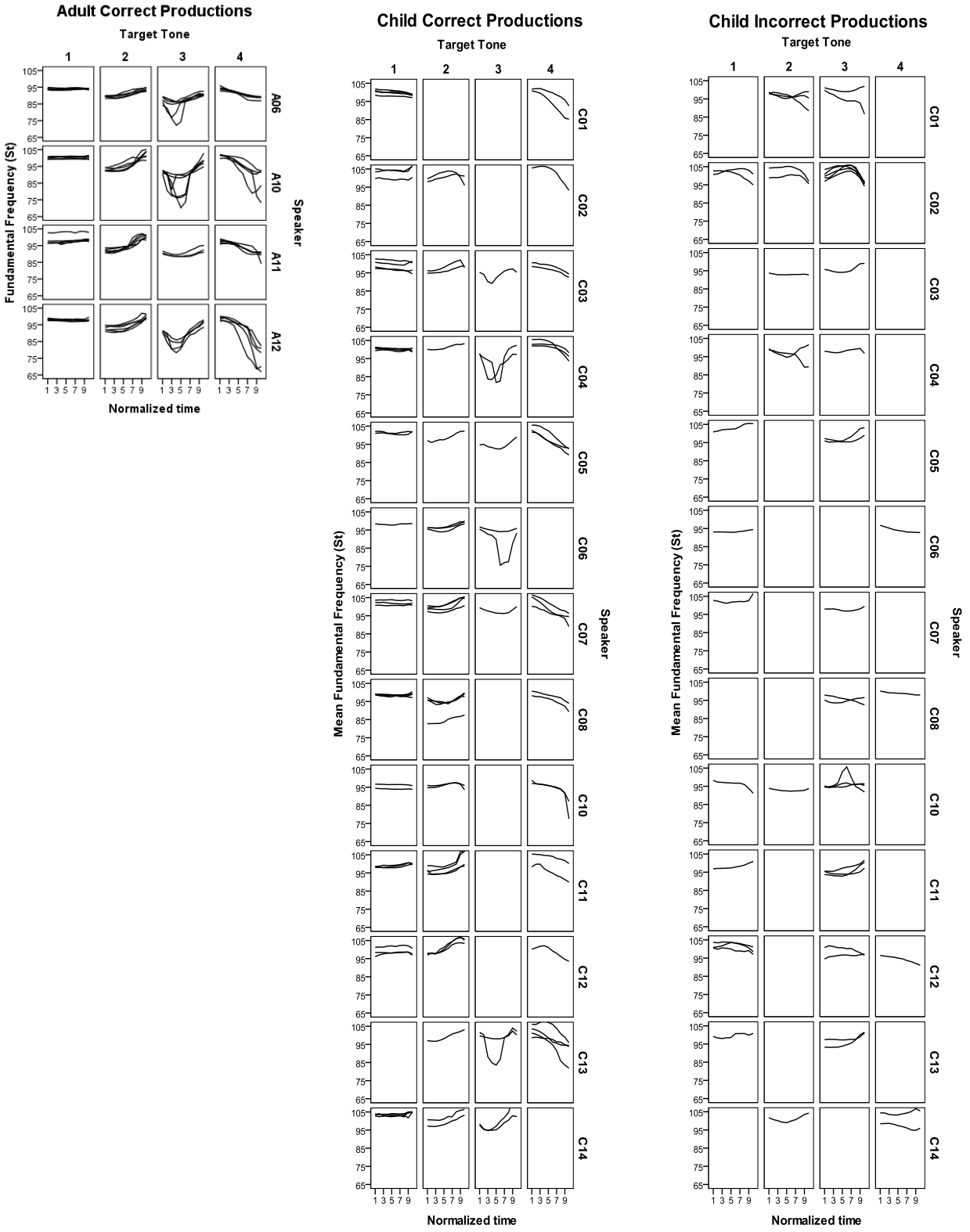

Figure 2. F0 plots of the tones produced by all speakers by accuracy group.

