Mothers do not enhance tonal contrasts in child-directed speech: Perceptual and acoustic evidence from child-directed Mandarin lexical tones

Puisan Wong[a)]

Division of Speech and Hearing Sciences, The University of Hong Kong, Room 757, Meng Wah Complex, Pokfulam, Hong Kong

(Received 24 December 2017; revised 18 April 2018; accepted 20 April 2018; published online 29 May 2018)

Prosodically, child-directed speech typically has a higher pitch and more varied pitch contours. Studies that have examined acoustic differences between child-directed and adult-directed vowels and consonants have reported mixed results and proposed two hypotheses explaining the function of the acoustic modifications in child-directed speech. The hyperarticulation hypothesis suggests that mothers enhance the phonemic contrasts in child-directed speech to facilitate speech and language acquisition in children. The pragmatic hypothesis claims that the acoustic differences between child-directed and adult-directed speech result from mothers' expression of affective emotions towards young children. In tone languages, pitch is used at the syllable level to make lexical contrasts and at the utterance level to serve pragmatic functions. This study compared the perceptual clarity and acoustic characteristics of adult-directed and child-directed Mandarin tones to test the two hypotheses. 1648 child-directed and adult-directed tones produced by 20 mothers in monosyllabic and disyllabic words were low-pass filtered to eliminate segmental information and presented to five judges for tone identification. Child-directed tones were identified with poorer accuracy than adult-directed tones. Acoustic analysis revealed that child-directed tones, regardless of tone type, were produced with higher pitch and more positive slopes than adult-directed tones. The findings did not support the hyperarticulation hypothesis. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5037092

[SHF] Pages: 3169–3183

I. INTRODUCTION

People in various cultures and countries intuitively use a special register, called child-directed speech, when talking to young children. This speech register is different from adult-directed speech in various acoustic aspects. Studies that have found expanded vowel space in child-directed speech have proposed the hyperarticulation hypothesis, which suggests that mothers enhance the phonetic contrasts in phonemes to facilitate speech and language acquisition in children (Kuhl *et al.*, 1997; Liu *et al.*, 2003; Martin *et al.*, 2015). Studies that found smaller vowel space or more overlapping phonetic categories among the vowels in child-directed speech have proposed the pragmatic hypothesis, which suggests that the acoustic modifications in the child-directed register are the result of mothers' expression of affective emotions (e.g., happiness, love, care) to engage or maintain the attention of young children (Benders, 2013; Englund and Behne, 2006; McMurray *et al.*, 2013). This study aimed to test whether mothers hyperarticulate the tones in child-directed speech by comparing the acoustic characteristics and perceptual clarity of Mandarin tones in child-directed and adult-directed speech.

A. Acoustic characteristics of prosodic properties of child-directed speech

Previous studies that have examined the prosodic characteristics of child-directed speech suggest some universal acoustic characteristics in child-directed speech. Adults and children alike tend to produce speech with higher pitch (Fernald and Simon, 1984; Garnica, 1977), larger pitch range (Cooper and Aslin, 1994; Fernald and Kuhl, 1987), and more varied prosody (Fisher and Tokura, 1996; McMurray *et al.*, 2013) when speaking to young children. These prosodic characteristics have been reported in a range of languages, such as English, Italian, French, German, and Mandarin (Fernald *et al.*, 1989; Grieser and Kuhl, 1988).

B. Acoustic characteristics of phonetic segments in child-directed speech

Most studies that have examined the acoustic characteristics of the phonetic segments in child-directed speech have focused on vowel productions and compared the vowel space formed by the three corner vowels—/i, u, a/—in adult-directed and child-directed speech. Larger vowel spaces in child-directed speech than in adult-directed speech have been reported in a number of languages including British English, American English, Russian, Swedish, and Mandarin, leading authors to suggest that adults hyperarticulate the vowels in child-directed speech to facilitate vowel discrimination and acquisition in young children (Burnham *et al.*, 2002; Kuhl *et al.*, 1997; Liu *et al.*, 2003; Uther *et al.*, 2007). More compellingly, some studies have found positive correlations between the size of mothers' vowel space and their children's speech and language ability. Liu and colleagues reported that the size of the vowel space of

[a)]Electronic mail: pswResearch@gmail.com

Mandarin-speaking mothers' child-directed speech to their children aged 6 to 12 months during play was not only positively associated with their children's speech discrimination ability at the time of the experiment (Liu *et al.*, 2003), but was also positively correlated to the children's receptive vocabulary at age 5 years (Liu, 2014). Hartman *et al.* (2017) found similar results with English-speaking mothers and reported that the vowel space of the mothers' child-directed speech to their 18-month-old children during unstructured play sessions predicted the children's vocabulary scores in three norm-referenced vocabulary tests when the children reached 2 years of age.

However, not all studies that have measured vowel space in child-directed and adult-directed speech support the hyperarticulation account. Benders (2013) and Englund and Behne (2006) found smaller vowel spaces in child-directed speech than in adult-directed speech in Dutch and Australian English. Studies that have examined the acoustic properties of a larger set of vowels, including both the corner vowels and the internal vowels, have found more overlapping vowel categories and larger within-category variability in child-directed speech, showing little evidence of hyperarticulation of vowels (Cristia and Seidl, 2014; McMurray *et al.*, 2013). Given that formant frequencies are related to fundamental frequency (F0)/ pitch and, therefore, changes in prosody may result in formant frequency changes, these studies proposed a pragmatic account, suggesting that the expanded vowel spaces observed in child-directed speech were not a result of mothers' phonetic enhancement of vowel contrasts but rather a by-product of mothers' use of prosody to express affective emotions to children in child-directed speech (Benders, 2013; McMurray *et al.*, 2013; Trainor *et al.*, 2000). Using a machine learning paradigm, Martin *et al.* (2015) further challenged the hyperarticulation hypothesis by showing that vowels in child-directed speech were identified with lower accuracy than vowels in adult-directed speech.

Several studies that have compared voice onset times in voiced and voiceless stop consonants in child-directed and adult-directed speech have also found inconsistent results. Some studies reported no difference in voice onset time between voiced and voiceless stops in child-directed and adult-directed speech (Malsheen, 1980), some found shorter voice onset times in child-directed speech (Sundberg and Lacerda, 1999), some found significantly longer voice onset times in voiceless stops in child-directed speech (Englund, 2005), and still others found longer voice onset time in both child-directed and adult-directed speech or more overlapping voice-onset time between voiced and voiceless stops in child-directed speech than in adult-directed speech (Baran *et al.*, 1977; Sundberg and Lacerda, 1999; McMurray *et al.*, 2013). Bohn (2013) compared vowel space, voice onset time of stops, amplitude peak for sibilant fricatives, and syllable, vowel, and fricative durations in Danish child-directed and adult-directed speech. Contrary to expectation, Bohn found longer syllable durations and more contrastive spectral properties in the vowels and consonants in adult-directed speech than in child-directed speech. Moreover, no consistent relation was found between the acoustic properties in mothers' child-directed speech and children's receptive vocabulary size. These findings pose additional challenges to the hyperarticulation hypothesis.

C. Phonemic and prosodic use of pitch in tone languages

Examination of the use of pitch in child-directed and adult-directed speech in tone languages may provide unique insights into the hyperarticulation and pragmatic hypotheses of child-directed speech. Pitch in tone languages is used for both phonemic and prosodic purposes. At the syllabic level, pitch is used to make lexical contrasts, while at the utterance level, it is used to express emotions and other pragmatic functions. It was predicted that if the primary purpose of child-directed speech was to enhance the acoustic contrasts of tones, the pitch contours of child-directed tones would be closer to the underlying pitch targets for the tones (e.g., the high level tone would be produced with higher pitch and smaller pitch range), and the tones in child-directed speech would be clearer and perceived with higher accuracy than tones in adult-directed speech. On the other hand, if the purpose of child-directed speech was to serve pragmatic functions, the prosodic use of pitch may compromise the tone identities, resulting in lower perceptual accuracy of tones in child-directed speech than in adult-directed speech.

D. Acoustic properties of child-directed lexical tones

Several acoustic studies have examined child-directed lexical tones. One of them supported the pragmatic hypothesis. Papoušek and Hwang (1991) asked six Mandarin speakers to imagine that they were talking to a two-month-old infant, an adult and a foreign student, and record two sets of sentences with one to ten syllables in a variety of sentence structures, such as statements, questions, requests and exclamatory sentences. Acoustic analysis showed that maximum, minimum and mean F0 values were higher and syllable and utterance duration was longer in child-directed speech than in adult-directed speech, but the range of F0 was comparable in the two speech registers. More rising was found in the tones at the final positions of the utterances in child-directed speech than in adult-directed speech. Based on visual inspection of the pitch contours, Papoušek and Hwang reported that the tone contours were less distinctive in child-directed speech than in adult-directed speech. They concluded that pitch modifications in child-directed speech primarily serve prosodic functions and compromise tone identities. However, no perceptual judgment was carried out to verify the findings.

Other acoustic studies on child-directed tones have reported contradictory results and supported the hyperarticulation hypothesis. Liu *et al.* (2007) specifically tested whether Mandarin tone contours were distorted in child-directed speech. Sixteen mothers of 10- to 12-month-old children were provided with a set of toys and pictures that represented 12 disyllabic words with the four Mandarin tones in the first syllable followed by Tone 1 (T1, High Level) in the second syllable. Child-directed speech was collected when the mothers were playing with the children with the set of toys and pictures, while adult-directed speech was collected when the mothers described to the experimenter their children's interest using the target words. Based on the findings that tone duration was longer, mean F0 was higher, and F0 range was larger in child-directed speech than adult-directed speech, while the order of mean F0, F0 range, and tone duration of the four tones was maintained in child-directed and adult-directed speech, the authors concluded that mothers enhanced tonal contrasts without distorting the tonal contours. Liu *et al.* (2009) reported similar results using the same stimuli and methods to measure mothers' tone productions to children when the children were age 1 year and again when the children were 5 years old.

