JASA ARTICLE

Acquisition of Mandarin tones by Canadian first graders: Effect of prior exposure to tonal and non-tonal languages

Ana Meckelborg,[1] Mimi Luu,[1] Theresa Nguyen,[1] Youran Lin,[1] Fangfang Li,[2] and Karen Pollock[1,a)]
[1]Department of Communication Sciences and Disorders, University of Alberta, Edmonton, Alberta T6G 2G4, Canada
[2]Department of Psychology, University of Lethbridge, Lethbridge, Alberta, Canada

ABSTRACT:
This study examines the tone productions of school-aged children with and without a tonal language background who are learning Mandarin as a second language (L2) or heritage language in Mandarin-English bilingual schools in Western Canada. Tones are frequently identified as one of the most challenging aspects of phonology for Mandarin L2 learners to acquire. In this study, tone productions of bilingual children from three home language backgrounds, English, Cantonese, and Mandarin Chinese, were compared for transcribed accuracy using mixed effects logistic regression. In addition, the fundamental frequency contours of correct tone productions were fitted with generalized additive mixed models to analyse the acoustic differences between groups. Error patterns were also analysed for possible Cantonese substitutions. Our results suggest that children with a Cantonese background are more accurate in tone productions than children with an English language background, but they also made more errors than their peers with a Mandarin language background. These findings suggest that a tonal language background could result in positive transfer among school-age children who are in the early stages of learning Mandarin as an L2.
 https://doi.org/10.1121/10.0024985

(Received 18 June 2023; revised 4 February 2024; accepted 6 February 2024; published online 23 February 2024)
[Editor: James F. Lynch] Pages: 1608–1623

I. INTRODUCTION

Lexical tone is a critical prosodic feature that provides meaning contrast in languages such as Mandarin Chinese. Despite the early mastery of tones by children who acquire tonal languages as native speakers, those who learn these languages later in life as a second language (L2) often find tones challenging. The degree of such challenge, however, may vary depending on whether the L2 learners' first language (L1) has tones or not. Existing theoretical frameworks of L2 speech acquisition, such as the speech learning model (SLM) (Flege, 1995; Flege and Bohn, 2021), do not clearly stipulate the developmental process of L2 in relation to L1 in the suprasegmental domain. The perceptual assimilation model (PAM) (Best and Tyler, 2007; So and Best, 2010) states that the way L2 suprasegmental features such as tones can be assimilated into existing L1 categories is based on similarities between tone pairs in the two languages, the same way that L2 listeners assimilate segmental categories. However, non-native speakers' perceptual mapping patterns is only one factor that influences L2 acquisition.

Empirical research has found mixed results ranging from facilitatory, to inhibitory, and to no effects of a tonal L1 background [e.g., Francis *et al.* (2008); Hallé *et al.* (2004); Hao (2012); Lee *et al.* (1996); So and Best (2010, 2014); and Wayland and Guion (2004)]. These varied results are likely due to different aspects of the L2 examined (i.e., production vs perception), the specific acoustic-perceptual characteristics of tones in L1 and L2, the relative complexity of L1 and L2 tone systems, and the quantity and quality of L2 input. It is worth noting that a large body of literature on L2 tone acquisition concerns the effect of linguistic experience on tonal perception (Hallé *et al.*, 2004; Peng *et al.*, 2012; Reid *et al.*, 2015; Wu *et al.*, 2019). For example, So and Best (2010) investigated the Mandarin lexical tone perception in naive listeners of four language backgrounds, Mandarin native, Hong Kong Cantonese, Japanese, and English. They found that the lack of linguistic experience with tones in English speakers resulted in reduced tonal perceptual sensitivity, but the tone identification errors were largely consistent across the four language groups, lending no support to the facilitatory effect of prior tonal language background.

Production studies of bilingual tonal acquisition mostly focus on adult L2 speakers and may not reflect how children, who have a less refined L1 and more neuroplasticity, learn a tonal L2. The few studies on bilingual children's tone acquisition focus on children of diverse backgrounds and acquisition contexts and utilize different methodologies. For example, in a study with Cantonese-L1 children in Australia (age range = 2;0–5;7) learning English as a L2, children produced tones with accuracy comparable to that of monolingual Cantonese-L1 peers in Hong Kong (Holm and Dodd, 2006). Similarly, Mok and Lee (2018) utilised the CHILDES corpus to investigate the tonal production patterns of two simultaneous Cantonese-English bilingual children in Hong Kong (age range 2;0–2;6) and found that they also had tone accuracy rates similar to those of monolingual

a)Email: pollock@ualberta.ca

Cantonese peers. In addition to comparable tone production accuracy, however, the tones produced by some of the Cantonese-English bilingual children in the study of Mok and Lee (2018) exhibited a high-low template, a pattern resembling the English trochaic (stressed-unstressed) pattern. These two studies suggest the overpowering influence of home language input in bilingual children's tonal acquisition when children are very young as well as how tonal acquisition in the L2 may be impacted by a non-tonal L1.

In slightly older bilingual children, one might expect a greater impact of L1 on L2 tone production. Yao and colleagues (Yao *et al*., 2020) examined children of slightly older age who were Urdu-L1/Cantonese-L2 bilinguals (ages 4;5–6;6) in mainstream Cantonese preschools in Hong Kong. They produced Cantonese tones less accurately compared with monolingual Cantonese peers, again revealing the influence of a non-tonal language, in this case Urdu, on tonal acquisition of the target language in bilingual children. In an observational report on the phonetic errors of Cantonese-L1 children (4–12 years of age) learning Mandarin in schools in Hong Kong, Duan *et al*. (2022) noted that tone substitutions were prominent among the speech errors. However, a detailed analysis of the possible impact of Cantonese tones on the production of Mandarin tones was not provided. In spite of highly varied participant characteristics and learning context and research methodologies, these studies point to the importance of age of acquisition, home language, societal language, acquisition style, the social status of the two languages, and interaction between the two languages in bilingual children's tonal acquisition.

The current study adds to the diverse literature on children's tonal acquisition by presenting another unique population of children who are in Grade 1 of a two-way bilingual Mandarin-English program in Canada. Three subgroups of children from different home language backgrounds (Mandarin, English, and Cantonese) are included to examine the influence of tonal (Cantonese or Mandarin) and non-tonal (English) experience on the acquisition of Mandarin tones. Compared with previous studies, which focused solely on the impact of a non-tonal language on children's acquisition of tones, our study includes children from a tonal language background (Cantonese) that is different from the target language (Mandarin) as well as a non-tonal language background (English). This design extends previous research to identify the important effect of tonal language background for bilingual tone acquisition. See Appendix B for a comparison of the present study with previous literature on child bilingual tonal productions.

Theoretically, the current study contributes in several ways to several long-standing questions in the field of L2 phonetic learning. First, what is the role of age of acquisition if children are not exposed to L2 right after birth and are acquiring a tonal L2 that is not the socially dominant language of the community? Children in this study are exposed to a socially minority language (Mandarin) relatively early in life, usually by age 5, when enrolled in kindergarten, but not as early as those children reported in the previous studies. Will this age be early enough for a successful tonal learning experience? Both the speech learning model revised (SLM-r) (Flege and Bohn, 2021) and the critical theory hypothesis (Lenneberg, 1967) stress the importance of early exposure of L2 input in gaining native-like proficiency of sound production. The age advantage has also been robustly demonstrated by large-scale studies of immigrants who arrive at the destination country at varying time points of their life. For example, Flege and colleagues (1995) asked 240 Italians who immigrated to Canada to produce English sentences and then had native speakers rate the degree of foreign accent of these sentence productions. They found a strong correlation between speakers' degree of foreign accent and their age of arrival in Canada.

The key difference between immigrants learning the language spoken in their new community and our child learners is the social status of the language being learned, which in turn affects the quality and quantity of input. Children enrolled in bilingual education programs typically learn a socially minority language that does not have support and resources from the family, community, or other environments outside the school context. Even children with language support at home [i.e., heritage language (HL) learners] experience reduced opportunities to hear and practice the language with peers and in the broader community. The minority status of a language has many implications for learning including but not limited to reduced motivation, compromised input quantity and quality, and the lack of opportunities to practice the language. Thus, the age advantage usually found in the immigrant population may be less apparent in this type of child learner.

Second, how well can children acquire a language in a less organic environment? The specific learning style of bilingual education in this population of children offers more exposure to the minority language than traditional language classes, but children still primarily receive input from teachers and have mixed language interactions with peers. Their use of the minority language is also constrained by the curriculum, unlike the multi-speaker and multi-faceted environments that they were exposed to when acquiring their L1. The special features of this population of bilingual children's learning context provide an interesting case for identifying the specific roles played by age of acquisition, target language's social status, and the complex range of input quantity and quality resulting from different amount of language experience that children receive from home, school, and the society in phonetic acquisition.