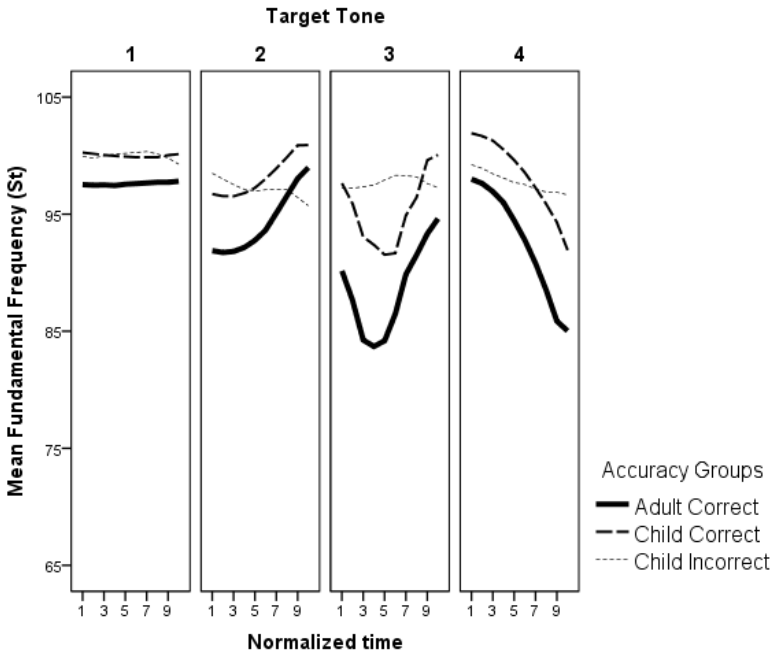

Figure 3. Average F0 plots of the four tones by accuracy group

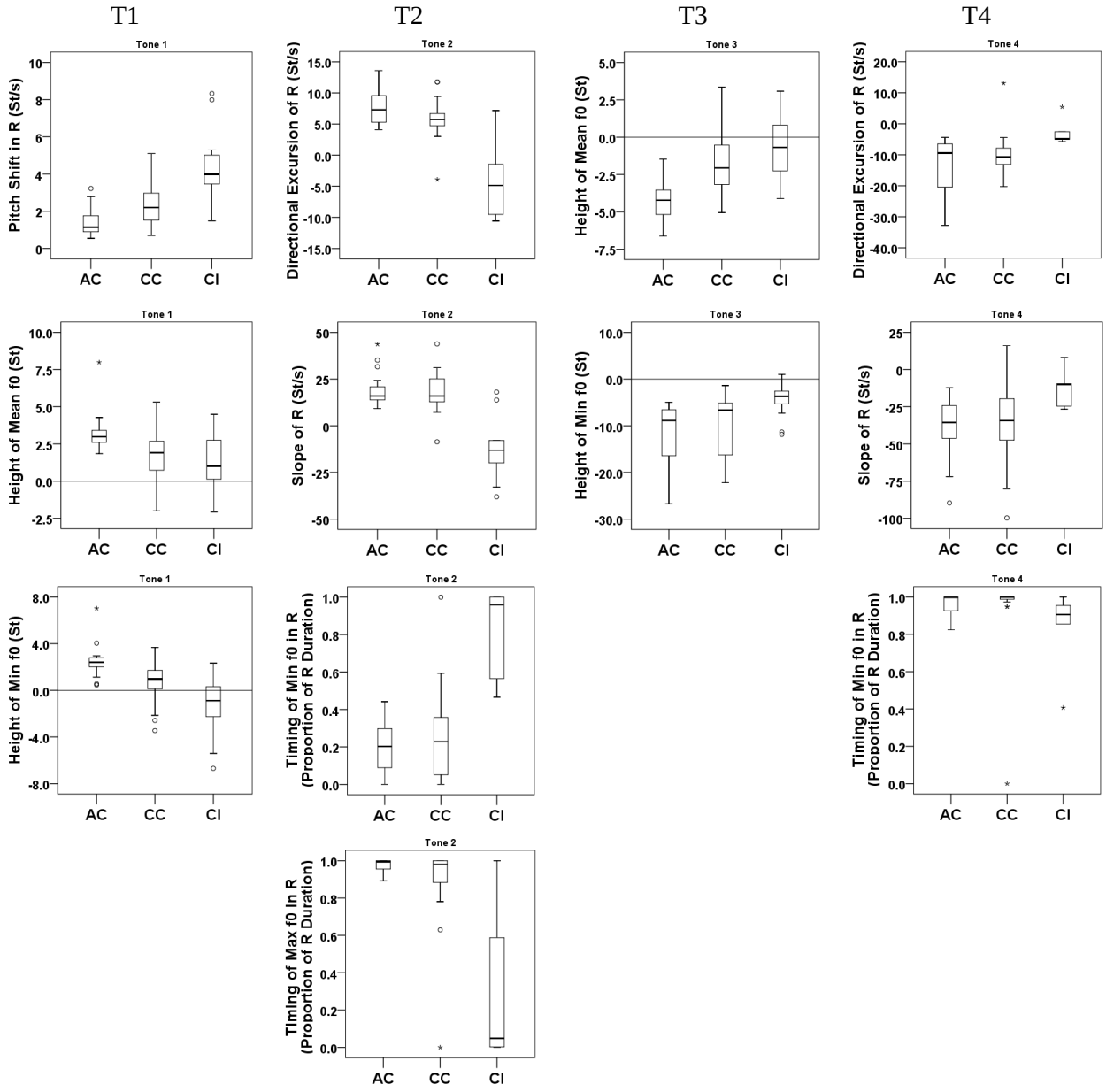

Figure 4. Box plots of the distributions of the acoustic parameters in the three accuracy groups. The dotted horizontal lines at 0 represent mean f0 of all tones produced by the speaker (Speaker Mean). “AC”, “CC” and “CI” represent “Adult Correct”, “Child Correct” and “Child Incorrect”, respectively.