Rattanasone *et al.* (2013) also found similar results with Cantonese child-directed speech using different acoustic methods. Twenty-two mothers were asked to produce three monosyllabic words in Tone 1 [T1, High Level, hearing level (HL)], Tone 2 (T2, High Rising, HR), and Tone 4 (T4, Low Falling, LF) in Cantonese while talking to an adult and while playing with their children with a set of provided toys. Data were collected at three time points when the children were between 3 to 12 months of age. F0 at the onset and offset of each tone production were measured and plotted against each other. The centroids of the clusters of the F0 onset-offset functions of the three tones were connected to form two tone triangles, one for child-directed speech and one for adult-directed speech. The height, size, and the length of the sides of the tone triangles were compared. The results showed higher pitch, larger pitch range, and significantly larger tone triangles in child-directed speech than in adult-directed speech. The authors concluded that mothers hyperarticulated the tones when talking to young children.

E. Possible reasons for the discrepant findings in previous studies on child-directed tones

Several reasons may explain the mixed results in previous acoustic studies on child-directed lexical tones. First, other than Papoušek and Hwang (1991), who asked participants to read sentences in child-directed and adult-directed speech, all other acoustic studies used different activities for collecting tones in child-directed and adult-directed speech and did not control the prosodic context of tone productions in the two speech registers. It has been well established that pitch levels and shapes of tones are affected by different prosodic contexts, such as type of utterance, position of the tone in an utterance, and emotion. Ma *et al.* (2006) and Vance (1976) reported that the pitch level of tones dropped continuously from the beginning to the end of a statement. Ma *et al.* (2006) also demonstrated that all Cantonese tones, including the level tones, had a rising pitch contour in the final position of questions. Several studies found that happy emotions were expressed with higher pitch, larger pitch range, and longer durations (Quam and Swingley, 2012; Singh *et al.*, 2002; Trainor *et al.*, 2000). Most studies that examined tone productions in child-directed and adult-directed speech collected adult-directed tones from interviews and child-directed tones from playing with children. Typically, conversations at interviews are conveyed in neutral emotions, whereas conversations during playing with young children may involve different emotions (e.g., excitement, happiness) and pragmatic functions (e.g., questions, requests). Thus, the acoustic differences of the tones between adult-directed and child-directed speech reported in these studies could have been confounded by differences in prosodic contexts, utterance positions, and emotional expressions arising from the use of different activities and utterances in the two conditions.

Second, other than Papoušek and Hwang (1991), who asked participants to produce the same set of sentences in child-directed and adult-directed speech, studies on child-directed lexical tones have not controlled the phonetic contexts of the tones in the two registers. Xu (1997, 2001) showed that pitch contours of the same tone vary greatly depending on phonetic context. To illustrate, Mandarin has four lexical tones. Tone 1 (T1, H), Tone 2 (T2, R), Tone 3 (T3, L), and Tone 4 (T4, F) have a high and level, rising, dipping, and high falling pitch contour, respectively. Figure 1 shows the pitch contours of the four Mandarin tones in monosyllabic isolated words. Figure 2 (adapted from Fig. 6 in Xu, 1997) presents the pitch contours of the same tones in disyllabic words with each panel showing the same tone in the first syllable followed by the four different tones in the second syllable. No contour for T3 (L) followed by another T3 (L) are shown, because this tone combination undergoes a tone sandhi rule and is produced as T2 (R) followed by T3 (L). Figure 3 (adapted from Fig. 3 in Xu, 1997) shows the same 15 disyllabic pitch contours, but reorganized by the tone in the second syllable. As presented, T3 (L) has a

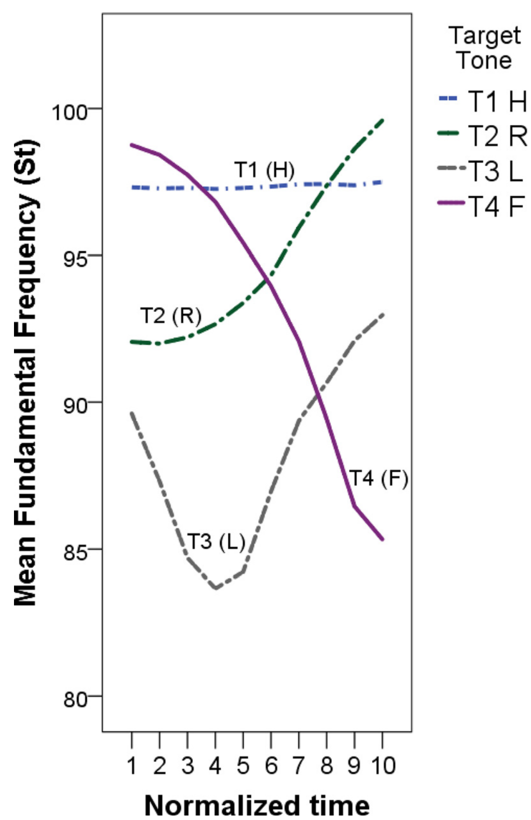

FIG. 1. (Color online) Pitch contours of Mandarin tones in monosyllabic words.

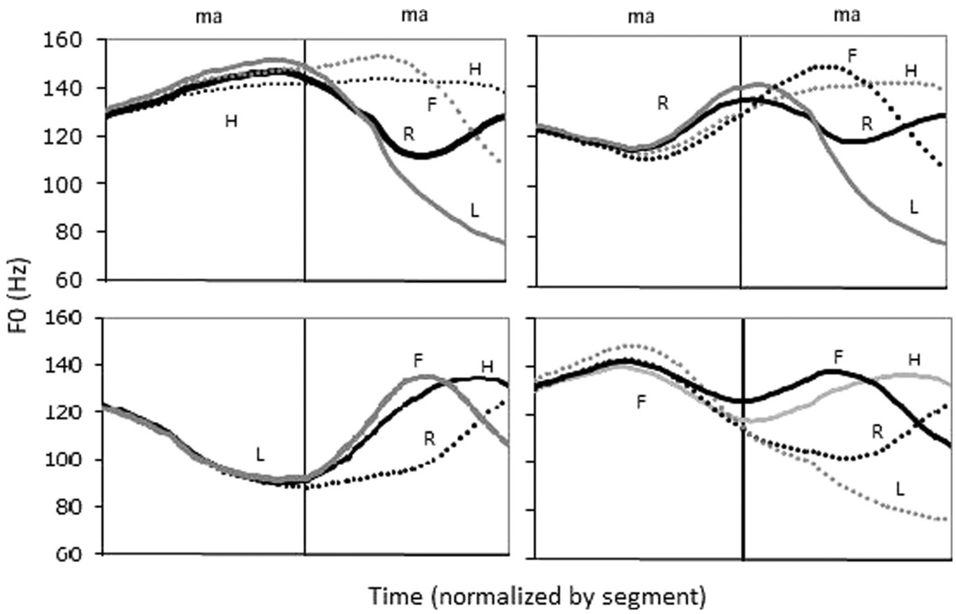

FIG. 2. Pitch contours of Mandarin tones in disyllabic words with the same tone in the first syllable. Adapted from Fig. 6 in Xu (1997) with permission from Elsevier.

dipping contour in isolation (Fig. 1) but a low falling contour in other contexts. Note that when the four tones are produced in disyllabic contexts, the F0 contours of the same tone in the second syllable vary greatly and are affected by the tone in the preceding syllable (see the contours, particularly in the first half of the syllable, in the second syllable in Fig. 3). Thus, the onset F0, the mean F0, minimum F0, maximum F0, and initial F0 of the same tone with different preceding tones are very different. Comparing acoustic measurements of tones produced in different phonetic contexts in child-directed and adult-directed speech could therefore confound the data.

Third, the acoustic parameters selected in previous studies to measure the phonetic differences of the tones between child-directed speech and adult-directed speech and to determine whether the tones were hyperarticulated in child-directed speech might not be the perceptual target for tone identification. Fu and Zeng (2000) have demonstrated that the primary and sufficient cue for tone identification is the pitch contour. In Mandarin, each of the four tones has a unique contour shape. Thus, the shape of the pitch contour is a more important cue for tone perception than the pitch level of the contour (Whalen and Xu, 1992). Because the pitch at the beginning of a syllable is influenced by various factors such as the initial consonant (Wong and Xu, 2007; Xu and Xu, 2003) and the preceding tone (Xu, 2001), the pitch contour in the initial portion of the syllable is less reliable for tone identification than the pitch contour towards the end of the syllable (Khouw and Ciocca, 2007; Xu, 1997, 2001; Xu and Wang, 2001). Thus, the perceptual target or the reliable cue for tone identification lies in the second half of the syllable (Khouw and Ciocca, 2007; Xu, 1997, 2001; Xu and Wang, 2001).

