Testing the hyper-articulation and prosodic hypotheses of child-directed speech: Insights from the perceptual and acoustic characteristics of child-directed Cantonese tones

Puisan Wong[1] and Kelly Wing Sum Ng[2]

[1]Corresponding author. Division of Speech and Hearing Sciences, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong. Email: pswResearch@gmail.com

[2]Division of Speech and Hearing Sciences, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong. Email: kellynws@connect.hku.hk

Acknowledgements

This research was supported by funding from the General Research Fund (GRF) of the Research Grant Council of Hong Kong to the first author.

Abstract

Purpose: The function of child-directed speech has been debated for decades. This study examined the perceptual and acoustic characteristics of child-directed and adult-directed Cantonese tones to test the hyper-articulation and prosodic hypotheses that have been proposed to account for the acoustic modifications in child-directed speech.

Method: Sixty-two mother-child dyads participated in the study. The mothers verbally labelled 30 pictures in monosyllabic isolated words and in the final position of a carrier sentence to the experimenter and their one- to five-year-old children. The 8634 adult-directed and child-directed productions were low-pass filtered to eliminate lexical information and presented to five judges for tone identification. Acoustic analysis was performed on the productions.

Results: Acoustically, child-directed tones were produced with elevated pitch and the pitch level decreased as the child's age increased. Acoustic contrasts between phonetically similar and more confusing tones were not enhanced in child-directed speech, and, unexpectedly, child-directed tones were identified with lower accuracy than adult-directed tones. The perceptual errors of child-directed tones mirrored the errors found in identifying tones excised from sentence final position, which had a pitch lowering effect on the tones. The lower perceptual accuracy, the lack of enhanced acoustic contrasts in confusing tone pairs, and the similarities in the error patterns in identifying tones in child-directed speech and tones in utterance final position suggest that the acoustic modifications in child-directed tones are prosodic effects serving pragmatic purposes.

Conclusion: The findings reject the hyper-articulation hypothesis and support the prosodic hypothesis of child-directed speech.

Introduction

Adults and children from different cultures instinctively adopt a special speech register, called child-directed speech (CDS), when talking to young children (Fernald, et al., 1989; Shatz & Gelman, 1973). Prosodically, CDS is produced with higher pitch (Fernald & Simon, 1984; Garnica, 1977), larger pitch range (Cooper & Aslin, 1994; Fernald & Kuhl, 1987), and greater pitch variations (Fernald, 1989; Fisher & Tokura, 1996) than adult-directed speech (ADS). These prosodic characteristics appear to be universal and have been reported in different languages (Fernald et al., 1989; Grieser & Kuhl, 1988). The function of the acoustic modifications in CDS is unclear. The hyper-articulation hypothesis proposes that adults enhance the acoustic contrasts of phonemes in CDS to facilitate speech and language acquisition in children, while the prosodic hypothesis suggests that adults modify the acoustic signal in CDS for pragmatic purposes such as expressing affective emotions (e.g., Benders, 2013) and regulating children's attention (see Fernald, 1992). This study tested these two hypotheses by comparing adult-directed and child-directed Cantonese tones.

Evidence for the Hyper-articulation Hypothesis for Child-directed Speech

Most of the supporting evidence for the hyper-articulation hypothesis derived from studies that compared the vowel space formed by the formant frequencies of three corner vowels /i, u, a/ in CDS and ADS (Kuhl et al., 1997; Liu, Tsao, & Kuhl, 2009). Larger vowel space in CDS has been found across non-tonal (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997), tonal (Liu, Tsao, & Kuhl, 2009) and pitch-accent (Andruski, Kuhl, & Hayashi, 1999) languages. These studies suggested that mothers exaggerated the vowels in CDS to increase the acoustic distance between vowel categories, making the vowel contrasts more salient for children's learning.

Studies that examined the relationship between mothers' use of CDS and their children's language ability offered compelling supports to the hyper-articulation hypothesis. Liu, Kuhl, & Tsao (2003) reported that the size of the vowel space of mothers' speech to their children below one year of age correlated positively to the children's ability to discriminate consonants (Liu, Kuhl, & Tsao, 2003) and the children's expressive language ability when they turned five years old (Liu, 2014). Supported by the finding that two-year-old children learned words better in CDS than in ADS (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011), these studies strongly suggested that mothers hyper-articulate the phonemes in CDS for didactic purposes.

Challenges to the Hyper-articulation Hypothesis

Yet, the hyper-articulation hypothesis was challenged by studies that measured not only the corner vowels but also the inner vowels in the vowel space. Cristia and Seidl (2014) and McMurray, Kovack-Lesh, Goodwin, and McEchron (2013) found that although the corner vowels were expanded in CDS compared to ADS, not all interior vowels in CDS shifted in the directions that enhanced phonetic contrasts. The formant measurements of some inner vowels clustered closer together in the vowel space and larger acoustic variability of the same vowel was found in CDS than in ADS, leading to substantial overlaps in the vowel categories in CDS compared to ADS.

No study has compared the perceptual accuracy of the segmental phonemes produced in ADS and CDS. A couple of studies used statistical models and computer algorithms to examine the discriminability of the phonemes in ADS and CDS. McMurray et al (2013) collected AD and CD vowel productions using a book reading task and employed a statistic model to identify the vowels. Vowels in CDS were more difficult to identify than in ADS.

Martin et al., (2015) used a computer algorithm to discriminate Japanese AD and CD phonemes and found similar results.

More importantly, not all studies found expanded vowel space in CDS. Benders (2013), Dodane and Al-Tamimi (2007), and Englund and Behne (2006) reported reduced vowel space in CDS in English, Dutch, French, and Japanese. Mixed findings in studies that examined differences in voice onset time of stop consonants in CDS and ADS (Englund, 2005; Sundberg & Lacerda, 1999; McMurray et al., 2013) further challenged the hyper-articulation hypothesis.

The Prosodic Hypothesis of Child-directed Speech

The inconsistent findings in the aforementioned studies casted doubts on the proposal that mothers enhance the phonetic contrasts in CDS, leading to the proposal of the prosodic hypothesis which postulates that mothers acoustically modify their speech to their children for pragmatic purposes, such as expression of affects and regulating attention (Trainor & Desjardins, 2002; Uther et al., 2007). The observed expanded vowel space in CDS is only a by-product of the affective prosody in CDS (Benders, 2013; McMurray et al., 2013).

The prosodic hypothesis has been further supported by the acoustic similarities between affective speech and CDS. Like CDS, affective speech tends to have higher pitch and greater pitch range (Singh, Morgan, & Best, 2002). When participants were asked to express target utterances emotionally in ADS, the acoustic features of ADS resembled those in CDS (Trainor, Austin, & Desjardins, 2000). Singh et al. (2002) tested children’s responses to mothers’ happy and neutral ADS and CDS. They found that infants demonstrated significantly longer looking time when listening to happy CDS than neutral ADS and to happy ADS than neutral CDS, showing that it was the happy emotions that resulted in significantly longer looking time in infants. The findings also suggested that the acoustic modifications observed in CDS may serve affective rather than didactic purposes.

Observations of heightened pitch levels in speech directed to foreigners and pets also supported the prosodic function of CDS. Uther et al. (2007) compared the prosody in CDS, ADS and foreigner-directed speech. Significantly higher pitch was found in CDS than in foreigner-directed speech or ADS. The authors concluded that adults did not raise the pitch in CDS for didactic purposes; otherwise pitch should be higher in both CDS and foreigner-directed speech. Burnham et al. (2002) incorporated pet-directed speech into their study. They hypothesized that speakers would not exaggerate their speech for didactic purposes when talking to pets, but would raise their pitch levels for affective expressions. The results showed that the pitch in both CDS and pet-directed speech were higher than in ADS, with no difference in the pitch levels in CDS and pet-directed speech, supporting that acoustic modifications in CDS serve pragmatic rather than didactic functions.

Acoustic Modifications in Child-directed Lexical Tones

The vast majority of research on CDS focused on the segmental characteristics of non-tonal languages and few examined tonal languages. In tonal languages, pitch serves both phonemic and prosodic purposes. Pitch patterns within a syllable, called lexical tones, play the same role as the vowels and consonants in marking lexical contrasts. Thus, words produced with the same segmental structure (i.e., the same syllable) but different pitch contours in the syllable convey different meanings. Pitch is also used at the utterance level to serve pragmatic purposes such as indicating questions and statements and expression of emotions (Xu, Kelly, & Smillie, 2013). Therefore, the use of pitch in tonal languages provides unique opportunities for testing the hyper-articulation and prosodic hypotheses. If the hyper-articulation hypothesis is true, pitch modifications in CDS would enhance the tonal contrasts. If the prosodic hypothesis is true, mothers would modify pitch at the utterance level which would not enhance tonal contrasts and may compromise the identity of lexical tones. If both hypotheses are true, an enhancement in the tonal contrasts and additional pitch modifications that do not serve the enhancement of tonal contrasts would be observed.

Several studies have examined the acoustic features of child-directed Mandarin lexical tones. Mandarin has four lexical tones (i.e., High-level, High-rising, Low-dipping and High-falling) (Howie, 1976) and the primary and sufficient cue for tone identification is the level and shape of the fundamental frequency (f0) or pitch contour (Fu & Zeng, 2000; Xu & Wang, 2001). Liu et al. (2007, 2009) measured adult-directed and child-directed Mandarin tones and found higher mean F0, larger F0 range and longer duration in CDS than ADS. Based on the findings, they claimed that mothers hyper-articulated the tones in CDS. However, the same acoustic characteristics could also result from affective or emotional expressions. Because no acoustic measures were performed on the shapes of the tones, which are the distinctive perceptual cues of Mandarin tones (Xu & Wang, 2001), and no perceptual evaluation was performed on the tones, it was unclear whether the contrasts of the tones in CDS were enhanced.

Also, tone productions in ADS and CDS in the study may have been confounded by the differences in the prosodic and phonetic contexts of the tones produced in CDS and ADS because in the study, child-directed tones were collected from spontaneous productions of the mothers when they were playing with their children with some selected toys, whereas adult-directed tones were recorded while the mothers were talking about their children’s interest to the experimenter. The same tones or target words produced in the two speech registers were likely produced in different utterance positions (e.g., isolation, utterance initial medial and final), emotional contexts (e.g., expression of excitement, sadness, happiness during play), phonetic contexts (e.g., coarticulation of tones in different tonal contexts), and sentence structures (e.g., questions, statement). These contexts have been found to influence the F0 of tones (Xu, 1997, 2001; Ma, Ciocca, & Whitehill, 2006). It is, therefore, unclear whether the observed acoustic differences were attributed to the different phonetic and prosodic contexts in the speech samples or differences in the two speech register used for different conversational partners.

Wong (2018) examined the perceptual and acoustic properties of child-directed and adult-directed Mandarin tones in monosyllabic and disyllabic words produced in isolation in a picture reading task. The acoustic data confirmed that mothers produced the tones in CDS with significantly higher pitch levels and more positive pitch slopes than in ADS. Given that even the high level and the falling tones were also produced with more positive slopes in CDS than in ADS, the acoustic modifications in CDS did not appear to enhance the tonal contrasts and seemed to support the prosodic account. More importantly, child-directed tones were identified with significantly lower accuracy than adult-directed tones. Thus, the findings did not support that mothers hyper-articulated the tones in CDS. However, given that Mandarin has a simpler tonal system and each of the four tones has distinct pitch contours, it is possible that Mandarin-speaking mothers do not need to hyper-articulate the tones in Mandarin to enhance tonal contrasts.

