Manuscript Details

Manuscript number	PHONETICS_2017_191_R4
Title	Development of tonal discrimination in young heritage speakers of Cantonese
Article type	Research Paper

Abstract

This study uses the Perceptual Assimilation Model for Suprasegmentals (PAM-S) (So & Best, 2008, 2010), supported by the assumptions of the L2 Intonation Learning theory (LILt, Mennen, 2015), to investigate how young heritage speakers of Cantonese living in the United States acquired Cantonese tones. Sixty-seven heritage speakers, aged 5–11, were tested on their perception of Cantonese tonal contrasts using an ABX discrimination task. They were compared to 64 peers 5–12 in Hong Kong, where Cantonese is spoken as the majority language but English is also acquired from a young age. Two pairs of tones were tested: Tones 2 (mid rising) and 5 (low rising), which have similar pitch heights and contours, and Tones 1 (high level) and 4 (low falling), which have a larger phonetic contrast. As predicted, the heritage speakers were more accurate in discriminating between the more distinct pair of tones than between the more similar pair. They also scored lower than their peers from Hong Kong in both contrast conditions. Age of testing predicted accuracy for both groups, and Chinese literacy also had a significant effect for the heritage speakers. The potential lack of the Tone 2–5 contrast in the heritage speakers' input is discussed as an explanation for these findings. This study illustrates the divergence in heritage speakers' phonological development compared to majority language speakers, and shows the relevance of the PAM-S and LILt to the heritage language context.

Keywords	Heritage language; Cantonese; tonal acquisition; bilingual speech perception; Perceptual Assimilation Model for Suprasegmentals; L2 Intonation Learning theory

Corresponding Author	Ting Yan Rachel Kan
Corresponding Author's Institution	University of Essex
Order of Authors	Ting Yan Rachel Kan, Monika Schmid

Suggested reviewers	Conxita Lleo, Charles Chang, Naomi Nagy

Submission Files Included in this PDF

File Name [File Type]

Cover.docx [Cover Letter]

Response.docx [Response to Reviewers]

Highlights.docx [Highlights]

Manuscript.pdf [Manuscript File]

figure1.eps [Figure]

figure2.eps [Figure]

Submission Files Not Included in this PDF

File Name [File Type]

figure3.png [Figure]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Rachel Kan
Monika Schmid
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom

28 November 2018

Taehong Cho
Hanyang University
Seoul
The Republic of Korea

Dear Professor Cho,

Thank you very much for giving us the opportunity to revise and resubmit our research article 'Development of tonal discrimination in young heritage speakers of Cantonese' (Ref: PHONETICS_2017_191_R4) to The Journal of Phonetics. We are grateful for the comments from the editors and we have addressed them as listed in the Response file, and we hope you find the revised manuscript to be publishable in The Journal of Phonetics. Thank you very much for your consideration.

Yours sincerely,

Rachel Kan
Monika Schmid

Editor comments

· On the first page, the impression is given that perception of tones is "better" than production of tones in heritage speaker groups. But then on p. 6 starting on line 322, contradictory information appears. Could this be clarified / corrected?

> The contradiction in adults could be either because there are only a limited number of studies on adults in Chinese tones (so the pattern isn't clear yet), or because the perception of consonants/vowels are acquired differently compared to tones. P.1 (para 2) now shows that not all studies find HSs to be monolingual-like.

· On line 333 of page 6, I would suggest "supported by LiLt" rather than "together with LiLt", so more in line with previous statements.

> This has been changed.

· Same page, what is meant by an i-category? Please explain for readers unfamiliar with this term.

> A definition has been added to Section 1.2, para 2. "I-category" is replaced by intonational categories in the rest of the paper.

· P. 7, line 395 – data ***were***, not data ***was***

> This has been changed.

· Footnote on p. 7: IPA transcription: bɜːrd

> This has been changed.

· P. 8: "since syllables in Cantonese are not meaningful in all tones", what do you mean by this? Syllables aren't "in tones".

> This has been changed. This means that some syllables don't correspond to a real Cantonese word when carrying a certain tone.

· Same paragraph on the same page. Confusing, could you clarify? Tonemes are always "meaningful" as per their definition.

> This has been changed (also in other mentions of 'tonemes').

· First full sentence on top of p. 9 - You have written out thirty-eight, but then n=36. This is contradictory. The sentences is also confusing. Do you mean Mandarin was an additional language, or do you mean that they spoke one or more additional languages *on top*of Cantonese, English and Mandarin?

> Mandarin is also counted. This line has been rewritten and the correct numbers included.

· First full paragraph on p. 9 – Please clarify slightly more, e.g. add to the final sentence of this paragraph something like: "so the thought is that the HSs will have received more English input, potentially at a younger age, than the HK speakers."

This has been added.

· In the following paragraph, are these standard deviations really correct? Shouldn't it be 20%, and not .20%?

Yes, this has been changed.

· Same paragraph - Very interesting! So it's actually how much Cantonese *they* are speaking, not how much is being spoken *to them*. This could be discussed more in the discussion in the appropriate section, particularly relevant as you have conducted a perception task, not a production task.

This is added to Section 4.0, p.16.

· Bottom of page 9 - These are different to the control set of stimuli you previously described. Please explain and / or correct.

The description at the top of p.7 has been corrected.

· Line 671 p. 12 - ***were*** not ***was***

This has been changed.

· P. 13, line 728 – please write out ***10*** (=ten) at beginning of sentence

This has been changed.

· Line 752 - ***regressions were*** - write in plural not singular

This has been changed.

· Line 772 – "the" is missing, i.e. "that THE difference"

This has been changed.

· Table 8 – could literacy be entered in this model as well?

Literacy was not entered in this model because all the subjects in this group are 'literate' (they would have been learning how to read/write Chinese since kindergarten).

· P. 16, starting lines 924: However, you didn't find that AOA was a significant factor, so regardless of the interpretation of AOA (input and / or neural plasticity), it would be important to emphasise at this point that - at least during childhood (which is when the subjects were assessed) - the effects of neural plasticity and / or input were not apparent.

