JSLHR

Research Article

Assessing the Link Between Perception and Production in Cantonese Tone Acquisition

Peggy Pik Ki Mok,[a] Holly Sze Ho Fung,[a] and Vivian Guo Li[a]

Purpose: Previous studies showed early production precedes late perception in Cantonese tone acquisition, contrary to the general principle that perception precedes production in child language. How tone production and perception are linked in 1st language acquisition remains largely unknown. Our study revisited the acquisition of tone in Cantonese-speaking children, exploring the possible link between production and perception in 1st language acquisition.
Method: One hundred eleven Cantonese-speaking children aged between 2;0 and 6;0 (years;months) and 10 adolescent reference speakers participated in tone production and perception experiments. Production materials with 30 monosyllabic words were transcribed in filtered and unfiltered conditions by 2 native judges. Perception accuracy was based on a 2-alternative forced-choice task with pictures covering all possible tone pair contrasts.
Results: Children's accuracy of production and perception of all the 6 Cantonese tones was still not adultlike by age 6;0. Both production and perception accuracies matured with age. A weak positive link was found between the 2 accuracies. Mother's native language contributed to children's production accuracy.
Conclusions: Our findings show that production and perception abilities are associated in tone acquisition. Further study is needed to explore factors affecting production accuracy in children.
Supplemental Material: https://doi.org/10.23641/asha.7960826

The link between perception and production is a perennial question in speech acquisition research. Although various relationships between the two aspects of speech communication are possible, most second language speech acquisition theories posit that accurate perception precedes accurate production (Best, 1995; Escudero, 2009; Flege, 1995). In first language acquisition, numerous studies have demonstrated that infants can perceive many phonetic contrasts, segmental (e.g., Werker & Tees, 1983, 1984) and suprasegmental (Singh & Fu, 2016) alike, well before they can produce them accurately. Phenomena such as "fis" (Berko & Brown, 1960) and "tum" (Butler, 1980) are good illustrations of the general principle that accurate perception usually precedes accurate production in learning one's native language (L1). For example, English-speaking children could distinguish "fis" and "fish" in perception, but they could only say "fis" in production. Nevertheless, the existing literature on the acquisition of Cantonese tone suggests an opposite pattern. Cantonese monolingual children were shown to produce all lexical tones accurately by age 2;6 (years;months; So & Dodd, 1995; To, Cheung, & McLeod, 2013b), whereas perception studies revealed that they could not distinguish all tones correctly until age 10 years (Ciocca & Lui, 2003). Such large discrepancy calls for a systematic reevaluation of the acquisition of Cantonese tones by monolingual children. Our study addressed this issue with both production and perception data from the same group of 111 children spanning ages 2;1–6;0. Our data provide strong empirical evidence to assess the link between production and perception in first language acquisition of lexical tone.

The Acquisition of Cantonese Tones

The use of lexical tone (T) is a prominent phonological characteristic of Cantonese. Each syllable (roughly equivalent to a morpheme) carries a tone. Cantonese has a complex tone system. There are six lexical tones based on pitch contrast alone: T1 [55] high level, T2 [25] high rising, T3 [33] midlevel, T4 [21] low falling, T5 [23] low rising, and T6 [22] low level (Bauer & Benedict, 1997; Fok-Chan, 1974). The numbers in [ ] represent the relative starting and ending pitch

[a]The Chinese University of Hong Kong, Shatin, Hong Kong
Correspondence to Peggy Mok: peggymok@cuhk.edu.hk

Editor-in-Chief: Julie Liss
Editor: Bharath Chandrasekaran
Received November 21, 2017
Revision received February 6, 2018
Accepted November 14, 2018
https://doi.org/10.1044/2018_JSLHR-S-17-0430

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

levels of each tone, with 5 being the highest and 1 being the lowest pitch level of a speaker's normal pitch range (Chao, 1930, 1947). T1 is well separated from the other five tones by being at the highest pitch level, whereas the other five tones occupy the low to midpitch range. They are differentiated mostly in the second half of the syllable. The six lexical tones can be divided into two registers: T1, T2, and T3 in high register and T4, T5, and T6 in low register (Yip, 2002). The six tones appear in open syllables or syllables with nasal codas [-m, -n, -ŋ]. There are three allotones, which are traditionally called the *entering tones* in Chinese phonology. They only appear in syllables with unreleased stop codas [-p, -t, -k]: T7 [5] high stopped, T8 [3] midstopped, and T9 [2] low stopped. They are much shorter in duration and are considered allotones of the three corresponding unstopped level tones T1, T3, and T6, respectively (Bauer & Benedict, 1997; Chao, 1947).

The complex Cantonese tone system is undergoing changes in recent years in that some similar tone pairs began to merge. Mok, Zuo, and Wong (2013) reported some young Cantonese speakers in Hong Kong may not clearly distinguish the two rising tones (T2 vs. T5), the two level tones (T3 vs. T6), and the low-falling and low-level tones (T4 vs. T6) in their production and perception. The merging is in an incipient stage, as these speakers still had six tone categories.

One possible reason for the tone-merging phenomenon may be due to the dynamic demographic composition and language contact in Hong Kong due to cross-border marriage over the past few decades. Children born to cross-border marriage may be subjected to imperfect tone acquisition because their parents may not speak standard Hong Kong Cantonese. Both the quantity and quality of tone input to these children are likely to affect their acquisition of the complex Cantonese tones, particularly the difficult tone pairs identified above. Nevertheless, so far, no study has investigated how this factor may influence Cantonese tone acquisition.

Despite the complexity of the Cantonese tone system, several earlier studies have collectively shown that Cantonese monolingual children could produce all the six tones accurately very early, by age 2;0 (So & Dodd, 1995; Tse, 1978) or 2;6 (To et al., 2013b). The longitudinal conversational data of one child aged between 1;3 and 2;6 in Tse (1978) and four children aged 1;2–2;0 in So and Dodd (1995) illustrated that they have acquired all Cantonese tones by age 2;0. Tse divided the acquisition of tone production in three stages: In Stage 1 (1;2–1;4), T1 [55] and T4 [21] were acquired; in Stage 2 (1;5–1;8), T3 [33], T2 [25], and the three allotones were acquired; in Stage 3 (1,9), T5 [23] and T6 [22] were acquired. The duration of acquiring the first to the last tone spanned only a period of 8 months. The four children in So and Dodd had very similar pattern of order and rate of acquisition. They reported that the children acquired T1 [55] and T3 [33] first, then T2 [25] and the three allotones. Two children acquired T6 [22] before T4 [21] and T5 [23], whereas one child showed the opposite pattern. Another child acquired these three tones simultaneously. Their data show that all four children had acquired the Cantonese tones by age 2;0, although the specific order of acquisition may differ.

Cross-sectional data of many more children with elicited production present a similar picture. So and Dodd (1995) tested 268 Cantonese-speaking children aged 2;0–6;0. Using two examples of each tone, they found that only two children made tone errors, one 4-year-old made two errors and a 5-year-old made three errors. They concluded that by age 2;0, most children had mastered the tonal contrasts in Cantonese. The large-scale study by To et al. (2013b), which tested 1,726 children aged 2;4–12;4, also echoes their findings. Using two words for each tone including the "entering tones," To et al. found that, for the youngest age group (2;4–2;9, 104 children), the averaged production accuracy was already at ceiling (M = 98.02%, SD = 5.19%). As there were no data before age 2;4 in their study, they concluded that tone acquisition was complete by age 2;6.

Both So and Dodd (1995) and To et al. (2013b) found that Cantonese-speaking children finished acquiring tones well before consonants and vowels. So and Dodd reported that the children in their study completed the acquisition of syllable-initial consonants by age 5;0 and syllable-final consonants by age 4;6. All the vowels were used contrastively by 90% of children in all age groups, including the youngest one (2;0–2;5). To et al. reported that all 19 initial consonants were acquired by age 6;0. Most final consonants were acquired by age 5;0, although some final consonants were not acquired even by the oldest age (11;7). Vowels were acquired by age 5;0, and diphthongs were acquired by age 4;0.

The above studies showing early acquisition of all the six tones were based on transcription data of children's production. Although there were perception studies showing that even infants could distinguish simple tone contrasts (e.g., Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Singh & Fu, 2016), unfortunately, there are no perception data of Cantonese tone contrasting all tone pairs by children under age 3;0 in the literature. The ability of infants learning tone languages to maintain sensitivity to acoustic differences between simple stimuli (e.g., contrasting just two tones) in their first year of life is not the same as the ability to distinguish all possible tonal contrasts in authentic situations related to meaning later in life. Thus, we still do not know if children so young (around ages 2;0–2;6) could distinguish all the tones related to meaning correctly or not.