A. Mandarin tones

Like many languages in the world, Mandarin utilizes lexical tones—syllable-level shifts in the fundamental frequency (f0) contour—to draw meaningful distinctions between words (Singh and Fu, 2016). An ability to both perceive and produce these tones is integral to learning the language because a change in a tone signifies a change in meaning. Mandarin has four basic tones, typically referred to as T1-T4, which are shown in Fig. 1. These can be described numerically using a 5-level system developed by Chao (1968). This system divides the vocal range of a speaker into levels (1 = lowest pitch, 5 = highest pitch), and uses combinations of digits to describe pitch levels at syllable onset, offset, or other major shape transitions throughout the duration of the syllable. Each tone has a distinct pitch contour: level, rising, dipping, or falling. There are also secondary features that distinguish tones, such as duration (Blicher *et al.*, 1990; Xu, 1997) and vocal register, but the current study will focus on the primary feature of f0:

$$F0 = (\log_{10} x_i - \log_{10} x_{min}) / (\log_{10} x_{max} - \log_{10} x_{min}). \quad (1)$$

B. Timing of Mandarin tone acquisition

Given the important role that tones play in conveying meaning, it was long accepted that native Mandarin speakers acquire tones early, even before consonants and vowels, through previous transcription-based studies (Hao, 2012; Zhu and Dodd, 2000; Li and Thompson, 1977; So and Dodd, 1995). More recently, acoustic analysis has confirmed that tone contrasts emerge early in production. However, even for native Mandarin speakers, it takes time to refine the acoustic realizations of tones (Wong, 2012; Wong 2013; Xu Rattanasone *et al.*, 2018; Yang, 2023). A recent study found that primary tone cues (i.e., f0 contours) could take up to seven years to mature, and that secondary cues (i.e., duration, amplitude, vocal register, etc.) may continue to develop into adulthood (Nari *et al.*, 2021).

Moreover, not all tones are acquired at the same rate. Reportedly, tone acquisition is limited by the maturation of speech motor control (Wong, 2008). Recent studies by Wong and colleagues that extensively investigated tonal acquisition have found that tones appeared to be acquired following the order of T4, T1, T2, T3, from earliest to latest, which presumably reflects an increase in articulatory complexity (Wong, 2012, 2013). Similar developmental patterns were found for both children growing up in tonal language environments such as Taiwan and for Mandarin-speaking children growing up in the United States with less exposure to ambient tonal language environments (Wong *et al.*, 2005). Although there exists some disagreement on the order of tonal acquisition in native speakers (Miracle, 1989; Shen, 1989; Wang, 1995; Wong, 2012), the general consensus is that tones are a challenging aspect for native speakers of English who are learning Mandarin as an L2 (Chen, 1997; Hao, 2012; Shen, 1989; Winke, 2007; Zhang, 2010).

C. Cantonese tones

Cantonese is also a tonal language, but unlike Mandarin, it specifically utilizes different tones depending on syllable composition. Although Cantonese includes several dialects with subtle tonal variations among them, we chose to focus on the Standard Cantonese tones from Hong Kong (Bauer and Benedict, 1997; Mok *et al.*, 2013), which is also a variety of Cantonese that children in this study are mostly frequently exposed to at home. Hong Kong Cantonese includes six tones on syllables that do not end in the plosives /p/, /t/, or /k/. These tones are illustrated in Fig. 2. The abbreviations of C1–C6 for Cantonese tones are used in order to distinguish them from Mandarin tones (T1–T4). Some speakers of Cantonese use a high-falling (52 or 53) tone, which is considered a variation of C1 for some speakers (Hao, 2012; Hashimoto 1972) or another tone altogether (Bauer and Benedict, 1997). Cantonese also includes "stopped" tones, which are variations on the six main tones, for syllables that end in /p/, /t/, or /k/ (Hao, 2012). These stopped tones are not phonologically contrastive with the non-stopped counterparts and are therefore not included in Fig. 2.

According to Wong and colleagues (2017), the order of tonal acquisition for children who acquire Cantonese as their first language is C5, C4, C1, C2, C3, C6, from earliest to latest. By age 2;6, Cantonese-speaking children produce tones with few errors (To *et al.*, 2013; Tse, 1978), indicating that, like in Mandarin, tones are acquired early. However, there is also evidence that children under age 6 do not produce tones as well as adults, specifically that they confuse the level tones (C1, C3, and C6), the rising tones (C2 and C5), as well as C4 and C6 (Barrey and Blamey, 2004). For Cantonese-Mandarin bilingual speakers in Hong Kong who are Cantonese-dominant, research suggests that they acquire

Tone	Pitch	Contour
T1	55	Level
T2	35	Rising
T3	214	Dipping
T4	51	Falling

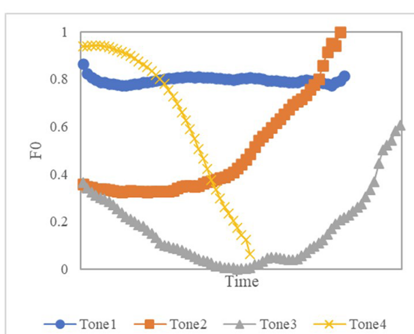

FIG. 1. (Color online) Mandarin tones. F0 contours on the right were produced by a native speaker of Mandarin who was a Canadian graduate student born and raised in Northern China. F0 values were normalized using Eq. (1) (Zhang, 2018).

Tone	Pitch	Contour
C1	55/53	high-level
C2	25	high-rising
C3	33	mid-level
C4	21	mid-low falling
C5	23	mid-low rising
C6	22	mid-low level

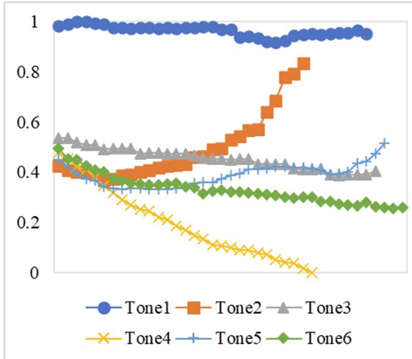

FIG. 2. (Color online) Cantonese tones. f0 contours on the right were produced by a native speaker of Cantonese who was an undergraduate student born and raised in Hong Kong. F0 values were normalized using Eq. (1).

Mandarin tones by age 3 (Law and So, 2006), but adult L2 learners of Mandarin with a Cantonese L1 are not able to identify Mandarin tones significantly better than learners with an English L1 (Hao, 2012). No research to date has examined the L2 acquisition of Mandarin tones by Cantonese speakers in an environment where Mandarin and Cantonese are both minority languages, as in the current study.

D. Transfer effects

Transfer effects are common in bilingual acquisition. Particularly, when there is a high degree of similarity between the two languages due to factors such as language relatedness, etymological alignment, and cognate effects, the acquisition can be made easier or more difficult (Costa *et al.*, 2005; Dijkstra *et al.*, 2010; Meziane and MacLeod, 2020; Wu *et al.*, 2019). With specific regard to second language speech acquisition, an L1 with phonological characteristics similar to the L2 could result in a positive transfer, such as greater tone accuracy for speakers of a tonal L1 compared to those with a dissimilar non-tonal L1. An L1 and L2 with conflicting phonology may also result in a negative transfer effect, such as confusions among perceptually similar tones between languages.

With regard to the effect of a tonal L1 on one's ability to distinguish tones in an L2, previous findings have been mixed. When distinguishing between tones in an unfamiliar language, speakers of a tonal language performed better than English speakers in some cases (Lee *et al.*, 1996; Wayland and Guion, 2004), but equally as well as English speakers in other cases (Francis *et al.*, 2008; Lee *et al.*, 1996; So and Best, 2010). When examining monolingual adult learners, Lee and colleagues (1996) found that Cantonese native speakers were better at correctly discriminating between Mandarin tones than English native speakers, but Mandarin and English native speakers performed similarly when discriminating Cantonese tones. In studies where speakers of tonal and non-tonal languages distinguished between tones of a language they were learning as an L2, speakers of a tonal L1 sometimes performed equally as well as speakers of a non-tonal L1 (Hao, 2012). One study examined the tone identification accuracy of early Hmong-English bilinguals and English monolinguals who were learning Mandarin as an L2 and found that Hmong speakers actually performed worse than English speakers (Wang, 2006). Though they are both tonal languages, there is minimal similarity between Hmong and Mandarin tones, and learners in this study were very early learners of Mandarin. Both of these factors may have contributed to low tone accuracy among speakers of Hmong in this study. Much of the research to date has focused on perception as a means to assess tonal knowledge. However, there is some limited research that directly examines tone production.

E. Transfer effects on L2 tone production

Some studies note specific error patterns in the Mandarin-L2 tone productions of learners with a tonal L1. For example, for speakers of Cantonese, confusions occurred between T1 and T4 and between T2 and T3 (Hao, 2012; So and Best, 2010). This may be due to the similarity of these tones to Cantonese C1 and C2, respectively (Gu, 2016; So and Best, 2010; Zong, 2000). Similarly, Li *et al.* (2017) investigated Cantonese-L1 children's perception and found that they have difficulty identifying Mandarin T2-T3 and T1-T4 pairs. As noted in the introduction, production studies in children are particularly lacking. Among the few studies on children's tonal production, most are transcription-based and examine the impact of a non-tonal language on tonal production [e.g., Mok and Lee (2018) and Yao *et al.* (2020)]. To our knowledge, the only study that addressed the impact of a tonal L1 on L2 tone production was an observational report of Cantonese-L1 children learning Mandarin in Hong Kong (Duan *et al.*, 2022). Although tone errors were reportedly common, they were described in terms of mispronunciations of one Mandarin tone for another, and possible transfer effects from Cantonese to Mandarin tones were not explored.