This has been added to the paragraph.

· Following paragraph: You found that the HK and HS parents spoke similar amounts of Cantonese with the children, but that there was a significant difference between *how much* Cantonese the children *spoke back* at them. So your results may simply indicate that with regard to perception, how much the language is *actually spoken* doesn't influence results (at least as long as a certain amount is spoken).

If speaking the language doesn’t influence results, wouldn’t we see no relation between output and perception? (e.g. large difference in output but same discrimination accuracy, or no difference in output but different discrimination accuracy) But now the results show a group difference in both output and discrimination, but no difference in input. Would this somehow show that output, but not input, is related to perception? We leave it for future work to consider what the relative roles of input and output in perceptive and productive abilities are. In referring to the fact that there was a difference in output but not input, we now point out that this can reflect that even when parents of the two groups do not provide different input (as far as the questionnaire shows), the HSs produce less Cantonese even at home, as reflective of the stronger influence of English in the US compared to in Hong Kong.

· P. 18, first full paragraph. Actually, the youngest HK speakers were 3 years of age.

This has been corrected (in the methodology), all participants were above 5 years of age.

· Same paragraph, line 1036 – “the two groups had more different scores” – please rephrase

This has been rephrased: ‘there was bigger difference in scores between the two groups in the Similar category’.

- Young heritage speakers discriminated Cantonese tones with low accuracy.
- They scored significantly lower compared to majority language speaker peers.
- Both groups discriminated distinct tone pairs more accurately than similar ones.
- There was an overall improvement with age of testing in both groups.

Development of tonal discrimination in young heritage speakers of Cantonese

Rachel T. Y. Kan[ab*], Monika S. Schmid[a]

*Corresponding author: rtykan@eduhk.hk, Tel: +852 2948 7692
[a] University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom
[b] Current address: The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong

Abstract
This study uses the Perceptual Assimilation Model for Suprasegmentals (PAM-S) (So & Best, 2008, 2010), supported by the assumptions of the L2 Intonation Learning theory (LILt, Mennen, 2015), to investigate how young heritage speakers of Cantonese living in the United States acquired Cantonese tones. Sixty-seven heritage speakers, aged 5–11, were tested on their perception of Cantonese tonal contrasts using an ABX discrimination task. They were compared to 64 peers aged 5–12 in Hong Kong, where Cantonese is spoken as the majority language but English is also acquired from a young age. Two pairs of tones were tested: Tones 2 (mid rising) and 5 (low rising), which have similar pitch heights and contours, and Tones 1 (high level) and 4 (low falling), which have a larger phonetic contrast. As predicted, the heritage speakers were more accurate in discriminating between the more distinct pair of tones than between the more similar pair. They also scored lower than their peers from Hong Kong in both contrast conditions. Age of testing predicted accuracy for both groups, and Chinese literacy also had a significant effect for the heritage speakers. The potential lack of the Tone 2–5 contrast in the heritage speakers' input is discussed as an explanation for these findings. This study illustrates the divergence in heritage speakers' phonological development compared to majority language speakers, and shows the relevance of the PAM-S and LILt to the heritage language context.

Keywords
Heritage language, Cantonese, tonal acquisition, bilingual speech perception, Perceptual Assimilation Model for Suprasegmentals, L2 Intonation Learning theory

Development of tonal discrimination in young heritage speakers of Cantonese

1.0 Introduction

Heritage speakers (HSs) are bilinguals who grew up speaking a minority language at home, while acquiring the majority language spoken in society at school or from the community (Montrul, 2008; Valdés, 2001). There are different definitions for HSs, but the majority of linguistic research focusses on HSs from immigrant backgrounds (Fishman, 2001; Montrul, 2016). Even if there are other HSs in the same local community, the heritage language (HL) tends to not be supported, at least not widely, in the host country (Rothman, 2009). HSs might be exposed mainly to the HL in the first few years of their lives, as they interact with other speakers mainly within the home environment. However, as they generally use more and more of the majority language at school and with peers, low levels of input and opportunity for use can affect the development of the HL.

1.1 HL phonology

The ability to discriminate and produce contrastive sounds is fundamental to identifying, comprehending, and producing words (Kuhl, 2004; Werker, Byers-Heinlein, & Fennell, 2009). The knowledge of the phonetics and phonology of a native language develops early in infants and children (Jusczyk, Houston, & Newsome, 1999; Kuhl, 1985). In terms of perception, some studies have found that HSs perform very similarly to monolinguals (Kim, 2016; Lukyanchenko & Gor, 2011), while others have found the two groups to be significantly different (e.g. So, 2000; Yang, 2015). HSs benefit from, among other things, exposure to the language from a very young age, and in certain specific populations, HSs have been documented as more accurate in perceiving and producing sounds of their HL compared to second language (L2) learners of that same language (e.g. Au, Knightly, Jun, & Oh, 2002; Boomershine, 2013; Chang, Yao, Haynes, & Rhodes, 2011; Knightly, Jun, Oh, & Au, 2003; Oh, Jun, Knightly, & Au, 2003). Similar benefits are observed in adoptees, who are exposed to their birth language only briefly (e.g. Choi, Broersma, & Cutler, 2017; Oh, Au, & Jun, 2010; Zhou, 2015, but see Pallier et al., 2003; Ventureyra & Pallier, 2004; Ventureyra, Pallier, & Yoo, 2004 on first language loss). These studies suggest that early exposure leads to the acquisition of some aspects of phonetics/phonology, which persist into later childhood and even adulthood.

However, where production is concerned, not all HSs attain monolingual-like phonetic abilities in their HL (e.g. Godson, 2004; Rao, 2015; Ronquest, 2013). In some cases, the HL is produced with characteristics of the majority language (Godson, 2004). Research so far suggests that divergence in the HL in comparison to monolingual speakers occurs only on a phonetic level, and phonemic contrasts in the HL are maintained (Chang et al., 2011; Tse, 2016). HSs' non-monolingual-like phonetic abilities, at least in production, are therefore likely a by-product of bilingualism (e.g. Bosch, Costa, & Sebastián-Gallés, 2000; Flege, Schirru, & MacKay, 2003).