Nonetheless, perception studies with children aged 3;0 above using experimental methods demonstrate that the acquisition of the Cantonese tone system takes much longer to complete. Using the syllable /ji/ with different lexical tones, both Ching (1984) and Ciocca and Lui (2003) found that the tones were only acquired by the age of 10 years. Nevertheless, some words with the syllable /ji/ were not familiar to young children. Using stimuli that were more familiar to children, Lee, Chan, Lam, van Hasselt, and Tong (2015) found that the perceptual performance of 6-year-old children was mostly similar to that of adults, but they were still not adultlike for some difficult contrasts. Small tone perceptual improvement was found between ages 6;0 and 10;0. Children at age 10;0 were adultlike in all contrasts,

concurring the findings of Ciocca and Lui (2003). These perception studies all found that the merging tone pairs, T2/T5, T3/T6, and T4/T6 (Mok et al., 2013), were difficult for children to distinguish.

The large discrepancy in the age of acquisition between production studies using transcription data and perception studies using experiments is very noteworthy, particularly in terms of the general principle that comprehension precedes production in child language development mentioned above. It is really puzzling to find that production studies show very early mastery of Cantonese tones whereas perception studies reveal a much slower process. Moreover, these studies were conducted by independent researchers, but they all point to the same conclusion: Early production precedes late perception. A logical question to ask is how could the Cantonese children produce the complex tones correctly if they could not distinguish them accurately in the first place?

Some likely reasons for the large discrepancy in the two types of studies are methodological in nature. Perception studies were based on all possible tone pairs, whereas the production studies used only very few words for picture naming. The accuracy in production may be inflated as a result. In addition, the criteria for tone acquisition were not consistent in previous studies. Tse (1978) did not mention explicitly the criteria used in defining the acquisition of tone. So and Dodd (1995) defined the acquisition of a particular tonal category when it was "used contrastively on at least 50% of opportunities or correctly on 90% of opportunities," but no further explanation was given. Perception studies usually compared children's performance with that of adults (Ciocca & Lui, 2003; Lee et al., 2015).

Another methodological issue is that, as pointed out rightly by Wong, Schwartz, and Jenkins (2005), transcription data can be easily biased by lexical, semantic, syntactic, and contextual cues, which may create tone expectations that can influence transcription accuracy by native judges. In order to minimize these influences, Wong and colleagues (Wong, 2013; Wong et al., 2005) used a novel method to do transcription. They low-pass filtered children's Mandarin production at 500 Hz to remove segmental information so the transcribers could only rely on the pitch information to judge the accuracy of children's Mandarin tones. They found that, contrary to previous findings using transcription data, which also showed an early acquisition of Mandarin tones by the age of 2 years (Li & Thompson, 1977; Zhu, 2002; Zhu & Dodd, 2000), the acquisition of Mandarin tones is much more protracted. Children as old as 5 years old still did not produce any of the four monosyllabic tones with adultlike accuracy. In terms of perception, Wong et al. (2005) also showed that Mandarin children at age 3;0 still had difficulty in perceiving the dipping tone T3. Moreover, even children's Mandarin tones that were correctly categorized by adult transcribers were still phonetically different from those produced by adult speakers (Wong, 2012). Wong's work using acoustic methods show that the acquisition of Mandarin tone is much more complicated than what has been suggested by studies using simple transcription data. Similar discrepancy between transcription and acoustic findings in the acquisition of segments can be found in other studies as well (see review in Munson, Edwards, & Beckman, 2012).

There are strong parallels between the aforementioned studies on Cantonese and Mandarin tone acquisition. Studies using simple transcription data of natural productions showed very early acquisition of all tones (by around age 2 years), whereas perception studies and studies using transcription of filtered speech showed a much more protracted course of acquisition. Therefore, using more rigorous data collection methods, it is likely to find that Cantonese children have not fully acquired the tone system by age 2;6, contrary to the conclusions of previous studies. Indeed, this is the case. In a recent study, adopting the same method they used for Mandarin tones (transcription with filtered materials and acoustic analysis), Wong, Fu, and Cheung (2017) demonstrated that 3-year-old Cantonese children still had not fully acquired Cantonese tones in both production and perception. Their data were limited to only 20 children in one age group (3;1–3;11), so many questions about Cantonese tone acquisition still remained unanswered. Their even more recent study (Wong & Leung, 2018) used the same method to investigate Cantonese tone production and perception by 4- to 6-year-old children and found a similar conclusion that Cantonese tones were not acquired early. However, Wong only did transcription using filtered materials, so it is difficult to compare her data with previous studies showing early acquisition. Our study reevaluated Cantonese tone acquisition, including children of a wider age range (2;1–6:0) using both filtered and unfiltered materials, hoping to get a more comprehensive understanding of the acquisition process.

In addition, previous studies investigated tone production (So & Dodd, 1995; To et al., 2013b) and tone perception (Ciocca & Lui, 2003; Lee et al., 2015; Lee, Chiu, & Hasselt, 2002) separately, making it impossible to compare children's production and perception performance because data were collected from different children. The critical relationship between production and perception in phonological development is still unsettled (Clark & Hecht, 1983; Polka, Rvachew, & Mattock, 2008; Vihman, 2014, 2017). Previous perception studies usually involved young infants focusing on their abilities to discriminate some simple sound contrasts, whereas production studies necessarily involved older children with better articulatory control. Recently, there are some studies investigating the link between perception and production in children's acquisition of consonants, notably English /r/ (e.g., Idemaru & Holt, 2013; McAllister Byun & Tiede, 2017), but mixed results were presented. Very little data have been reported on the link between children's tone production and perception. Wong found no correlation between tone production accuracy and tone perception accuracy for both Mandarin and Cantonese, but there were only 13 Mandarin children (Wong et al., 2005) and 20 Cantonese children (Wong et al., 2017) in one age band (3 years old) in their studies. Thus, the null findings could be due to the lack of statistical power. When they included more children (n = 48) with a wider age range (4–6 years old), they found a weak

correlation (R^2 = .194) between Cantonese tone perception and production with all tones combined (Wong & Leung, 2018). Clearly, more data are needed to address this issue in children's first language acquisition. The production and perception data from the same 111 children ranging from ages 2;1 to 6;0 in our study allow us to confirm if there is any positive correlation between the two and if perception accuracy can predict production accuracy, providing valuable information on phonological development of tone and also contributing to the link between perception and production in first language acquisition research in general.

This Study

Our study revisited the acquisition of Cantonese tones by monolingual children aged 2;1–6;0. We did not include children under age 2;0 because it would be very difficult to conduct our perception experiment (picture identification) with them and that they may not know all the words in the production materials. We collected both production and perception data cross-sectionally from the same group of children who were divided into eight narrow age bands (every 6 months) for more refined assessment of their performance with age. Previous perception studies on Cantonese tones only worked with children aged 3;0 (Lee et al., 2015) or 4;0 (Ciocca & Lui, 2003) above. We filled in the gap by providing new tone perception data between ages 2;1 and 3;0. In addition to having new perception data, the production data between ages 2;1 and 3;0 is crucial because they allow us to compare their production performance with those in previous studies showing early acquisition of Cantonese tones by age 2;0 or 2;6.

Wong et al. (Wong et al., 2017, 2005; Wong & Leung, 2018) applied a low-pass filter at 500 Hz to their production data for more stringent transcription. Nevertheless, the judges in Wong et al. only transcribed the filtered materials, so it was unclear how production accuracy would vary depending on judgment criteria, which made it hard to compare Wong et al.'s data with previous studies using transcription of natural data by one judge. In addition, listening to filtered materials is not the same as speech perception (Mok & Zuo, 2012). Therefore, we followed Wong et al. in applying a low-pass filter to the production data for transcription, but our native judges transcribed both the natural and filtered production data. This allows us to compare our data with previous studies using simple transcription data and also to assess how the removal of segmental contexts by filtering affects the judgment of production accuracy of the children.

Only a small portion of production data was cross-checked in previous transcription-based studies (10% in So & Dodd, 1995; 7.5% in To et al., 2013b). Although they got high interrater reliability, there would still be a portion without agreement if the whole set of data was considered. Thus, in order to enhance the reliability of the data, two native Cantonese-speaking judges listened to all production materials in our study, both natural and filtered. This allows us to assess production accuracy under different criteria and also to evaluate the validity of these criteria. Our rigorously collected data help us to examine the link between perception and production in tone acquisition comprehensively.

Method

Participants

One hundred eleven Hong Kong Cantonese-speaking children aged 2;1–6;0 with no reported speech, hearing, or learning impairment participated in the experiment. They were children in five local kindergartens cum nurseries. They all spoke Cantonese as their first language at home, but the language background questionnaires revealed that 24% of the children's fathers and 36% of the mothers were not native speakers of Hong Kong Cantonese. Native speakers of other varieties of Cantonese (e.g., Guangzhou) were not counted as native Hong Kong Cantonese speakers. The breakdown of participants according to age, sex, and parents' first languages can be found in Table 1. Ten native Cantonese-speaking adolescents aged 15;9–16;7 with no perceived tone merge in their production[1] were recruited as reference speakers for comparison. Previous studies demonstrated that adultlike perceptual patterns were found at age 10 years (Ciocca & Lui, 2003; Lee et al., 2015) and that because our adolescent speakers were screened for their tone production accuracy, they could safely serve as reference speakers in our project. Adolescent speakers were used as reference in a previous study on Cantonese tone as well (Khouw & Ciocca, 2007).