None of the previous studies has compared the shapes of the tone contours in adult-directed and child-directed speech or measured perceptual judgments on adult-directed and child-directed tones. The conclusions that mothers hyperarticulated tones were based on larger tone space (Rattanasone

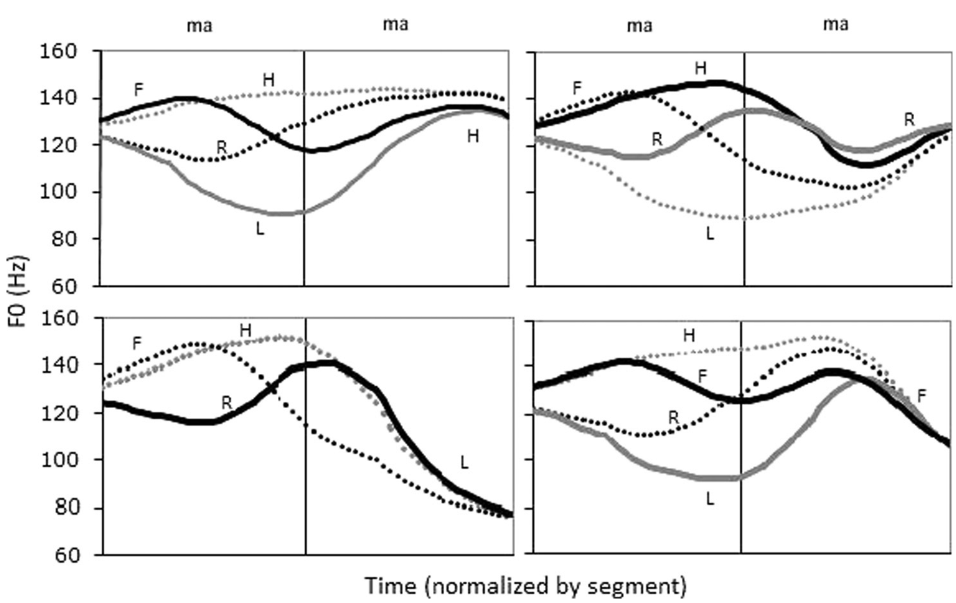

FIG. 3. Pitch contours of Mandarin tones in disyllabic words with the same tone in the second syllable. Adapted from Fig. 3 in Xu (1997) with permission from Elsevier.

et al., 2013), larger pitch range, and higher initial, mean, minimum, and maximum F0 (Liu et al., 2007, Liu et al., 2009) in child-directed and adult-directed speech. However, these acoustic parameters are not the perceptual targets for tone identification (Xu, 1997, 2001; Xu and Wang, 2001) and can vary in different phonetic and prosodic contexts.

The goal of this study was to test the hyperarticulation hypothesis of child-directed speech by examining whether mothers enhanced tonal contrasts in Mandarin tones when speaking to young children. Mandarin was selected because it is the most widely spoken tone language in the world. To control for the effect of phonetic and prosodic context on tone productions, mother-and-child and mother-and-adult dyads were engaged in a picture book reading task. To control for phonetic and prosodic effect on pitch, only target monosyllabic and disyllabic words produced in isolation were analyzed. Acoustic analysis was performed on the three different phonetic contexts of tone productions [i.e., tones in monosyllabic words, tones in the first syllable of disyllabic words (disyllabic tones), and tones in the second syllable of disyllabic words]. Both pitch levels and pitch shapes were compared in adult-directed and child-directed speech. To measure the most reliable perceptual cue of tones and to avoid pitch fluctuations at the beginning of the tone contours due to phonetic effects, acoustic comparisons were focused on the pitch contours of the tones in the second half of the syllable (i.e., the pitch target). To provide detailed information on the acoustic modifications in child-directed and adult-directed tones, the acoustic properties of the tones in child-directed and adult-directed speech were examined. Perceptual accuracy of the tones in child-directed speech and adult-directed speech was compared to determine whether the acoustic modifications in child-directed tones resulted in phonemic enhancement. The specific aims were (1) to characterize the acoustic properties of monosyllabic and disyllabic Mandarin tones in child-directed speech and (2) to investigate whether mothers hyperarticulated the Mandarin tones in child-directed speech to facilitate tone perception.

II. METHOD

The methods largely followed the procedures in Wong and Leung (2018), and Wong et al. (2017) and were approved by the Human Research Protection Program at the City University of New York and the Behavioral and Social Sciences Institutional Review Board at the Ohio State University.

A. Participants

Twenty mother-child dyads participated in the study. The mothers were native Mandarin speakers between 28 and 48 years of age. Mandarin was their strongest and home language and the language they used exclusively with their children. All mothers were from middle-class families. Two of them received an associate's degree, while the other mothers attained a bachelor's degree or higher.

The 20 children (6M and 14F) were between the age of 2;5 and 5;4. Children in this age range were selected because previous studies reported that children in this age range have not acquired the production of monosyllabic (Wong et al., 2005; Wong, 2012a; Wong, 2012b; Wong, 2013) or disyllabic Mandarin tones (Wong and Strange, 2017). All children passed a hearing screening at 1000, 2000, and 4000 Hz at 20 dB HL under headphones using conditioned play audiometry (American Speech-Language-Hearing Association, 1997) and a language screening using the Language Disorder Scale of Preschoolers (LDSP, 學前兒童語言障礙評量表; Lin and Lin, 1994).

B. Stimuli

Table I lists the 38 target words. There were eight monosyllabic words (2 words $\times$ 4 tones) and 30 disyllabic words (2 words $\times$ 15 disyllabic combinations). The words were depicted in color pictures and put into two picture books with different randomized orders.

C. Procedures for tone production collection

Each mother-child dyad participated in a one-hour session in a quiet room. Mothers filled out a questionnaire about their language background and their children's developmental history and language use. One of the two picture books was randomly assigned to them. Mothers were told to label the 38 pictures to the experimenter twice. Then they were asked to read the picture book with their children and name the same pictures two times to their children. During the process, children tended to imitate after the mother, and the child-and-mother dyads sometimes interacted and talked about the pictures. Their productions were audio recorded with 16 bit and 44.1 kHz sampling rate, and only target words produced by the mothers in isolation were selected for analysis.

1. Judges

To determine the clarity of the tones produced by the mothers in the two registers, five (3F, 2M) native Mandarin speakers between the age of 19 and 28 years who used Mandarin as their first and home language rated the tones produced by the mothers. The judges passed a hearing screening and reported no concern or difficulty in speech and language. Three of the judges were undergraduate students, and two were postgraduate students.

2. Stimuli for tone judgment

To familiarize the judges with the tone judgment procedures and to make sure that the judges were able to identify the tones in filtered stimuli, 42 training stimuli were prepared for two training blocks. 12 of them were monosyllabic nonsense words (3 words $\times$ 4 tones) and 30 were disyllabic nonsense words (15 tone combinations $\times$ 2 words). The words were recorded by a female Mandarin-speaker who did not participate in the study.

The mothers' adult-directed and child-directed monosyllabic and disyllabic productions were used for experimental stimuli. To reduce the number of judgment sessions, two productions of each monosyllabic word and the first

TABLE I. Word stimuli.

	Target tone	Chinese word	English meaning	IPA	Pinyin
Monosyllabic Words	T1 (H)	燈	lamp	/təŋ/	deng1
		花	flower	/xwa/	hua1
	T2 (R)	門	door	/məŋ/	men2
		魚	fish	/ɥy/	yu2
	T3 (L)	馬	horse	/ma/	ma3
		碗	bowl	/wan/	wan3
	T4 (F)	大	big	/ta/	da4
		熱	hot	/ɹɤ/	re4
Disyllabic Words	T11	烏龜	turtle	/wu kwei/	wu1 gui1
		西瓜	watermelon	/ɕi kwa/	xi1 gua1
	T12	公園	park	/kuŋ ɥɛn/	gong1 yuan2
		刷牙	brushing one's teeth	/ʂwa ja/	shua1 ya2
	T13	喝水	drinking water	/xɤ ʂwei/	he1 shui3
		鉛筆	pencil	/tɕ^{h}jɛn pi/	qian1 bi3
	T14	雞蛋	egg	/tɕi tan/	ji1 dan4
		拉鏈	zipper	/la ljɛn/	la1 lian4
	T21	毛巾	towel	/mɑu tɕin/	mao2 jin1
		牙膏	toothpaste	/ja kɑu/	ya2 gao1
	T22	廚房	kitchen	/tʂ^{h}u fɑŋ/	chu2 fang2
		蝴蝶	butterfly	/xu the/	hu2 die2
	T23	牛奶	milk	/njou nai/	niu2 nai3
		蘋果	apple	/p^{h}jəŋ kwo/	ping2 guo3
	T24	螃蟹	crab	/p^{h}ɑŋ ɕje/	pang2 xie4
		玩具	toy	/wan tɕy/	wan2 ju4
	T31	餅乾	biscuit	/pjəŋ kan/	bing3 gan1
		剪刀	scissors	/tɕjɛn tɑu/	jian3 dao1
	T32	草莓	strawberry	/ts^{h}ɑu mei/	cao3 mei2
		恐龍	dinosaur	/k^{h}uŋ luŋ/	kong3 long2
	T34	手套	gloves	/ʂou t^{h}ɑu/	shou3 tao4
		眼鏡	glasses	/jɛn tɕjəŋ/	yan3 jing4
	T41	蛋糕	cake	/tan kɑu/	dan4 gao1
		麵包	bread	/mjɛn pɑu/	mian4 bao1
	T42	麵條	noodles	/mjɛn t^{h}jɑu/	mian4 tiao1
		氣球	ballon	/tɕ^{h}i tɕ^{h}jou/	qi4 qiu2
	T43	報紙	newspaper	/pɑu tʂɹ/	bao4 zhi3
		電腦	computer	/tjɛn nɑu/	dian4 nao3
	T44	大象	elephant	/ta ɕjɑŋ/	da4 xiang4
		電話	telephone	/tjɛn xwa/	dian4 hua4

production of the disyllabic words were used for tone judgment. Non-target words, noisy recordings, recordings with overlapping voices, productions with voice break, blank recordings, and words not produced in isolation were excluded. There were a total of 1648 usable productions, including 316 and 204 monosyllabic adult-directed and child-directed productions, respectively, and 596 and 532 disyllabic adult-directed and child-directed productions, respectively. The productions were blocked by register (i.e., adult-directed/child-directed) and speaker to assist the judges' normalization of the speaker's pitch range. Thus, there were 80 blocks of experimental stimuli [20 speakers × 2 syllable contexts (i.e., monosyllabic/ disyllabic context) × 2 registers (adult-directed/child-directed)].