References

Boersma, P., & Weenink, D. (1992). *Praat version 4.1.6*. University of Amsterdam, Netherlands:

Chao, Y. R. (1973/1951). The cantian idiolect: An analysis of the Chinese spoken by a twenty-eight-month-old child. In C. A. Ferguson, & D. I. Slobin (Eds.), *Studies of child language development* (pp. 13-33). New York: Holt, Rinehart & Winston.

Clumeck, H. (1977). Topics in the acquisition of Mandarin phonology: A case study. *Papers and Reports on Child Language Development, 14*(December), 37-73.

Clumeck, H. (1980). The acquisition of tone. In G. H. Yeni-Komshian, J. F. Kavanaugh & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. production* (pp. 257-275). New York: Academic Press.

Clumeck, H. V. (1977). *Studies in the acquisition of Mandarin phonology.* Unpublished doctoral dissertation, University of California, Berkerley.,

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York: Oxford University Press.

Erickson, D. M. (1976). *A physiological analysis of the tones of Thai.* Unpublished Ph.D., The University of Connecticut, United States -- Connecticut.

Erickson, D. (2011). Thai tones revisited. *Journal of the Phonetic Society of Japan,* 15, 1-9.

Erickson, D. (1993). Laryngeal muscle activity in connection with Thai tones. *Research Institute of Logopedics and Phoniatrics Annual Bulletin, 27,* 135-149.

Fu, Q. J., & Zeng, F. G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing, 5*, 45-57.

Gårding, E., Kratochvil, P., Svantesson, J., & Zhang, J. (1986). Tone 4 and tone 3 discrimination in modern standard Chinese. *Language & Speech, 29*(3), 281-293.

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

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica, 33*, 353-367.

Honda, K. (1995). Laryngeal and extra-laryngeal mechanisms of F0 control. In K. S. Harris, F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues : For Katherine Safford Harris* (pp. 215-232)

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech, 42*(4), 401-411.

House, A., S, & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America, 25*(1), 105-113.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones* Cambridge University Press New York.

Hua, Z. (2002). *Phonological development in specific context: Studies of chinese-speaking children*. Clevedon, England: Multilingual Matters Limited.

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

Koike, K. J. M., & Asp, C. W. (1981). Tennessee test of rhythm and intonation patterns. *Journal of Speech and Hearing Disorders, 46*(1), 81-87.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*(2), 185-199.

Loeb, D. F., & Allen, G. D. (1993). Preschoolers' imitation of intonation contours. *Journal of Speech and Hearing Research, 36*(1), 4-13.

Luo, X., & Fu, Q. J. (2004). Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *Journal of the Acoustical Society of America, 116*(6), 3659-3667.

Massaro, D. W., Cohen, M. M., & Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13*, 267-290.

Nolan, F. (2003). "Intonational equivalence: An experimental evaluation of pitch scales," in The 15th International Congress of Phonetic Sciences (Barcelona), pp. 771-774.

Ohala, J. J., & Ewan, W. G. (1973). Speed of pitch change. *The Journal of the Acoustical Society of America, 53*(1), 345.

Sagart, L., Hallé, P., Boysson-Bardies, B. d., & Arabia-Guidet, C. (1986). Tone production in modern standard Chinese : An electromyographic investigation. *Cahiers De Linguistique - Asie Orientale, 15*(2), 205-221.

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics, 18*(3)

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

Snow, D. (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language, and Hearing Research, 41*(3), 576-587.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics, 7*, 71-79.

Tagliaferri, B. (2005). *Paradigm* [Perception Research Systems Inc.] Retrieved from www.perceptionresearchsystems.com

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *The Journal of the Acoustical Society of America, 70*(2), 350-355.

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

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research, 48*(5), 1065-1079.

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

Wong, P., & Strange, W. (submitted). Development of disyllabic Mandarin lexical tone production in Mandarin-speaking children residing in the U.S.

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

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*(4), 179-203.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics, 27*(1), 55-105.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America, 111*, 1399-1413.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33*(4), 319-337.

Xu, Y. (2005-2010). *ProsodyPro.praat*, from http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/.

Xu, Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics, 37*(4), 502-520.