The coarse-grained description of tonal production from transcription-based and observational studies prevents a detailed and objective evaluation of the influence of tonal and non-tonal languages on tonal productions. Instrumental analysis, such as acoustic analysis of children's speech production, could provide more insights regarding how children gradually develop their tone production skills that cannot be easily categorized as accurate or not (Nari *et al.*, 2021; Wong, 2008; Wong *et al.*, 2017). Ample research on children's speech production has demonstrated the utility and benefits of using acoustic analysis in addition to transcription-based analysis (Li and Munson, 2016; Scobbie *et al.*, 2000; Simon, 2010). Without such objective and fine-grained measurements, the transfer effects of a tonal L2 in children's tone productions are still unclear.

To address gaps in the research about the effects of tonal language background on the acquisition of Mandarin tones, our study examines the tone productions of children with different home language backgrounds. Children from Cantonese- and English-speaking families, who are learning Mandarin as an L2, are compared with children from Mandarin-speaking families who are learning Mandarin as an HL. The children all attend a Mandarin-English bilingual school in an English-dominant community in western Canada. By comparing the Cantonese- and English-speaking students' tone productions with those of children who are Mandarin heritage speakers, we aim to address the following research questions:

(1) Do children with a tonal language background (i.e. Cantonese) produce Mandarin tones more accurately than children with a non-tonal language background (i.e. English)? A higher rate of tone production accuracy from the Cantonese group would indicate an advantage of the tonal background and an associated positive transfer effect.
(2) Are there acoustic patterns that suggest subtle transfer effects that are less perceptually salient?

II. METHODS

A. Participants

The present study is a part of a larger project that examines the speech development of children who are enrolled in bilingual education programs. All participants from this study attended a publicly funded Chinese (Mandarin) Bilingual program in Edmonton, Alberta, Canada. Offered in fifteen schools across the city, this long-standing program provides an opportunity for students to learn in both Mandarin and English from Kindergarten through high school (ECBEA, 2024; Liu, 2020). At the elementary school level (Kindergarten through Grade 6), 50% of instruction is provided in Mandarin and 50% in English. There are no requirements for children to speak any Mandarin prior to admission if they are enrolled in Kindergarten or Grade 1. This subsequently allows for students with a wide range of language backgrounds to attend the program. Some children speak English as a first language (L1) and have no prior experience with Mandarin, some are heritage Mandarin speakers who also speak English outside of their home, and others have a different Chinese language, such as Cantonese, as their L1.

Participants were recruited through voluntary sign-up by their parents. Each parent completed a questionnaire about the language environment of their child. The questionnaire was adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian *et al.*, 2007) and the Alberta Language Environment Questionnaire (ALEQ) (Paradis, 2011). It was made available in both English and Chinese, and in both online and paper versions.

A total of 38 participants were identified from the larger pool of participants, based on their home language backgrounds as reported in the parent questionnaires. Among them, 11 children were from a Mandarin background (6 girls), 11 from a Cantonese background (5 girls), and 16 from an English background (9 girls). All children heard their home language since birth. All participants were in Grade 1, to ensure a focus on new learners. Only children who began learning Mandarin after age 4;0 were included in the Cantonese and English groups, to ensure that they were learning Mandarin after they had already gained some mastery of their native language. This also allowed for the inclusion of children who started learning Mandarin in kindergarten and those who attended a Mandarin preschool. The majority of children (87.5%) entered the bilingual program in kindergarten, thus having at least one year of formal education in Mandarin. All children, regardless of their home language background, received the same mix of Mandarin and English instruction at school.

Children in both the Cantonese and English groups began hearing Mandarin at an average of 4 years 11 months, as compared to the Mandarin group who heard Mandarin from infancy. However, it is important to note that because Edmonton is an English-dominant city, students in the Mandarin and Cantonese groups, who did not receive substantial English input from their caregivers, still grew up in an English-dominant environment outside of the home. Considering this mix of English and a tonal language, these two groups are considered to be HL speakers of Mandarin and Cantonese, respectively. Given their later onset of exposure to Mandarin, both the Cantonese HL and English L1 children are Mandarin L2 learners. For simplicity in reporting, these groups will hereafter be referred to as the English, Cantonese, and Mandarin groups.

Table I provides participant information, including the students' age, the age when they started to hear each language on a regular basis at home, and the current amount of exposure to each language at home. The Cantonese group had greater and earlier exposure to English than the Mandarin group, which can be explained by the history of Chinese immigration in Canada, where a large proportion of the earlier immigrant waves was Cantonese-speaking (Duff and Doherty, 2019). These immigrant families might have incorporated English into their family environment to a larger extent than more recent Mandarin-speaking immigrants. In addition to age of exposure to each language, parents were also asked to rank order their child's

TABLE I. Participant information.

Group	English (English L1, Mandarin L2)	Cantonese (Cantonese HL, Mandarin L2)	Mandarin (Mandarin HL)
Age in months Mean (SD)	78.44 (3.74)	79.64 (3.44)	77.73 (2.87)
English age of onset in months Mean (SD)	0.19 (0.40)	0.60 (14.45)	32.09 (24.76)
Cantonese age of onset in months Mean (SD)	N/A	0	N/A
Mandarin age of onset in months Mean (SD)	59.45 (4.23)	59.45 (4.20)	1.64 (5.43)
Current percentage of exposure to English at home[a] Mean (SD)	94.72 (11.72)	52.27 (30.96)	14.77 (20.42)
Current percentage of exposure to Cantonese at home[a] Mean (SD)	0	45.23 (29.43)	0
Current percentage of exposure to Mandarin at home[a] Mean (SD)	0.59 (1.36)	2.5 (4.98)	85.22 (20.42)
Current percentage of children reported by parents as English-dominant[b]	100	54.55	81.82

[a]Percentage of time that the language is spoken by the two primary caregivers.
[b]Note that language dominance does not always align with the child's first acquired language (L1).

dominance in each language at the time of the study. As shown in Table I, many parents listed English as their child's current dominant language, indicating that, in spite of significant exposure to the HL at home, these children have also been strongly influenced by the language of the broader community.

B. Procedures

As part of the larger study, a single-word elicitation task was adapted from Zhao and Bernhardt (2012) and Zhu (2002). It contained 72 Mandarin words and included at least five examples of each tone and at least one example of each disyllabic tone series. Target words were accompanied by pictures. Productions were elicited by questions such as "what is this?" and "what is this person doing?" If the child did not produce a word spontaneously, the examiner used a hierarchy of prompts that included (1) repeating the question, (2) providing the first sound, and (3) whispering the first syllable. If the child produced the target word after any of these prompts, it was regarded as spontaneous. If not, the examiner provided an imitative model, and the child's production was regarded as imitative.

Examiners were native speakers of Mandarin. The assessments were conducted in a quiet room at school. Speech samples were recorded using a Zoom H1n digital recorder with a Pro Lavalier JK MIC-J 055 unidirectional cardioid condenser microphone clipped at the front of the child's shirt.

The speech samples were processed in PHON (Hedlund and Rose, 2020) by a team of four native Mandarin-speakers (hereafter, transcribers). Transcribers segmented the recordings and coded whether each word was produced spontaneously or imitatively. Tones were transcribed in the 5-level numeric convention (Chao, 1968): T1 [55], T2 [35], T3 [214], T3 (semi-sandhi) [21], T4 [51], or uncategorizable [7]. The semi-sandhi [21] is a positional allophone of T3 and should not occur in single-syllable production (Xu Rattanasone *et al.*, 2018). The transcribers held three sessions to practice consensus transcription and weekly meetings to recalibrate their transcription standards. Among the thirty-eight selected participants, 11 (29%) were transcribed by two transcribers and compared for consistency. The inter-transcriber reliability of tone judgements was 90%.

For the present study, eight monosyllabic words with a high rate of spontaneity (between 86% and 98%) were selected from the word elicitation task to represent the four lexical tones in Mandarin, two examples for each tone. See Table II for the list of selected words. Though children were not asked to produce any Cantonese words in this study, the Cantonese translations of the Mandarin target words are included in Table II to show the relative similarity in the phonetic realization of many of these words. The similarity of segments and/or tones in some of the target words reflects the shared historical origins of Mandarin and Cantonese.

Table II. Also includes the transcription of each word in both Pinyin and Jyutping. These are romanization systems for capturing the describing phonetic realization in Mandarin and Cantonese, respectively. Pinyin is the most widely accepted method of Mandarin romanization, whereas Jyutping is one of several methods of Cantonese romanization and was developed by the Linguistic Society of Hong Kong (2024).

Each sound file was exported from PHON to PRAAT (Boersma and Weenink, 2024) for acoustic analysis. Two researchers manually marked TextGrid boundaries at the start of the vowel and the end of the voicing pulses, compared their decisions for consistency, and reached a consensus. Initial consonants were not included in the analysis as they do not contain adequate tone information (Cheng, 1966; Kratochvil and Chao, 1970; Wong, 2012), while nasal

TABLE II. Target words for the four lexical tones in Mandarin.