Following the methods used in previous studies such as Wong and Chan (2018) and Wong *et al.* (2017), to control for lexical bias in tone judgment, all training and adult-directed tone production stimuli were low-pass filtered at 400 Hz to eliminate the segmental information in the productions. Because child-directed stimuli usually were produced with higher fundamental frequencies, they were low pass filtered at 500 Hz. To control for relative intensity of tones, the filtered stimuli were normalized to 70 dB.

D. Procedures for tone judgment

Tone rating was carried out in a quiet room. The judges attended eleven 1–2 h sessions to rate the tones at their own pace. All experimental trials were presented using a customized MATLAB program. Monosyllabic tones were rated before disyllabic tones. The training block was presented before the experimental blocks. Adult-directed and child-directed experimental blocks were randomized, but each session had a mixed of adult-directed and child-directed blocks. The trials within the experimental and training blocks were presented in random order. The judges listened to the stimuli over headphones and selected the tone by clicking the corresponding button on the screen out of four buttons marked with the tone numbers and visual displays of the tone contours. Because the judges needed to rate the tones on filtered stimuli without lexical support, they were allowed to listen to the stimuli as many times as they needed to. All judges re-rated six blocks of the monosyllabic productions and three blocks of disyllabic production to allow determination of intrajudge reliability.

E. Acoustic analysis

Acoustic analyses were performed on the recorded adult-directed and child-directed tones produced by the mothers. The method largely followed the methods in Wong (2012a), Wong and Chan (2018), and Wong *et al.* (2017), which have been found to successfully discriminate correct and incorrect tone productions (Wong, 2012a; Wong *et al.*, 2017) and normalize the intrinsic pitch differences among speakers of different ages (Wong, 2012a; Wong *et al.*, 2017) and genders (Wong and Chan, 2018).

1. Segmentation and vocal pulse checking

Segmentation was performed on unfiltered stimuli. Each syllable was manually segmented into two sections: the initial section and the tone section using PRAAT (Boersma and Weenink, 2014). The initial section began at the onset of the initial voiced and voiceless consonants and ended at the end of the fourth regular vocal cycle (i.e., glottal pulse) in the waveform of the syllable, which was also the beginning of the tone section. The tone section ended at the beginning of the last four vocal cycles in the syllable. Leaving four cycles before and after the tone section was to avoid pitch lowering when getting mean pitch at tone onset and offset in the tone section. Thus, the initial section included initial voiceless consonants, the irregular vocal cycles preceding the tone, if any, and the first four regular vocal cycles of the vowel or the initial voice consonants. The tone section included all the vocal cycles in the syllable except the first and the last four vocal cycles.

2. Acoustic parameters

Segmentations obtained from the unfiltered stimuli were applied to the filtered sound files for pitch analysis. Each marking of the vocal cycles in the tone section generated by PRAAT was manually checked for accuracy. Duration of the tone section and the F0 at each 1/20 of the tone section were computed using a customized script (Prosody Pro 3.1, Xu 2005–2016). The F0 measurements were converted to semitones (St), a logarithmic pitch scale in which equal increments represent roughly equivalent sensational response by the human ear (Russo and Thompson, 2005). Ten acoustic parameters were selected (Table II). Six of them were for characterizing the pitch level of the tone contours. They included the initial pitch (i.e., pitch at time point 1), the pitch at the mid-point (i.e., pitch at time point 11, which was also the initial pitch of the pitch target), and the final pitch (i.e., pitch at time point 20) of the whole tone contour, and the minimum, maximum, and mean pitch in the second half of the syllable (i.e., between time points 11–20), where the pitch target (i.e., the reliable cue for tone perception) lies. To adjust for individual differences in intrinsic pitch, the six pitch level measurements were converted to six "Pitch Height" measures by computing the pitch level relative to the speaker's vocal pitch [i.e., subtracting the mean pitch of all the tones produced by the same speaker (i.e., the Speaker Mean Pitch) from the pitch level measurements]. This method has been shown to successfully normalize the pitch differences between children and adults (Wong, 2012a;

TABLE II. Definition of acoustic terms and acoustic parameters. Note: "St" stands for semi-tone; "ms" stands for millisecond; 100% indicates that the parameters were measured with reference to the whole pitch section; 50% indicates that the parameters were measured from the second half of the syllable.

Acoustic parameter		Definition
Speaker Mean Pitch (St)		Average pitch of the adult-directed productions of the speaker
Pitch Height measures	Initial Pitch Height 100% (St)	Pitch at tone onset (i.e., time point 1) minus Speaker Mean Pitch.
	Mid Pitch Height 100% (St)	Pitch at the midpoint of the tone (i.e., time point 11) minus Speaker Mean Pitch. This was also the pitch height at the onset of the pitch target in the second half of the tone.
	Final Pitch Height 100% (St)	Pitch at tone offset (i.e., time point 20) minus Speaker Mean Pitch
	Min Pitch Height 50% (St)	Minimum pitch in the second half of the tone minus Speaker Mean Pitch
	Max Pitch Height 50% (St)	Maximum pitch in the second half of the tone minus Speaker Mean Pitch
	Mean Pitch Height 50% (St)	Mean Pitch in the second half of the tone minus Speaker Mean Pitch
Pitch Range 50% (St)		The absolute value of the difference between Max Pitch 50% and Min Pitch 50%. This was to index the degree of pitch fluctuation in the tonal target.
Pitch Range 100% (St)		The absolute value of the difference between the maximum and minimum pitch in the whole tone contour. This was to index the degree of pitch fluctuation in the whole tone contour.
Slope 50% (St/ms)		Directional Excursion in the second half of the tone divided by the duration between Max Pitch 50% and Min Pitch 50%. This measured the speed of pitch change.
Tone Duration 100% (ms)		Duration of the voiced section in the syllable, including all the pitch cycles in the vowels and the voiced consonants

Wong *et al.*, 2017) and male and female speakers (Wong and Chan, 2018). For example, "Initial Pitch Height" was obtained by measuring the initial F0 of the tone, converting it to semi tones, and then subtracting the Speaker Mean Pitch from it. A negative value indicated that the initial pitch of the tone was lower than the mean pitch of the speaker. In addition, the pitch range in the second half of the tonal contour was measured to examine the amount of pitch fluctuation in the tone target. It was calculated by subtracting the minimum pitch from the maximum pitch in the second half of the syllable. The pitch range of the whole tone contour was computed by subtracting the minimum pitch from the maximum pitch of the whole tone contour. This was to measure pitch fluctuation in the tone contour in the syllable. The pitch slope, which was calculated by dividing the Pitch Range in the second half of the syllable by the duration between the maximum and minimum pitch in the second half of the syllable, was used to compare the shapes of the tone targets. A positive slope denoted a rising pitch contour, and a negative slope indicated a falling pitch contour. Larger positive slopes represented steeper rising pitch, while more negative slopes indicated steeper falling pitch. The duration of the whole tone contour was also measured. Table II lists the acoustic parameters and their definitions.

III. RESULTS

A. Acoustic characteristics of child-directed tones

In this section, first, pitch contours of the tones in child-directed and adult-directed speech produced in different syllable contexts, including monosyllabic words, and the first and second syllables of disyllabic words were compared visually. Second, statistical tests were performed to compare the ten acoustic parameters listed in Table II in different syllable contexts and the two speech registers. Figures 4–6 show the time-normalized average pitch contours of monosyllabic and disyllabic Mandarin tones in child-directed speech and adult-directed speech. By visual inspection, in most cases, the contours of child-directed and adult-directed tones appear to parallel each other.

Statistical tests were performed to determine whether the acoustic measures of the tones were different in different syllable contexts and different speech registers. For each of the acoustic parameters listed in Table II, except Speaker Mean Pitch, a three-way repeated measure analysis of variance (ANOVA) was performed using syllable context, register, and tone as within-subject factors and the acoustic measure as the dependent variable. As expected, significant main effects of tone were found in all comparisons (all $ps < 0.001$, $\eta p^2 = 0.815$–0.928), indicating that the acoustic parameters differed for different tones. Table III shows the results of the main effects of syllable, the main effects of register, the interaction effects between register and tone, and the interaction effects between register and syllable after correcting for multiple comparisons. No significant interaction between syllable, tone and register was found, and the results are, therefore, not included in the table.