Materials

The production experiment was a picture-naming task. Previous studies used only two words for each tone. We used five words for each tone to increase the validity of our results. Our materials included 30 monosyllabic Cantonese words (6 tones × 5 words), each represented by a colored picture. Nineteen of the words were adopted from the Hong Kong Cantonese Articulation Test (HKCAT; Cheung, Ng, & To, 2006), and 11 words were supplemented by ourselves because there were insufficient words for certain tones (especially T5 and T6) in the original test. All the words were familiar concepts to children, for example, "花 *flower,* 跑 *run,* 褲 *trousers,* 門 *door*" (HKCAT) and "眼 *eye,* 飯 *cooked rice,* 喊 *cry,* 麵 *noodle*" (supplemented).

The perception materials consisted of 60 questions (15 tone pairs × 4 questions), each complemented with two black-and-white illustrations representing two monosyllabic

[1]To ensure that they could clearly distinguish the six tones in their production, they were recorded producing the syllables /fɐn/, /jɐn/, /ji/, and /si/ in six tones (all are real words in Cantonese). Their production was independently listened to by two phonetically trained native speakers of Cantonese. The 10 reference speakers were screened out from 22 speakers who attempted the test. The reference speakers were all born in Hong Kong and spoke Cantonese as their native language. Nevertheless, we did not know their parents' language background because these speakers were participants of another study.

Table 1. Participants' background information and interrater agreement on unfiltered and filtered materials for different age groups.

Age group (years;months)	Sex	*n*	Father's L1	*n*	Mother's L1	*n*	Total	Interrater agreement: Unfiltered condition	Interrater agreement: Filtered condition
2;1–2;6	F	5	HK Cantonese	9	HK Cantonese	7	13	0.755	0.673
	M	8	Others	4	Others	6			
2;7–3;0	F	7	HK Cantonese	15	HK Cantonese	12	18	0.787	0.730
	M	11	Others	3	Others	6			
3;1–3;6	F	7	HK Cantonese	11	HK Cantonese	10	15	0.781	0.691
	M	8	Others	4	Others	5			
3;7–4;0	F	9	HK Cantonese	10	HK Cantonese	5	12	0.774	0.640
	M	3	Others	2	Others	7			
4;1–4;6	F	7	HK Cantonese	13	HK Cantonese	13	17	0.826	0.716
	M	10	Others	4	Others	4			
4;7–5;0	F	6	HK Cantonese	11	HK Cantonese	10	15	0.871	0.833
	M	9	Others	4	Others	5			
5;1–5;6	F	4	HK Cantonese	8	HK Cantonese	9	11	0.917	0.715
	M	7	Others	3	Others	2			
5;7–6;0	F	7	HK Cantonese	7	HK Cantonese	5	10	0.848	0.710
	M	3	Others	3	Others	5			
Reference	F	5	HK Cantonese	NA	HK Cantonese	NA	10	0.945	0.832
	M	5	Others	NA	Others	NA			

Note. L1 = native language; HK = Hong Kong; NA = not applicable.

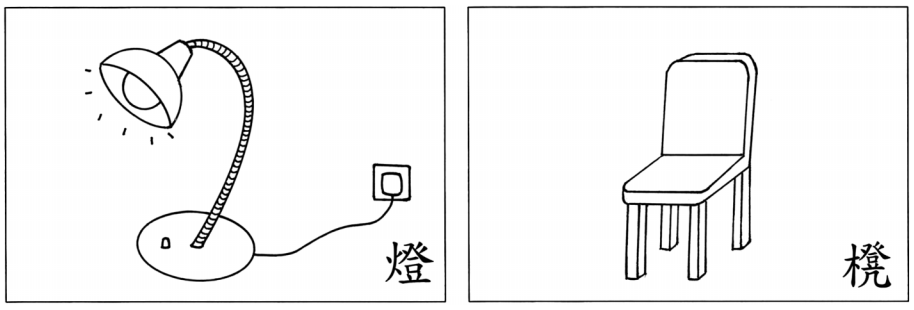
Cantonese words contrasting only in their tones. The materials were adopted from the Cantonese Tone Identification Test (Lee, 2012), but without the two distractor items in each question. Figure 1 gives an example of the perception materials. Other examples of the Cantonese Tone Identification Test word pairs include "糖 *candy* (T2) vs. 湯 *soup* (T1), 肥 *fat* (T4) vs. 飛 *fly* (T1), 錫 *kiss* (T3) vs. 石 *pebble* (T6), 滿 *full* (T5) vs. 門 *door* (T4)." The distractors were excluded because it was found in a pilot test that children under age 4 years had difficulty concentrating throughout the whole perception task if four illustrations instead of two were presented for each question.

Procedure

Parental consent was sought before the children were allowed to participate in the experiments. The parents were also asked to fill in a questionnaire about their own language background and the children's use of language in various contexts. The experiment was conducted individually in quiet rooms in the participating kindergartens cum nurseries and was administered by phonetically trained native speakers of Cantonese who were able to distinguish the six lexical tones clearly in both their production and perception. Older children could usually finish both the production and perception tasks in one session (~30 to 45 min with breaks), whereas younger children often did the production and perception tasks in separate sessions or even on different days to avoid fatigue and fussiness—and sometimes due to logistic arrangement of the kindergartens. Because of this reason, data of eight children in the youngest age group (2;1–2;6) and one child in the 4;7–5;0 age group were not included in our study because they did the production and perception tasks in different age bands. Eleven children in the 4;7–5;0 age group were excluded due to environmental noise during the experiments. These children were not included in Table 1.

In the production task, participants were recorded naming each colored picture twice. We followed the question instructions in HKCAT to elicit production (questions

Figure 1. Example of perception materials based on the Cantonese Tone Identification Test. The pictures represent the contrast between T1 (left, "lamp" /teng/ 55) and T3 (right, "chair" /teng/ 33).



such as *ni1 go3 hai6 me1 aa3?* "What is this?"). Similar questions were used for supplemented material. The children would produce the target words in isolation. They would be prompted to say the target word again for the second recording. In case of failure to produce a target word despite hints given, they would be asked to repeat after the experimenters. This happened occasionally for the two youngest age groups (2;1–3;0). As for the perception task, participants were asked to choose the correct illustration from a pair in a booklet after listening to the experimenters' production of the target word of a question in the carrier phrase *bin1 jat1 go3 hai6 __ aa3?* ("Which one is __?"). The children would point to the correct picture. If they did not respond, the experimenters would further ask them *hai6 ni1 go3 ding6 hai6 go2 go3 aa3* ("Is it this one or that one?"). Each of the 60 questions was tested on twice: Upon completion of all the questions, they were asked a second time—this time with the other word in each minimal pair being the target. Each correct response scored one mark and wrong response zero. The reason for testing on both items in a minimal pair was to avoid selection bias, under which children were inclined to choose from a pair the item they knew rather than the one they did not, regardless of which one the target was. If each pair was tested only once, those questions with one item more frequently used or acquired earlier than the other might have inflated correction rates among younger children if the target words happened to be the more familiar items, and vice versa.

Data Analysis

Each child's production of the target words was transcribed independently by two phonetically trained native Cantonese speakers with and without low-pass filtered at 500 Hz. Wong (Wong, 2013; Wong et al., 2017; Wong & Leung, 2018) used five and even 10 (Wong et al., 2005) judges in her studies and only considered tokens identified by all judges correctly to be correct tokens. We found that to be overly stringent, because this method only includes tokens that were very clearly produced. We simplified her method to two judges. This also avoided the need to do complicated statistical analyses on the interrater reliability by multiple judges.

For the filtered condition, transcription was done blindly as segmental information was removed by filtering, that is, target words of the filtered materials were unavailable to the judges and they only listened to the pitch contours. A third research assistant played the filtered sound files to the two judges who were seated back-to-back and marked down their judgments. It was impossible to transcribe the unfiltered materials blindly because the judges knew which words the children were producing just by listening to them. Before transcribing each child's materials, a short unfiltered recording of the exchange between the child and the experimenters was listened to for familiarization of the child's lexical pitch range. Tokens that could not be perceived as any of the six tones were marked "uncategorized." Transcription of the reference speakers' materials followed the same procedures, except with a lower filter threshold at 400 Hz because a 500-Hz threshold would render some of the filtered speech intelligible. The 500-Hz and 400-Hz filter thresholds were also used by Wong (Wong, 2013; Wong et al., 2017). Subsequently, production accuracy was calculated in four conditions based on (a) whether the speaker's materials were filtered (unfiltered/filtered) and (b) whether the intended tones were identified by one or both transcribers (one judge, more lenient/two judges, more stringent).

For each perception question, the scores from both attempts were averaged. Perception accuracy for each tone was then calculated as the mean score of all questions contrasting it. Instead of arbitrarily setting up an accuracy threshold for a tone to be considered acquired, we followed the practice of previous perception studies by comparing children's performance with that of the reference speakers.