There is a great variation at the individual level produced by the interaction of various factors, such as age of arrival (AOA) of HSs born in the home country, quantity and quality of input, and sociolinguistic factors (Polinsky & Kagan, 2007). For example, HSs with later AOAs/age of first exposure to the L2 are more target-like in the phonological production of their first language (L1)/HL, compared to speakers with earlier AOAs (Flores & Rato, 2016; Godson, 2004). In particular, it appears that speakers arriving before age 12 are more likely to acquire target-like HL speech perception (Ahn, Chang, deKeyser, & Lee-Ellis, 2017). More input and output, including through being taught in the HL at school, has also been found to be advantageous for HL production (Hakuta & D'Andrea, 1992; Rao, 2015; Oh et al., 2003). Since some input providers will undergo more extensive phonetic attrition than others, the input received by individual HSs can be qualitatively different (Chang et al., 2011). Sociolinguistic factors, such as language preference, have also been considered (Kupisch et al., 2014). Although the above studies focus on production, they show that the attainment of HL phonology varies according to speakers' background and behaviour.

In bilingualism research, suprasegmental features have not received as much attention as segmentals. The acquisition of tones is also understudied, as the languages with the most L2 speakers (e.g. English, Spanish) do not have tones. Speakers of non-tone L1s struggle to learn tone L2s, because they are not habituated to attending to cues relevant to tones (e.g. Hallé, Chang, & Best, 2004; Wang, Behne, Jongman, & Sereno, 2004), whereas speakers of tone L1s have an advantage in acquiring a tone L2 (Wayland & Guion, 2004). Speakers of different languages also attend differently to tonal cues such as pitch height and tone contour (Fok-Chan, 1974; Tse, 1973), depending on which ones are relevant to their L1. Therefore, speakers who are more sensitive to the relevant cues perceive tones more accurately (Francis, Ciocca, Ma, & Fenn, 2008; Gandour, 1983; Wayland & Guion, 2004).

1.2 Theoretical frameworks for HL tone acquisition

Theoretical frameworks for HL phonology include the Native Language Magnet model (NLM, Kuhl, 1994) and the Speech Learning Model (SLM, Flege, 1995, 2007), as well as the Perceptual Assimilation Model for Suprasegmentals (PAM-S) (So & Best, 2008, 2010) and the L2 Intonation Learning theory (LILt, Mennen, 2015) that target specifically suprasegmentals. The PAM-S has been applied to the learning of L2 tones with considerable success (e.g. Reid et al., 2015; So, 2012; So & Best, 2010, 2011, 2014), and the current study endeavours to extend it to HL tones.

The PAM-S proposes that L2 intonational categories (prosodic categories such as tones and intonation) are perceptually assimilated to L1 categories. Different assimilation types, based on phonetic similarities between the two languages, predict how well L2 contrasts are perceived. In categorised assimilation, an L2 category corresponds to an L1 category, whereas in uncategorised assimilation, an L2 category is mapped onto more than one L1 category. There are six assimilation types:

(1) Two-Category Assimilation (TC), where two non-native categories assimilate separately to two native ones; (2) Single-Category Assimilation (SC), where two non-native categories assimilate equally to one native category; (3) Category-Goodness Assimilation (CG), where two non-native categories assimilate unequally to one native category; (4) Uncategorized-Categorised Pair Assimilation (UC), where one non-native category is uncategorised and another assimilates to a native category; (5) Uncategorised-Uncategorised Assimilation (UU), where non-native categories undergo uncategorised assimilation; (6) Non-Assimilable (NA), where two non-native categories are perceived as non-speech sounds (Best, 1995). Among the three types involving categorised assimilation, the best discrimination is predicted for TC followed by CG, while poor discrimination is predicted for SC (Best, 1995; Best, McRoberts, & Goodell, 2001).

The predictions of the PAM-S have been supported by studies on tones. For example, Mandarin Tones 1 (high level) and 4 (high falling) were predicted to undergo SC (Single-Category) Assimilation in Cantonese speakers, both to the Cantonese Tone 1 (high level), and indeed L1 Cantonese speakers learning Mandarin showed poor discrimination of this tone pair (adults: Hao, 2012; So & Best, 2010; children: Li, To, & Ng, 2017). PAM-S has also been applied to the acquisition of L2 phonology that differed significantly from the L1, for example with So (2012) finding English L1 speakers assimilating Mandarin tones to English intonational categories (e.g. Tone 1 to Flat Pitch, Tone 4 to Statement). However, not all predicted assimilation types have been found (e.g. Hao, 2012; Li et al., 2017). For example, Mandarin Tones 2 (mid-rising) and 3 (mid-falling-rising) were predicted to undergo UC (Uncategorised-Categorised Pair) Assimilation, with Mandarin Tone 2 matched to the Cantonese Tones 2 (mid rising) and/or 3 (mid level), and Mandarin Tone 3 matched to Cantonese Tones 4 (low falling) or 5 (low rising). Cantonese speakers were expected to show relatively accurate discrimination of this tone pair, but the results showed the opposite (Hao, 2012; Li et al., 2017).

For HL phonology, the PAM-S has a great potential in explaining HSs' perceptive ability and allowing comparisons between languages with different prosodic features (as shown most recently in Ahn et al., 2017). However, the directionality of assimilation (i.e. of L2 to L1) cannot be readily extended to the majority language-HL pair, since the relationship and dynamics between the two pairs of languages are not the same. To account for potential assimilation of the HL to the L2, PAM-S can be supplemented by the L2 Intonation Learning theory (LILt, Mennen, 2015). LILt is concerned with L2 intonation production, and predicts learners' difficulty according to cross-language differences along four dimensions of intonation: systematic, realisational, semantic, and frequency. Two of its assumptions specify a role for AOA and linguistic experience – important factors already identified for HL acquisition – and could be applied together with the PAM-S. First, earlier AOA or age of first exposure to the L2 is hypothesised to predict more target-like L2 intonation. In addition, production of L2 intonation becomes more target-like with increasing experience in the L2. Combined with the PAM-S framework, the assumptions of LILt could mean that HSs with earlier AOAs and more experience in the majority language are more likely to assimilate towards majority language categories, resulting in less target-like HL phonology. The current study tests the predictions of PAM-S, supported by the assumptions of LILt, by investigating the acquisition of heritage Cantonese tones.