Significant main effects of register were found for all acoustic measures except duration of the tone contour. Other than tone duration, all the acoustic measures were significantly higher or larger in child-directed speech than in adult-directed speech. Though significant interactions between syllable and register were found for seven of the ten acoustic measures, results of pairwise comparisons consistently showed significantly higher or larger values in child-directed speech than in adult-directed speech in all syllable positions. Taken together, the findings indicated no tone duration difference between child-directed and adult-directed speech in any syllable position. For all other acoustic parameters except pitch range, though the magnitude of difference between child-directed speech and adult-directed speech may not be the same in different syllable positions, acoustic measures were higher or larger in child-directed speech than in adult-directed speech in all syllable positions. Pairwise comparisons showed that these acoustic parameters differed significantly between adult-directed and child-directed speech in some syllable contexts but not in others.

No significant interaction was found between register and tone for seven of the ten acoustic parameters, indicating that these acoustic parameters, including pitch slope and pitch range, were significantly higher and larger in child-

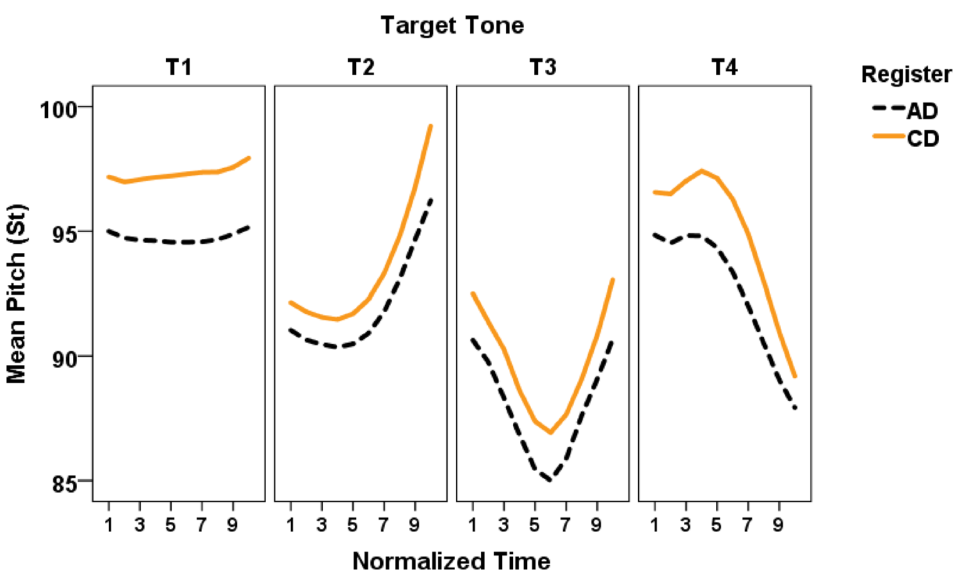

FIG. 4. (Color online) Mean pitch contours of adult-directed and child-directed tones in monosyllabic words.

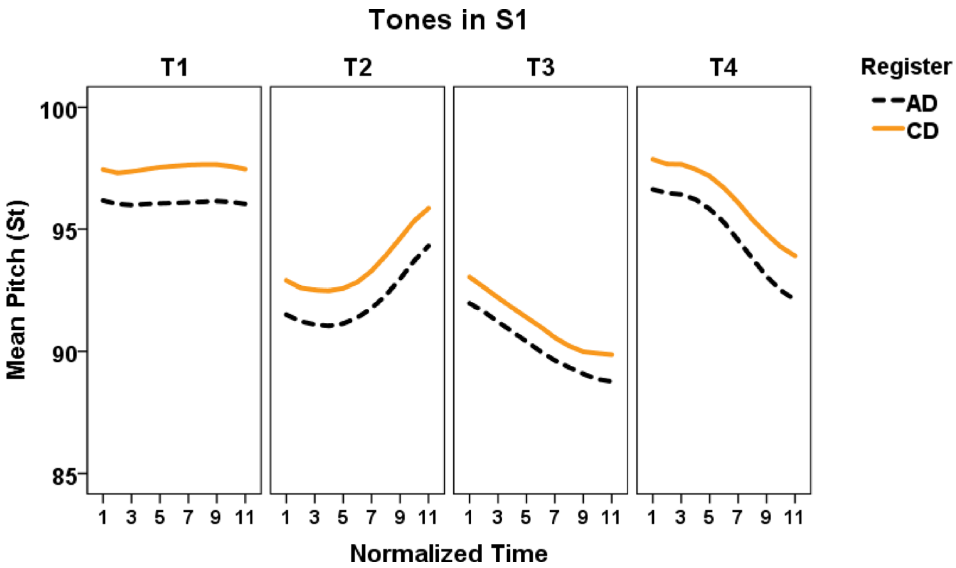

FIG. 5. (Color online) Mean pitch contours of adult-directed and child-directed tones in the first syllable of disyllabic words.

directed speech than in adult-directed speech, regardless of tone type. Significant interactions were found between register and tone on the other three acoustic parameters (i.e., initial pitch height, mid pitch height, and duration). However, pairwise comparisons after correcting for multiple comparisons showed that all tones had higher initial and mid pitch in child-directed speech than in adult-directed speech, while no significant duration difference was found between child-directed and adult-directed speech for any tone. Together, the findings showed that regardless of tone type, tone duration did not differ significantly between child-directed and adult-directed speech, and that all tones, regardless of their tone level or tone shape, were produced with higher pitch, larger pitch range and more positive slopes in child-directed speech than in adult-directed speech.

No interaction among syllable, tone and register was found in any of the ten acoustic parameters.

B. Effect of syllable contexts

As shown in Table III, there was a significant main effect of syllable on all ten acoustic parameters, indicating that the acoustic characteristics of the tones differed significantly in different syllable contexts. Results of pairwise comparisons with Bonferroni adjustments for multiple comparisons showed a general trend for acoustic measures to be higher or larger in monosyllabic words than in the first and second syllables in disyllabic words, though the initial pitch height and pitch range of the whole tone contours were not significantly different in monosyllabic words and in the second syllable of disyllabic words.

C. Clarity and error patterns of adult-directed and child-directed tones

To determine if mothers produced the tones more clearly in child-directed speech, perceptual accuracy of the tones in child-directed and adult-directed speech in different syllable contexts were compared. First, kappa statistics were used to determine the level of agreement among the five judges, between each pair of judges, and within each judge on their tone ratings. Next, perceptual accuracy and error patterns for the tones in child-directed and adult-directed speech in different syllable contexts were examined. Finally,

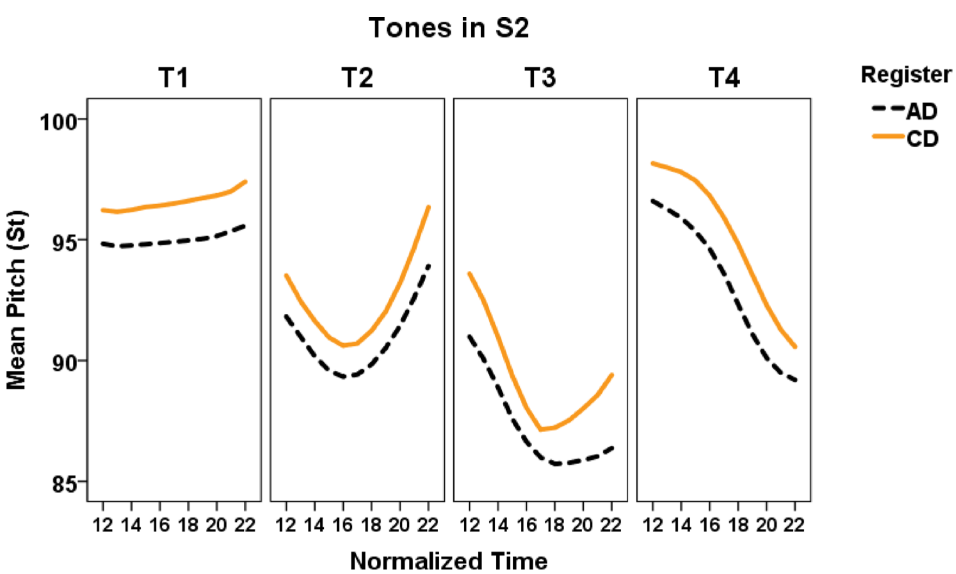

FIG. 6. (Color online) Mean pitch contours of adult-directed and child-directed tones in the second syllable of disyllabic words.

TABLE III. Acoustic characteristics and perceptual accuracy of adult-directed and child-directed tones. "100%" indicates acoustic measurements taken from the whole tone contour, while "50%" indicates measurements obtained from the second half of the tone contour. "MT," "S1," "S2," "ADS," and "CDS" stand for "tones in monosyllabic words," "the tone in the first syllable of disyllabic words," "the tone in the second syllable of disyllabic words," "adult-directed speech," and "child-directed speech," respectively. "n.s." indicates a non-significant result. There was no significant interaction effect of register × tone × syllable on tone accuracy or any of the ten acoustic parameters.