Cantonese is particularly appropriate for testing the two hypotheses of CDS because it has a more complex tonal system with tonal contrasts that are confusing even for some native Cantonese speakers (see Wong & Chan, 2018, for details). Phonemically, Cantonese has six tones: Tone 1 (T1, High-Level, HL), Tone 2 (T2, High-Rising, HR), Tone 3 (T3, Mid-Level, ML), Tone 4 (T4, Low-Falling, LF), Tone 5 (T5, Low-Rising, LR), and Tone 6 (T6, Low-Level, LL). Unlike Mandarin tones which are mostly contrasted by F0/pitch shapes, Cantonese tones are contrasted by both F0 /pitch levels and F0/pitch shapes. There are three level tones (T1 (HL), T3 (ML), T6 (LL)), two rising tones (T2 (HR), T5 (LR)) and one falling tone (T4 (LF)) in Cantonese. Figure 1 shows the F0 contours of the six tones produced by nine female speakers in 718 monosyllabic words (Wong & Chan, 2018). Two pairs of Cantonese tones, namely HR (T2) and LR (T5), and ML (T3) and LL (T6), have similar pitch shapes and relative small differences in pitch levels (Fig 1). Even native Cantonese-speaking adults have particular difficulties with these two pairs of tones (Ciocca & Lui, 2003; Lee, Chan, Lam, van Hasselt, & Tong, 2015; Wong & Leung, 2018) and some native speakers merge these pairs of tones and do not differentiate them in their perception and/or production (Kei, Smith, So, Lau, & Capell, 2002; Mok, Zuo, & Wong, 2013). Given the lower perceptual accuracy in these Cantonese tone pairs and because Cantonese-speaking children as old as six years of age have not mastered the perception or production of the six Cantonese tones in monosyllabic words (Wong, Fu & Cheung, 2017; Wong & Leung, 2018), if the hyper-articulation hypothesis is right, mothers would produce Cantonese tones, particularly the more confusing tone pairs, more distinctively in CDS to facilitate the discrimination of the tones in young children. It was, therefore, predicted that the acoustic differences of the confusing tone pairs would be larger in CDS than in ADS and higher perceptual accuracy would be found for child-directed tones than adult-directed tones. On the other hand, if the prosodic hypothesis is correct, mothers would use more varied pitch to express pragmatic functions such as affections when speaking to children. Thus, the change of pitch at the prosodic level may affect the F0 contours of the lexical tones at the syllable level, leading to poorer identification of the tones in CDS. Also, contrastive acoustic differences of the confusing tone pairs would not be systematically enhanced in CDS.

Figure 1. Fundamental frequency contours of Cantonese tones spoken by female speakers

Few studies have tested the hyper-articulation hypothesis of CDS using Cantonese tones. Rattanasone, Burnham and Reilly (2013) compared the tone space of the three corner tones (T1 (HL), T2 (HR) and T4 (LF)) in CDS and ADS and found significantly larger tone space in CDS than in ADS. They claimed that Cantonese-speaking mothers exaggerated the acoustic differences in tones for didactic purposes, supporting the hyper-articulation hypothesis. However, tone space formed by the onset and offset frequencies of the corner tones may not represent meaningful acoustic cues for perceiving Cantonese tones. First, tone contours at the onset of a syllable are affected by various factors such as the initial consonants (Xu & Xu, 2003) and phonetic contexts (Xu, 1997, 2001; Wong & Strange, 2017) and are not reliable cues for tone discrimination. The perceptual targets and reliable cues for lexical tones are the F0/pitch levels and shapes towards the end of the syllable (Xu, 1997; Whitehill, Ciocca, & Chow, 2000), which were not measured in the study. Second, the three corner tones are among the most easily identifiable tones in Cantonese (Ciocca & Lui, 2003; Lee et al., 2015). Third, the study did not control for the phonetic and prosodic variations in the tones produced in child-directed and adult-directed conditions. Therefore, the acoustic differences measured may be reflective of effects of contextual differences rather than effects of CDS on tones.

Developmental Changes of Child-directed Speech

Limited research has investigated the developmental trend in the acoustic modifications in CDS. The current findings show that as the children get older, the effect of CDS decreases. Warren-Leubecker and Bohannon (1984) found that English-speaking mothers produced higher pitch in their speech to two-year-olds than to five-year-olds. Similarly, Amano, Nakatani, and Kondo (2006) found that Japanese mothers of children between birth and five years of age lowered the pitch in CDS when speaking to older than to younger children. Only one study examined developmental changes in CDS in tone languages. Liu et al. (2009) found higher pitch level, larger tone space and larger pitch range in Mandarin-speaking mothers' speech to two-year-olds than to 5-year old children, which in turn were significantly different than in ADS. However, the studies did not examine the developmental changes of tone shapes, which were important acoustic cues for Mandarin tones. Therefore, the purpose of the developmental acoustic differences in CDS remains unclear.

The Present Study

Without perceptual judgment of CDS and ADS, it remains unclear whether the observed acoustic differences in CDS enhanced the phonetic contrasts of the speech sounds.

This study tested the hyper-articulation and prosodic hypotheses of CDS by examining both the perceptual and acoustic characteristics of monosyllabic Cantonese tone productions in ADS and CDS in mothers of one- to five-year-old children using a picture reading task. It was hypothesized that if the hyper-articulation hypothesis is true, more distinctive acoustic cues for the tones would be found in CDS, possibly leading to better perceptual identification of the tones in CDS than in ADS, particularly for the more confusing tone pairs. Also, the acoustic distance of the cues for discriminating the more confusing tones would be larger in CDS than in ADS. The degree of acoustic enhancement of the distinctive acoustic cues would decrease with the increase of the child's age. If the prosodic hypothesis is correct, pitch would be used for pragmatic purposes and could affect the pitch contours of tones, leading to lower perceptual accuracy in child-directed tones. Acoustic modifications would be found in child-directed tones but the acoustic differences may not enhance the tonal contrasts of the tones. No study has compared the perceptual accuracy of CDS and ADS. To further determine whether the observed differences in CDS were prosodic effects, perceptual and acoustic characteristics of tones in CDS and in sentence final position were compared. Previous acoustic studies found that pitch of the tones was heightened in CDS (Liu et al., 2007; 2009) but lowered in sentence final position (Ma et al., 2006; Vance, 1976). It was hypothesized that if pitch modifications in child-directed tones are prosodic effects, perceptual and acoustic similarities would be found in child-directed tones and tones in sentence final position. The research questions included (1) What are the acoustic characteristics of child-directed Cantonese tones? (2) What are the perceptual characteristics of child-directed tones? (3) Do mothers hyper-articulate the tones in child-directed speech? (4) Are the perceptual characteristics of child-directed Cantonese tones indicative of prosodic effects?

Methods

The protocol used in this study was approved by the Human Research Ethics Committee of the University of Hong Kong.

Adult Participants

Forty-eight Cantonese-speaking mothers ($Mean_{Age}$ = 35.59 years, SD_{Age} = 3.76 years) (Appendix A) provided written informed consent for the participation of themselves and their children. Twelve mothers participated with two of their children and one mother participated with three children. Altogether, there were 62 mother-child dyads. Mothers filled out a questionnaire, and passed a Cantonese tones perception test (CanTIT) (Mean $score_{Form A}$ = 29.85 out of 30 points, $SD_{Form A}$ = 0.40 points) and a pure tone hearing screening test at 500 Hz, 1K Hz, 2K Hz, 4K Hz and 8K Hz at 20 dB HL bilaterally under headphones using pure tone audiometry. All mothers reported that Cantonese was their first and strongest language and the language they used with their children. All mothers were born and raised in Hong Kong, except M29, who was born in Guangdong province in China and spoke Xin Xing Cantonese, a Cantonese dialect with the same tonal system as Hong Kong Cantonese, before migrating to Hong Kong at nine years old. She was not excluded in this study because she passed the Hong Kong Cantonese tones perceptual test (CanTIT: Form A) with the maximum score and she was not an outliner in any of the measures used in this study. Seven mothers knew other Chinese dialects but reported that they never used them to speak to their children.

Child Participants

Sixty-two children (Appendix A) provided assent to participation. Sixteen of them were one-year-olds, sixteen were two-year-olds, fifteen were three-year-olds and fifteen were five-year-olds. All children were normally developing with no remarkable developmental history. One- to five-year-old children were administered the Cantonese Child Developmental Inventory (CDI), designed for these age groups. Six-year-old children were conducted the Cantonese Oral Language Deficiency Early Identification Test for Pre-primary Children (學前兒童粵語表達能力識別測驗) (Po Leung Kuk, 2012). No child scored at or below one standard deviation below the mean in the age-appropriate language screening test.

Stimuli

Forty-two Cantonese monosyllabic words were selected (Appendix B). Twelve words with entering tones, which are reduced tones occur in close syllables with final stop consonants and are conventionally considered as the allotones of T1 (HL), T3 (ML) and T6 (LL). These tones were included in the picture naming task and tone judgement but were excluded for further analysis because of the drastic acoustic differences on the tone shapes and duration from their full tone counterparts (Wong & Chan, 2018). The 30 words were mostly selected based on their familiarity to young children and whether the target words could be presented clearly in pictures. All tones had at least one word with a long vowel and a diphthong. All tones also had a word with a high vowel, except T2. Of the 30 words with full tones, 18 (3 words x 6 tones) were familiar words produced by 90% or more of 30-

month-old children growing up in Hong Kong (Tardif, Fletcher, Liang, & Kaciroti, 2009). Twelve words (2 words x 6 tones) were not found in the vocabulary of pre-school children (Tardif et al., 2009). Three additional words were included for three practice trials before the experimental trials.

Procedure

Each of the 62 mother-child dyads attended a ninety-minute session either in a quiet room in their home or in a sound-attenuated recording room in the University of Hong Kong. The mothers verbally labelled 45 pictures representing the practice and target words two times to the experimenter and to their children. In the first block, mothers produced the monosyllabic words in isolation. In the second block, they produced the tones in the final position of the carrier sentence “聽我讀 /t^{h}εŋ1 ŋɔ2 tʊk6/ [Listen to me saying ___]. Mothers with more than one child repeated the procedures with each of their children.

Perceptual Judgment of tones

Sound files from two child-mother dyads (M46, M52) and sound files in isolated word condition of a mother (M55) were excluded due to corruption, leaving 8634 sound files for tone judgment. Target words in sentence final condition were excised and saved in individual files. Because lexical biases may affect accurate identification of phonetic differences (Davis & Johnsrude, 2007; Oller & Eilers, 1975), ADS were low-pass filtered at 400 Hz, whereas CDS were low-pass filtered at 500 Hz so that pitch information of the word was retained while lexical information was removed (Wong, 2012a; 2013). CDS were low-pass filtered at a higher cut-off frequency because it is usually produced with high pitch. The tone productions were blocked by speakers and speech registers. The four conditions of tone productions (i.e., isolated tones in ADS, isolated tones in CDS, tones in sentence final in ADS, tones in sentence final in CDS) of the same speaker were put in four experimental blocks in the same experiment. Thus there were 60 experiments. Each had 4 blocks of 42 tone productions, except that there were only two blocks of tones for M55.