1.3 Lexical tones in Cantonese

Cantonese is the majority language spoken in Hong Kong and some areas of Guangdong Province and Guangxi Province in China. Varieties of Cantonese are spoken within these regions, but all are mutually intelligible. In Hong Kong Cantonese (HKCAN), there are six lexical tones (Bauer & Benedict, 1997). This paper uses Jyutping as the Cantonese romanisation system, and Cantonese tones will be referred to by their Jyutping number. Lexical tones are used in Cantonese to distinguish words. For example, 丘 *jau1* means 'hill', 柚 *jau2* means 'grapefruit', 游 *jau4* means 'swim', and 有 *jau5* means 'have'.

The different tones are distinguished by their relative pitch and contour (Fok-Chan, 1974; see Table 1). For example, Tone 6 (low level) has a 'low' pitch level relative to other tones, and maintains a 'level' pitch throughout the duration of the syllable. Most tonemes (phonemes with tone as a contrastive feature) have multiple meanings, but not all syllables correspond to meaningful Cantonese words when carrying each of the six tones (e.g. *ziu* is meaningful when carrying Tones 1, 2, 3, and 6, but not Tones 4 and 5). In addition, the six tones do not occur with equal frequency. For example, level tones occur more frequently than falling tones, which in turn are more frequent than rising tones (Leung, Law, & Fung, 2004).

Table 1 ***Tones in HKCAN***

Tone number	1	2	3	4	5	6
Description	high level	mid rising	mid level	low falling	low rising	low level

Guangzhou is the capital city of the Guangdong Province, so Guangzhou Cantonese (GZCAN) is considered the main variant of Cantonese other than HKCAN.[1] The same tones are used in GZCAN and HKCAN, and the pitch range of different tones is similar across the two variants (Wu, 2006). However, GZCAN differs from HKCAN in that it has two variants of Tone 1 (high level and high falling), although more recent studies have found that younger GZCAN speakers do not use the high falling tone as often as older speakers (Bauer, 1998; So, 1996; Wu, 2006). In contrast, while earlier HKCAN speakers used both high level and high falling tones (e.g. Chao,

[1] The varieties of Cantonese spoken in the main cities/regions of Guangdong Province differ slightly from one another, but are more similar to GZCAN than HKCAN due to the history of separation of Hong Kong from mainland China. Here, 'GZCAN' is used to include these other varieties for the sake of simplicity.

1947), by the 1990s most of them only used it non-contrastively or not at all (Bauer & Benedict, 1997; So, 1996). The high falling tone was no longer reported in studies of Cantonese acquisition by children in Hong Kong (So & Dodd, 1995; Tse, 1991).[2] Another notable change in progress is the near- or full-merging of Tones 2 (mid rising) and 5 (low rising) in adult speakers, mostly in production (HKCAN: Bauer, Cheung, & Cheung, 2003; Fok-Chan, 1974; Kei, Smyth, So, Lau, & Capell, 2002; Mok & Wong, 2010; Mok, Zuo, & Wong, 2013; GZCAN: Ou, 2012). The most recent data indicates that some speakers no longer distinguish between these two tones (HKCAN and heritage Cantonese in Canada: Soo & Monahan, 2017) or produce the contrast (Zhuhai Cantonese: Zhang, 2018), but the rate of merging is not the same across all varieties (e.g. near-complete merging in Macau, and partial merging in HKCAN speakers aged 16–35, Zhang, 2018).[3]

Cantonese also has intonation patterns, which are similar to those of other languages (e.g. English, Bauer & Benedict, 1997), but relevant pitch change is often applied only to the final tone in a sentence, or is appended to it (Ma, Ciocca, & Whitehill, 2006; Xu & Mok, 2011). For example, tones in initial and medial positions of questions retain their canonical forms, but a rising pattern can be imposed on tones in final positions (Ma et al., 2006).

The two following sections introduce Cantonese tone acquisition in Hong Kong speakers and HSs respectively. Studies on Mandarin tones are also referred to, particularly since heritage Cantonese is under-explored; Mandarin is the national language of China and has four main lexical tones.

1.4 Cantonese tone acquisition in Hong Kong speakers

Language-specific tonal categories in Cantonese emerge as early as at 9 months, and the production of tonal contrasts is generally evident by age 2 (So & Dodd, 1995; Yeung, Chen, & Werker, 2013). Some studies found that children do not perform at an adult-like level until between 3–6 years, (Lee, Chan, Lam, van Hasselt, & Tong, 2015; Wong, Fu, & Cheung, 2017), or even 9–10 years (Ching, 1984; Ciocca & Lui, 2003), especially when the tasks used are cognitively demanding or require knowledge of written forms. Cantonese- and Mandarin-speaking children acquire level tones before contour tones (Li & Thompson, 1977; So & Dodd, 1995), and acoustically distinct tones before more similar tones (Ching, 1984; Ciocca & Lui, 2003).[4] Ciocca and Lui (2003) suggested that the frequency of occurrence of tone pairs was not an important determiner of the order of acquiring tone contrasts, based on their analysis using the combined frequency of the two tones in target pairs. These results demonstrate that although tones emerge and can be produced at an early age, development continues in later childhood, especially in terms of complex processing of tonal knowledge. (See also Wong, 2013 and Wong, Schwartz, & Jenkins, 2005 for similar conclusions about tone acquisition in Mandarin.)