Tone	Character	Mandarin Pinyin	Mandarin IPA	Cantonese Jyutping	Cantonese IPA	Meaning
T1	三	sān	[san^{55}]	saam1	[saːm̚^{55}]	three
T1	八	bā	[pa^{55}]	baat3	[baːt̚^{33}]	eight
T2	蓝	lán	[lan^{35}]	laam4	[laːm^{21}]	blue
T2	鱼	yú	[y^{35}]	jyu2	[jy^{25}]	fish
T3	五	wǔ	[u^{214}]	ng5	[ŋ̩^{25}]	five
T3	狗	gǒu	[ko͡ʊ^{214}]	gau2	[kɐ͡u^{25}]	dog
T4	二	èr	[ɐ˞^{51}]	ji6	[ji^{22}]	two
T4	绿	lǜ	[ly^{51}]	luk6	[lʊk̚^{22}]	green

codas were included because they form part of the rime (Cheng, 1966; Howie, 1974).

The f0 was extracted from the TextGrid boundaries of the speech samples using a script (Lin, 2020), which divided each sample into ten equal time intervals and reported a mean f0 in hertz (Hz) for each interval. Data were visually inspected for improbable pitch curves and missing information. Extremely low or irregular f0 values, such as those found in creaky voice (Keating *et al.*, 2015), cannot be detected in PRAAT, so these values were calculated manually or marked with a floor value of 65 Hz. The floor value was chosen because it was below both the speaking range of Cantonese-English bilingual children (Ng *et al.*, 2010) and the lowest values that were accurately detected in PRAAT. The erroneous f0 values above 65 Hz were recalculated manually, by measuring cycles per second within the section of time that corresponded to the data point in question.

To investigate whether children's Mandarin productions were influenced by Cantonese tones, two native Cantonese speakers listened to each incorrect Mandarin tone production to determine whether it could be categorized as a Cantonese tone. Each production was transcribed to be either a Cantonese tone (C1–6) or not a Cantonese tone (X for uncategorizable). Both transcribers were blind to the group (Cantonese, Mandarin, or English) in which each participant belonged. Because of the small number of productions, transcriptions were discussed by the transcribers and any disagreements resolved through consensus.

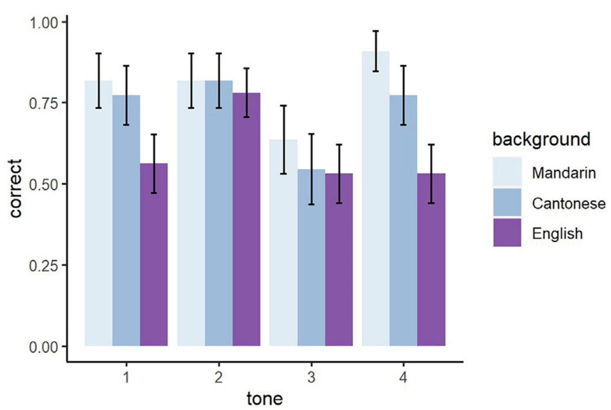

FIG. 3. (Color online) Percentage of correct Mandarin tone productions in Grade 1 students with three language backgrounds.

III. RESULTS

A. Accuracy of tone production

Figure 3 summarizes the percentage of Mandarin tones produced correctly for each group and tone. The tones in order from highest to lowest accuracy were as follows: T2 > T4 > T1 > T3.

A mixed effects logistic regression was conducted using R to analyse the differences in accuracy between language backgrounds (see Fig. 3 and Table III). Within this model, the dependent variable was accuracy, independent fixed effect variables were language background and tone, and individual participant was included as a random intercept. In addition, a by-participant random slope was included that varies with accuracy to further account for repeated measures. Model comparisons based on AICs indicated no significant contribution of the interaction between background and tone, although the inspection of raw data across backgrounds and tones seems to suggest lower accuracy rates for T1 and T4 among the English group compared to the other two groups. The Mandarin group was used as the reference level, to which the Cantonese and English groups were compared. The English group had a significantly lower percentage of tones correct (61.30%) than the Mandarin HL speakers (81.71%) ($p = 0.006$). Though the Cantonese group had lower accuracy (74.82%) than the Mandarin group on average, this difference was not significant. A pairwise comparison using the emmeans package indicated no significant difference between the Cantonese and English groups ($p = 0.186$), although there was a large difference in their mean estimated accuracy rates. The overall accuracy of T3

TABLE III. Results of linear mixed effects logistic regression.

Independent variable	Estimate	Standard error	Z value	P value
Background (Mandarin as a reference)				
English	−1.04	0.38	−2.73	0.006[a]
Cantonese	−0.41	0.41	−0.98	0.323
Tone (T1 as a reference)				
T2	0.62	0.41	1.50	0.133
T3	−0.63	0.37	−1.69	0.090
T4	0.07	0.38	0.19	0.847

[a]$p < 0.05$.

Error Tone	T1			T2			T3			T4			X		
Group[a]	E	C	M	E	C	M	E	C	M	E	C	M	E	C	M
Target Tone T1	-	-	-	2/14 14.3%	0/5 0.0%	0/4 0.0%	2/14 14.3%	0/5 0.0%	0/4 0.0%	6/14 42.9%	**4/5 80.0%**	0/4 0.0%	4/14 28.6%	1/5 20.0%	**4/4 100.0%**
Target Tone T2	0/7 0.0%	0/4 0.0%	0/4 0.0%	-	-	-	**5/7 71.4%**	1/4 25.0%	**2/4 50.0%**	1/7 14.3%	0/4 0.0%	0/4 0.0%	1/7 14.3%	**3/4 75.0%**	**2/4 50.0%**
Target Tone T3	2/15 13.3%	0/10 0.0%	2/8 25.0%	**9/15 60.0%**	4/10 40.0%	0/8 0.0%	1/15 * 6.7%	1/10 * 10.0%	**4/8* 50.0%**	0/15 0.0%	0/10 0.0%	1/8 12.5%	3/15 20.0%	**5/10 50.0%**	1/8 12.5%
Target Tone T4	4/15 26.7%	1/5 20.0%	0/2 0.0%	5/15 33.3%	0/5 0.0%	0/2 0.0%	3/15 20.0%	0/5 0.0%	2/2 100.0%	-	-	-	3/15 20.0%	**4/5 80.0%**	0/2 0.0%

[a] E = English group; C = Cantonese group; M = Mandarin group; X = not categorizable as a Mandarin tone
* Realized as semi-sandhi

FIG. 4. Total number of Mandarin substitutions and uncategorizable errors in each group, as coded by Mandarin judges. For each group and target tone [e.g., English, T1], data is presented in fraction (error type / total errors) and percentage of total errors.

was lower than other tones, with a result trending toward significance ($p = 0.09$). A pairwise comparison indicated a significantly lower accuracy rate of T3 (59.11%) than T2 (83.44%) ($p = 0.012$) but not than the other tones, although the mean estimated accuracy rate of T3 also appeared much lower than those of T1 (73.12%) and T4 (74.55%).

B. Error analysis

During initial transcription, the actual contour of incorrect productions was coded as T1–T4. T3 [214] that was produced as [21] (semi-sandhi) was coded as incorrect, as semi-sandhi only occurs when T3 is followed by another tone (Xu Rattanasone *et al.*, 2018). The rates of specific Mandarin substitutions and uncategorizable errors across groups are shown in Fig. 4.

In Fig. 4, several outstanding substitution patterns (≥50% out of four or more errors) were bolded. First, T1 errors produced by the Cantonese group were often transcribed as T4 (80% out of the 5 errors). Second, T2 errors produced by English and Mandarin groups were often transcribed as T3 (71.4% out of 7 errors and 50% out of 4 errors, respectively), and T3 errors produced by the English group were often transcribed as T2 (60% out of 15 errors). This bidirectional T2-T3 confusion also occurred in the Cantonese group, although the percentages of such error patterns were lower than the other two groups. Finally, T3 errors produced by the Mandarin group were often transcribed as a semi-sandhi [21], which is an allophonic variation of T3 in di- or multisyllabic productions that should not apply in monosyllabic productions (Xu Rattanasone *et al.*, 2018).

In addition to these substitution patterns, there were a number of uncategorizable productions in both the Mandarin group (T1 and T2 targets) and the Cantonese group (T2, T3, and T4 targets). The higher rates of uncategorizable productions, particularly in the Cantonese group, could be underlied by Cantonese tone substitutions which were not captured by the Mandarin transcribers. Therefore, all error productions were further examined for possible Cantonese tone substitutions. The Cantonese group had no more Cantonese tone substitutions (20.83%) than the Mandarin (22.22%) or English (25.49%) group (see Table IV and Fig. 5). Visual analysis of specific tone substitutions in the Cantonese group did not reveal any distinct error patterns.

C. Acoustic analysis of correct tone productions

The f0 contour of correct tone productions was also analysed to determine whether different language backgrounds resulted in noticeably different patterns of correct tone production. To analyse the shape of the correct tones,

TABLE IV. Number and percent of Cantonese tone substitutions as a proportion of total errors.