Five native Cantonese-speaking undergraduate students (aged from 20 to 23 years old) born and raised in Hong Kong reporting Cantonese as their first and strongest language served as judges. None reported having any difficulties with speech or language. They listened to the tones in the 60 experiments in multiple 1-2 hour sessions. Blocks of tone productions and productions within blocks were presented via headphones in random orders. Each judge identified the tones by typing the tone number representing the tones. All judges re-rated productions of six mothers of children in different age groups (10% of all production) to determine intra-judge reliability.

Acoustic Analysis of tones

Following the methods in Wong (2012b) and Wong, Fu, & Cheung, (2017) the recorded tones were manually segmented into the initial section, pitch section and final section using Praat (Boersma & Weenink, 2014). The initial section started from the beginning of the articulation of the word to the end of the first regular vocal pulse. The final section started from the beginning of the last regular vocal pulse to the end of the articulation. The pitch section included all the regular vocal pulses throughout the productions except the first and the last vocal cycle. The markings of the vocal pulses generated by Praat were checked for error and corrected manually.

The pitch section was divided into 20 intervals equal in duration. To better reflect pitch perception, F0 was converted to the psycho-acoustic scale of semitones (St) using 1 Hz as reference frequency (Nolan, 2003). Because F0 contours of tones in the first half of the syllable are affected by factors such as the preceding tone (Wong & Strange, 2017; Xu & Liu, 2006) and the aspiration feature of initial consonants (Wong & Xu, 2007; Xu & Xu, 2003), and because reliable cues and pitch target for tone identification occur towards the end of the syllable (Xu, 1997; Xu & Wang, 2001; Whitehill, Ciocca, and Chow, 2000), acoustic comparisons were focused on the tone contours in the last 50% of the syllable (Wong, Fu & Cheung, 2017). Following Wong and Chan (2018) and Wong, Fu and Cheung (2017), eight acoustic parameters that characterized the pitch levels and pitch shapes of the tones were measured. To normalize for individual differences on vocal pitch, measured pitch values were converted to pitch height values by subtracting the mean pitch across all productions of the speaker from the measured pitch (Wong, 2012b; Wong et al., 2017) This method has been proved to successfully normalize tones produced by speakers with different pitch ranges such as male and female, and children and adults (Wong, Fu, & Cheung, 2017;Wong & Chan, 2018). Table 1 provides description of the acoustic terms and the eight acoustic parameters used.

---- Table 1 ----

Results

In the following analyses, tones in isolated words in ADS (AD-Iso), which represent tone productions in neutral contexts, are used as reference. Acoustic and perceptual differences between AD-Iso and tones produced in isolation in CDS (CD-Iso) represent effect of the child-directed register on tones, while acoustic and perceptual differences between AD-Iso and tones excised from sentence final position in ADS (AD-SentF) represent effect of sentence final prosody on tones. Tone productions in CDS in sentence final position (CD-SentF) is affected by the combined effect of sentence final prosody and the child-directed register. They were included in the testing procedure to balance the tasks in ADS and CDS, and were included in the statistical models. However, detailed analysis was not performed on this condition because of the impossibility to separate the two effects (sentence final prosody and CDS) on the tones. Data from M55 were discarded for further analysis due to missing data on isolated word productions.

Mothers who participated in the study with more than one child were included in the analyses because they provided information on the effect of children's age on CDS by the same mothers and because analyses on a subset of the data with only one child for each mother showed the same statistical results and correlational patterns, except that there was no significant differences in slope 50% for T5 in AD-Iso and CD-Iso and that the Mid Pitch Height of T4 (LF) was significantly higher in AD-Iso than in AD-SentF ($p < .001$, $r = .272$).

Inter-judge Reliability and Intra-judge Reliability in Tone Judgement

The judges were highly reliable and consistent in Cantonese tone identification. Kappa statistics, which take into account agreement by chance, were used to examine the level of agreement among the five judges in their ratings of all the tones produced by the mothers, and the level of agreement of their ratings and re-ratings of the tones produced by six of the mothers (about 10% of all productions). According to the conventional interpretation of the kappa coefficients (Landis and Koch, 1977; Posner et al., 1990), results of Cohen's Kappa (κ) revealed substantial inter-judge reliability between each pair of judges, κ ranged from 0.718 to 0.785. Fleiss's Kappa (κ) of 0.749 also showed substantial agreement among all the judges. Cohen's Kappa for intra-judge reliability ranged from 0.777 to 0.838, for the five judges, indicating substantial to almost perfect intra-judge agreement.

Acoustic Characteristics of Child-directed Cantonese Tones

Effect of the child-directed register on the six tones. Figure 2 shows the mean pitch contours of the six tones in isolated words and in sentence final position in CDS and ADS. As shown, the pitch contours of the tones in CDS (CD-Iso and CD-SentF) were higher than in ADS (AD-Iso and AD-SentF) with the tonal contours in sentence final position lower than in isolated productions in both AD and CD conditions. The essentially parallel contours of the same tone in most of the conditions, except in sentence final conditions suggested that the shapes of the tone contours were by and large maintained across the conditions.

A four-way repeated measure ANOVA (register (AD, CD) x prosodic context (isolation, sentence final) x word familiarity (familiar, unfamiliar) x tones (6 tones)) was performed for each of the eight acoustic parameters (see Table 1). The results showed main effects of registers, prosodic contexts, and target tones on all acoustic parameters (all ps < .001) except that main effects of register and prosodic context were not significantly different for slopes. To examine the effect of CDS on the acoustic measurements of each tone, post-hoc pairwise comparisons were performed to compare the acoustic differences of each tones in CD-Iso and AD-Iso. All the six pitch height parameters were significantly higher in CD-Iso than in AD-Iso for all the six tones (Tables 2-7), indicating that CDS had pitch heightening effects on the six tones. Durations were all longer in CD-Iso than in AD-Iso. The slopes in CD-Iso and AD-Iso were comparable for T3 (ML), T4 (LF) and T6 (LL) (Table 2-7), but the pitch contour of the two rising tones, T2 (HR) and T5 (LR), rose more sharply in CD-Iso than in AD-Iso. For T1 (HL), the slope of the pitch contour fell more sharply in CD-Iso than in AD-Iso. These findings showed that tones in CDS were produced with higher pitch levels and longer durations. Some of the tones in CDS maintained the same tone shapes as in ADS. Of the tones that exhibited significantly different contour shapes in CDS, the modified pitch contour shapes may deviate from the expected pitch shape of the tones (e.g., High level tone fell significantly more sharply in CDS).

Figure 2. Mean pitch contours of the six tones in isolated words and in sentence final position in child-directed and adult-directed speech

----- Table 2-7 ------

Developmental changes of the acoustic parameters of the tones in CDS. Pearson's correlations were performed between the age of the children and the values of the eight acoustic parameters in the four production contexts. The results in Table 8 showed no significant developmental changes of the acoustic measures in the two AD conditions. Significant negative correlations were found between age and the six pitch height measures in the CD-Iso condition, except for Mid Pitch Height for T4 (LF) and Min Pitch Height for T4 (LF) and T5 (LR), indicating that the pitch level of mothers' tones in CDS decreased as the child's age increased. Yet, no significant correlation was found between age of children and tone duration or pitch slopes in CDS. The findings showed that mother mainly modified the pitch levels but not the pitch shapes or tone duration when speaking to children of different ages, and the degree of modifications decreased as the child got older. The correlations in CD-SentF were similar to those in CD-Iso (Table 8), but with smaller effects, suggesting that the sentence final context may have offset some of the effects of CDS.

---- Table 8 ----

Perceptual Characteristics of Child-directed Cantonese Tones

Perceptual accuracy of tones in CD-Iso and AD-Iso was compared to examine the effect of child-directed register on Cantonese tone identification accuracy. Figure 3 shows the perceived accuracy of the tones in word isolation and sentence final position in adult-directed and child-directed speech. Tables 9-12 show the identification accuracy and error patterns of the tones in the four production contexts by the judges. Tables 9 and 10 show the percent correct and substitution patterns of the six tones in AD-Iso and CD-Iso, respectively. Because the data violated the assumptions for parametric statistics, tone accuracy values were converted to rational arcsine units (RAUs), which approximated percentage values but have the statistical properties of the arcsine transformation (Studebaker, 1985; Studebaker, McDaniel, & Sherbecoe, 1995). A three-way ANOVA (Prosodic context x tone x register) was performed on the RAU scores. The results showed significant main effect of tone ($F(3.97, 229.99) = 122.01$, $p<.001$, $\eta p^2 = .68$), significant interaction effects of prosodic context x tone ($F(4.04, 234.21) = 3.84$, $p=.005$, $\eta p^2 = .062$), prosodic context x register ($F(1, 58) = 17.25$, $p < .001$, $\eta p^2 = .23$), and tone x register ($F(4.278, 248.118) = 11.076$, p , .001, $\eta p^2 = .16$). Pairwise comparisons with Bonferroni adjustments for multiple comparisons showed lower overall tone accuracy in CD-Iso than in AD-Iso ($p = .001$, Table 13). Among the six tones, T5 (LR) and T6 (LL) were perceived with lower accuracy in CD-Iso than in AD-Iso (ps < .001, Table 13). Error patterns in Tables 9 and 10 show that more T5 (LR) productions were perceived as T2 (HR) and more T6 (LL) were perceived as T3 (ML) in CD-Iso. These findings revealed that tones were perceived with lower accuracy in CDS than in ADS, with more tones with lower pitch levels being perceived as tones with similar pitch shapes but higher pitch levels (i.e., T5 (LR) → T2 (HR), T6 (LL) → T3 (ML)).

Figure 3. Perceptual accuracy of the six tones and with all tones collapsed in different prosodic contexts in child-directed and adult-directed speech.

Notes. "AD-Iso" stands for adult-directed tones produced in isolation. "CD-Iso" stands for child-directed tones produced in isolation. "AD-SentF" stands for adult-directed tones excised from sentence final position. "CD-SentF" stands for child-directed tones excised from sentence final position.

----- Tables 9-12 -----

----- Table 13 -----

Do mothers hyper-articulate the tones in CDS?

To determine whether mothers hyper-articulated the tones in CDS, the acoustic differences in mean pitch levels and the slopes of the tones in the two confusing tones pairs (e.g., slope of T2 (HR) minus slope of T5 (LR), and pitch level of T3 (ML) minus pitch level of T6 (LL)) in AD-Iso and CD-Iso were compared. Pitch levels and slopes were selected because they are phonemically contrastive for the confusing tones. No significant differences were found in the four paired sample t-tests (ps=.269 to .745), suggesting that the pitch level and pitch shape differences in the confusing tones were comparable in AD-Iso and in CD-Iso and no exaggeration of the acoustic differences was found in the two speech registers.

Are the perceptual characteristics of child-directed Cantonese tones indicative of prosodic effects?