Results

The interrater reliability between the two native judges for unfiltered and filtered materials can be found in Table 1. As expected, the agreement for filtered materials was lower than that for unfiltered materials across age groups. The agreement of adolescent reference data in the unfiltered condition was very comparable to other studies (0.945). In both conditions, the agreement generally increased with age (although reversions were also found), which indicated that the tone productions of younger children were more varied and harder to judge.

The overall mean production accuracy under four conditions is given in Table 2 (left panel). One-sample t tests revealed that the accuracy of all age groups in all judgment conditions were significantly above chance level (1/6 = 16.7%), $p < .001$ (full statistical details can be found in Supplemental Material S1). The effects of filtering and having two judges are obvious. The accuracy in the filtered conditions is markedly lower than that in the unfiltered conditions across age groups. Having two judges agreed on the same production also lowers the accuracy considerably. Interestingly, the differences of these conditions are larger for the child participants than for the adolescent reference, which indicate that segmental contexts were more influential in tone judgments for children and that there were more variations in children's tone production.

If we only consider the transcription of one judge in the unfiltered condition (like previous studies), we probably would conclude that Cantonese children have acquired the tones early by age 3;0 (mean production accuracy of 90% in the 2;7–3;0 age group). Also using unfiltered materials, the children's accuracy agreed by two judges still did not meet that of the reference speakers by age 6;0 (although the 5;1–5;6 age group was closer). It is interesting to note that, in the unfiltered-two-judges and the filtered-one-judge conditions, there was a general increase of production accuracy with age. However, in the unfiltered-one-judge (most lenient) and filtered-two-judges (most stringent) conditions, production accuracy levels off: reaching ceiling early in the

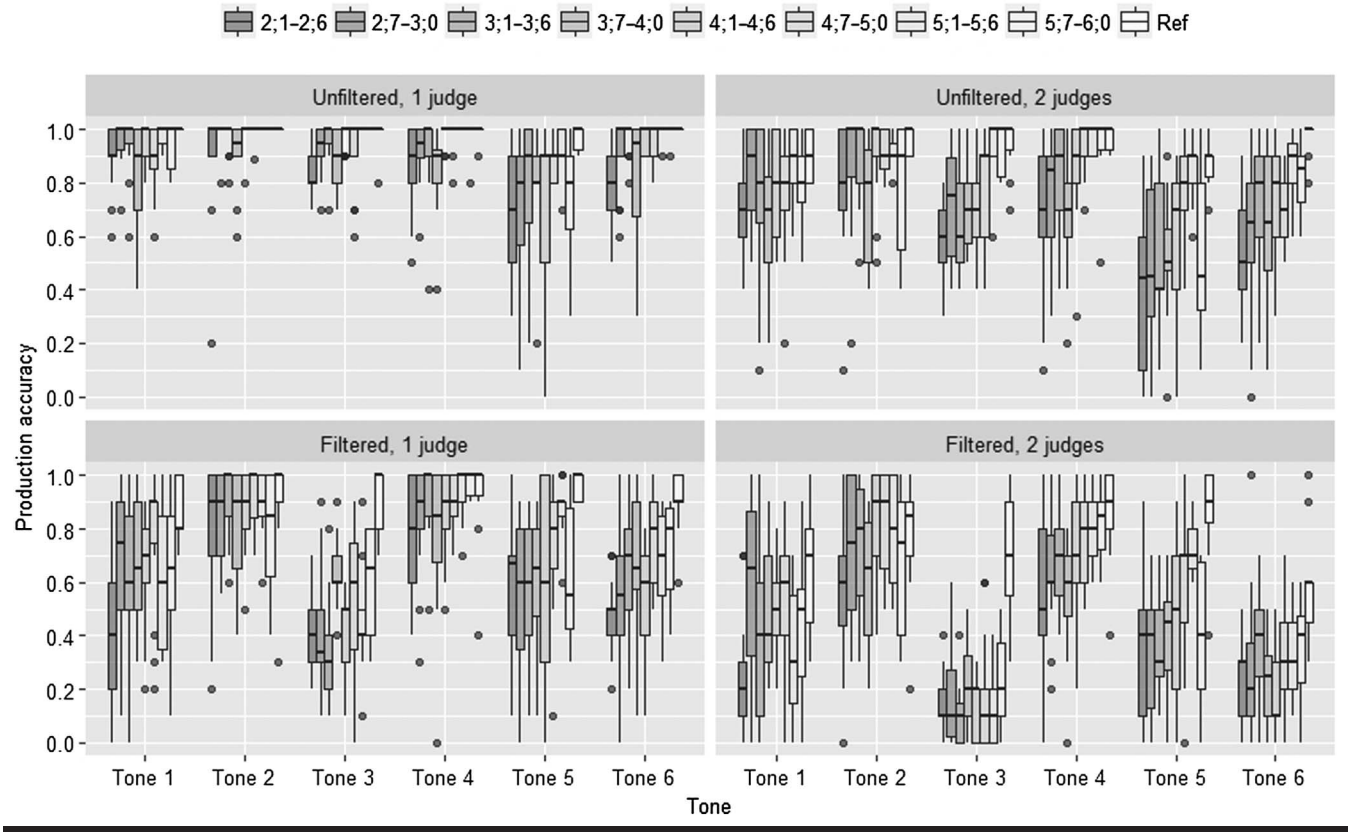
Table 2. Overall mean production accuracy and standard deviation in four judgment conditions and mean perception accuracy by age.

		Production accuracy (chance = 1/6)									
		Unfiltered				Filtered				Perception accuracy (chance = 1/2)	
		1 judge		2 judges		1 judge		2 judges			
Age	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
2;1–2;6	13	0.82	0.09	0.61	0.12	0.57	0.10	0.36	0.11	0.60	0.09
2;7–3;0	18	0.90	0.07	0.71	0.12	0.65	0.10	0.46	0.08	0.64	0.04
3;1–3;6	15	0.93	0.09	0.73	0.11	0.67	0.11	0.45	0.13	0.70	0.07
3;7–4;0	12	0.85	0.10	0.66	0.11	0.70	0.12	0.42	0.09	0.75	0.10
4;1–4;6	17	0.93	0.06	0.76	0.11	0.67	0.10	0.47	0.10	0.84	0.06
4;7–5;0	15	0.94	0.04	0.81	0.07	0.77	0.09	0.53	0.10	0.90	0.06
5;1–5;6	11	0.98	0.02	0.90	0.06	0.74	0.07	0.52	0.08	0.89	0.07
5;7–6;0	10	0.94	0.05	0.80	0.09	0.73	0.09	0.49	0.09	0.90	0.04
Reference	10	0.99	0.01	0.94	0.04	0.90	0.05	0.74	0.11	0.99	0.01

most lenient condition but hovering around 50% in the most stringent condition. Even the accuracy of the adolescent reference speakers, who were confirmed not merging any of the tones, was only 0.74 in the most stringent condition. This suggests that neither of these two extremes was realistic.

Figure 2 shows the production accuracy of individual tones under the four judgment conditions. Each panel in Figure 2 shows one judgment condition. The six tones were arranged on the horizontal axis; accuracy was on the vertical axis. The boxplots represent data for each age group, from the youngest (2;1–2;6) on the left to the reference speakers on the right. Some patterns can be observed. First, except Tone 5, the other tones were at or near ceiling accuracy in the most lenient condition (top left panel). Tone 5 was also judged to be less accurately produced in other conditions too. The accuracy difference between the most stringent condition (filtered-two-judges, bottom right) and the other conditions is more obvious for the level tones (T3 and T6,

Figure 2. Mean production accuracy of individual tones by age in four judgment conditions.



to a lesser extent T1), even for the reference speakers. The boxplots for T3 and T6 of the filtered-two-judges condition were much lower than the respective boxplots in other conditions. This was probably due to the fact that it was much harder to assess whether a level pitch contour was a T3 [33] or T6 [22] in isolation without segmental contexts, given the minimal difference in pitch between the two tones (as little as 20 Hz in adult female speech). A small change in pitch range during the production experiment of individual items could easily influence the judgments of the transcribers. It appears that the accuracy in the four conditions concur most for T2, followed by T4, showing generally higher accuracies than other tones across conditions. This probably was due to the more prominent pitch contour of these two tones (an obvious rise toward the end in T2 and the only falling contour in T4). These distinct contour cues made them more identifiable even in the filtered condition.

The overall mean perception accuracy by age can be found in Table 2 (right panel). It should be noted that the perception accuracy in Table 2 should not be directly compared with the production accuracy in the same table as the chance levels of the two sets of data are different. Again, one-sample *t* tests revealed that the perception accuracy of all age groups were significantly higher than chance level (1/2 = 50%), $p < .001$ (full statistical details can be found in Supplemental Material S1). There is a clear pattern of perception accuracy increasing with age. Such pattern can also be found in the perception accuracy of individual tones by different age groups shown in Figure 3. The reference speakers were at ceiling for all six tones. Except T1, the children in the oldest age group (5;7–6;0) still fell short of adult accuracy in perception. In addition to perceptual development, such increase can also be due to maturation that older kids are better with general task performance. It is interesting to note that the age of 4 years appears to be a watershed in tone perception development. Before age 4;0, there was a more steady increase in perception accuracy. The perception development seemed to have slowed down after age 4;0 with reduced improvement.