Tone	Mandarin group	Cantonese group	English group
T1	0/4	0/5	3/14
	0.0%	0.0%	21.4%
T2	2/4	0/4	0/7
	50.0%	0.0%	0.0%
T3	2/8	4/10	7/15
	25.0%	40.0%	46.7%
T4	0/2	1/5	3/15
	0.0%	20.0%	20.0%
Total	4/18	5/24	13/51
	22.2%	20.8%	25.5%

https://doi.org/10.1121/10.0024985

		Errors Coded as Cantonese Tones																		Uncategorizable Errors		
Error Type		C1			C2			C3			C4			C5			C6					
Group		E	C	M	E	C	M	E	C	M	E	C	M	E	C	M	E	C	M	E	C	M
Target Words	sān	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	2	0	4
	bā	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
	lán	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	yú	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
	wǔ	2	0	0	0	1	0	0	0	0	1	0	1	4	2	0	0	0	0	1	3	2
	gǒu	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	2	1
	èr	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0
	lǜ	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0

FIG. 5. Cantonese tone substitutions and uncategorizable errors for all groups by word token.

the data were modeled using generalized additive mixed models (GAMMs). A GAMM model generates flexible non-linear regression curves called "smooths," so unlike a typical linear regression, it accounts for dynamic variation within the data over time (Soskúthy, 2021). This is particularly useful for the analysis of tones, where differences between groups can vary at multiple points within a contour. Like other mixed modeling methods, it also recognizes that interactions between variables and random factors can both account for variation within the data.

Correct tone productions were analysed with a GAMM model using an R package itsadug (van Rij *et al.*, 2022) and an R function gamm_hacks (Soskúthy, 2021), including the fixed effects of background and tone and the random effect of word. The f0 values of each contour were normalized based on the maximum and minimum values of each participant's productions using Eq. (1), in order to better control for differences between speakers. By comparing models using the compareML function, a model with the interaction effect between tone and background was selected due to significantly lower Akaike information criterion (AIC) compared to the model with only the fixed effect of tone (AIC difference = 100.56) and the model with only the fixed effect of background (AIC difference = 92.89). The selected interactive model was corrected for autoregression and explained 34.3% of variance in the data (R2 = 0.325, fREML = –1432.8). The random effect of word was insignificant (F = 0.069, p = 0.207). The full statistics of the model are presented in Appendix A.

The results from this analysis are presented in Fig. 6. There are several methods of significance testing when using GAMMs, including visual analysis of the amount of overlap between 95% confidence intervals (Soskúthy, 2021). This type of analysis is useful for capturing the difference between groups at specific points along a contour. In the present dataset, a significant difference is visible in the contour of T3 within the Mandarin group, indicating that the Cantonese and English groups did not dip as low as the Mandarin group in their productions, even though they were still perceived as correct. Group comparisons using the plot_diff function indicated that the significant difference between the English and Mandarin groups was present from normalized time point 4 to 10 (67% of the full contour), and the significant difference between the Cantonese and Mandarin groups was present from time point 3.18 to 10 (76% of the full contour), which both involved the valley and rising sections of T3. No significant difference was detected between English and Cantonese groups' T3 productions. Correct T1, T2, and T4 were all produced in a relatively consistent manner regardless of language background although significant differences with low amplitudes were incidentally detected. For example, there is a significant difference between English and Mandarin groups' T1 contours across 88% of the full contour, where the English group had a slightly higher F0 contour overall. These differences, although statistically significant, are not likely to be linguistically significant compared to the differences in T3.

IV. DISCUSSION

A. Tone accuracy

The first aim of this study was to determine whether children with a tonal language background (e.g., Cantonese) produce Mandarin tones more accurately than children with a non-tonal L1 (e.g., English), in comparison to their peers who speak Mandarin as a HL. Similar to previous research on the accuracy of Mandarin tone discrimination in adult Cantonese speakers (Lee *et al.*, 1996) the results of this study found that children who spoke Cantonese as a HL had higher average tone accuracy than children with an English background. Despite the differences in the tonal systems of Cantonese and Mandarin, previous knowledge of a tonal language appeared to aid children in the Cantonese group to produce Mandarin tones at a similar level of accuracy to their peers that speak Mandarin as a HL.

This advantage can be explained from two perspectives: perceptual sensitivity to phonetic features and categorical assimilation between L1 and L2. The high tone accuracy within the Cantonese group also reinforces the concept of category precision in SLM-r (Flege and Bohn, 2021). SLM-r suggests that speakers who have a refined knowledge of the sound categories in their L1 can more easily notice

JASA
https://doi.org/10.1121/10.0024985

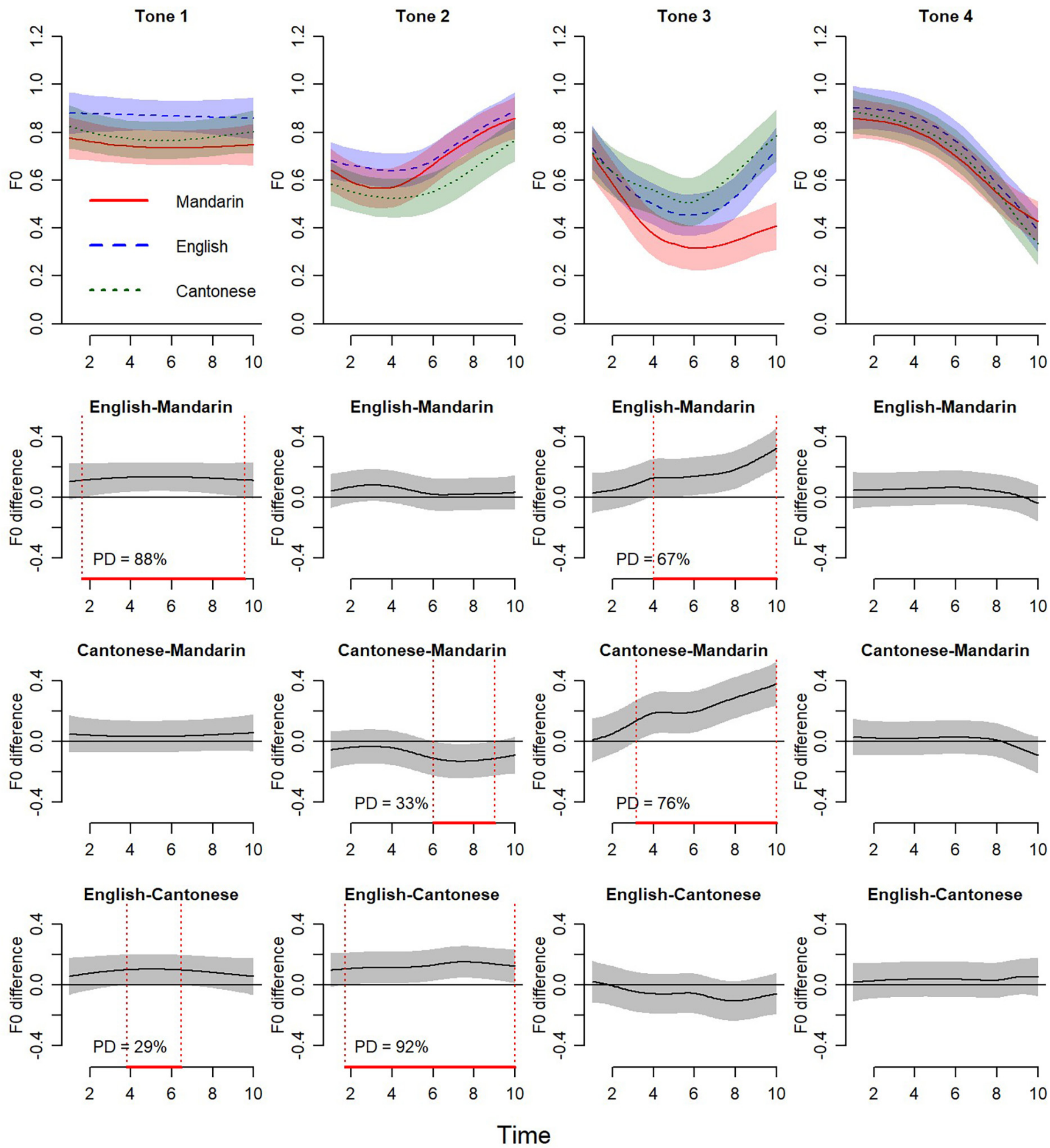

FIG. 6. (Color online) GAMM estimated average correct tone contours by language background and tone target, accompanied by difference plots between contours. Shading around each line indicates a 95% confidence interval. Each tone target is presented in a column. The top row includes the GAMM estimated contours by three language backgrounds, Mandarin group in red, English group in blue, and Cantonese group in green. The following three rows present the difference contours between background groups, respectively. Sections of significant difference are marked with red vertical dashed lines. The amplitudes of differences can be observed from the plots, and the proportions of difference (PD) over the *x* axis of time are indicated in the plots where applicable.

distinct features of L2 sounds than those with unrefined L1 sound categories. When extended to tone category refinement, this perspective also explains the high tone accuracy of the Cantonese group. Cantonese HL speakers have refined L1 tone categories, while native English speakers have no tone categories. The L1 tone categories allow Cantonese speakers to develop tone categories more quickly than English speakers, even considering these speakers have had the same quantity and quality of Mandarin L2 exposure.

Previous research also indicates that Cantonese speakers perform better than English speakers at perceiving tone contrasts in Mandarin (Lee *et al.*, 1996), suggesting that L1 tone system complexity plays a role. The Cantonese tone system is more complex than that of Mandarin, which gives Cantonese speakers an advantage when discriminating between Mandarin tones, but not Mandarin speakers discriminating between Cantonese tones. This perspective also explains the high tone production accuracy within the Cantonese group of the current study, as children who speak Cantonese have a complex L1 tone system, which makes the simpler L2 system relatively easier to acquire.