To answer the fourth question, first, acoustic characteristics of tones in AD-SentF were compared to those in AD-Iso to confirm whether sentence final prosody had pitch lowering effects on the tones produced by the mothers. Then, the perceptual accuracies of the tones in AD-Iso and AD-SentF contexts were compared to determine the effect of pitch lowering on the perceptual accuracy of tones. Lastly, effects of sentence final prosody on tone perception were compared to the effects of CDS on tone perception to examine whether the observed perceptual differences in CDS was affected by prosodic effects.

Effects of sentence-final prosody on the acoustics of the tones. As mentioned above, results of the four-way repeated measure ANOVA (register x prosodic context x word familiarity x tones) showed significant main effects of registers and prosodic contexts on all the acoustic measurements (all ps < .001), except for slopes. Pairwise comparisons with Bonferroni adjustment for multiple comparisons were conducted to compare the acoustic measures in AD-SentF and AD-Iso. The results in Tables 14-19 showed that with the exception of the Mid Pitch Height in T4, all the six pitch height measures were significantly lower in AD-SentF than in AD-Iso, confirming a pitch lowering effect of sentence-final prosody on tones. These findings indicated that pitch levels of tones were lower in sentence-final position than in isolation, whereas the duration and pitch slopes were maintained.

----- Tables 14-19 ----

Effects of pitch lowering on tone perceptual accuracy. Tables 9 and 11 present the accuracy scores and error patterns of the perceptual accuracy of the tones in AD-Iso and AD-SentF, respectively. Results of the three-way ANOVA (prosodic context x tone x register) showed that the overall tone accuracy in AD-SentF was lower than in AD-Iso, (p = .004, Table 13), indicating more perception errors in identifying tones in sentence final position, a pitch lowering context. Pairwise comparisons were performed to compare the accuracy of each tone in AD-SentF and AD-Iso to determine which of the six tones were affected by the pitch-lowering context. The results showed that T2 (HR) and T3 (ML) were identified with significantly lower accuracy in AD-SentF than in AD-Iso (ps = .011 and = .003, respectively, Table 8), indicating that sentence-final pitch lowering prosody negatively affects the perception of these two tones. Table 11 shows that more productions of T2 (HR) were perceived as T5 (LR) and more productions of T3 (ML) were perceived as T6 (LL) in AD-SentF than in AD-Iso, suggesting that under pitch lowering prosodic effects, tones with relatively higher pitch levels were more likely to be mis-categorized as tones with similar contour shapes but lower pitch levels (i.e., T2 (HR) → T5 (LR), and T3 (ML) → T6 (LL)).

Similarities and differences between the effects of CDS and sentence final prosody on tones. Acoustically, CDS had a pitch heightening effect on the six pitch height measures (Tables 2-7) while sentence final prosody had a pitch lowering effect on the pitch heights of the tones (Tables 14-19). Perceptually, the perceptual rates of tones in CD-Iso (Table 10) and AD-SentF (Table 11) were lower than in AD-Iso (Tables 9), particularly in the two confusing tone pairs (i.e., T2 (HR) - T5 (LR), and T3 (ML) – T6 (LL) (Tables 9-12, 13 and results presented above). Comparing to the neutral condition (AD-Iso), tones in AD-SentF were produced with lower pitch (Table 14) and more tones with higher pitch levels were misperceived as tones with similar pitch shapes but lower pitch levels (i.e., more T2 (HR) → T5 (LR) and more T3 (ML) → T6 (LL) (Tables 9, 11), leading to more errors in T2 (HR) and T3 (ML) in AD-SentF than in AD-Iso. Taken together both CDS and sentence final prosody had an adverse effect on tone perception accuracy, but the effects were in opposite directions, with CDS having larger effects on tones with lower pitch and sentence final prosody having larger effects on tones with higher pitch.

Discussion

This study was set out to test the hyper-articulation and prosodic hypotheses on the function of the acoustic modifications observed in CDS by examining child-directed and adult-directed Cantonese tones.

Acoustic characteristics of Child-directed Cantonese Tones

The first research question examined the acoustic characteristic of child-directed Cantonese-tones. The results confirmed that mothers acoustically modified their tones when speaking to young children. All the tones were produced with higher pitch levels and longer durations in CDS than in ADS. However, the shapes of the tone contour changed in various directions in CDS and the changes did not enhance tone identification. Three of the tones maintained the same pitch shapes in ADS-Iso and CDS-Iso. However, T1 (HL) was produced with a pitch contour dropping significantly more sharply than the level pitch contour of T1 (HL) in AD-Iso. The two rising tones were produced with steeper rising slopes in CD-Iso than in AD-Iso, but such acoustic modifications did not enhance tone perception accuracy in CDS, as indicated by the significantly lower overall tone identification accuracy in CDS, and in particular, lower tone identification accuracy for T5 (LR) and T6 (LL) in CD-Iso than in AD-Iso.

Consistent developmental changes were observed in the acoustic properties of child-directed Cantonese tones. Mothers decreased the pitch level of the tones as the age of the child increased. Tone duration and tone shape in CDS, however, did not change with the child's age. These findings were consistent with the findings in Garnica (1977), who reported that the pitch height and pitch range in mothers' speech directed to two-year-old children were higher than to five-year-olds. Stern, Spieker, Barnett and MacKain (1983) who examined the acoustic properties of CDS in mothers' speech also reported similar developmental changes. They found that mothers' pitch range to four- month-old children was larger than to new-borns and 12- and 14-month old children. Amano et al. (2006), however, reported shorter developmental changes in Japanese mothers. They found that mothers' pitch level decreased when they addressed infants between 0-1;6, but the pitch level did not differ from that in ADS when they talked to children older than 1;7. Because the study examined mothers' productions recorded from their daily conversations with their children without controlling the phonetic, prosodic and emotional contexts, the findings may be affected by the variations in phonetic and prosodic contexts in the tones produced to different children

Heightened pitch levels appear to be a universal feature of CDS. Studies that compared pitch levels in adult-directed and CDS consistently found higher pitch in child-directed productions than adult-directed productions, and the finding have been reported in a range of non-tonal languages, such as English (Lam & Kitamura, 2012), German (Fernald & Mazzie, 1991) and Norwegian (Englund & Behne, 2006), and in several tonal languages such as Mandarin (Liu et al., 2009) and Cantonese (Ng, 2016; Rattanasone et al., 2013).

Perceptual characteristics of child-directed Cantonese tones

Listeners made significantly more perception errors in child-directed tones than adult-directed tones, particularly in the tones of the two easily confusing tone pairs (T2 (HR) vs. T5 (LR) and T3 (ML) vs. T6 (LL)). T5 (HR) and T6 (LL) were identified significantly poorer in CDS than in ADS despite the fact that CDS were low-pass filtered at a higher cut-off frequency. The error patterns showed that tones with lower pitch levels were more likely to be identified as tones with similar tone shapes but higher pitch levels in CDS. Thus, more T5 (LR) was mis-perceived as T2 (HR) and more T6 (LL) was mis-perceived as T3 (ML) in CDS than in ADS. Very few previous studies compared the perceptual accuracy of CDS and ADS. Those that compared perceptual accuracies in ADS and CDS consistently found lower accuracy in CDS. Using a similar research design, Wong, (2018) reported lower identification accuracy in child-directed monosyllabic and disyllabic Mandarin tones. Using statistical and computer models to categorize the phonemes based on the acoustic measurements obtained from CDS and ADS, Martin et al., (2014) and McMurray et al (2013) also reported poorer identification of phonemes in CDS than in ADS.

Do mothers enhance the phonetic contrasts of the more confusing Cantonese Tones in CDS?

The findings of this study did not show evidence that mothers hyper-articulated the tones in CDS. The most compelling evidence was that child-directed tones were identified with lower accuracy than ADS. Also, no significant difference in the acoustic distance in the pitch levels and the pitch shapes of the two easily confused tone pairs was found in ADS and CDS. Moreover, though acoustic modifications were found in child-directed Cantonese tones, the major changes involved heightening of pitch levels. There were some changes in the pitch shapes of the tones in CDS. However, the changes sometimes moved away from the expected canonical form of the tones (e.g., a steeper falling contour in T1 (HL) in CDS than in ADS) and the modified contour shapes of the tones in CDS did not enhance the perceptual accuracy of the tones (e.g., steeper rising contours in T5 (LR) in CDS resulted in poorer identification of T5 (LR) in CDS.

The findings in this study are inconsistent with the findings in previous studies that supported hyper-articulation in child-directed tones (e.g., Liu et al., 2003, 2007, 2009; Rattanasone et al. (2013). Differences in the research design might explain the discrepancies in the findings between these studies and those in the present study. First, all previous studies that examined child-directed lexical tones did not control the speech production contexts. ADS was typically collected in a formal interview with the mothers while CDS was usually collected during mothers' free play with the children. As indicated above, pitch levels and contours of tones are affected by prosodic, phonetic and emotional contexts. Thus, the pitch measures obtained in these samples may have been confounded by these effects. This study, however, tightly controlled the production contexts in ADS and CDS. Only tones produced in monosyllabic words in isolation in the same picture reading task in ADS and CDS were compared. Second, in previous studies, limited acoustic parameters were measured and the acoustic parameters selected might not be relevant to tone identification. Liu et al. (2007, 2009) based their findings on the comparisons of mean F0, F0 range, tone duration and F0 turning points of the four Mandarin tones. No comparison was made on the tone shapes, the perceptual cues for Mandarin tones. The conclusion that mothers hyper-articulated the tones was based on the significantly higher pitch in CDS than in ADS, and the same order of pitch duration and pitch height in child-directed and adult-directed tones. However, these measurements could also be explained by prosodic differences in the productions in the two registers. Rattanasone et al. (2013) compared the pitch at the onset and offset of three of the Cantonese tones (T1 (HL), T2 (HR) and T4 (LF)). These acoustic parameters were not the perceptual cues for the identification of the tones (Khouw & Ciocca, 2007; Xu & Wang, 2001). The present study, however, sampled a number of acoustic parameters including mean pitch, the pitch levels at five critical points, and the contour shape of the tones from the second half of the tone contours that have been found to provide more reliable cues for tone identification (Khouw & Ciocca, 2007; Xu & Wang, 2001). Acoustic contrasts in the primary cues for distinguishing easily confused tones were also compared to determine whether mothers exaggerated the acoustic distance of the more confusing tones. Third, none of the previous studies that concluded that mothers enhanced the acoustic cues of the tones in CDS examined tone perception accuracy in CDS. Rattanasone et al. (2013) examined vowel space and tone space of CDS in Cantonese. They found contradictory findings between the two measures. Mothers increased the tone space but did not increase the vowel space in CDS. Based on the expanded tone space in CDS, the authors concluded that mothers hyper-articulated the tones. However, tone space only took into account the area formed by the F0 onset and offset of T1 (HL), T2 (HR) and T4 (LF) (Barry & Blamey, 2004). Larger tone space does not necessarily mean that the tones are more perceptually distinct. In the present study, despite the findings of acoustic modifications in the tones, the perceptual results showed that these acoustic changes did not lead to better identification of the tones and, therefore, rejected the hyper-articulation hypothesis.

Are the perceptual characteristics of Cantonese tones indicative of prosodic effects?