In order to substantiate the above observation based on averaged data that the age of 4 years appears to be a watershed in tone perception development, we examined the data of individual tone pairs. The slowing down of perception development after age 4;0 can be seen in individual tone pairs as well. The perception accuracy data were grouped yearly instead of half-yearly in Figure 4 for visual clarity and fewer comparisons. A two-way mixed analysis of variance with yearly age group and tone pair as fixed factors reveals significant main effects for both age group, $F(1, 4) = 90.783$, $p < .0001$, $\eta^2 = .759$, and tone pair, $F(10.766, 1238.094) = 49.610$, $p < .0001$, $\eta^2 = .301$ (Greenhouse–Geisser corrected), and a significant interaction between them, $F(56, 1610) = 3.069$, $p < .0001$, $\eta^2 = .096$. Multiple comparisons with Bonferroni corrections were conducted to further examine the interaction. Because of space limit, a summary table with the *p* values and full statistical results of the multiple comparisons can be found in Supplemental Material S1.

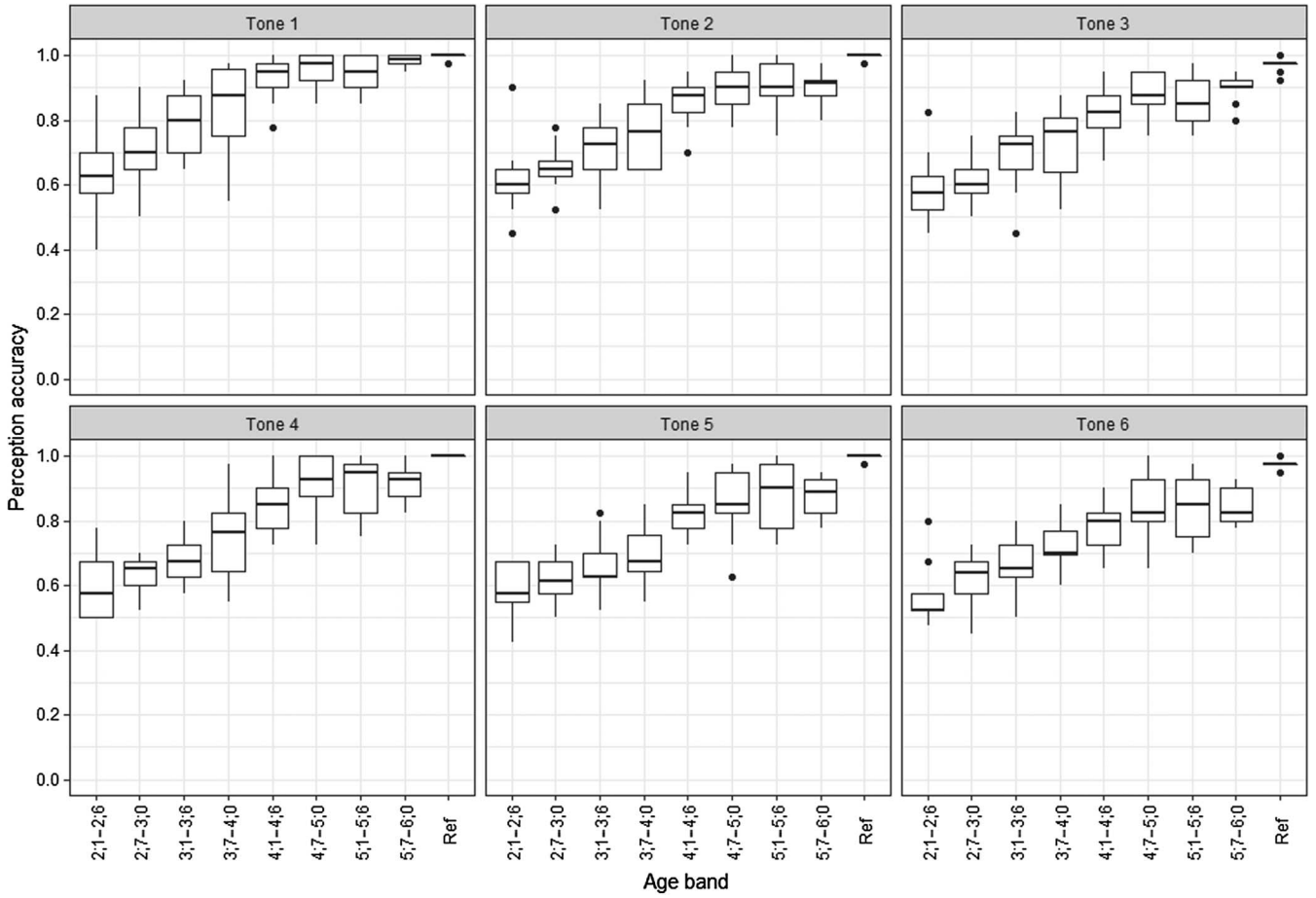
To highlight the main patterns, there was an obvious improvement between ages 2;1 and 4;0 (the two lines were wider apart), whereas the data nearly overlapped for ages 4;1–6;0. Multiple comparisons between younger children (ages 2;1–3;0 and 3;1–4;0 separately) and older children (ages 4;1–5;0 and 5;1–6;0 separately) were mostly significant (generally better than $p < .001$). The comparisons among older children (ages 4;1–5;0 vs. 5;1–6;0) and between older children and the reference speakers were mostly nonsignificant. These comparisons confirm that the age of 4 years acts as a watershed in tone perception development and that perceptual development of tone has slowed down after age 4;0.

There were three noticeable dips in perception accuracy of the child data in Figure 4: T2T5, T3T6, and to a lesser extent T4T6. These pairs were acoustically very similar and were identified to be merging among young Cantonese speakers (Mok et al., 2013). Those three pairs were also the most difficult to discriminate by children and adults in Lee et al. (2015). In our data, all comparisons among children were nonsignificant for T3T6 (unlike for most other tone pairs reported above), whereas those between the reference speakers and children (four yearly groups) were all significant ($p < .002$ or better). The reference speakers were very accurate in tone perception, as they were screened for their production accuracy, but still, there was also a slight dip for T3T6 for them. This again illustrates the difficulty in distinguishing these two similar level tones. For T2T5, the differences among younger children (ages 2;1–3;0 vs. 3;1–4;0) and among older children (ages 4;1–5;0 vs. 5;1–6;0) were not significant, whereas those between younger children and older children were mostly significant ($p < .01$ or better). The reference speakers were significantly better than all children (all *p*s < .001 or better). For T4T6, younger children (ages 2;1–3;0 and 3;1–4;0 separately) were significantly worse than older children and the reference speakers, whereas the reference speakers were better than children aged 4;1–5;0 ($p < .001$) but not significantly better than the oldest children.

Both production and perception data demonstrate a general increase of accuracy with age, albeit the pattern being much stronger for the perception data than for the production data, which depended on judgment conditions. A number of correlation analyses were conducted to assess the relationship with age. Table 3 (top panel) gives the results of the Spearman's rho analyses. Unexpectedly, for production data, all four conditions significantly correlated with age, despite the observed leveling off of production accuracy in two conditions discussed above. The strength of correlation differed among the four conditions though, with the unfiltered-two-judges condition most strongly correlated with age ($r = .548$). The perception data also correlated very strongly with age ($r = .834$), as observed in Figure 3. These data clearly suggest a normal developmental pattern in tone acquisition: maturity with age and incomplete acquisition by age 6;0, contrary to findings in previous tone acquisition studies showing very early acquisition.

Finally, as both the production and perception data were collected from the same groups of children, we can

Figure 3. Perception accuracy of individual tones by age.