There were also patterns approaching significance that emerged in the accuracy of T3. Children were more likely to produce T3 incorrectly regardless of language background. Previous research indicates that T3 takes the longest to acquire among native speakers (Wong, 2012, 2013), and the current results indicate that school-age L2 learners also have challenges with this tone. The possible reasons for these findings are discussed in more detail below.

B. Influence of Cantonese in tone errors

The second aim of this study was to investigate the specific transfer effects of Cantonese when learning Mandarin as a second language. All tone productions were first examined by Mandarin transcribers. The only outstanding substitution pattern for the Cantonese group was T4 for T1. This could be explained by the free variations in Cantonese, C1 [55/53], that is, the Cantonese group assimilated both Mandarin T1 [55] and T4 [51] to C1 (Hao, 2012; So and Best, 2010). In contrast, in Mandarin, a restrained pitch shift is an important acoustic feature for T1 as a high-level tone (Wong, 2012). Therefore, the pattern of T4 substitution for T1 in the Cantonese group could be related to the Cantonese tone system. However, there were only five T1 errors produced by the Cantonese group in total. Thus, the underlying mechanism of the pattern remains inconclusive.

Moreover, pattern analyses revealed bidirectional confusion between T2 and T3. The same pattern was documented among adult L2 learners (Hao, 2012) and explained by the phonetic similarities between T2 and T3, both having a dipping section and a rising section in their pitch contours [see Fig. 1 and So and Best (2010)]. Meanwhile, the occurrence rates of this pattern were higher in the Mandarin and English groups than the Cantonese group. Therefore, it could be explained by universal phonetic similarities as opposed to Cantonese influence.

Furthermore, students in the Mandarin group produced a contextually inappropriate variant of T3 (semi-sandhi) in monosyllabic words. This pattern was not as pronounced in the Cantonese or English group, so it did not suggest a particular influence of Cantonese. Low pitch is the key feature of T3 and is usually more challenging to articulate (Wong, 2012; Xu, 1997). Thus, the Mandarin group's production of a low falling semi-tone may signal an intermediate level of Mandarin speech development.

The transcriptions revealed a large number of uncategorizable productions in the Cantonese group, which indicated that the error productions warranted further examination by Cantonese transcribers for potential substitutions of Cantonese tones. The results of the error analysis in this study also found that within the productions judged as incorrect, there were not many Cantonese tone substitutions within the Cantonese group as judged by native Cantonese listeners. As a higher percentage of Cantonese-like errors were made in the Mandarin and English groups, merely by chance (refer to Table IV), Cantonese-like tone errors in the Cantonese group cannot be attributed to language background. Thus, there appears to be no negative transfer from a Cantonese language background in the acquisition of Mandarin tones.

Previous studies found that adult native speakers of Cantonese had specific tonal confusions when learning Mandarin as an L2 (Hao, 2012; So and Best, 2010). This was found in very few T4 for T1 productions by the Cantonese group, but no further Cantonese substitution patterns were identified. The SLM-r suggests that L1 tone substitution is a characteristic of early language learners (Flege and Bohn, 2021). The lack of Cantonese tone substitutions in the current study suggests that the children in the Cantonese group have received enough quality Mandarin exposure to develop beyond the early L2 learning stage.

C. Covert differences in correct tone contours

Through a visual analysis of the GAMMs, the present study found that correct T1, T2, and T4 productions were all produced the same regardless of language background. In contrast, the T3 productions judged as correct by Mandarin speakers did differ depending on language background. Specifically, the English and Cantonese groups exhibited a shallower dip in their correct T3 productions compared to the Mandarin group. This indicates that though they are still judged as correct, children with a Cantonese tonal background do not have T3 productions that are as refined as those of children with a Mandarin background. This is consistent with findings that very young Mandarin speakers also produce a shallower dip in their correct T3 productions than adult speakers (Wong, 2012).

The cause for this finding in the T3 dip and for the low T3 accuracy across groups reported earlier could be due to the unique features of this tone. For example, Cantonese and Mandarin share tones that have rising, falling, and flat intonation, but the dipping tone is unique to Mandarin. Similarly, though English does not have tones, it does contain rising and falling intonation within prosodic structures but does not have a distinct dipping intonation. The presence of rising and falling contours in speakers' language backgrounds could result in positive transfer on T2 and T4 that does not occur in T3. In addition, the challenge may also be exacerbated by the two-part structure of T3. The coordinated motor movements that are required for falling then rising may make it more challenging than tones with a single movement (Wong, 2012). Furthermore, research suggests that the low, non-rising section of F0 contour is a distinctive feature of T3 (Blicher *et al.*, 1990) and there has been a trend to emphasize the low feature among native speakers' productions in the past century (Yang, 2023). The interplay between the novel dipping shape and its phonetic complexity could result in low accuracy and naive productions.

Another reason for the shallow dip in T3 and the relatively high rate of T2 substitution may result from tone sandhi, which creates a complex relationship between T2 and T3. The full sandhi rule in Mandarin states that when two T3 words occur in a row, the first is produced as a T2 (Xu Rattanasone, 2018). As suggested by Lee-Kim (2021), new learners may be confused by this relationship, resulting in T3 productions that resemble T2, and are therefore flatter than adult native speaker T3 productions. However, with more exposure to Mandarin, this error pattern may diminish. Further research may concentrate on older students to investigate how T3 knowledge matures over time in a bilingual school program.

D. Empirical and theoretical contributions

Our study presents a special population of children with varying home language backgrounds who are enrolled in a Mandarin-English bilingual school to learn a minority tonal language in Canada. To our knowledge, such a population has not been examined before. Previous empirical studies on bilingual tonal production in children are very limited and mostly focused either on younger children or on children who acquire a non-tonal language along with a tonal language in a home or preschool setting (Holm and Dodd, 2006; Mok and Lee, 2018; Yao *et al.*, 2020). Our study is the first of this kind comparing the effect of a tonal versus a non-tonal home language when learning a tonal target language in a bilingual education setting.

This study also makes significant contributions to theories of bilingual speech production by adding to the sparse literature of child bilingual learners of a tonal language. First, the fact that Grade 1 children who learn Mandarin as a L2 are not completely accurate in their tone productions suggests that age 5 is not early enough and/or the less organic Mandarin environment of a school setting compromises the age effect that we see clearly in adult speakers who learn a socially majority L2. Second, the transfer effect that has been identified by the current study supports one of the major tenets of SLM, which stipulates that both languages share a common acoustic space for phonetic learning. The confirmed positive transfer effect in our study also corroborates the PAM model on its extension to the prosodic domain.

E. Limitations

Though the present study contributes to the understanding of L1, HL, and L2 interactions and has specific implications for L2 acquisition of tones, some limitations must be noted. First, the sample size was relatively small, given the size of the school population from which participants were recruited and the strict criteria for inclusion in each of the three groups. In addition, information about language background was drawn from parent reports, so there could be inaccuracies or missing information. In the Cantonese group, many reports stated that children had minimal exposure to Mandarin prior to admittance in the bilingual program. However, it is difficult to truly isolate Cantonese within the multicultural environment in Canada. For instance, some parents of the children in the Cantonese group may have had previous education in Mandarin, watch television shows in Mandarin, or participate in community events that include both Cantonese and Mandarin speakers, so Cantonese-speaking children could have had some passive exposure to Mandarin before starting school.

Another limitation exists in the number and type of target words chosen for analysis. Words that were chosen were monosyllabic, and most children could produce them spontaneously. However, many of these words were also phonetically similar in Mandarin and Cantonese (Table II). This similarity, including the presence of cognates or near cognates, could have facilitated specific positive transfer or more broadly contributed to the Cantonese group's learning proficiency. The Cantonese group may have lower tone accuracy when producing words that have distinct phonetic structures in Mandarin and Cantonese. The simple (CV(n)) monosyllabic shape of the words may also have contributed to the high accuracy rate. Future research may seek to examine children's performance on disyllabic words and/or to investigate the impact of a controlled set of cognates versus non-cognates on tone accuracy.

A further limitation arose in the error analysis due to the lack of context about pitch. Cantonese C1, C3, and C5 have the same shape but different height, but the judges listened to single word productions. There may be cases where the shape was similar to the Mandarin target T1, but the height of the production was unknown due to missing information about the speaker's relative pitch. Because judges had limited context, it is possible that some perceptual judgements could be inaccurate.

V. CONCLUSIONS

In summary, our research found the following:

(1) The Cantonese group had a comparable percentage of accurate tone productions to the Mandarin group, suggesting positive transfer from Cantonese children's tonal language background.
(2) Within incorrect productions, the Cantonese group did not make a larger number of Cantonese tonal errors than other groups. This indicates that there was no identifiable negative transfer from a tonal language background in the error patterns of school-aged children.
(3) The Cantonese and English groups had a shallower dip in their correct T3 productions than the Mandarin group. This indicates that though these tones were perceived as correct, the T3 productions of the Cantonese and English groups still need further refinement to match those of the Mandarin group.