To better answer the question, this study compared the acoustic and perceptual characteristics of child-directed tones to tones affected by pitch-lowering effects in sentence final position. Acoustically, similar to those reported in previous studies (e.g., Ma, Ciocca & Whitehill, 2004; Vance, 1976; Liu et al., 2007), this study found that sentence final prosody had pitch lowering effect on Cantonese tones, while the child-directed register had pitch heightening effect on tones. Perceptually, similar error patterns were found in identifying tones affected by pitch lowering effect in sentence final and tones affected by pitch heightening effect in CDS. Tones in these two contexts were identified with lower accuracy than adult-directed tones in isolation. Identification accuracy of the easily confusing tone pairs (T2 (HR) vs. T5 (LR), and T3 (ML) vs. T6 (LL)), which had similar pitch shapes and relatively small differences in the pitch levels, were particularly more difficult in these two contexts. The main differences in tone perceptual accuracy were that in the pitch lowering context in sentence final position, the tones in the confusing tone pairs with higher pitch levels (i.e., T2 (HR) and T3 (ML)) were more likely to be mistaken as the tone counterpart that had similar pitch shape but lower pitch levels (i.e., T2 (HR) being perceived as T5 (LR) and T3 (ML) being perceived as T6 (LL)). Yet, in pitch heightening contexts in CDS, the opposite occurred. Tones in the confusing tone pairs with lower pitch levels were more likely to be mistaken as the tone counterpart that had similar pitch shape but higher pitch levels (i.e., T5 (LR) being perceived as T2 (HR) and T6 (LL) being perceived as T3 (ML)). The findings suggested that the perceptual characteristics of child-directed tones could be the effects of the pitch raising prosodic contexts in CDS.

A number of studies have investigated the relationship between pitch and emotions. Happy emotions (Xu et al., 2013) and positive moods (Singh et al., 2002; Trainor et al., 2000) are expressed in high pitch, regardless of the speech register (Singh et al., 2002; Trainor et al., 2000). Thus, the raising of pitch in CDS may be mother's expression of positive affects towards their children in CDS. If the purpose of the acoustic modifications in CDS was to serve pragmatic functions, it was not surprising that the pitch changes in CDS went in various directions and can sometimes deviated from the expected pitch levels and pitch shapes of the tones. It may also explain the inconsistent findings in the vowel space and voice onset time of stop consonants found in CDS.

To examine the pitch lowering prosodic effect on tone perception so as to understand the possible effects of pitch raising on tone perception in CDS, this study determined accuracy of tones excised from word final position in sentences. Though the findings showed that tones that were taken from sentence final were identified with lower accuracy rates than tones produced in isolation, because tone perception is adjusted by the pitch level of the tones in the immediate contexts (i.e., preceding and following the target tone) (Wong & Diehl, 2003), the accuracy rates of the tones are expected to be higher when they are presented in the original sentence with information about the pitch levels of the preceding tones.

It is less clear how well the perceptual accuracy of CD tones in isolation can be improved without linguistic cues. Native speakers have no difficulties normalizing tones produced by speakers of very different intrinsic pitch (e.g., male and female, Wong & Chan, 2018). However, unlike tones in sentence final position, which differed from tones in isolation mostly in pitch level only, CD tones differed from AD tones in pitch heights, pitch shapes and durations, and the direction of change was not as consistent as in AD tones produced in sentence final position (compare Tables 5 vs. 6 and Tables 2 vs. 6). Yet, tone ambiguity is expected to be largely resolved in real life situations given the support of linguistic (e.g., lexical status, and semantic and syntactic contexts) and environmental cues.

Conclusion

In conclusion, this study investigated the function of CDS using both perceptual and acoustic analyses of child-directed and adult-directed Cantonese tones. Mothers elevate the pitch levels when speaking to young children. However, they did not enhance the distance of the acoustic cues contrasting the more confusing tones. Overall, tone perception accuracy was lower in CDS than in ADS. The major errors in the perception of child-directed tones involve the mis-perception of tones of lower pitch levels as tones with similar pitch shapes but higher pitch levels. These results do not support the hyper-articulation hypothesis of CDS. The results appear to support the pragmatic hypotheses of CDS because there is strong and consistent evidence that mothers do acoustically modify their speech when talking to young children. However, these acoustic changes do not always enhance the tonal contrasts in CDS and may sometimes lead to poorer tone identification in CDS. The acoustic characteristics in CDS resemble those in affectionate speech and mirror the effects of sentence final prosody. Patterns of tone perception accuracy in CD tones also mirror those due to pitch lowering effect of sentence final prosody.

References

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *The Journal of the Acoustical Society of America, 119*(3), 1636-1647.

Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *The Journal of the Acoustical Society of America, 105*(2), 1095-1096.

Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *The Journal of the Acoustical Society of America, 116*(3), 1739-1748.

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development, 36*(4), 847-862.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer (Version 5. 3. 79) [Computer software]. Available from http://www.praat.org/

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

Chao, Y. R. (1947). *Cantonese primer*. Cambridge, MA: Harvard University Press.

Ciocca, V., & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders, 1*(2), 141-147.

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

Cristia, A., & Seidl, A. (2014). The hyper-articulation hypothesis of infant-directed speech. *Journal of Child Language, 41*(4), 913-934.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1), 132-147.

Dodane, C., & Al-Tamimi, J. (2007, August). An acoustic comparison of vowel systems in adultdirected-speech and CDS: Evidence from French, English & Japanese. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)* (pp. 1573-1576). Saarbrücken: Saarland University.

Englund, K. (2005). Voice onset time in infant directed speech over the first six months. First Language 25, 220–34.

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

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60*(6), 1497-1510.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jürgens, & M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262-282). Cambridge: Cambridge University Press.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279-293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477-501.

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

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(1), 45-57.

Garnica, O. K. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 63-88). Cambridge: Cambridge University Press.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*(1), 14-20.

Howie, J. M. (1976). Acoustical studies of Mandarin vowels and tones (Vol. 6). Cambridge University Press. Chapter 1, pp. 1-31.

Kei, J., Smith, V., So, L. K., Lau, C. C., & Capell, K. (2002). Assessing the accuracy of production of Cantonese lexical tones: A comparison between perceptual judgement and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing, 7*(1), 25-38.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. Journal of Phonetics, 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., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684-686.

Lam, C., & Kitamura, C. (2010). Maternal interactions with a hearing and hearing-impaired twin: Similarities and differences in speech input, interaction quality, and word production. *Journal of Speech, Language, and Hearing Research, 53*(3), 543-555.

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyper-articulation. *Developmental Science, 15*(2), 212-221.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Lee, K. Y., Chan, K. T., Lam, J. H., van Hasselt, C. A., & Tong, M. C. (2015). Lexical tone perception in native speakers of Cantonese. *International Journal of Speech-Language Pathology, 17*(1), 53-62.

Liu, H. M. (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 Education and Development, 25*(8), 1103-1117.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1-F10.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43*(4), 912-917.

Liu, H. M., Tsao, F. M., & 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. *Journal of Child Language, 36*(4), 909-922.

Ma, J. K., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. The Journal of the Acoustical Society of America, 120(6), 3978-3987..

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, *7*(3), 185-201.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. Psychological science, 26(3), 341-347.

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

Mok, P. P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language variation and change, 25*(3), 341-370.

Ng, H. Y. (2016) *Mothers do not enhance phonemic contrasts of Cantonese lexical tones with didactic intent in infant-directed speech*. (Unpublished bachelor dissertation). The University of Hong Kong, Hong Kong.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 771-774). Barcelona, Spain.

Oller, D. K., & Eilers, R. E. (1975). Phonetic expectation and transcription validity. *Phonetica*, *31*(3-4), 288-304.

Po Leung Kuk. (2012). *Cantonese oral language deficiency early identification test for pre-primary children*. Hong Kong: Po Leung Kuk District-based Speech Therapy Team.

Posner, K. L., Sampson, P. D., Caplan, R. A., Ward, R. J., and Cheney, F. W. (1990). Measuring interrater reliability among multiple raters: an example of methods for nominal data. Stat. Med. 9, 1103–1115.

Rattanasone, N. X., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics, 41*(5), 332-343.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development*, *38*(5), 1-38.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3*(3), 365-394.

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1-15.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. Journal of Speech and Hearing Research, 28(3), 455-462.

Studebaker, G. A., McDaniel, D. M., & Sherbecoe, R. L. (1995). Evaluating relative speech recognition performance using the proficiency factor and rationalized arcsine differences. Journal of American Academy of Audiology, 6, 173-182.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. Phonetica, 56(3-4), 186-199.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*, *36*(5), 1115-1144.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11*(3), 188-195.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335-340.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication, 49*(1), 2-7.

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica, 33*(5), 368-392.

Warren-Leubecker, A., & Bohannon III, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, *55*(4), 1379-1385.

Wong, P. C., & Diehl, R. L. (2003). Perceptual normalization for inter-and intratalker variation in Cantonese level tones. Journal of Speech, Language, and Hearing Research, 46(2), 413-421.

Wong, P. (2012a). Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and in the United States: Interjudge reliability and perceptual results. *Journal of Speech, Language, and Hearing Research*, 55(5), 1423-1437.

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

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, 133(1), 434-443.

Wong, P., & Chan, H. (2018). Acoustic Characteristics of Highly Distinguishable Cantonese Tones. Journal of Acoustical Society of America

Wong, P., Fu, W. M., & 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. Frontiers in Psychology. Published online: August 29, 2017.

Wong, P., & Leung, C. T. (2018) Suprasegmental features are not acquired early: Perception and production of monosyllabic Cantonese lexical tones in four- to six-year-old children. *Journal of Speech, Language, and Hearing Research*.

Wong, P., & Strange, W. (2017). Phonetic complexity affects children’s Mandarin tone production accuracy in disyllabic words: A perceptual study. PLOS One. Published online: August 14, 2017.

Wong, P. (2018). Mothers do not enhance tonal contrasts in child-directed speech: Perceptual and acoustic evidence from child-directed Mandarin lexical tones, The Journal of the Acoustical Society of America.

Wong, Y. W., & Xu, Y. (2007). Consonantal perturbation of f0 contours of Cantonese tones. In *Proceedings of The 16th International Congress of Phonetic Sciences* (pp. 1293-1296). Saarbrucken, Germany: Universität des Saarlandes.

Whitehill, T. L., Ciocca, V., & Chow, D. T. Y. (2000). Acoustic analysis of lexical tone contrasts in dysarthria. *Journal of Medical Speech-Language Pathology, 8*(4), 337-344.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(2), 165-181.

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

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, *17*, 1-31.

Xu, Y. (2004) Understanding tone from the perspective of production and perception. Language and Linguistics, 5, 757-797

Xu, Y., Kelly, A., & Smillie, C. (2013). Emotional expressions as communicative signals. In S. Hancil & D. Hirst (Eds.), *Prosody and Iconicity* (pp. 33-60). Amsterdam: John Benjamins Publishing Company.

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics, 18*(1), 125-159.

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

Figure Caption

Figure 1. Fundamental frequency contours of Cantonese tones.