1.5 HS acquisition of Cantonese tones

Some studies have examined acquisition of Cantonese tones in HSs, and found HSs, at least very young ones, to perceive Cantonese tones like speakers in Hong Kong. J. K.-P. Tse (1978)

---

[2] Because of such findings, in this study, only the high level variant of Tone 1 was tested, and not the high falling variant. Any possibility that participants in either language group used the high falling tone was not expected to affect the results. In the tone discrimination task, Tone 1 (high level) was always paired with Tone 4 (low falling), and accurate discrimination was predicted. Participants who also used the high falling variant of Tone 1 would still be able to discriminate accurately between Tone 1 (high level) and Tone 4 (low falling).

[3] An anonymous reviewer has pointed out recent research on Cantonese tones and the authors would like to thank them for their constructive advice.

[4] In general, acoustically more similar tones are difficult to distinguish compared to more salient contrasts, for both toddlers and adults (Shi, Gao, Achim, & Li, 2017; Singh, Hui, Chan, & Golinkoff, 2014; So & Best, 2010; Tsao, 2008).

and S.-M. Tse (1982) are case studies of HSs aged 0;1-2;8, and both showed that young speakers developed similarly to Hong Kong speakers, particularly if they received a high level of Cantonese exposure. Although the participants were living in Taiwan and Australia respectively at the time of testing, they are often used as examples of acquiring Cantonese in a Cantonese-speaking environment because the input up to the time of study was almost exclusively in Cantonese. Participants in both studies were shown to have mastered the production of Cantonese tones at the same age as reported for Hong Kong children in the literature. In addition, the order of acquiring tones was also similar. (See Chang, 2016 for similar results showing Korean HSs to be indistinguishable from native speakers in their HL perception.)

In terms of production, there are indications that young HSs acquire tones more slowly. For example, Wong (2012), testing Mandarin, found that although HSs aged 3 were acquiring tones in the same order as peers in Hong Kong, they were less accurate in producing some of the tones. Delays were reported for a Cantonese-English bilingual child in Hong Kong who was dominant in English (Law, 2006). Transfer from the L2 was also observed, for example with the falling intonation of English statements changing the pitch level of target tones. In general, bilingual children show some phonological delays compared to monolinguals when their exposure and use of the tested language is lower (Law & So, 2006).

In adult HSs, divergence from monolingual speakers is found in both the perception and production of tones (So, 2000; Mandarin: Yang, 2015). Some studies compared HSs to L2 learners, and found HSs to be more monolingual-like in acquiring HL tones (Chang & Yao, 2016; Yang, 2015). Both So and Yang found later AOA to be associated with more monolingual-like performance in the L1/HL. More generally, research on other HLs has found a role for AOA, amount of HL exposure and formal HL learning, and the degree of phonological similarity between the HL and the majority language in the acquisition of HL phonology (e.g. Ahn et al., 2017; Rao, 2015; Stoehr, Benders, Van Hell, & Fikkert, 2017). Therefore, aside from individual differences in cognitive and perceptive abilities (e.g. Bowles, Chang, & Karuzis, 2016; Chang & Bowles, 2015), personal background factors have also been examined as predictors of HL acquisition.

1.6 Research question and hypotheses

The overall aim of this study is to test whether the PAM-S supported by LILt predicts how HSs acquire Cantonese tones. Their predictions for the results are explained in this section.

The PAM-S restated for HLs proposes that the perception of HL contrasts is constrained by phonological and phonetic properties of the HL and the majority language. Depending on how the HL categories relate to the majority language ones, those that are more likely to assimilate will not form target-like categories, resulting in poor discrimination. English has four intonational categories categories: Flat Pitch (with the same pitch level throughout), Question (rising pitch level), Statement (gentle fall in pitch level), and Exclamation (steep fall) (So & Best, 2014). As for Cantonese, since intonation has a limited effect on the pitch levels of tones, in this study the categories for Cantonese will include only tones.

The first research question asks whether HSs' accuracy in discriminating different pairs of tones can be predicted by the PAM-S. We test the discrimination of two pairs of tones hypothesised to undergo TC Assimilation and SC Assimilation respectively. The correspondence between the Cantonese and English categories is based on phonetic similarities and differences (Best et al., 2001). In the first pair, Tone 1 (high level) corresponds to Flat Pitch due to its level pitch contour, while Tone 4 (low falling) corresponds to Statement due to its gentle fall in pitch level. Since these two tones correspond to two separate English intonational categories, good discrimination is predicted. In the second pair, both Tones 2 (mid rising) and 5 (low rising) correspond to Question due to their rising contour, and since they both assimilate to the same category (SC Assimilation), poor discrimination is predicted. Therefore, Tones 1 and 4 (the

'Distinct' pair) are hypothesised to be discriminated more accurately than Tones 2 and 5 (the 'Similar' pair) (Best, 1993, 1995).

As a control, pairs of stimuli differing in nucleus and rime, but sharing the same onset and tone were also tested (Table 2). The contrast between the control pairs are (in IPA) /ɐu/-/eŋ/ and /ɐm/-/œːŋ/. The contrast in coda (empty or /m/ vs. /ŋ/) exists in English, and between each pair, the vowels in the nucleus do not have overlapping articulatory features. Therefore, TC Assimilation is expected and good discrimination is predicted.[5] The discrimination of the control stimuli is predicted to be as accurate as for the distinct pair, as they share the same assimilation type (i.e. TC).

The second research question asks how the occurrence of assimilation can be determined in HSs. Instead of equating the HL to either the 'native language' or the 'L2' of the PAM-S, we propose that AOA and linguistic experience – as put forward in LILt – determine the degree of assimilation of the HL to the L2: the earlier the AOA (or first exposure to the majority language) and the less the HL is used, the less established the HL categories and the more likely HL categories assimilate to the majority language, resulting in poorer discrimination of HL contrasts overall. In contrast, late AOA and high levels of HL use lead to target-like HL perception (as has been shown in studies such as Ahn et al., 2017; So, 2000; Stoehr et al., 2017). The role of these two predictors, and other related indicators, will also be tested.