Further directions for study include the addition of disyllabic words to determine the accuracy of words with tone interactions such as tone sandhi or semi-sandhi. Words that have distinct Mandarin and Cantonese realizations could also be examined, to better understand the influence of a tonal language background on the realization of unique L2 words. The tone productions of children from older grades could also be examined, to see whether tone accuracy improves and how correct tones are refined after more extensive exposure to Mandarin in the school setting. Alternatively, the productions of all three groups could be compared to adult speakers of Mandarin, such as teachers, to examine the differences in accuracy and correct tone contours that exist between school-age heritage speakers of Mandarin and adult Mandarin speakers. These questions could provide useful information about tone development on a variety of levels.

In sum, the results of this study suggest that a tonal language background such as Cantonese has a positive influence on Mandarin tone acquisition. As there is limited research in the area of tone acquisition in bilingual schools, further studies may seek to examine the effect of language background as children grow older. It is possible that the gap in tone accuracy between children of different language backgrounds will diminish as children gain more experience with Mandarin.

Through an examination of the learning process of tones by L2 Mandarin speakers, the current study may also inform the expectations and approach of teachers in bilingual schools. Children who do not have prior knowledge of a tonal language may need more exposure to tones before they achieve mastery in this area, or even explicit teaching about tones. Teachers may also expect that children without previous tonal language exposure will naturally take longer to produce tones accurately. This knowledge will support the communication expectations and language learning of students enrolled in bilingual programs.

ACKNOWLEDGMENTS

This paper draws on research supported by the Social Sciences and Humanities Research Council (SSHRC) under the Insight Grant No. [435–2017-1086]. Special thanks to Minjia Tao and Binqing Yu for their work on data collection, to Xiaozhu Chen, Nan Xing, and Lujia Yang for their work on transcription, as well as the children, teachers, and principals in the participating bilingual schools. F.L. and K.P. are co-senior authors.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

This research was approved by the Research Ethics Board at the University of Alberta (Pro00075638). Informed written consent was obtained from the parents of all participants, and children provided verbal or written assent.

DATA AVAILABILITY

Data available on request from the authors.

APPENDIX A:

Model statistics of the GAMM with tone and background interaction are shown in Table V. Each contour of interaction effect is indexed in the model in order to make direct comparisons between contours. Parametric (linear) terms and smooth (nonlinear) terms are presented.

TABLE V. Model statistics of the GAMM with tone and background interaction.

Model statistics: $R^2 = 0.325$, deviance explained = 34.3%, fREML = −1432.8, n = 2110

	Parametric coefficients			Smooth terms		
	estimate	t	p	est. df	F	p
Tone1.Mandarin	0.746486	22.47897	0.000[a]	1.763007	0.719311	0.552
Tone2.Mandarin	−0.05873	−1.2407	0.215	4.229307	6.284018	0.000[a]
Tone3.Mandarin	−0.3248	−6.4352	0.000[a]	4.393974	10.09298	0.000[a]
Tone4.Mandarin	−0.0634	−1.37349	0.170	3.586689	14.77357	0.000[a]
Tone1.Cantonese	0.038954	0.841623	0.400	2.181968	1.12604	0.386
Tone2.Cantonese	−0.14318	−3.02763	0.002[b]	3.451362	5.651087	0.000[a]
Tone3.Cantonese	−0.12207	−2.3139	0.021[c]	4.585498	5.520122	0.000[a]
Tone4.Cantonese	−0.06195	−1.29193	0.197	3.535856	21.17918	0.000[a]
Tone1.English	0.122467	2.723574	0.007[b]	1.000042	0.134133	0.714
Tone2.English	−0.01823	−0.41362	0.679	4.012958	6.755235	0.000[a]
Tone3.English	−0.17597	−3.6565	0.000[a]	4.905367	9.978711	0.000[a]
Tone4.English	−0.025	−0.52124	0.602	3.798771	17.86734	0.000[a]
Word (random effect)				4.267159	0.068538	0.207

[a]$p < 0.001$.
[b]$p < 0.01$.
[c]$p < 0.05$.

https://doi.org/10.1121/10.0024985

TABLE VI. Comparison of current study with previous studies on bilingual children's tonal productions.

Study	Target language (status in community)	Participants	Onset of exposure to target language	Dominant language at time of study	Current language exposure at home	Societal language	Acquisition style (for target language)
The present study	Mandarin (minority)	38 children (6 yrs of age) attending Mandarin- English bilingual school in Canada Three subgroups with home language background in Mandarin, Cantonese, or English	Mandarin subgroup: < 1 yr Cantonese subgroup: > 4 yrs English subgroup: > 4 yrs	Mostly English dominant Mandarin subgroup: 82% Cantonese subgroup: 55% English subgroup: 100%	Mandarin subgroup: Predominantly Mandarin (tonal) Cantonese subgroup: Cantonese (tonal) and English (non-tonal) English subgroup: English (non-tonal)	English (non-tonal)	Mandarin subgroup: home and bilingual school Cantonese subgroup: mostly bilingual school English subgroup: bilingual school only
Duan *et al*. (2022)	Mandarin (minority)	9 Cantonese-L1 children (4–12 yrs of age) learning Mandarin at school in Hong Kong	Not reported	Cantonese (also proficient in English)	Not reported	Cantonese (tonal) and English (non-tonal)	Mandarin language classes in general education program
Holm and Dodd (2006)	Cantonese (minority)	40 Cantonese-L1/ English-L2 children (2;2–5;7) in Australia; plus 16 children of similar age and background from Dodd *et al*. (1996)	Since attending daycare	Cantonese	Cantonese (tonal)	English (non-tonal)	Home language exposure since birth
Mok and Lee (2018)	Cantonese (majority)	5 Simultaneous Cantonese-English bilingual children at 2;0 and 2;6 (longitudinal) in Hong Kong Compared with age-matched monolingual Cantonese children (3 longitudinal, 10 children cross-sectional) in Hong Kong	Since birth	4 Cantonese dominant and 1 English dominant	Mothers speaking Cantonese (tonal) and fathers speaking English (non-tonal)	Cantonese (tonal) and English (non-tonal)	Bilingual home exposure (one parent – one language model)
Yao *et al*. (2020)	Cantonese (majority)	21 Urdu-L1/Cantonese-L2 children (4–6 yrs of age) attending mainstream Cantonese preschool in Hong Kong Compared to 20 Cantonese-L1 children (4–6 yrs of age) in Hong Kong	Since preschool	Urdu	Urdu (non-tonal)	Cantonese (tonal)	Cantonese preschool

APPENDIX B:

Comparison of current study with previous studies on bilingual children's tonal productions is shown in Table VI.

Barrey, J. G., and Blamey, P. J. (**2004**). "The acoustic analysis of tone differentiation as a means of assessing tone production in speakers of Cantonese," J. Acoust. Soc. Am. **116**(3), 1739–1748.

Bauer, R., and Benedict, P. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception," in *Language Experience in Second Language Teaching*, edited by O. Bohn and M. J. Munro (John Benjamins, Amsterdam).

Blicher, D., Diehl, R., and Cohen, L. (**1990**). "Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement," J. Phon. **18**(1), 37–49.

Boersma, P., and Weenink, D. (**2024**). "Praat: Doing phonetics by computer" (version 6.4.05) [computer program], http://www.praat.org/ (Last viewed January 27, 2024).

Chao, Y. R. (**1968**). *A Grammar of Spoken Chinese* (University of California Press, Berkley).

Chen, Q. (**1997**). "Toward a sequential approach for tonal error analysis," J. Chin. Lang. Teachers Assoc. **32**(1), 21–39.

Cheng, R. (**1966**). "Mandarin phonological structure," J. Ling. *2*(2), 135–158.

Costa, A., Santesteban, M., and Cano, A. (**2005**). "On the facilitatory effects of cognates in bilingual speech production," Brain Language **94**, 94–103.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (**2010**). "How cross-language similarity and task demands affect cognate recognition," J. Mem. Lang. **62**(3), 284–301.

Dodd, B., So, L., and Li, W. (**1996**). "Symptoms of disorder without impairment: The written and spoken errors of bilinguals," in *Evaluating Theories of Language*, edited by B. Dodd, R. Campbell, and L. Worrall (Whurr, London), pp. 119–139.

Duan, R., Chen, J., and Zhang, Y. (**2022**). "Errors in Putonghua acquisition and teaching strategies for Hong Kong children," Mod. Ling. **10**(8), 1772–1777 [In Chinese].

Duff, P., and Doherty, L. (**2019**). "Learning 'Chinese' as a heritage language: Challenges, issues and ways forward," in *The Routledge Handbook of Chinese Applied Linguistics*, edited by C.-R. Huang, Z. Jing-Schmidt, and B. Meisterernst (Routledge, New York), pp. 149–164.

Edmonton Chinese Bilingual Education Association (ECBEA) (**2024**). "About the program," https://www.ecbea.org/about-the-program/ (Last viewed June 17, 2023).

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, London), pp. 233–277.

Flege, J. E., and Bohn, O. S. (**2021**). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by S. Chen (Cambridge University Press, Cambridge), pp. 3–83.

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995**). "Effects of age of second-language learning on the production of English consonants," Speech Commun. **16**(1), 1–26.

Francis, A., Ciocca, V., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," J. Phonetics **36**(2), 268–294.