Note. 1(HL), 2 (HR), 3 (ML), 4 (LF), 5 (LR), 6 (LL) stand for T1 (High Level), T2 (High Rising), T3 (Mid Level), T4 (Low Falling), T5 (Low Rising), and T6 (Low Level), respectively.

Figure 2. Mean pitch contours of the six tones in isolated words and in sentence final position in child-directed and adult-directed speech

Notes. 1(HL), 2 (HR), 3 (ML), 4 (LF), 5 (LR), 6 (LL) stand for T1 (High Level), T2 (High Rising), T3 (Mid Level), T4 (Low Falling), T5 (Low Rising), and T6 (Low Level), respectively. AD and CD stand for adult-directed and child-directed, respectively.

Figure 3. Perceptual accuracy of the six tones in different prosodic contexts in child-directed and adult-directed speech.

Notes. "AD-Iso" stands for adult-directed tones produced in isolation. "CD-Iso" stands for child-directed tones produced in isolation. "AD-SentF" stands for adult-directed tones excised from sentence final position. "CD-SentF" stands for child-directed tones excised from sentence final position.

Table 1

Description of Acoustic Terms and Acoustic Parameters

<table>
<tr><th colspan="2">Acoustic term and parameter</th><th>Description</th></tr>
<tr><td colspan="2">Speaker Mean Pitch (St)</td><td>Mean pitch cross all AD productions of the speaker</td></tr>
<tr><td colspan="2">Pitch Height (St)</td><td>Pitch level relative to the mean pitch of the speaker (i.e., measured pitch value minus speaker mean pitch), negative values indicate that the measured pitch is lower than the Speaker Mean Pitch, positive values indicate that the measured pitch is higher than the Speaker Mean Pitch.</td></tr>
<tr><td colspan="2">Pitch Target</td><td>The final 50% of the tone contours</td></tr>
<tr><td rowspan="6">Pitch Height measures</td><td>Initial Pitch Height 100% (St)</td><td>Pitch at tone onset (i.e., time point 1) of the production relative to the mean pitch of the speaker (i.e., minus Speaker Mean Pitch)</td></tr>
<tr><td>Final Pitch Height 100% (St)</td><td>Pitch at tone offset (i.e., time point 20) minus Speaker Mean Pitch</td></tr>
<tr><td>Mid Pitch Height 100% (St)</td><td>Pitch at the midpoint of the tone (i.e., time point 11) minus Speaker Mean Pitch. This is also the pitch height at the onset of the pitch target (i.e., the final 50% of the tone).</td></tr>
<tr><td>Min Pitch Height 50% (St)</td><td>Minimum pitch in the second half of the tone minus Speaker Mean Pitch</td></tr>
<tr><td>Max Pitch Height 50% (St)</td><td>Maximum pitch in the second half of the tone minus Speaker Mean Pitch</td></tr>
<tr><td>Mean Pitch Height 50% (St)</td><td>Mean Pitch in the second half of the tone minus Speaker Mean Pitch</td></tr>
<tr><td colspan="2">Tone Duration 100% (ms)</td><td>Duration of the voiced section in the syllable, including all the pitch cycles in the vowels and the voiced consonants</td></tr>
<tr><td colspan="2">Slope 50% (St/ms)</td><td>(Max Pitch Height 50% - Min Pitch Height 50%) divided by the duration between Max Pitch Height 50% and Min Pitch Height 50%, positive values indicate rising slopes and negative values indicate falling slopes. This measures the pitch shapes.</td></tr>
</table>

Note. “St” stands for semi-tone, ‘ms’ stands for millisecond, ‘AD’ stands for adult-directed.

Table 2

Acoustic Differences in Adult-directed and Child-directed Tones Produced in Isolated Monosyllabic Words – Tone 1 (High Level)

Acoustic Parameters[a]	Tone 1 (High Level)		
	Result	**p-value**	**d**
Initial Pitch Height 100% (St)	CD > AD**	<0.001	0.438
Final Pitch Height 100% (St)	CD > AD**	<0.001	-0.450
Mid Pitch Height 100% (St)	CD > AD**	<0.001	-0.340
Min Pitch Height 50% (St)	CD > AD**	<0.001	-0.343
Max Pitch Height 50% (St)	CD > AD**	<0.001	-0.452
Mean Pitch Height 50% (St)	CD > AD**	<0.001	-0.425
Tone Duration 100% (ms)	CD > AD**	<0.001	-0.270
Slope 50% (St/ms)	CD > AD**[b]	0.009	0.431

Notes. [a]Refer to Table 1 for the definitions of the acoustic parameters
[b]CD with a larger value in the positive direction, indicating steeper rising contours
'AD' stands for adult-directed. 'CD' stands for child-directed. 'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 3

Acoustic Differences in Adult-directed and Child-directed Tones Produced in Isolated Monosyllabic Words – Tone 2 (High Rising)

Acoustic Parameters[a]	**Tone 2 (High Rising)**		
	Result	**p-value**	**d**
Initial Pitch Height 100% (St)	CD > AD**	<0.001	-0.359
Final Pitch Height 100% (St)	CD > AD**	<0.001	-0.343
Mid Pitch Height 100% (St)	CD > AD**	<0.001	-0.424
Min Pitch Height 50% (St)	CD > AD**	<0.001	-0.309
Max Pitch Height 50% (St)	CD > AD**	<0.001	-0.438
Mean Pitch Height 50% (St)	CD > AD**	<0.001	-0.399
Tone Duration 100% (ms)	CD > AD**	<0.001	-0.252
Slope 50% (St/ms)	CD > AD**[b]	0.001	-0.267

Note. [a]Refer to Table 1 for the definitions of the acoustic parameters
[b]CD with a larger value in the positive direction, indicating steeper rising contours
'AD' stands for adult-directed. 'CD' stands for child-directed. 'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 4

Acoustic Differences in Adult-directed and Child-directed Tones Produced in Isolated Monosyllabic Words – Tone 3 (Mid Level)

Acoustic Parameters[a]	**Tone 3 (Mid Level)**		
	Result	**p-value**	**d**
Initial Pitch Height 100% (St)	CD > AD**	<0.001	-0.414
Final Pitch Height 100% (St)	CD > AD**	<0.001	-0.416
Mid Pitch Height 100% (St)	CD > AD**	<0.001	-0.389
Min Pitch Height 50% (St)	CD > AD**	<0.001	-0.312
Max Pitch Height 50% (St)	CD > AD**	<0.001	-0.467
Mean Pitch Height 50% (St)	CD > AD**	<0.001	-0.413
Tone Duration 100% (ms)	CD > AD**	<0.001	-0.253
Slope 50% (St/ms)	CD = AD	0.225	0.152

Note. [a]Refer to Table 1 for the definitions of the acoustic parameters
'AD' stands for adult-directed. 'CD' stands for child-directed. 'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 5

Acoustic Differences in Adult-directed and Child-directed Tones Produced in Isolated Monosyllabic Words – Tone 4 (Low Falling)

Acoustic Parameters[a]	Tone 4 (Low Falling)		
	Result	**p-value**	**d**
Initial Pitch Height 100% (St)	CD > AD**	<0.001	-0.414
Final Pitch Height 100% (St)	CD > AD**	<0.001	-0.444
Mid Pitch Height 100% (St)	CD > AD**	0.001	-0.315
Min Pitch Height 50% (St)	CD > AD**	0.001	-0.281
Max Pitch Height 50% (St)	CD > AD**	<0.001	-0.486
Mean Pitch Height 50% (St)	CD > AD**	<0.001	-0.428
Tone Duration 100% (ms)	CD > AD**	<0.001	-0.345
Slope 50% (St/ms)	CD = AD	0.897	-0.010

Note. [a]Refer to Table 1 for the definitions of the acoustic parameters
'AD' stands for adult-directed. 'CD' stands for child-directed. 'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 6

Acoustic Differences in Adult-directed and Child-directed Tones Produced in Isolated Monosyllabic Words – Tone 5 (Low Rising)

Acoustic Parameters[a]	**Tone 5 (Low Rising)**		
	Result	**p-value**	**d**
Initial Pitch Height 100% (St)	CD > AD**	<0.001	-0.340
Final Pitch Height 100% (St)	CD > AD**	<0.001	-0.338
Mid Pitch Height 100% (St)	CD > AD**	<0.001	-0.370
Min Pitch Height 50% (St)	CD > AD**	<0.001	-0.286
Max Pitch Height 50% (St)	CD > AD**	<0.001	-0.400
Mean Pitch Height 50% (St)	CD > AD**	<0.001	-0.357
Tone Duration 100% (ms)	CD > AD**	<0.001	-0.256
Slope 50% (St/ms)	CD > AD*[b]	0.044	-0.391

Note. [a]Refer to Table 1 for the definitions of the acoustic parameters
[b]CD with a larger value in the positive direction, indicating steeper rising contours
'AD' stands for adult-directed. 'CD' stands for child-directed. 'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 7

Acoustic Differences in Adult-directed and Child-directed Tones Produced in Isolated Monosyllabic Words – Tone 6 (Low Level)

Acoustic Parameters[a]	**Tone 6 (Low Level)**		
	Result	**p-value**	**d**
Initial Pitch Height 100% (St)	CD > AD**	<0.001	-0.361
Final Pitch Height 100% (St)	CD > AD**	<0.001	-0.486
Mid Pitch Height 100% (St)	CD > AD**	<0.001	-0.420
Min Pitch Height 50% (St)	CD > AD**	<0.001	-0.440
Max Pitch Height 50% (St)	CD > AD**	<0.001	-0.484
Mean Pitch Height 50% (St)	CD > AD**	<0.001	-0.493
Tone Duration 100% (ms)	CD > AD**	<0.001	-0.279
Slope 50% (St/ms)	CD = AD	0.063	0.331

Note. [a]Refer to Table 1 for the definitions of the acoustic parameters
‘AD’ stands for adult-directed. ‘CD’ stands for child-directed. ‘St’ stands for semi-tone. ‘ms’ stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 8