Effect	Acoustic parameters	F-test	Pairwise comparison[a]
Main effect of syllable	Initial Pitch Height 100%	$F(1.16, 22.02) = 25.62$ $p < 0.001$, $\eta p^2 = 0.57$	MT = S2 > S1 ps < 0.001
	Mid Pitch Height 100%	$F(2, 38) = 18.85$ $p < 0.001$, $\eta p^2 = 0.50$	MT > S1 = S2 ps = < 0.001 − 0.001
	Final Pitch Height 100%	$F(1.20, 22.88) = 223.53$ $p < 0.001$, $\eta p^2 = 0.92$	MT > S1 > S2 ps = < 0.001 − 0.001
	Min Pitch Height 50%	$F(1.23, 23.30) = 15.01$ $p < 0.001$, $\eta p^2 = .44$	MT = S1 > S2 ps < 0.001
	Max Pitch Height 50%	$F(1.22, 23.09) = 294.32$ $p < 0.001$, $\eta p^2 = 0.94$	MT > S2 > S1 ps = < 0.001 − 0.023
	Mean Pitch Height 50%	$F(1.25, 23.72) = 126.80$ $p < 0.001$, $\eta p^2 = 0.87$	MT > S2 > S1 ps = < 0.001 − 0.001
	Pitch Range 50%	$F(2, 38) = 148.78$ $p < 0.001$, $\eta p^2 = 0.89$	MT > S2 > S1 ps < 0.001
	Pitch Range 100%	$F(2, 38) = 128.17$ $p < 0.001$, $\eta p^2 = 0.87$	MT = S2 > S1 ps < 0.001
	Slope 50%	$F(2, 38) = 25.05$ $p < 0.001$, $\eta p^2 = 0.57$	MT > S1 > S2 ps = < 0.001 − 0.007
	Duration 100%	$F(2, 38) = 196.27$ $p < 0.001$, $\eta p^2 = 0.91$	MT > S1 > S2 ps = 0.010 − 0.006
	Tone Accuracy	$F(1.51,28.70) = 3.53$ $p = 0.039$, $\eta p^2 = 0.16$	n.s.
Main effect of register	Initial Pitch Height 100%	$F(1, 19) = 65.87$ $p < 0.001$, $\eta p^2 = 0.78$	CDS > ADS p < 0.001
	Mid Pitch Height 100%	$F(1, 19) = 59.01$ $p < 0.001$, $\eta p^2 = 0.76$	CDS > ADS p < 0.001
	Final Pitch Height 100%	$F(1, 19) = 46.94$ $p < 0.001$, $\eta p^2 = 0.71$	CDS > ADS p < 0.001
	Min Pitch Height 50%	$F(1, 19) = 38.27$ $p < 0.001$, $\eta p^2 = 0.67$	CDS > ADS p < 0.001
	Max Pitch Height 50%	$F(1, 19) = 53.79$ $p < 0.001$, $\eta p^2 = 0.74$	CDS > ADS p < 0.001
	Mean Pitch Height 50%	$F(1, 19) = 49.52$ $p < 0.001$, $\eta p^2 = 0.72$	CDS > ADS p < 0.001
	Pitch Range 50%	$F(1, 19) = 5.71$ $p = 0.027$, $\eta p^2 = 0.23$	CDS > ADS p = 0.027
	Pitch Range 100%	$F(1, 19) = 8.70$ $p = 0.008$, $\eta p^2 = 0.31$	CDS > ADS p = 0.008
	Slope 50%	$F(1, 19) = 10.05$ $p = 0.005$, $\eta p^2 = 0.35$	CDS > ADS p = 0.005
	Duration 100%	n.s.	n.s.
	Tone Accuracy	$F(1,19) = 15.13$ $p = 0.001$, $\eta p^2 = 0.44$	ADS > CDS p = 0.001
Register × syllable	Initial Pitch Height 100%	$F(2, 38) = 9.12$ $p = 0.001$, $\eta p^2 = 0.32$	CDS > ADS p < 0.001 for all syllables
	Mid Pitch Height 100%	$F(2, 38) = 6.53$ $p = 0.004$, $\eta p^2 = 0.26$	CDS > ADS p < 0.001 for all syllables
	Final Pitch Height 100%	$F(2, 38) = 6.12$ $p = 0.005$, $\eta p^2 = 0.24$	CDS > ADS p < 0.001 for all syllables
	Min Pitch Height 50%	n.s.	n.s.
	Max Pitch Height 50%	$F(2, 38) = 12.72$ $p < 0.001$, $\eta p^2 = 0.40$	CDS > ADS p < 0.001 for all syllables
	Mean Pitch Height 50%	$F(2, 38) = 6.42$ $p = 0.004$, $\eta p^2 = 0.25$	CDS > ADS p < 0.001 for all syllables
	Pitch Range 50%	$F(1.39, 26.35) = 4.48$ $p = 0.033$, $\eta p^2 = 0.19$	CDS > ADS p = 0.025 for MT

TABLE III. *Continued*

Effect	Acoustic parameters	F-test	Pairwise comparison[a]
			n.s. in S1 and S2
	Pitch Range 100%	$F(2, 38) = 3.45$ $p = 0.042$, $\eta p^2 = 0.15$	CDS > ADS $p = 0.041$ for MT $p = 0.001$, for S2. n.s. in S1
	Slope 50%	n.s.	n.s.
	Duration 100%	n.s.	n.s.
	Tone Accuracy	$F(2,38) = 3.48$ $p = 0.041$, $\eta p^2 = 0.15$	ADS > CDS $p < 0.001$ for MT $p = 0.012$ in S1 n.s. in S2
Register × tone	Initial Pitch Height 100%	$F(1.94, 36.78) = 3.89$ $p = 0.031$, $\eta p^2 = 0.17$	CDS > ADS $p < 0.001$ for all tones
	Mid Pitch Height 100%	$F(3, 57) = 4.65$ $p = 0.006$, $\eta p^2 = 0.20$	CDS > ADS $p < 0.001$ for all tones
	Final Pitch Height 100%	n.s.	n.s.
	Min Pitch Height 50%	n.s.	n.s.
	Max Pitch Height 50%	n.s.	n.s.
	Mean Pitch Height 50%	n.s.	n.s.
	Pitch Range 50%	n.s.	n.s.
	Pitch Range 100%	n.s.	n.s.
	Slope 50%	n.s.	n.s.
	Duration 100%	$F(3, 57) = 3.66$ $p = 0.017$, $\eta p^2 = 0.16$	n.s. for all tones
	Tone Accuracy	n.s.	n.s.

[a]All p-values have been adjusted with the Bonferroni correction for multiple comparisons. Only p-values of significant differences were reported.

statistical tests were performed to determine if the observed differences reach statistical significance.

1. Inter-judge and intra-judge reliability

Kappa statistics adjusted for chance were used to determine inter-judge and intra-judge reliability. Fless' kappas were 0.921 and 0.907 for the five judges' ratings of the mothers' monosyllabic and disyllabic tones, respectively, indicating that as a group, the five judges reached almost perfect reliability on their tone ratings. Cohen's kappas for each pair of judges ranged from 0.872 to 0.956 for monosyllabic tones, and 0.874 to 0.942 for disyllabic tones, showing nearly perfect agreement between all pairs of judges. The judges also demonstrated almost perfect intra-judge reliability, with Cohen's kappas for each judge's first and second ratings of the same productions ranging from 0.962 to 0.987 for monosyllabic productions and 0.931 to 0.974 for disyllabic productions.

2. Perceptual accuracy of child-directed and adult-directed tones

Tone production clarity was defined as the judges' correct identification of the four tones. Table IV shows the accuracy and error patterns for tone identification in adult-directed and child-directed speech in different syllable contexts. The underlined numbers indicate error responses that constituted more than 5% of the total number of trials of the tone. Overall, mothers' tones appeared to have more errors and a greater variety of error patterns in child-directed speech than in adult-directed speech.

Statistical analysis was performed to determine if tone accuracy differed significantly between child-directed and adult-directed speech. Because the data violated assumptions for parametric statistics due to ceiling performance, scores were converted to rational arcsine units (RAUs), which are close to percentages but have the statistical characteristics of the arcsine transformation (Studebaker, 1985). A three-way ANOVA (syllable × tone × register) was performed on the transformed accuracy scores. Table III shows the major results. There were significant main effects of syllable [$F(1.51, 28.70) = 3.53$, $p = 0.039$, $\eta p^2 = 0.16$] and register [$F(1,19) = 15.13$, $p = 0.001$, $\eta p^2 = 0.44$] as well as a significant two-way interaction between syllable and register [$F(2,38) = 3.48$, $p = 0.041$, $\eta p^2 = 0.15$] for tone accuracy. Surprisingly, overall, tone accuracy was significantly lower in child-directed speech than in adult-directed speech. Pairwise comparisons with Bonferroni corrections for multiple comparisons revealed lower tone accuracy in child-directed speech than in adult-directed speech in monosyllabic words and in the first syllable of disyllabic words, while the difference was not significant in the second syllable of disyllabic words. Figure 7 shows the tone accuracy rates in child-directed and adult-directed speech in the three contexts. No significant interaction between register and tone was found, suggesting that regardless of tone type, tone accuracy was lower in child-directed speech than in adult-directed speech. No significant interaction effect among syllable context, tone and register was found.

TABLE IV. Confusion matrixes of tone production by register and syllable. Correct identifications are in boldface. Errors that constituted more than 5% of the trials are underlined.