The third research question asks whether any observed between-group differences are consistent with the PAM-S and LILt. As explained above, it is expected that the HSs will have a larger range of AOA and Cantonese experience, and therefore a larger range of perceptive ability. They will also be less accurate in discriminating the similar tone pair. Therefore, it is predicted that the HSs will have overall a larger range of scores and lower accuracy compared to Hong Kong (HK) participants born and raised in a Cantonese-majority language environment. On the other hand, no assimilation is predicted for the HK group (Table 2), although this does not imply the same accuracy level across all three contrast conditions: in light of the reports on different rates of merging between Tones 2 (mid rising) and 5 (low rising), some HK speakers may show lower accuracy on the similar tone pair, if the input that they receive lacks that contrast or if the contrast is less evident.

Participants aged 5–12 are tested, in order to observe how Cantonese tones continue to develop in later childhood, and to be able to examine how well PAM-S and LILt apply to developing phonology. Children in New York City (where data were collected) begin kindergarten in the calendar year they turn five years old, so HSs aged five who are already in school will have shifted from a mainly Cantonese environment to receiving large amounts of English input at school. On the other hand, peers in Hong Kong live in a Cantonese-speaking environment and learn English at school, but they are not immersed in English in the same way as HSs. Therefore in selecting this age range, the comparison between the two groups captures the period of time when the language exposure of the two groups begins to diverge more dramatically. Although previous research showed delays or other divergences in young HSs, especially those with early AOAs, when compared to monolinguals (e.g. Ahn et al., 2017; Law & So, 2006), the contrast of the two young bilingual groups in this study allows the investigation of the effects of different amounts of exposure on tone acquisition, and also of whether young bilinguals' tone discrimination can be predicted.

---

[5] Although the Cantonese vowels of the control stimuli, including the diphthong, are not all used in (General American) English, their closest equivalents can illustrate the contrast between the nucleuses of the control stimuli. For example: /ɐu/ - cow [kaʊ~kæʊ]; /eŋ/ - sing [sɪŋ]; /ɐm/ - *sum* [sʌm] or *kingdom* [ˈkɪŋdəm]; /œːŋ/ - there is no close equivalent in English, but imagine *bird* [bɜːrd] ending with /ŋ/.

Since development over the age range is expected, and it has been suggested that frequency does not play a main role in determining the order of acquiring tonal contrasts (Ciocca & Lui, 2003), frequency is also calculated here to ensure that it does not relate strongly to the results of this study. However, a different method to Ciocca and Lui's is proposed: since some syllables in Cantonese do not correspond to a meaningful Cantonese word when carrying certain tones, the (in)ability to discriminate between a given pair of tones can be considered relevant only when the syllable in question is meaningful when carrying both tones of the pair; if the syllable is meaningful when carrying only one of the two tones, then regardless of whether the listener can discriminate between the two tones, the same number of words that match the produced sound is available to the listener. Therefore, frequency in this study will be calculated by the occurrence of syllables that are meaningful when carrying each tone of the target pairs.

The hypotheses of this study are summarised in the following table:

Table 2 ***Summary of hypotheses***

Tone pair	Tone	Corresponding English intonational category	Assimilation type	Predicted accuracy in discriminating tone pair	
				HS	HK
Distinct	1 (high level) 4 (low falling)	Flat pitch Statement	Two-Category (TC)	High	High
Similar	2 (mid rising) 5 (low rising)	Question Question	Single-Category (SC)	Low	High, but may be lower than on the Distinct pair
Control	--	--	Two-Category (TC)	High	High

2.0 Methodology[6]

2.1 Participants

Sixty-seven HSs were recruited from three primary schools in New York City. They were taking part in an after-school programme in their schools that was provided by the local Chinese association, and they lived in neighbourhoods with a relatively high proportion of Chinese-speaking inhabitants. The children were identified as speaking predominantly Cantonese at home by the programme staff, which was confirmed by a survey distributed to their parents. Participants were aged 5;3–11;4 (mean = 8;7, SD = 1;7, see also Table A.1). The majority of participants were born in the United States (*n* = 46). Of those born outside the United States, 17 were born in mainland China, three in Hong Kong, and one in Mexico. Age of arrival (AOA) in the United States ranged from 1;6–9;3 (mean = 4;7, SD = 2;4). Apart from Cantonese and English, participants also spoke Mandarin, Taishanese (Hoisanwaa), and Teochew.[7] Twenty-six of the HSs reported being literate in Chinese, which is defined in this study as being able to read and write at least some simple text, and not just a few words. Of these 26, 13 acquired Chinese literacy through instruction in Cantonese, three through instruction in Mandarin, and ten through a mix of both. Thirty-nine reported never having visited Hong Kong/China, while 28 reported one or more visits.

Sixty-four children in Hong Kong (HK) were tested as a control group. They were recruited from local primary schools and through informal networks. Participants were aged 5;3–12;4 (mean = 9;3, SD = 1;10, see also Table A.1). Sixty-three HK participants were born in Hong Kong, and

[6] Ethical approval for this study was obtained from the Social Sciences Faculty Ethics Sub-Committee, University of Essex. Written consent for children's participation and the use of all collected data was obtained from participants' parents before testing took place.

[7] Mandarin, Taishanese, and Teochew all have tone systems, but the number of tones in each is different and there is no systematic correspondence with the Cantonese system.

one was born in mainland China and moved to Hong Kong when he was five months old. Forty-five participants were taught mostly or always in Cantonese at school, ten were taught half in Cantonese and half in English, and nine mostly or always in English. Fifty-one participants reported acquiring Chinese literacy through instruction in Cantonese, while the rest acquired it through instruction in Mandarin. Seventeen participants reported speaking one or more language(s) other than Cantonese and English; of these, 13 spoke Mandarin.