Gu, W. (**2016**). "Tone and intonation in Mandarin utterances by HK Cantonese L2 Learners," J. School Chin. Lang. Culture Nanjing Normal Univ. *2*, 123–132 (in Chinese).

Hallé, P. A., Chang, Y.-C., and Best, C. T. (**2004**). "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," J. Phon. **32**(3), 395–421.

Hao, Y. (**2012**). "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," J. Phon. **40**(2), 269–279.

Hashimoto, O. K. (**1972**). *Phonology of Cantonese* (Cambridge University Press, Cambridge).

Hedlund, G., and Rose, Y. (**2020**). *Phon 3.1*. [computer program], https://phon.ca (Last viewed June 17, 2023).

Holm, A., and Dodd, B. (**2006**). "Phonological development and disorder of bilingual children acquiring Cantonese and English," in *Phonological Development and Disorder in Children: A Multilingual Perspective*, edited by H. Zhu and B. Dodd (Multilingual Matters, Bristol), pp. 286–325.

Howie, J. (**1974**). "On the domain of tone in Mandarin," Phonetica **30**(3), 129–148.

Keating, P., Garellek, M., and Kreiman, J. (**2015**). "Acoustic properties of different kinds of creaky voice," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*.

Kratochvil, P., and Chao, Y. R. (**1970**). "Review of a grammar of spoken Chinese," Language **46**(2), 513–524.

Law, N., and So, L. (**2006**). "The relationship of phonological development and language dominance in bilingual Cantonese-Putonghua children," Int. J. Bilingualism **10**(4), 405–428.

Lee, Y., Vakoch, D., and Wurm, L. (**1996**). "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison," J. Psycholinguist. Res. *25* (5), 527–542.

Lee-Kim, S. I. (**2021**). "Development of Mandarin tones and segments by Korean learners: From naive listeners to novice learners," J. Phon. **86**, 101036.

Lenneberg, E. H. (**1967**). *Biological Foundations of Language* (Wiley, New York).

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

Li, F., and Munson, B. (**2016**). "The development of voiceless sibilant fricatives in Putonghua-speaking children," J. Speech. Lang. Hear. Res. **59**(4), 699–712.

Li, X., To, C. K. S., and Ng, M. L. (**2017**). "Effects of L1 tone on perception of L2 tone—A study of Mandarin tone learning by native Cantonese children," Bilingualism **20**(3), 549–560.

Lin, Y. (**2020**). "A Praat script for f0 extraction," https://youran-lin.github.io/blog_posts/20200513_f0PraatScript.html (Last viewed April 7, 2022).

Linguistic Society of Hong Kong (**2024**). "The jyutping scheme," https://web.archive.org/web/20130426050642/http://www.lshk.org/node/47 (Last viewed February 13, 2024).

Liu, W. (**2020**). "Success factors for a Mandarin bilingual program: An auto ethnographic case study," CSL **55**(3), 208–229.

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (**2007**). "The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," J. Speech. Lang. Hear. Res. **50**(4), 940–967.

Meziane, R. S., and MacLeod, A. A. N. (**2020**). "The application of bilingual phonological learning models to early second language development," in *Child Bilingualism and Second Language Learning: Multidisciplinary Perspectives*, edited by F. Li, K. Pollock, and R. Gibb (Benjamins, Amsterdam), pp. 9–27.

Miracle, W. C. (**1989**). "Tone production of American students of Chinese: A preliminary acoustic study," J. Chin. Lang. Teachers Assoc. **24**(3), 49–65.

Mok, P., Zuo, D., and Wong, P. W. Y. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Var. Change **25**(3), 341–370.

Mok, P. P. K., and Lee, A. (**2018**). "The acquisition of lexical tones by Cantonese–English bilingual children," J. Child Lang. **45**(6), 1357–1376.

Nari, R., Aoju, C., and Kuang, J. (**2021**). "Going beyond F0: The acquisition of Mandarin tones," J. Child Lang. **48**(2), 387–398.

Ng, M., Hsueh, G., and Sam Leung, C. (**2010**). "Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children," Int. J. Speech-Lang. Pathol. **12**(3), 230–236.

Paradis, J. (**2011**). "Individual differences in child English second language acquisition," Ling. Appr. Biling. **1**(3), 213–237.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J., and Wang, W. S.-Y. (**2012**). "The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems," J. Speech. Lang. Hear. Res. *55*, 579–595.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., and Best, C. T. (**2015**). "Perceptual assimilation of lexical tone: The roles of language experience and visual information," Atten. Percept. Psychophys. *77*, 571–591.

Scobbie, J. E., Gibbon, F., Hardcastle, W. J., and Fletcher, P. (**2000**). "Covert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*, edited by M. Broe and J. Pierrehumbert (Cambridge University Press, Cambridge), pp. 194–203.

Shen, X. (**1989**). "Toward a register approach in teaching Mandarin tones," J. Chin. Lang. Teachers Assoc. **24**(3), 27–47.

Simon, E. (**2010**). "Child L2 development: A longitudinal case study on Voice Onset Times in word-initial stops," J. Child Lang. **37**(1), 159–173.

Singh, L., and Fu, C. S. (**2016**). "A new view of language development: The acquisition of lexical tone," Child Dev. **87**(3), 834–854.

So, C., and Best, C. (**2010**). "Cross-linguistic perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," Lang. Speech **53**(2), 273–293.

So, C., and Best, C. (**2014**). "Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories," Stud. Second Lang. Acquis. **36**(2), 195–221.

So, L., and Dodd, B. (**1995**). "The acquisition of phonology by Cantonese-speaking children," J. Child Lang. **22**(3), 473–495.

Soskúthy, M. (**2021**). "Evaluating generalized additive mixed modeling strategies for dynamic speech analysis," J. Phon. **84**, 101017.

To, C. K., Cheung, P. S., and McLeod, S. (**2013**). "A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones," J. Speech. Lang. Hear. Res. **56**(1), 103–122.

Tse, J. K. P. (**1978**). "Tone acquisition in Cantonese: A longitudinal case study," J. Child Lang. *5*(2), 191–204.

van Rij, J., Wieling, M., Baayen, R., and van Rijn, H. (**2022**). "itsadug: Interpreting time series and autocorrelated data using GAMMs," R package version 2.4.1.

Wang, H. S. (**1995**). *Experimental Studies in Taiwanese Phonology* (Crane, Taipei City).

Wang, X. (**2006**). "Perception of L2 tones: L1 lexical tone experience may not help," in *Proceedings of Speech Prosody 2006*, edited by R. Hoffman and H. Mixdorff, ISCA Archive.

Wayland, R., and Guion, S. (**2004**). "Training English and Chinese listeners to perceive Thai tones: A preliminary report," Language Learn. **54**(4), 681–712.

Winke, P. (**2007**). "Tuning into tones: The effect of L1 background on L2 Chinese learners' tonal production," J. Chin. Lang. Teachers Assoc. **42**(3), 21–55.

Wong, P. (**2008**). "Development of lexical tone production in disyllabic words by 2- to 6-year-old Mandarin-speaking children," Ph.D. thesis, City University of New York, New York, available at https://academicworks.cuny.edu/gc_etds/4748/.

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

Wong, P. (**2013**). "Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan," J. Acoust. Soc. Am. **133**(1), 434–443.

Wong, P., Fu, W. M., and Cheung, E. Y. (**2017**). "Cantonese-speaking children do not acquire tone perception before tone production—A perceptual and acoustic study of three-year-olds' monosyllabic tones," Front. Psychol. **8**, 1–15.

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

Wu, J., Chen, Y., van Heuven, V. J., and Schiller, N. O. (**2019**). "Dynamic effect of tonal similarity in bilingual auditory lexical processing," Lang., Cognit. Neurosci. **34**(5), 580–598.

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

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., and Demuth, K. (**2018**). "Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like," Front. Psychol. **9**, 817.

Yang, C. (**2023**). "Tone change in mandarin: Comparing acoustic analyses from the early 20th century to today," in *Proceedings of the 20th International Congress of Phonetic Sciences*, edited by R. Skarnitz and J. Volin (Guarant International, Prague, Czechia), pp. 823–831.

Yao, Y., Chan, A., Fung, R., Wu, W. L., Leung, N., Lee, S., and Luo, J. (**2020**). "Cantonese tone production in pre-school Urdu–Cantonese bilingual minority children," Int. J. Biling. **24**(4), 767–782.

Zhang, H. (**2010**). "Phonological universals and tonal acquisition," J. Chin. Lang. Teachers Assoc. **45**(1), 39–65.

Zhang, J. (**2018**). "A comparison of tone normalization methods for language variation research," in *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong, https://aclanthology.org/Y18-1095 (Last viewed March 5, 2022).

Zhao, J., and Bernhardt, B. M. (**2012**). *Mandarin Single-Word Elicitation Tool for Phonology*, University of British Columbia, http://phonodevelopment.sites.olt.ubc.ca/mandarin-picture-elicitation_2012 (Last viewed November 22, 2018).

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

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

Zong, S. (**2000**). "Findings report on the negative transfer of local accent in Putonghua acquisition of college students in the Cantonese-speaking region," J. Dongguan Inst. Technol. *7*(1), 86–98 (in Chinese).