Correlations between the Acoustic Parameters and Children's Age

Condition	Acoustic Parameters [a]	T1 (HL)	T2 (HR)	T3 (ML)	T4 (LF)	T5 (LR)	T6 (LL)
Adult-directed tones in isolation	Initial Pitch Height 100%	-0.04	-0.01	-0.01	-0.11	0.05	-0.04
	Final Pitch Height 100%	-0.07	-0.03	0.01	-0.13	0.09	-0.08
	Mid Pitch Height 100%	0.01	0.11	0.06	0.01	0.13	0.04
	Min Pitch Height 50%	0.01	-0.03	0.07	-0.03	0.08	-0.01
	Max Pitch Height 50%	-0.09	0.12	0.03	-0.09	0.09	0.10
	Mean Pitch Height 50%	-0.06	0.18	0.02	-0.10	0.09	-0.01
	Tone Duration 100%	-0.01	0.09	0.13	0.09	0.07	0.20
	Slope 50%	0.14	0.06	-0.01	0.19	-0.11	0.16
Adult-directed tones in sentence final position	Initial Pitch Height 100%	0.13	-0.16	-0.08	-0.11	-0.05	0.08
	Final Pitch Height 100%	0.04	-0.11	-0.13	-0.08	0.02	-0.14
	Mid Pitch Height 100%	0.04	0.03	0.04	0.10	0.14	0.13
	Min Pitch Height 50%	0.02	-0.12	-0.03	0.10	0.02	0.05
	Max Pitch Height 50%	0.07	0.02	-0.05	-0.04	0.07	0.03
	Mean Pitch Height 50%	0.01	-0.09	-0.09	0.04	0.05	0.01
	Tone Duration 100%	-0.03	0.05	0.08	-0.04	0.07	0.05
	Slope 50%	0.06	0.07	0.16	0.23	0.06	0.13
Child-directed tones in isolation	Initial Pitch Height 100%	-0.49**	-0.27*	-0.53**	-0.32*	-0.37**	-0.35**
	Final Pitch Height 100%	-0.56**	-0.37**	-0.50**	-0.36**	-0.30*	-0.45**
	Mid Pitch Height 100%	-0.57**	-0.45**	-0.44**	-0.12	-0.36**	-0.32*
	Min Pitch Height 50%	-0.57**	-0.33*	-0.40**	-0.15	-0.26	-0.36**
	Max Pitch Height 50%	-0.57**	-0.48**	-0.52**	-0.35**	-0.41**	-0.42**
	Mean Pitch Height 50%	-0.60**	-0.52**	-0.50**	-0.30*	-0.37**	-0.43**
	Tone Duration 100%	-0.01	-0.09	-0.13	-0.16	-0.94	-0.10
	Slope 50%	0.14	-0.19	0.06	0.02	-0.21	0.17
Child-directed tones in sentence final position	Initial Pitch Height 100%	-0.28*	-0.31*	-0.23	-0.29*	-0.21	-0.21
	Final Pitch Height 100%	-0.30*	-0.38**	-0.31*	0.31*	-0.30*	-0.37*
	Mid Pitch Height 100%	-0.26*	-0.12	-0.34**	-0.06	-0.09	-0.28*
	Min Pitch Height 50%	-0.26*	-0.34**	-0.34**	-0.11	-0.27*	-0.31*
	Max Pitch Height 50%	-0.30*	-0.16	-0.34**	-0.26*	-0.13	-0.34**
	Mean Pitch Height 50%	-0.31*	-0.29*	-0.35**	-0.20	-0.26*	-0.36**
	Tone Duration 100%	-0.09	-0.01	-0.07	-0.11	0.01	-0.05
	Slope 50%	-0.02	0.10	-0.05	0.17	0.09	0.08

Note. = .05 significance level ** = .01 significance level*

[a] Refer to Table 1 for the definitions of the acoustic parameters

Dark shaded cells mark significant correlations with large effects. Medium shaded cells mark significant correlations with medium effects. Light shaded cells mark significant correlations with small effects

Child-directed tones in sentence final position represent tones affected by both sentence final prosody and child-directed register

Table 9

Confusion Matrix of Tone Production in Adult-directed Tones in Isolated Words

Target tones	**Judges' responses (%)**					
	T1	**T2**	**T3**	**T4**	**T5**	**T6**
T1 (HL)	**87**	0	12	0	0	1
T2 (HR)	0	**96**	0	0	4	0
T3 (ML)	3	0	**73**	1	2	20
T4 (LF)	0	0	1	**91**	1	8
T5 (LR)	0	6	2	0	**91**	2
T6 (LL)	0	0	18	3	2	**77**

Notes. Correct identifications are marked by bold cells. Underlined cells mark error patterns that constitute more than 10% of the trials. T1, T2, T3, T4, T5, and T6 stand for Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, and Tone 6, respectively. ‘HL’ stands for High Level. ‘HR’ stands for High Rising. ‘ML’ stands for Mid Level. ‘LF’ stands for Low Falling. ‘LR’ stands for Low Rising. ‘LL’ stands for Low Level.

Table 10

Confusion Matrix of Tone Production in Child-directed Tones in Isolated Words

Target tones	**Judges' responses (%)**					
	T1	**T2**	**T3**	**T4**	**T5**	**T6**
T1 (HL)	**87**	0	10	0	0	1
T2 (HR)	0	**98**	0	0	2	0
T3 (ML)	15	0	**69**	2	2	13
T4 (LF)	0	0	1	**88**	2	8
T5 (LR)	0	15	2	1	**82**	1
T6 (LL)	4	0	29	5	2	**61**

Notes. Correct identifications are marked by bold numbers. Underlined numbers mark error patterns that constitute more than 10% of the trials. T1, T2, T3, T4, T5, and T6 stand for Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, and Tone 6, respectively. ‘HL’ stands for High Level. ‘HR’ stands for High Rising. ‘ML’ stands for Mid Level. ‘LF’ stands for Low Falling. ‘LR’ stands for Low Rising. ‘LL’ stands for Low Level.

Table 11

Confusion Matrix of Tone Production in Adult-directed Tones in Sentence Final Position

Target tones	**Judges' responses (%)**					
	T1	**T2**	**T3**	**T4**	**T5**	**T6**
T1 (HL)	**83**	1	14	0	0	2
T2 (HR)	0	**91**	0	1	8	0
T3 (ML)	1	0	**65**	2	2	30
T4 (LF)	0	1	0	**90**	3	5
T5 (LR)	0	4	2	1	**90**	3
T6 (LL)	0	0	10	10	4	**76**

Notes. Correct identifications are marked by bold numbers. Underlined numbers mark error patterns that constitute more than 10% of the trials. T1, T2, T3, T4, T5, and T6 stand for Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, and Tone 6, respectively. ‘HL’ stands for High Level. ‘HR’ stands for High Rising. ‘ML’ stands for Mid Level. ‘LF’ stands for Low Falling. ‘LR’ stands for Low Rising. ‘LL’ stands for Low Level.

Table 12

Confusion Matrix of Tone Production in Child-directed Tones in Sentence Final Position

Target tones	**Judges' responses (%)**					
	T1	**T2**	**T3**	**T4**	**T5**	**T6**
T1 (HL)	**90**	0	8	0	0	1
T2 (HR)	0	**97**	0	0	2	0
T3 (ML)	10	0	**68**	4	3	15
T4 (LF)	0	0	1	**91**	3	5
T5 (LR)	0	7	2	0	**90**	1
T6 (LL)	2	0	16	8	2	**71**

Notes. Correct identifications are marked by bold fonts. Underlined numbers mark error patterns that constitute more than 10% of the trials. Data in this table represent accuracy affected by both pitch lowering effect in sentence final position and pitch raising effect of the CD register and, were excluded for analyses. T1, T2, T3, T4, T5, and T6 stand for Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, and Tone 6, respectively. ‘HL’ stands for High Level. ‘HR’ stands for High Rising. ‘ML’ stands for Mid Level. ‘LF’ stands for Low Falling. ‘LR’ stands for Low Rising. ‘LL’ stands for Low Level.

Table 13

Differences in Tone Perceptual Accuracy in Different Speech Registers and Prosodic Contexts

Target Tone	Results	p	r
All Tones	CD-Iso < AD-Iso	0.001**	0.366
	AD-SentF < AD-Iso	0.004**	0.186
T1 (HL)	AD-SentF < AD-Iso	0.049*	0.150
T2 (HR)	AD-SentF < AD-Iso	0.011**	0.159
T3 (ML)	AD-SentF < AD-Iso	0.003**	0.272
T4 (LF)	n.s.	n.s.	n.s.
T5 (LR)	CD-Iso < AD-Iso	<0.001**	0.313
T6 (LL)	CD-Iso < AD-Iso	<0.001**	0.347

“AD-Iso” stands for “adult-directed tones produced in isolated words”.
“AD-SentF” stands for “adult-directed tones produced in sentence final position”.
“CD-Iso” stands for “child-directed production in isolated words”.
“<” stands for “with lower accuracy than”, “>” stands for “with higher accuracy than”
Dark shaded cells show effects of sentence final prosody on tone accuracy.
Light shaded cells show effects of child-directed register on tone accuracy.

T1, T2, T3, T4, T5, and T6 stand for Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, and Tone 6, respectively. ‘HL’ stands for High Level. ‘HR’ stands for High Rising. ‘ML’ stands for Mid Level. ‘LF’ stands for Low Falling. ‘LR’ stands for Low Rising. ‘LL’ stands for Low Level. ‘n.s.’ stands for non-significant.

Table 14

Acoustic differences of Adult-directed Tones Produced in Isolated Words and in Sentence Final Position – Tone 1 (High Level)

Acoustic Parameters[a]	**Tone 1 (High Level)**		
	Result	**p-value**	**r**
Initial Pitch Height 100% (St)	Iso > SentF**	<0.001	0.305
Final Pitch Height 100% (St)	Iso > SentF**	0.001	0.186
Mid Pitch Height 100% (St)	Iso > SentF**	<0.001	0.291
Min Pitch Height 50% (St)	Iso > SentF**	<0.001	0.254
Max Pitch Height 50% (St)	Iso > SentF**	<0.001	0.237
Mean Pitch Height 50% (St)	Iso > SentF**	<0.001	0.236
Tone Duration 100% (ms)	Iso = SentF	0.129	0.093
Slope 50% (St/ms)	SentF > Iso**[b]	0.005	0.273

Note. [a] Refer to Table 1 for the definitions of the acoustic parameters
[b] Slope of Tone 1 (High Level) in sentence final position had a larger value in the negative direction, indicating steeper falling contours.
"Iso" stands for tone production in isolation. "SentF" stands for tones produced in sentence final position.
'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 15

Acoustic differences of Adult-directed Tones Produced in Isolated Words and in Sentence Final Position – Tone 2 (High Rising)

Acoustic Parameters[a]	Tone 2 (High Rising)		
	Result	**p-value**	**r**
Initial Pitch Height 100% (St)	Iso > SentF**	<0.001	0.472
Final Pitch Height 100% (St)	Iso > SentF**	<0.001	0.255
Mid Pitch Height 100% (St)	Iso > SentF**	<0.001	0.267
Min Pitch Height 50% (St)	Iso > SentF**	<0.001	0.263
Max Pitch Height 50% (St)	Iso > SentF**	<0.001	0.287
Mean Pitch Height 50% (St)	Iso > SentF**	<0.001	0.494
Tone Duration 100% (ms)	Iso = SentF	0.121	0.100
Slope 50% (St/ms)	Iso = SentF	0.138	0.049

Note. [a] Refer to Table 1 for the definitions of the acoustic parameters
"Iso" stands for tone production in isolation. "SentF" stands for tones produced in sentence final position.
'St' stands for semi-tone. 'ms' stands for millisecond.
** = .05 significance level ** = .01 significance level*

Table 16

Acoustic differences of Adult-directed Tones Produced in Isolated Words and in Sentence Final Position – Tone 3 (Mid Level)

Acoustic Parameters[a]	Tone 3 (Mid Level)		
	Result	**p-value**	**r**
Initial Pitch Height 100% (St)	Iso > SentF**	<0.001	0.507
Final Pitch Height 100% (St)	Iso > SentF**	<0.001	0.52
Mid Pitch Height 100% (St)	Iso > SentF**	<0.001	0.302
Min Pitch Height 50% (St)	Iso > SentF**	<0.001	0.427
Max Pitch Height 50% (St)	Iso > SentF**	<0.001	0.427
Mean Pitch Height 50% (St)	Iso > SentF**	<0.001	0.495
Tone Duration 100% (ms)	Iso = SentF	0.166	0.048
Slope 50% (St/ms)	Iso = SentF	0.079	-0.099