Children from Hong Kong were selected as controls because like the HSs, they had been exposed to Cantonese from birth and used predominantly Cantonese at home. Also, both groups were bilinguals, and had been exposed to English at an early age; it was the majority language for HSs, while most HK speakers started learning English in kindergarten, and at the very latest when they started primary school. Therefore, the two groups are relatively comparable. However, the HSs might also have received more English input than the HK speakers because English use is more common in society in the United States.

The parents of the participants were all native speakers of Cantonese. Of the HSs' parents who responded, five were born in Hong Kong, 93 in mainland China (all in Cantonese-speaking regions where specified), one in Vietnam, and one in Mexico. Of the HK group's parents who responded, 61 were born in Hong Kong, 31 in mainland China (all in Cantonese-speaking regions where specified), two in Vietnam, two in Mexico, and one in Indonesia. In the case of the HSs, parents used Cantonese on average 84.42% of the time with the participant (SD = 21.31%), while the participants used Cantonese with them 78.84% of the time (SD = 23.69%). In the case of the HK speakers, parents used Cantonese with them 85.74% of the time (SD = 14.24%), and participants used Cantonese with their parents 88.87% of the time (SD = 13.38%). There was no difference between the HSs and the control group in terms of proportion of Cantonese use by either parent at home (ps > .05), but the HSs used Cantonese less often than the HK participants when speaking with their parents (with father: $t(105) = 2.26$, $p = .03$, with mother: $t(88.44) = 3.67$, $p < .001$). For the purpose of analysis, the proportions of Cantonese used by participants with each parent and vice versa were converted into a single score ('Cantonese experience') by taking the mean of the four measurements (Cronbach's $\alpha = .83$).

The parents of both groups rated themselves as highly proficient in Cantonese. These ratings were on a scale of 1–6, with 1 as not being able to understand or speak any Cantonese words, and 6 as being able to understand everything or speak fluently in all situations. On average, parents of the HSs scored 5.70 (SD = .77) and 5.69 (SD = .80) on listening and speaking respectively, while parents of the HK participants scored 5.80 (SD = .58) and 5.84 (SD = .51). There was no difference in self-ratings between the two groups of parents (ps > .05).

The HSs were overall younger than the HK participants ($t(125.27) = 2.742$, $p = .017$), so in order to remove possible confounding effects for between-group comparisons, age-matched subgroups were formed comprising 53 participants each. The mean age difference between each matched pair was .19 years (around 10 weeks), SD = .13 (around 7 weeks). The subgroups were used when comparing the two groups, but all participants were included for within-group analyses.

2.2 Discrimination task

Participants' perception of Cantonese tones was tested using an ABX discrimination task.

2.2.1 Stimulus

Two pairs of tones were tested. The first ('Distinct') pair consisted of Tones 1 (high level) and 4 (low falling). These two tones have different onset pitch and the distance between them increases throughout the syllable. The second ('Similar') pair of tones consisted of Tones 2 (mid rising) and 5 (low rising). Both tones have a low pitch onset and a rising contour. However, Tone 2 has a steeper gradient and rises to the high pitch level, while Tone 5 has a gentler gradient and ends at a middle pitch level (Matthews & Yip, 2001).

To create the stimuli, the two pairs of tones were combined with two onset-rime combinations, *tou* and *wai*, forming a set of four minimal pairs. A control set of stimuli consisted of items contrasting in nucleus and rime, but sharing the same (consonantal) onset and tone, namely /s/ or /l/ in Tones 2 (mid rising), 3 (mid level), 4 (low falling), and 5 (low rising). This resulted in a total of eight pairs of tonemes. These words were all high frequency words, as determined by their inclusion in the 'Hong Kong Chinese Lexical Lists for Primary Learning' (HKSAR Education Bureau, 2008). Finally, each syllable/word was prefixed with 呀 *aa3*, which is often used in Cantonese names or terms of address. The eight minimal pairs forming the stimuli set are listed in Table 3.

Table 3 ***List of stimuli according to contrast category***

Contrast category	Pairs of stimuli	
Distinct	呀威 *aa3wai1*	呀圍 *aa3wai4*
	呀滔 *aa3tou1*	呀圖 *aa3tou4*
Similar	呀喂 *aa3wai2*	呀偉 *aa3wai5*
	呀土 *aa3tou2*	呀肚 *aa3tou5*
Control	呀手 *aa3sau2*	呀醒 *aa3sing2*
	呀秀 *aa3sau3*	呀勝 *aa3sing3*
	呀林 *aa3lam4*	呀梁 *aa3loeng4*
	呀凜 *aa3lam5*	呀兩 *aa3loeng5*

The stimuli were produced by an adult female native speaker of HKCAN without any recent English immersion or frequent use. Figure 1 shows the pitch height and contour of Tones 1 (high level) and 4 (low falling) (the Distinct pair), and Figure 2 shows Tones 2 (mid rising) and 5 (low rising) (the Similar pair). These figures were produced using the four stimuli that contained the syllable *wai*. The two syllables of each stimuli are shown (*aa3* + target syllable), and the vertical dotted line at ~0.23s indicates where the second syllable begins. As an example, the first portion of the darker line in Figure 1 represents *aa3*, and the second portion represents *wai1*.

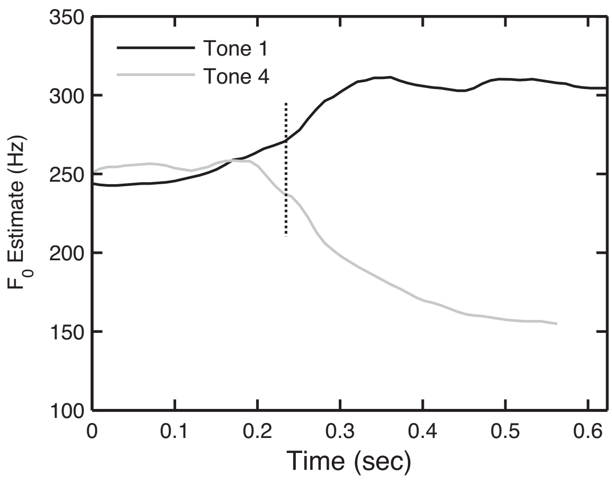


Figure 1 ***Pitch height and contour of the stimuli testing the Distinct pair of tones***