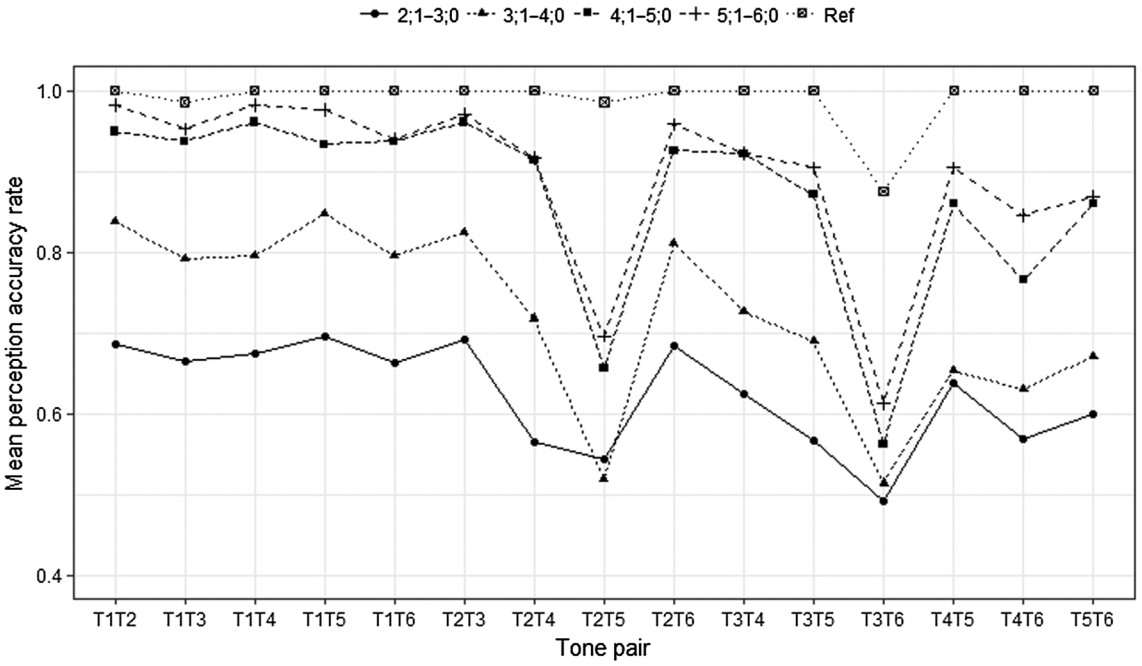


assess the relationship between these two important aspects in tone acquisition. A significant positive correlation is found between perception accuracy and production accuracy in all four conditions (see Table 3, bottom panel). This can be expected, given that both are found to strongly correlate with age above. We further asked if perception accuracy could predict production accuracy, that is, if perception preceded production in tone development. Figure 5 presents scatter plots showing the perception and production data in four conditions. Simple linear regressions with perception accuracy as predictor were conducted (see results in Figure 5). Perception accuracy significantly predicts production accuracy in all conditions, with the unfiltered-two-judges condition being the best model. Nevertheless, the R^2 is quite small (.2152), which indicates that perception ability was only one of the factors affecting children's tone production. Much of the variance in production cannot be explained by perception accuracy alone.

We turned to the language background of the children to explore possible factors, which may explain additional variance in the production data besides perception accuracy, as Table 1 shows that not all parents spoke Hong Kong Cantonese natively. Mixed effects logistic regressions were performed using the glmer function in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) on R (R Core Team, 2018) to assess the relationship between production accuracy and perception accuracy, child's age, and parents' L1. For each condition, three models were constructed. The base model consisted of each child's mean perception accuracy and age as fixed effects, as well as by-child and by-item random intercepts. The base model was compared with two other models that included one additional fixed factor, either father's L1 or mother's L1. The L1s of parents were binary categorical variables: being a native speaker of Hong Kong Cantonese or not. In the models, the parents' L1 was treatment coded with native speakers being the baseline condition. To attenuate multicollinearity, perception accuracy and age were *z*-score normalized. Inspection of variance inflation factors (VIFs) indicates that multicollinearity was moderate: VIFs for normalized perception and normalized age were around 3.3 (highest in all models was 3.5), and VIFs for other factors were smaller than 2, all of which were much lower than the common rejection threshold of 10.

In all four conditions, inclusion of father's L1 did not significantly improve the model fit. However, the inclusion of mother's L1 significantly improved the model fit

Figure 4. Mean perception accuracy according to tone pairs.



in the unfiltered-one-judge condition (change of deviance: 4.3102, $p = .038$) and the unfiltered-two-judges condition (change of deviance: 4.6179, $p = .032$). Results in Table 4 show that, compared with children whose mothers did not speak Hong Kong Cantonese natively, we can expect an increase of 0.492 and 0.302 in log odds ratio for correct production for the one-judge and two-judges conditions, respectively, for children with mothers being native speakers of Hong Kong Cantonese.

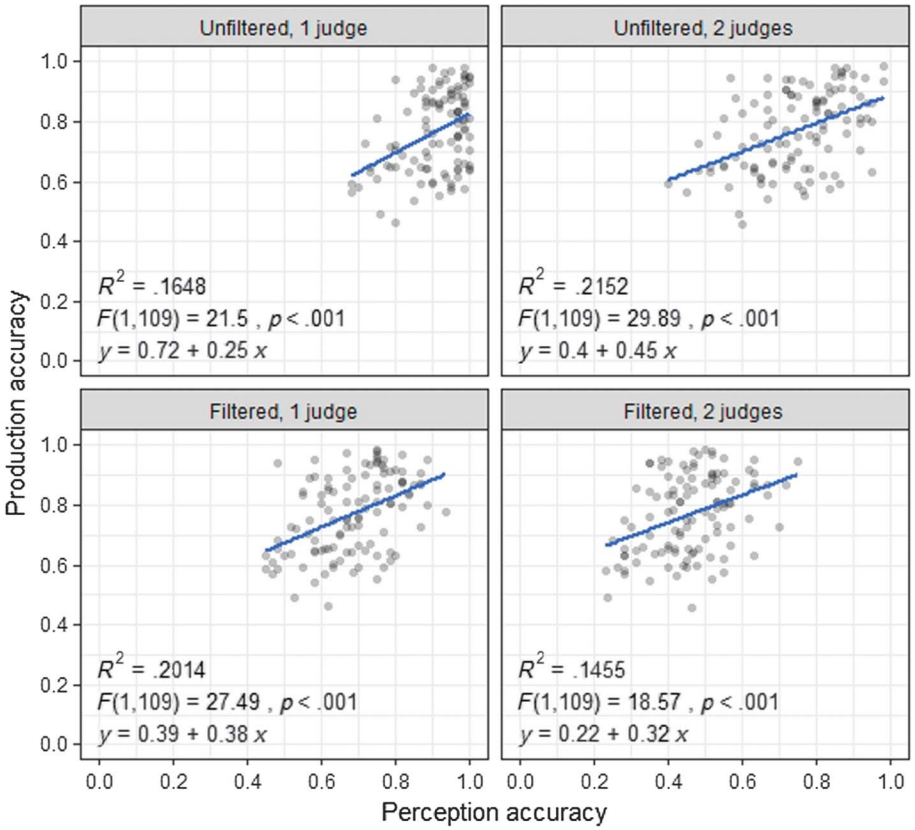
Discussion

Our study revisited the acquisition of Cantonese tones by monolingual children aged 2;1–6;0. Previous studies displayed large discrepancy between tone production and perception abilities: very early acquisition of production but relatively late acquisition of perception. Our findings reveal new understanding of the acquisition process. Although previous production studies using simple transcription data by one judge concluded that Cantonese children could produce all six tones correctly by age 2;0 or 2;6, our data demonstrated that production accuracy was still not adultlike by age 6;0 using more stringent judgment criteria. Even in the most lenient condition (unfiltered-one-judge), our data show that T5 was still not adultlike for the children at all age groups (see Figure 2). Contextual influence can have a strong impact on accuracy judgment. Our perception data concur with previous findings that perception accuracy matures with age beyond age 6;0. We additionally provided tone perception data from children between ages 2;1 and 3;0 and showed that the rates of perception improvement differ before and after age 4;0: improving more steadily before age 4;0 but improving more slowly after age 4;0. Although the number of participants in each half-yearly group was relatively small, which may render such conclusion tentative, the statistical results were based on yearly age groups

Table 3. Correlation analyses with age and with perception accuracy.

Metric	Analysis	Spearman's rho	*p*
Correlation with age	Production (unfiltered, 1 judge)	0.423	< .0001
	Production (unfiltered, 2 judges)	0.548	< .0001
	Production (filtered, 1 judge)	0.453	< .0001
	Production (filtered, 2 judges)	0.362	< .0001
	Perception	0.834	< .0001
Correlation with perception accuracy	Production (unfiltered, 1 judge)	0.345	< .0001
	Production (unfiltered, 2 judges)	0.448	< .0001
	Production (filtered, 1 judge)	0.441	< .0001
	Production (filtered, 2 judges)	0.346	< .0001

Figure 5. Scatter plots of perception and production data in four conditions with regression lines.



(combining two half-yearly groups for each year) with more participants, thus confirming that this conclusion is a valid one.

The production and perception data from the same 111 children in our study confirmed that both abilities correlate with age and that there is a positive link between the two in tone acquisition. Perception accuracy significantly predicts production accuracy, albeit the small amount of variance (R^2 = .2152) explained by perception ability alone. Mothers' L1 is another factor that can affect tone production accuracy of the children (more discussion on this below).

Table 4. Results of the mixed-effects logistic regressions.