Note. [a] Refer to Table 1 for the definitions of the acoustic parameters
"Iso" stands for tone production in isolation. "SentF" stands for tones produced in sentence final position.
** = .05 significance level ** = .01 significance level*

Table 17

Acoustic differences of Adult-directed Tones Produced in Isolated Words and in Sentence Final Position – Tone 4 (Low Falling)

Acoustic Parameters[a]	**Tone 4 (Low Falling)**		
	Result	**p-value**	**r**
Initial Pitch Height 100% (St)	Iso > SentF**	<0.001	0.563
Final Pitch Height 100% (St)	Iso > SentF**	<0.001	0.267
Mid Pitch Height 100% (St)	Iso = SentF	0.236	0.049
Min Pitch Height 50% (St)	Iso > SentF*	0.016	0.101
Max Pitch Height 50% (St)	Iso > SentF**	0.001	0.206
Mean Pitch Height 50% (St)	Iso > SentF**	0.004	0.157
Tone Duration 100% (ms)	Iso > SentF *	0.024	0.091
Slope 50% (St/ms)	Iso = SentF	0.147	-0.114

Note. [a] Refer to Table 1 for the definitions of the acoustic parameters
"Iso" stands for tone production in isolation. "SentF" stands for tones produced in sentence final position.
** = .05 significance level ** = .01 significance level*

Table 18

Acoustic differences of Adult-directed Tones Produced in Isolated Words and in Sentence Final Position – Tone 5 (Low Rising)

Acoustic Parameters[a]	**Tone 5 (Low Rising)**		
	Result	**p-value**	**r**
Initial Pitch Height 100% (St)	Iso > SentF**	<0.001	0.502
Final Pitch Height 100% (St)	Iso > SentF**	<0.001	0.396
Mid Pitch Height 100% (St)	Iso > SentF**	<0.001	0.514
Min Pitch Height 50% (St)	Iso > SentF**	<0.001	0.405
Max Pitch Height 50% (St)	Iso > SentF**	<0.001	0.558
Mean Pitch Height 50% (St)	Iso > SentF**	<0.001	0.562
Tone Duration 100% (ms)	Iso = SentF	0.195	0.045
Slope 50% (St/ms)	Iso = SentF	0.414	0.035

Note. [a] Refer to Table 1 for the definitions of the acoustic parameters
"Iso" stands for tone production in isolation. "SentF" stands for tones produced in sentence final position.
** = .05 significance level ** = .01 significance level*

Table 19

Acoustic differences of Adult-directed Tones Produced in Isolated Words and in Sentence Final Position – Tone 6 (Low Level)

Acoustic Parameters[a]	**Tone 6 (Low Level)**		
	Result	**p-value**	**r**
Initial Pitch Height 100% (St)	Iso > SentF**	<0.001	0.537
Final Pitch Height 100% (St)	Iso > SentF**	<0.001	0.571
Mid Pitch Height 100% (St)	Iso > SentF**	<0.001	0.333
Min Pitch Height 50% (St)	Iso > SentF**	<0.001	0.509
Max Pitch Height 50% (St)	Iso > SentF**	<0.001	0.395
Mean Pitch Height 50% (St)	Iso > SentF**	<0.001	0.508
Tone Duration 100% (ms)	Iso = SentF	0.582	0.046
Slope 50% (St/ms)	Iso = SentF	0.592	0.037

Note. [a] Refer to Table 1 for the definitions of the acoustic parameters
"Iso" stands for tone production in isolation. "SentF" stands for tones produced in sentence final position.
** = .05 significance level ** = .01 significance level*

Appendix A

Information of Participants

Mother Information			Child Information			Mother Information			Child Information		
Code	Age (Yrs)	Education Level	Age Group	Age	Gender	Code	Age (Yrs)	Education Level	Age Group	Age	Gender
1-year-old Child						3-year-old Child					
M01	31	High	1	1;2	F	M04[a]	36	Mid	3	3;5	M
M05[a]	36	Mid	1	1;8	M	M06[b]	36	High	3	3;11	F
M07[b]	36	High	1	1;2	F	M16[h]	37	High	3	3;6	F
M08[b]	36	High	1	1;2	F	M19	30	High	3	3;11	M
M12	30	High	1	1;8	M	M22	39	High	3	3;1	M
M13	28	Mid	1	1;7	F	M24	35	High	3	3;6	F
M14	36	High	1	1;2	F	M30[c, o]	36	High	3	3;5	M
M15	26	Mid	1	1;5	M	M32	NR	High	3	3;11	F
M20	36	High	1	1;1	M	M33	33	High	3	3;1	M
M29[o, p]	30	Low	1	1;7	M	M34[i]	NR	High	3	3;5	F
M31[c, o]	36	High	1	1;0	F	M37[j]	38	High	3	3;6	F
M41[d]	37	High	1	1;8	M	M43	40	High	3	3;3	M
M57[e]	NR	NR	1	1;3	F	M47	35	High	3	3;5	M
M60[f]	40	High	1	1;8	F	M50	NR	High	3	3;2	M
M61[o]	37	High	1	1;7	M	M59[f]	40	High	3	3;6	F
2-year-old Child						5-year-old Child					
M02	32	Mid	2	2;2	F	M18	41	NR	5	5;10	M
M03	36	High	2	2;0	F	M23	40	High	5	5;2	M
M09[g]	39	High	2	2;3	M	M27[l]	40	High	5	5;10	M
M09[g]	39	High	2	2;3	F	M28[l]	40	High	5	5;10	M
M11	24	Mid	2	2;7	M	M35[i]	NR	High	5	5;10	M
M17[h]	37	High	2	2;0	F	M36[i]	38	Mid	5	5;7	F
M21	35	High	2	2;3	M	M38[o]	34	High	5	5;2	F
M25	33	High	2	2;8	M	M39	40	High	5	5;1	M
M26[o]	34	High	2	2;1	M	M40[d]	37	NR	5	5;0	M
M44	33	High	2	2;4	M	M42[k]	40	NR	5	5;0	M
M48	32	High	2	2;5	M	M45[m]	36	High	5	5;9	F
M51	NR	High	2	2;0	M	M49	NR	High	5	5;11	F
M53	39	High	2	2;3	F	M54	41	High	5	5;11	F
M56	34	High	2	2;0	F	M55	36	High	5	5;4	F
M62[o]	33	High	2	2;6	M	M58[e]	NR	NR	5	5;0	F

Notes: Isolated Word Condition of M55 are excluded due to corrupted sound files.
Education Level: Low = Middle School or below; Mid = High School; High = Post-High School; NA = Not Applicable; NR = No response.

[a] M04 = M05
[b] M06 = M07 = M08
[c] M30 = M31
[d] M40 = M41
[e] M57 = M58
[f] M59 = M60
[g] M09 = M10
[h] M16 = M17
[i] M34 = M35
[j] M36 = M37
[k] M42 = M43
[l] M27 = M28
[m] M45 = M46
[n] M27 = M28
[o] Know another dialects
[p] Not born in Hong Kong

Appendix B

Stimuli

Word Familiarity	Tones	Chinese	English	IPA	Jyutping	Word Frequency[a]
High Familiar Words[a]	T1 (HL)	杯	Cup	/pui1/	Bui1	96%
		書	Book	/sy1/	Syu1	94%
		燈	Lamp	/tɐŋ1/	Dang1	94%
	T2 (HR)	狗	Dog	/kɐu2/	Gau2	96%
		糖	Candy	/t^{h}ɔŋ2/	Tong2	97%
		帽	Hat	/mou2/	Mou2	91%
	T3 (ML)	菜	Vegetables	/ts^{h}ɔi3/	Coi3	91%
		喊	Cry	/ham3/	Haam3	93%
		褲	Pants	/fu3/	Fu3	96%
	T4 (LF)	鞋	Shoes	/hai4/	Haai4	99%
		頭	Head	/t^{h}ɐu4/	Tau4	96%
		門	Door	/mun4/	Mun4	93%
	T5 (LR)	被	Duvet	/p^{h}ei5/	Pei5	93%
		雨	Rain	/jy5/	Jyu5	94%
		眼	Eyes	/ŋan5/	Ngaan5	97%
	T6 (LL)	飯	Rice	/fan6/	Faan6	93%
		麵	Noodles	/min6/	Min6	93%
		鼻	Nose	/pei6/	Bei6	96%
Low Familiar Words[b]	T1 (HL)	溪	Creek	/k^{h}ɐi1/	Kai1	N.A.
		鍋	Pot	/wɔ1/	Wo1	N.A.
	T2 (HR)	島	Island	/tou2/	Dou2	N.A.
		井	Well	/tsɛŋ2/	Zeng2	N.A.
	T3 (ML)	炭	Carbon	/t^{h}an3/	Taan3	N.A.
		炮	Cannon	/p^{h}au3/	Paau3	N.A.
	T4 (LF)	柴	Firewood	/ts^{h}ai4/	Caai4	N.A.
		矛	Spear	/mau4/	Maau4	N.A.
	T5 (LR)	盾	Shield	/t^{h}ɵn5/	Teon5	N.A.
		艇	Boat	/t^{h}ɛŋ5/	Teng5	N.A.
	T6 (LL)	雁	Goose	/ŋan6/	Ngaan6	N.A.
		艦	Ship	/lam6/	Laam6	N.A.
High Familiar Words[a]	T7 (HS)	筆	Pencil	/pɐt7/	Bat7	94%
		一	One	/jɐt7/	Jat7	97%
	T8 (MS)	錫	Kiss	/sɛk8/	Sek8	90%
		腳	Feet	/kœk8/	Goek8	96%
	T9 (LS)	熱	Hot	/jit9/	Jit9	94%
		襪	Socks	/mɐt9	Mat9	90%
Low Familiar Words[b]	T7 (HS)	穀	Grain	/kʊk7/	Guk7	N.A.
		禿	Bald	/t^{h}ʊk7/	Tuk7	N.A.
	T8 (MS)	殼	Shell	/hɔk8/	Hok8	N.A.
		塔	Tower	/t^{h}ap8/	Taap8	N.A.
	T9 (LS)	鑿	Chisel	/tsɔk9/	Zok9	N.A.
		墨	Ink	/mɐk9/	Mak9	N.A.

Notes. [a] Words produced by more than 90% of 30 month old children growing up in Hong Kong (Tardif et al., 2009).
[b]Words not found in Cantonese Communicative Development Inventory (CCDI) (Tardif et al., 2009)
T1, T2, T3, T4, T5, T6, T7, T8, and T9 stand for Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, Tone 6, Tone 7, Tone 8, and Tone 9, respectively. 'HL' stands for High Level. 'HR' stands for High Rising. 'ML' stands for Mid Level. 'LF' stands for Low Falling. 'LR' stands for Low Rising. 'LL' stands for Low Level. 'HS' stands for High Stop. 'MS' stands for Mid Stop. 'LS' stands for Low Stop. Entering tones (i.e., T7, T8 and T9) were excluded from analyses. 'IPA' stands for the International Phonetic Alphabet. 'N.A.' stands for not available.