5a: Adult-directed monosyllabic tones

Target tones	Judges' responses (%) T1	T2	T3	T4
T1	**95**	4	0	0
T2	0	**99**	1	0
T3	1	8	**88**	3
T4	1	0	0	**99**

5b: Child-directed monosyllabic tones

Target tones	Judges' responses (%) T1	T2	T3	T4
T1	**96**	4	0	0
T2	0	**97**	3	0
T3	1	13	**85**	1
T4	2	0	0	**98**

5c. Adult-directed tones in the first syllable of disyllabic words

Target tones	Judges' responses (%) T1	T2	T3	T4
T1	**96**	2	1	1
T2	0	**100**	0	0
T3	0	0	**100**	0
T4	3	0	1	**97**

5d. Child-directed tones in first syllable of disyllabic words

Target tones	Judges' responses (%) T1	T2	T3	T4
T1	**93**	6	1	1
T2	1	**97**	2	0
T3	0	0	**99**	1
T4	7	0	3	**89**

5e. Adult-directed tones in the second syllable of disyllabic words

Target tones	Judges' responses (%) T1	T2	T3	T4
T1	**97**	3	0	0
T2	0	**97**	3	0
T3	0	4	**90**	6
T4	0	0	1	**99**

5f. Child-directed tones in the second syllable of disyllabic words

Target tones	Judges' responses (%) T1	T2	T3	T4
T1	**95**	3	1	0
T2	1	**95**	4	0
T3	0	6	**86**	8
T4	1	0	0	**99**

IV. DISCUSSION

The aims of the study were to examine the acoustic characteristics of child-directed Mandarin tones and to determine whether mothers hyperarticulated these tones in child-directed speech to enhance contrasts among the tones.

A. Syllable position affects acoustics of tones

The significant main effect of syllable context on all acoustic measures of tones indicated that pitch height, pitch range, pitch shape, and tone duration were different in different syllable contexts. In general, these parameters were higher, larger, or longer in the first syllable of disyllabic words than in the second syllable. They were highest, largest, and longest in monosyllabic words. These findings suggest that syllable position affects the acoustics of tones (consistent with Wong and Strange, 2017 and Xu, 1997), and studies that compare the acoustics of tones in different experimental conditions should control for this effect. However, almost all previous studies have collected child-directed tones in mother-child play interactions and adult-directed tones in conversations with another adult (Liu *et al.*, 2007; Liu *et al.*, 2009; Rattanasone *et al.*, 2013). This lack of control of the effects of syllable and utterance position between the two registers may partly explain the discrepancies observed in the acoustic findings across different studies.

B. Acoustic characteristics of tones in child-directed speech

There was clear evidence that mothers acoustically modified the pitch contours of the tones when speaking to their children. Tones in child-directed speech were consistently produced with higher pitch and more positive slopes than tones in adult-directed speech. These findings were not affected by syllable position or tone type, because either no significant interaction was found, or when a significant interaction did occur, the acoustic measures were still higher and larger in child-directed speech than in adult-directed speech.

In general, tones were produced with larger pitch range in child-directed speech than in adult-directed speech. This

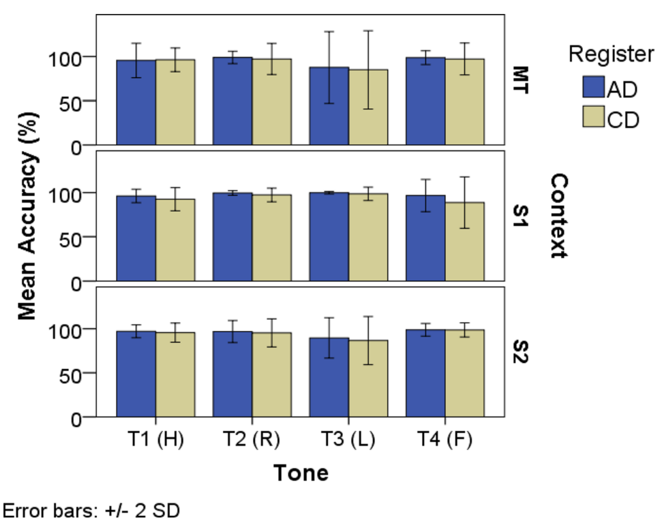

FIG. 7. (Color online) Accuracy of the four tones in adult-directed and child-directed speech in different production contexts. T1, T2, T3, and T4 stand for Tone 1, 2, 3, and 4, respectively. H, R, L, F stand for High, Rise, Low and Fall, respectively. MT stands for monosyllabic tone. S1 and S2 stand for tones in the first and second syllable, respectively.

difference was not affected by tone type but did interact with syllable position. The pitch range of the tonal targets was comparable in the first syllable in disyllabic words, while in the other two syllable positions, the pitch range was larger in child-directed speech than in adult-directed speech. In terms of tone duration, no difference was found between child-directed and adult-directed tones. All four tones were produced with comparable durations in child-directed and adult-directed speech in all syllable positions.

The finding that tones are produced with higher pitch in child-directed speech than in adult-directed speech is consistent with all previous studies on child-directed lexical tones (e.g., Liu *et al.*, 2007; Liu *et al.*, 2009; Papoušek and Hwang, 1991; Rattanasone *et al.*, 2013). With respect to tone duration, unlike previous studies which have consistently found that longer tone duration in child-directed speech than in adult-directed speech, this study found no significant effect of register on tone duration for any tone type or syllable position. These discrepancies between the present findings and those of most previous studies could be due to the fact that previous studies did not control syllable context, utterance structure, or pragmatic function in child-directed and adult-directed speech. Thus, observed duration differences between child-directed and adult-directed speech may have reflected differences in these factors. While Papoušek and Hwang (1991) did control for production context and reported longer syllable durations in child-directed speech, they measured utterance duration and not tone duration. More importantly, they computed syllable duration by dividing the duration of the whole utterance by the number of syllables in the utterance. It is therefore unclear whether the duration and number of pauses and the number of filler or interjections in the utterances were the same in child-directed and adult-directed utterances.

Mixed results have been reported in the literature on pitch range differences between child-directed and adult-directed speech. Papoušek and Hwang (1991) reported no pitch range differences between child-directed and adult-directed speech, while other studies reported larger pitch ranges in child-directed speech than in adult-directed speech (Cooper and Aslin, 1994; Fernald and Kuhl, 1987; Grieser and Kuhl, 1988). The findings of this study indicate that, overall, pitch range is larger in child-directed speech than in adult-directed speech, but that the differences depend on syllable context. In some syllable contexts, such as the first syllable of a disyllabic word, no difference was found in pitch range. Because Papoušek and Hwang (1991) did not perform analyses on the interaction between tone and syllable position, and other studies did not control the production contexts of tones in child-directed speech and adult-directed speech, it is difficult to compare the results. No previous studies have measured the shapes of the pitch contours in child-directed and adult-directed speech.

C. Evidence for hyper-articulation hypothesis of child-directed speech

This study found little evidence that mothers modified the acoustics of the tones to enhance phonetic contrasts among Mandarin tones. Previous studies examined acoustic differences between child-directed and adult-directed speech without examining the perceptual clarity of child-directed speech. Without a perceptual task, it is difficult to determine with confidence whether the observed acoustic differences between child-directed and adult-directed speech would lead to better or worse perception of the phonemes. This study examined both acoustic and perceptual differences between child-directed and adult-directed speech. The lower tone perceptual accuracy in child- versus adult-directed speech is strong evidence against the idea that child-directed speech is hyperarticulated. The fact that identification accuracy of the tones was affected by syllable position but not tone type further suggests that the acoustic modifications of the tones were not targeted at differentiating the four tones.

The acoustic findings were consistent with the poorer perceptual accuracy of the tones in child-directed speech than adult-directed speech. Regardless of tone types, all tones in child-directed speech were produced with higher pitch, more positive slope, and larger pitch range, suggesting that even tones that were expected to have lower pitch levels (e.g., T3, the low tone), smaller pitch range (e.g., T1, the high level tone), and more negative slopes (e.g., T4, the falling tone), were all produced with higher pitch, larger pitch range, and more positive slopes in child-directed speech than in adult-directed speech.

D. Evidence for pragmatic hypothesis of child-directed speech

The findings support the pragmatic hypothesis. Pragmatic meanings are expressed by different prosody, which involves pitch change and pitch contrasts at different parts of an utterance (Prom-on *et al.*, 2012). This study found a significant interaction between register and syllable, but no significant interaction between tone and register, for the various acoustic parameters, suggesting that syllable position rather than tone type affected the degree of acoustic modification of the tones in child-directed speech. Previous studies have reported different types of prosodic effects on the acoustics of tones in different utterance positions. For example, the pitch level of lexical tones in a statement has been reported to gradually decrease from the beginning to the end of an utterance, while the pitch contour of the six Cantonese tones, including the mid-level, low-falling, and low-level tones, all had rising contours at the ends of questions (Ma *et al.*, 2006).

Though the findings of this study suggest that the acoustic modifications in child-directed tones are likely a prosodic effect, the intent of these prosodic modifications is less clear. One possibility is that mothers use heightened pitch to express happy or affectionate emotions, which are usually conveyed with higher and more varied pitch (Collier and Hubbard, 1998; Yildirim *et al.*, 2004). Another possible communicative intent could be to regulate the child's attention. Young infants have been found to prefer listening to high-pitched utterances versus low-pitched utterances, even when all other aspects of the utterances are the same (Patterson *et al.* in Trainor *et al.*, 2000). They are not only more responsive to high-pitched child-directed speech in their native language (Cooper and Aslin, 1994; Cooper *et al.*, 1997), but also in foreign languages (Werker *et al.*, 1994), suggesting that it is the high pitch, rather than the content of the utterance, that draws young children's attention. Pitch has also been found to be used by mothers to capture and maintain the attention of infants (Fernald *et al.*, 1989; Stern *et al.*, 1982). When infants are looking away, mothers tend to use rising pitch contours to draw the infants' attention. When the infant is looking at the mother, the mother often uses a rise-fall contour to maintain the infant's attention, and when the infant is distressed, the mother usually speaks with slow falling pitch contours to comfort the infant (Fernald *et al.*, 1989; Stern *et al.*, 1982). Future studies that compare the use of pitch in emotional speech, in speech for attention regulation, and in child-directed speech can reveal more detailed information.