Condition	Deviance of adjusted model	Deviance of base model	Change of deviance	Predictor	Estimate	*SE*	*t* Value	Pr(>|*t*|)
Filtered, 1 judge	7,263.353	7,263.504	0.1510	(Intercept)	0.944	0.163	5.780	< .001
			$p(\chi^2)$ = .698	Perception_mean_scaled	0.139	0.097	1.436	.151
				Age_scaled	0.145	0.095	1.527	.127
				Mothers' language	–0.045	0.115	–0.389	.697
Filtered, 2 judges	7,709.891	7,710.741	0.8505	(Intercept)	–0.168	0.192	–0.872	.383
			$p(\chi^2)$ = .356	Perception_mean_scaled	0.124	0.091	1.361	.174
				Age_scaled	0.090	0.089	1.007	.314
				Mothers' language	–0.100	0.108	–0.925	.355
Unfiltered, 1 judge	3,298.318	3,302.628	4.3102	(Intercept)	3.185	0.203	15.709	< .001
			$p(\chi^2)$ = .038	Perception_mean_scaled	0.059	0.193	0.307	.759
				Age_scaled	0.435	0.192	2.266	.023
				Mothers' language	–0.492	0.231	–2.134	.033
Unfiltered, 2 judges	6,724.837	6,729.455	4.6179	(Intercept)	1.371	0.129	10.664	< .001
			$p(\chi^2)$ = .032	Perception_mean_scaled	0.014	0.116	0.120	.905
				Age_scaled	0.393	0.115	3.422	.001
				Mothers' language	–0.302	0.138	–2.188	.029

As discussed in the introduction, there are strong parallels between previous findings on Mandarin and Cantonese tone acquisition, both showing very early production. Scholars working on Mandarin tone acquisition rationalized the unusually early mastery by suggesting that "the acquisition of tones would be completed early, probably due to its capacity in differentiating lexical meaning and fulfilling children's communicative intentions" (Zhu, 2002, p. 45). Nevertheless, the same capacity is applicable to consonants and vowels as well. The main difference between tones and segments should lie in articulatory complexities and functional load. Simpler speech motor control needed for tone production (mainly laryngeal coordination vs. coordination involving various articulators in the oral cavity) and high functional load of tones (only a few tones vs. many segments) may render them easier for children to master. Nevertheless, our more stringent judgment criteria demonstrate that, even with these two properties, tone production accuracy was still not completely adultlike by age 6;0. Our findings "demystify" Cantonese tone acquisition by demonstrating that it indeed follows the general principles of first language acquisition that perception precedes (or at least goes hand in hand with) production, similar to segmental developments in both Mandarin and Cantonese, as well as other languages (Zhu & Dodd, 2006).

As there were four judgment criteria used in our study, which criterion can most faithfully reflect children's production performance? The unfiltered-one-judge condition used in previous studies was clearly too lenient and too easily subject to contextual influence, which led to the inaccurate conclusion of very early acquisition of tone. The most stringent criterion of filtered-two-judges condition proposed by Wong et al. (Wong et al., 2017, 2005; Wong & Leung, 2018) was not suitable either, because even adult nonmerging reference speakers were shown to have a relatively low accuracy (0.74; see Table 2). If we had followed Wong in using five judges, the accuracy rates would be even lower. In addition, filtering has a larger impact on level tones than contour tones (see Figure 2), which may unduly influence transcribers' judgments and data analysis. This problem was not revealed in Wong's studies on Mandarin because Mandarin tones are distinguished by pitch contour but not pitch height. Nevertheless, both pitch contour and pitch height are important features of Cantonese tones (Gandour, 1981, 1983). Our data illustrated this drawback clearly, as T3 and T6 were less affected in the unfiltered-two-judges condition. In Wong's recent studies in Cantonese (Wong et al., 2017), the perceived accuracy of T3 and T6 was also markedly lower for both adults and children (see Figure 3 in their study). They concluded that both adults and children produced these two tones less accurately than other tones without realizing that the reduced accuracy may also be an artifact caused by their method.

A faithful criterion should result in high accuracy for adult reference speakers and developmental improvement for children. The two remaining criteria fulfill this requirement and gave very comparable results, with unfiltered-two-judges yielding slightly higher accuracy. As listening to filtered materials is more about auditory perception than normal speech perception and that perceptual differences were found for these two types of materials with the same contours (Bidelman, Gandour, & Krishnan, 2010; Mok & Zuo, 2012), the unfiltered-two-judges condition should be a more realistic criterion that can be easily adopted. Either way, it is important to use more rigorous (but not overly stringent) methods to assess production accuracy.

On a related note, in addition to judgment data, acoustic analysis would be an ideal method to examine tone production. Many studies on segments have demonstrated that the time course of development in child's speech production is much more protracted when production accuracy is assessed by acoustic analysis than by transcription alone (Edwards & Beckman, 2008; Edwards, Beckman, & Munson, 2015; Munson et al., 2012). The same situation can be found in tone acquisition as well. Wong and colleagues (Wong, 2012; Wong et al., 2017) demonstrated that, even for Mandarin and Cantonese 3-year-old children whose tone production was judged to be accurate, the acoustic patterns of their tones were still different from those of adults. Acoustic analysis of the production data by both children and the reference speakers in our study is under way. It is probably impractical to expect children's production to be the same as those of adults, but if both were considered to be correctly produced by the native judges, it will be interesting to compare the differences in fine phonetic detail between the two groups of speakers and to examine the perceptual salience of these phonetic details. In addition, it would be possible for us to do more sophisticated analysis on both production and perception data (e.g., via individual differences multidimensional scaling using multivariate data; Chandrasekaran, Sampath, & Wong, 2010) to explore the dimensions underlying children's perceptual and acoustic space and compare them directly with acoustic data to investigate the production–perception link. This can give us a more in-depth understanding of the relationship between production and perception in first language acquisition.

Our tone perception data concur very well with Lee et al.'s (2015) study, not only in the overall pattern of continuous improvement from age 3;0 to 6;0 but also in the comparable percentages of average accuracy of different age groups (see Table 2 in our study and Table I in the study of Lee et al., 2015). Perception difficulty for individual tone pairs was also very similar (see Figure 4 in our study and Figure 2 in their study). This gives strong support that the perception accuracy of children aged 2;1–3;0 (around 60%), reported for the first time in our study, is trustworthy. Our perception experiment was a forced-choice identification task with only two pictures, that is, the chance level was 50%. The age of 2 years is probably the youngest age with which participation in a behavioral perception experiment like ours (picture identification) is possible. Younger children would need to be tested with preferential looking tasks, which data may not be directly comparable with ours. Although we do not have the data, it is reasonable to

predict that tone perception accuracy for children between ages 1;0 and 2;0 would probably be just around chance level. The findings of infants distinguishing tonal contrasts without meaning in their first year of life (e.g., Mattock & Burnham, 2006; Mattock et al., 2008) are not comparable with our data, as they were based on very simple contrasts (just two tones). This leads to an interesting question: How can young children, especially those between ages 1;0 and 3;0, understand adult Cantonese speech in which tonal variation is essential when their own tone perception ability is far from perfect?

One important consideration is that, in everyday interactions with children, it is uncommon to have the need to contrast monosyllabic minimal pairs in which tone is the only phonetic difference to distinguish meaning. Thus, even if their tone perception ability is weak, contexts of various sorts can help the children to understand what is going on. This would lessen the communication problems posed by poor tone perception ability, on the one hand, and would in turn mask their inability to distinguish the tones accurately, on the other.

This situation also relates to a key question in phonological acquisition: Do children acquire words first or do they acquire sounds (tones) first? Although infant perception studies demonstrate that perceptual narrowing found for segments can also be demonstrated for tones (Burnham & Mattock, 2007; Mattock & Burnham, 2006; Mattock et al., 2008; Yeung, Chen, & Werker, 2013), our perception data clearly indicate that children had not established the abstract tonal categories before they started using tones in their production. As mentioned above, previous studies showing early discrimination of tones by infants reviewed in Singh and Fu (2016) used only very simple tone contrasts (often just two tones). The ability of infants learning tone languages to maintain sensitivity to acoustic differences between simple stimuli in their first year of life is not the same as the ability to distinguish all possible tonal contrasts in authentic situations related to meaning later in life. Our tone data suggest that, together with previous studies on segments (e.g., Edwards et al., 2015; Redford, 2015; Vihman, DePaolis, & Keren-Portnoy, 2014), children were using whole-word patterns before abstract phonological categories emerged. They developed tonal categories alongside their use in production, notwithstanding how inaccurate that may be. The development of higher level phonological knowledge is a protracted process, equally for both segments (Munson et al., 2012) and tones. Children learn sounds within lexical contexts. In a recent target paper, Vihman (2017) argued that children learn both words and sounds concomitantly and that production may play a role in shaping perception. Her discussion was based on findings in segmental development in the first 2 years of life, but it applies equally well to tonal development in a later age as well.

If children were developing both perception and production abilities at the same time, as Vihman (2017) suggested, it is natural to expect a positive relationship between the two during language development. There were not many studies with both child perception and production data to assess this link, as perception studies often focus on perceptual narrowing of younger infants whereas production study often involved older children. Wong et al. (2017, 2005) found no relationship between perception and production in Mandarin and Cantonese tones with limited participants in one age band, whereas she found a weak correlation between the two in Cantonese (R^2 = .194) with more children (Wong & Leung, 2018). Our data clearly demonstrate that there is a significant positive link between perception and production accuracy in tone acquisition, supporting that perception and production develop concurrently. Nevertheless, the link is rather loose, as perception accuracy can only account for at most 22% of the variance in the production data, which corroborates Wong and Leung's results well. The low explanatory power should not come as a surprise though, because both abilities were still developing and that the developmental trajectories are not necessarily linear, as seen in both our production and perception data. Our findings concur with those in McAllister Byun and Tiede's (2017) study on the production and perception of English /r/ with older children. They also found a significant link between the two, but the link was weak and variable. All these findings suggest that a link between production and perception can be expected in children's first language acquisition, although the link may not be very robust as the two aspects involve abilities maturing at different speeds.

A question remains. What may explain the remaining approximately 80% of the variance in the production data unaccounted for by perception accuracy? The large-scale study by To, Cheung, and McLeod (2013a) has examined the effects of several demographic factors on Cantonese speech acquisition, for example, family income, parental education, the presence of siblings, and domestic helpers, but they were found to have minimal effect on speech acquisition. However, To et al. did not include family language background. Previous studies show that both the quantity and quality of input have an effect on language acquisition (Stevens, 2006; Unsworth, 2007). Moreover, an important factor affecting input quality is whether input providers speak a standard or nonstandard variety and whether they are native or nonnative speakers (Hulk & Cornips, 2006). Family language background may play a role in Cantonese tone acquisition. Indeed, it is what we found. Children whose mothers were native speakers of Hong Kong Cantonese produced tones slightly more accurately[2] than those with mothers speaking Cantonese nonnatively.

The dynamic demographic composition and language contact in Hong Kong due to cross-border marriage over the past few decades can explain the above findings. Children

[2]This conclusion is supported by the reduced deviance in the mixed-effects logistic regression analyses when mothers' native language was included (see Table 4). Earlier analyses using multiple regressions indicated that we can expect an increase of 4.4% and 4.9% in production accuracy for the unfiltered-one-judge and unfiltered-two-judges conditions, respectively, for children whose mothers were native speakers of Hong Kong Cantonese.

born to cross-border marriage may be subjected to imperfect tone acquisition. Cross-border marriage, which constitutes a significant portion of all marriages in Hong Kong (Government, 2007, 2010), often involved a Hong Kong man marrying a woman from mainland China. Over 10% of children born in Hong Kong in the last 20 years were from cross-border marriage. The numbers do not include those who were born in mainland China but grew up in Hong Kong, so the actual numbers are likely to be higher. Most of the cross-border mothers are mainland women who are not native speakers of standard Cantonese, although many of them can communicate in Cantonese with a strong accent. As a result, children of such families, even if they grow up in Hong Kong, are often exposed to deviant Cantonese tones of their primary caretakers in their early years. Both the quantity and quality of tone input to these children are likely to affect their acquisition of the complex Cantonese tones, particularly the difficult tone pairs identified above. Nevertheless, even family language background can only account for an additional few percentage of variance in children's production. There are still many unknown factors that need to be explored for a fuller picture of tone acquisition.

In conclusion, our study revisited tone acquisition in Cantonese-speaking children. We found that both production and perception mature with age and that Cantonese tones are still not fully acquired at age 6;0. There is a weak positive link between production and perception accuracy. Family language background in the form of the L1 of the primary caretaker is a factor contributing to children's tone acquisition, but further study is needed to understand tone acquisition more comprehensively. Our findings support the idea that children acquire words and phonology simultaneously in first language acquisition.

Acknowledgments

This study was supported by the Hong Kong SAR Government Research Grant Council General Research Fund 2015/16 Project Reference 14602715, awarded to the first author, and the funding support by the department of the authors. We would like to thank Mandy Cheung, Crystal Lee, and Mercy Wong for their help in data collection and analysis. We are grateful for the five kindergartens participating in our study: ELCHK Shatin Lutheran Kindergarten, Heng On Baptist Nursery School, Homantin Yang Memorial Methodist Pre-School, Lutheran Philip Hse Oi Lun Nursery School, and NTW&JWA Cheung Fat Estate Nursery School. We especially thank the parents and children for their participation. We also thank Carol To for her assistance in the earlier stages of the project.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin, NY: Mouton de Gruyter.

Berko, J., & Brown, R. (1960). Psycholinguistic research methods. In P. H. Mussen (Ed.), *Handbook of research methods in child development* (pp. 517–557). New York, NY: Wiley.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Bidelman, G. M., Gandour, J., & Krishnan, A. (2010). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23,* 425–434.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 259–280). Amsterdam, the Netherlands: John Benjamins.

Butler, S. (1980). The tum phenomenon. *Journal of Child Language, 7,* 428–429.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128,* 456–465.

Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique, 45,* 24–27.

Chao, Y. R. (1947). *Cantonese primer*. New York, NY: Greenwood Press.

Cheung, P. S. P., Ng, A., & To, C. K. S. (2006). *Hong Kong Cantonese Articulation Test*. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

Ching, Y. C. (1984). Lexical tone pattern learning in Cantonese children. *Language Learning and Communication, 3,* 317–334.

Ciocca, V., & Lui, J. (2003). The development of lexical tone perception in Cantonese. *Journal of Multilingual Communication Disorders, 1,* 141–147.

Clark, E., & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology, 34,* 325–349.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics, 22,* 937–956.

Edwards, J., Beckman, M. E., & Munson, B. (2015). Cross-language differences in acquisition. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 530–554). Hoboken, NJ: Wiley.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 151–190). Berlin, Germany: Mouton de Gruyter.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: Hong Kong University Press.

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics, 9*(1), 20–36.

Gandour, J. (1983). Tone perception in far eastern languages. *Journal of Phonetics, 11*(2), 149–175.

Government, Hong Kong SAR. (2007). *Demographic trends in Hong Kong 1981–2006*. Hong Kong: Census and Statistics Department.

Government, Hong Kong SAR. (2010). *The fertility trend in Hong Kong, 1981 to 2009*. Hong Kong: Census and Statistics Department.

Hulk, A. J. C., & Cornips, L. (2006). The acquisition of definite determiners in child L2 Dutch: Problems with neuter gender nouns. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition* (pp. 107–134). Amsterdam, the Netherlands: John Benjamins.

Idemaru, K., & Holt, L. L. (2013). The development trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America, 133,* 4232–4246.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*(1), 104–117. https://doi.org/10.1016/j.wocn.2005.10.003

Lee, K. Y. S. (2012). *The Cantonese Tone Identification Test (CANTIT)*. Hong Kong: Department of Otorhinolaryngology, Head & Neck Surgery, the Chinese University of Hong Kong.

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

Lee, K. Y. S., Chiu, S. N., & Hasselt, C. (2002). Tone perception ability of Cantonese-speaking children. *Language and Speech, 45*(Pt. 4), 387–406.

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

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10,* 241–265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106,* 1367–1381.

McAllister Byun, T., & Tiede, M. (2017). Perception–production relations in later development of American English rhotics. *PLOS ONE, 12*(2), e0172022.

Mok, P., & Zuo, D. (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *The Journal of the Acoustical Society of America, 132*(4), 2711–2720. https://doi.org/10.1121/1.4747010

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

Munson, B., Edwards, J., & Beckman, M. E. (2012). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 288–309). Oxford, England: Oxford University Press.

Polka, L., Rvachew, S., & Mattock, K. (2008). Experiential influences on speech perception and speech production in infancy. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 153–172). Oxford, England: Blackwell.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics, 53,* 141–152.

Singh, L., & Fu, C. S. L. (2016). A new view of language development: The acquisition of lexical tone. *Child Development, 87,* 834–854.

So, L. K. H., & Dodd, B. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language, 22*(3), 473–495.

Stevens, G. (2006). The age-length-onset problem in research on second language acquisition among immigrants. *Language Learning, 56,* 671–692.

To, C. K. S., Cheung, P. S. P., & McLeod, S. (2013a). The impact of extrinsic demographic factors on Cantonese speech acquisition. *Clinical Linguistics & Phonetics, 27,* 323–338.

To, C. K. S., Cheung, P. S. P., & McLeod, S. (2013b). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels and tones. *Journal of Speech, Language, and Hearing Research, 56,* 103–122.

Tse, J. K. P. (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language, 5,* 191–204.

Unsworth, S. (2007). *Age and input in early child bilingualism: The acquisition of grammatical gender in Dutch*. Paper presented at the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA), Montréal, Canada.

Vihman, M. (2014). *Phonological development: The first two years*. Chichester, West Sussex: Wiley.

Vihman, M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology, 108,* 1–27.

Vihman, M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning, 64*(Suppl. 2), 121–140.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of nonnative speech sounds. *Canadian Journal of Psychology, 37,* 278–286.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganisation in the first year of life. *Infant Behavior & Development, 7,* 49–63.

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

Wong, P. S. (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,* 434–443.

Wong, P. S., Fu, W. M., & Cheung, E. Y. L. (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, 8,* 1450.

Wong, P. S., & Leung, C. T. T. (2018). Suprasegmental features are not acquired early: Perception and production of monosyllabic Cantonese lexical tones in 4- to 6-year-old preschool children. *Journal of Speech, Language, and Hearing Research, 61,* 1070–1085.

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

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68,* 123–139.

Yip, M. (2002). *Tone*. Cambridge, England: Cambridge University Press.

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon, England: Multilingual Matters.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27,* 3–24.

Zhu, H., & Dodd, B. (Eds.). (2006). *Phonological development and disorders in children: A multilingual perspective*. Clevedon, England: Multilingual Matters.