In summary, mothers acoustically modify their tones when speaking to young children. When produced in the same syllable position and the same phonetic and prosodic contexts, child-directed Mandarin tones are characterized by higher pitch level and more positive slopes than the tones in adult-directed speech. In general, tones are also produced with a larger pitch range in child-directed speech than in adult-directed speech, although the differences are dependent on syllable position. Mothers do not expand the tone duration in child-directed speech. These acoustic modifications of child-directed tones applied similarly to all four tones, regardless of pitch level, pitch range, or pitch shape. Finally, tones in child-directed speech are identified with lower accuracy than the tones in adult-directed speech. The acoustic and perceptual findings do not support the hyperarticulation hypothesis of child-directed speech and appear to support the pragmatic hypothesis of child-directed speech.

ACKNOWLEDGMENTS

This research was supported by the Faculty Research Fund (FRF) of the Faculty of Education at The University of Hong Kong.

American Speech-Language-Hearing Association (**1997**). *Guidelines for Audiologic Screening*, American Speech-Language-Hearing Association, Rockville, MD.

Baran, J. A., Laufer, M. Z., and Daniloff, R. (**1977**). "Phonological contrastivity in conversation: A comparative study of voice onset time," J. Phonetics **5**, 339–350.

Benders, T. (**2013**). "Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not hyperarticulation intent," Infant Behav. Dev. **36**(4), 847–862.

Boersma, P., and Weenink, D. (**2014**). "Praat: Doing phonetics by computer" [computer program, version 5.4.00].

Bohn, O. S. (**2013**). "Acoustic characteristics of Danish infant directed speech," Proc. Meet. Acoust. **19**(1), 060055.

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (**2002**). "What's new, pussycat? On talking to babies and animals," Science **296**(5572), 1435–1435.

Collier, W. G., and Hubbard, T. L. (**1998**). "Judgments of happiness, brightness, speed and tempo change of auditory stimuli varying in pitch and tempo," Psychomusicol.: J. Res. Music Cogn. **17**(1–2), 36–55.

Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (**1997**). "The development of infants' preference for motherese," Infant Behav. Dev. **20**(4), 477–488.

Cooper, R. P., and Aslin, R. N. (**1994**). "Developmental differences in infant attention to the spectral properties of infant-directed speech," Child Dev. **65**(6), 1663–1677.

Cristia, A., and Seidl, A. (**2014**). "The hyperarticulation hypothesis of infant-directed speech," J. Child Lang. **41**(4), 913–934.

Englund, K., and Behne, D. (**2006**). "Changes in infant directed speech in the first six months," Infant Child Dev. **15**(2), 139–160.

Englund, K. T. (**2005**). "Voice onset time in infant directed speech over the first six months," First Lang. **25**, 219–234.

Fernald, A., and Kuhl, P. (**1987**). "Acoustic determinants of infant preference for motherese speech," Infant Behav. Dev. **10**(3), 279–293.

Fernald, A., and Simon, T. (**1984**). "Expanded intonation contours in mothers' speech to newborns," Dev. Psychol. **20**(1), 104–113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (**1989**). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," J. Child Lang. **16**(03), 477–501.

Fisher, C., and Tokura, H. (**1996**). "Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence," Child Dev. **67**(6), 3192–3218.

Fu, Q. J., and Zeng, F. G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pacific J. Speech Lang. Hear. **5**(1), 45–57.

Garnica, O. K. (**1977**). "Some prosodic and paralinguistic features of speech to young children," in *Talking to Children: Language Input and Acquisition*, edited by C. E. Snow and C. A. Ferguson (Cambridge University Press, Cambridge, England).

Grieser, D. L., and Kuhl, P. K. (**1988**). "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," Dev Psychol. **24**(1), 14–20.

Hartman, K. M., Ratner, N. B., and Newman, R. S. (**2017**). "Infant-directed speech (IDS) vowel clarity and child language outcomes," J. Child Lang. **44**(5), 1140–1162.

Khouw, E., and Ciocca, V. (**2007**). "Perceptual correlates of Cantonese tones," J. Phon. **35**(1), 104–117.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**(5326), 684–686.

Lin, B., and Lin, N. (**1994**). "學前兒童語言障礙評量表" ("Language disorder scale of preschoolers"), Department of Special Education, National Taiwan Normal University, Taipei, Taiwan.

Liu, H. (**2014**). "Lexical and acoustic features of maternal utterances addressing preverbal infants in picture book reading link to 5-year-old children's language development," Early Educ. Dev. **25**(8), 1103–1117.

Liu, H. M., Kuhl, P. K., and Tsao, F. M. (**2003**). "An association between mothers' speech clarity and infants' speech discrimination skills," Develop. Sci. **6**(3), F1–F10.

Liu, H. M., Tsao, F. M., and Kuhl, P. K. (**2007**). "Acoustic analysis of lexical tone in Mandarin infant-directed speech," Dev. Psychol. **43**(4), 912–917.

Liu, H. M., Tsao, F. M., and Kuhl, P. K. (**2009**). "Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study," J. Child Lang. **36**(04), 909–922.

Ma, J. K., Ciocca, V., and Whitehill, T. L. (**2006**). "Effect of intonation on Cantonese lexical tones," J. Acoust. Soc. Am. **120**(6), 3978–3987.

Malsheen, B. J. (**1980**). "Two hypotheses for phonetic clarification in the speech of mothers to children," in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic Press, New York), Vol. 2, pp. 173–184.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristia, A. (**2015**). "Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis," Psychol. Sci. **26**(3), 341–347.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., and McEchron, W. (**2013**). "Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?," Cognition **129**(2), 362–378.

Papoušek, M., and Hwang, S. F. C. (**1991**). "Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction," Appl. Psycholinguist. **12**(04), 481–504.

Prom-on, S., Liu, F., and Xu, Y. (**2012**). "Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling," J. Acoust. Soc. Am. **132**(1), 421–432.
Quam, C., and Swingley, D. (**2012**). "Development in children's interpretation of pitch cues to emotions," Child Dev. **83**(1), 236–250.
Rattanasone, N. X., Burnham, D., and Reilly, R. G. (**2013**). "Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age," J. Phon. **41**(5), 332–343.
Russo, F. A., and Thompson, W. F. (**2005**). "An interval size illusion: The influence of timbre on the perceived size of melodic intervals," Atten. Percept. Psycho. **67**(4), 559–568.
Singh, L., Morgan, J. L., and Best, C. T. (**2002**). "Infants' listening preferences: Baby talk or happy talk?," Infancy **3**(3), 365–394.
Stern, D. N., Spieker, S., and MacKain, K. (**1982**). "Intonation contours as signals in maternal speech to prelinguistic infants," Dev. Psychol. **18**(5), 727–735.
Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**(3), 455–462.
Sundberg, U., and Lacerda, F. (**1999**). "Voice onset time in speech to infants and adults," Phonetica **56**(3-4), 186–199.
Trainor, L. J., Austin, C. M., and Desjardins, R. N. (**2000**). "Is infant-directed speech prosody a result of the vocal expression of emotion?," Psychol. Sci. **11**(3), 188–195.
Uther, M., Knoll, M. A., and Burnham, D. (**2007**). "Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech," Speech Commun. **49**(1), 2–7.
Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica **33**(5), 368–392.
Werker, J. F., Pegg, J. E., and McLeod, P. J. (**1994**). "A cross-language investigation of infant preference for infant-directed communication," Infant Behav. Dev. **17**(3), 323–333.
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. (**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 3-year-olds growing up in Taiwan and in the United States: Interjudge reliability and perceptual results," J. Speech Lang. Hear. Res. **55**(5), 1423–1437.
Wong, P. (**2013**). "Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan," J. Acoust. Soc. Am. **133**(1), 434–443.
Wong, P., and Chan, H. (**2018**). "Acoustic characteristics of highly distinguishable Cantonese entering and non-entering tones," J. Acoust. Soc. Am. **143**(2), 765–779.
Wong, P., Fu, W. M., and Cheung, E. Y. (**2017**). "Cantonese-speaking children do not acquire tone perception before tone production—A perceptual and acoustic study of three-year-olds' monosyllabic tones," Front. Psychol. **8**, 1450.
Wong, P., and Leung, C. T. (**2018**). "Suprasegmental features are not acquired early: Perception and production of monosyllabic Cantonese lexical tones in four- to six-year-old pre-school children," J. Speech-Lang. Hear. Res. **61**, 1070–1085.
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. R. **48**(5), 1065–1079.
Wong, P., and Strange, W. (**2017**). "Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study," PloS One **12**(8), e0182337.
Wong, Y. W., and Xu, Y. (**2007**). "Consonantal perturbation of f0 contours of Cantonese tones," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrucken, pp. 1293–1296.
Xu, C. X., and Xu, Y. (**2003**). "Effects of consonant aspiration on Mandarin tones," J. Int. Phon. Assoc. **33**, 165–181.
Xu, Y. (**1997**). "Contextual tonal variations in Mandarin," J. Phon. **25**(1), 61–83.
Xu, Y. (**2001**). "Sources of tonal variations in connected speech," J. Chin. Ling. Monogr. Ser. **17**, 1–31.
Xu, Y., and Wang, Q. E. (**2001**). "Pitch targets and their realization: Evidence from Mandarin Chinese," Speech Commun. **33**, 319–337.
Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., and Narayanan, S. (**2004**). "An acoustic study of emotions expressed in speech," *International Conference of Spoken Language Processing* (ICSLP'04), pp. 2193–2196.