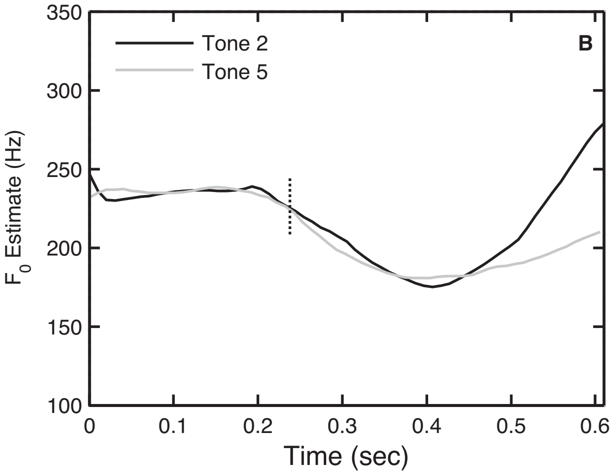


Figure 2 ***Pitch height and contour of the stimuli testing the Similar pair of tones***

Although co-articulation can change the pitch contour of the target (i.e. second) syllables (e.g. Chang & Bowles, 2015; Wong & Strange, 2017), the majority of Cantonese words are disyllabic and listeners use context to establish the pitch range and identify tones in different speakers' speech (e.g. Ma et al., 2006; Zhang, Peng, & Wang, 2012). Therefore, adding *aa3* before the target tones not only imitated authentic usage and contributed to the game-like nature of the task (see Section 2.2.2 below), but also offered participants a consistent context to help them perceive the target tones accurately. Figures 1 and 2 show that the target tones followed the canonical contours, and target syllables corresponded well to the respective English categories even after taking the pitch of *aa3* into account.

2.2.2 Trials

The task was presented using Opensesame (Mathôt, Schreij, & Theeuwes, 2012) on an 8" screen tablet with a pair of headphones. The task was framed as a game, where the participant helped a mother panda find a baby panda. A simple animation showed two baby pandas, one on each side of the screen. In each trial, the baby panda on the left 'uttered' the first item (A) of a stimulus pair, followed by the baby panda on the right uttering the other item (B) of the same stimulus pair. The baby pandas were shown to 'speak' by the appearance of a speech bubble on the screen next to the appropriate baby panda while the stimulus was played. Afterwards, the mother panda, who had a puzzled expression, appeared in the middle of the screen. She 'called' one of the baby pandas, which was shown by a speech bubble next to the mother panda while the target stimulus (X) was played. Participants were asked to find the baby panda that the mother panda was calling, by tapping on the correct side of the screen.

The following example illustrates a trial targeting the Distinct pair of tones:

(1) First baby panda: 呀威 *aa3wai1*
Second baby panda: 呀圍 *aa3wai4*
Mother panda: 呀圍 *aa3wai4*

The eight pairs of stimuli (Table 3) were each targeted four times, in a total of 32 trials. Since the left baby panda always 'spoke' first, the four configurations for each stimulus pair were obtained by targeting the two items of a pair twice each, once as the 'left' stimulus and once as the 'right' stimulus. For example, the stimulus pair XY would appear in four trials: XYX, XYY, YXX, YXY. The task was presented in two pseudo-randomised lists so that trials did not target the same contrast category consecutively. Two training trials (using a separate set of stimuli) were repeated until the participant provided 100% accurate responses.

2.2.3 Procedure

All HSs were tested in their schools, but in a room other than their own classrooms. The HK participants were either tested in a meeting room in their school, or in their homes. In order to minimise any feeling that the participants were being assessed, particularly for those tested in their schools, they were allowed to amuse themselves with computer games, books, or group activities before their session began. The testing session was introduced to the participants as a series of games. The discrimination task was the first task to be conducted.

2.3 Language background questionnaire (LBQ)

Data on participants' language background were collected via a questionnaire for parents written in Chinese, which was distributed at the end of testing. A shorter, oral version was also administered to participants, and their responses were used if their parents' questionnaires were not returned. The questionnaire was adapted from the BiLingual Language Experience Calculator (BiLEC) (Unsworth, 2013), and posed questions concerning children's family background (e.g. date and place of birth, parents' occupation and place of birth), and language background (e.g. Chinese literacy, languages spoken and age of first exposure). Current language use was also measured by asking for the proportion of Cantonese used between various family members, with teachers and fellow students, and during other activities such as reading and watching television.

2.4 Calculation of frequency

In order to find out whether the frequency of occurrence of tone pairs was related to how well they were discriminated, the frequency of syllables that are meaningful when carrying each of the tones of the two target pairs was counted in three sets of data: the Hong Kong Cantonese Corpus (HKCC, Luke & Wong, 2015), as well as the Hong Kong Cantonese Child Language Corpus (CANCORP, CHILDES version, Lee et al., 1996), with the utterances of Hong Kong children and Hong Kong adults examined separately. 'Meaningfulness' was determined using 'A Chinese Talking Syllabary of the Cantonese Dialect: An Electronic Depository' (Cantonese Pronunciation Electronic Dictionary Team, 1999), and frequency was calculated using PyCantonese (Lee, 2015).

3.0 Results

Descriptive statistics of the participants' scores (as the percentage of responses in which the target baby panda was accurately identified) are shown in Table 4. Trials with invalid responses (e.g. if participants tapped on the mother panda) were counted as inaccurate. The distribution of the participants' scores is summarised in Table 5. (Nobody scored lower than 10%.)

Table 4 ***Descriptive statistics for scores (%) by language group and contrast category***

Group	Contrast category	Mean	SD	Min	Max
HS	Control	89.65	16.08	31.25	100
	Distinct	80.78	22.75	25	100
	Similar	56.34	21.03	12.5	100
HK	Control	96.88	5.22	75	100
	Distinct	95.12	8.52	62.5	100
	Similar	89.45	12.24	50	100

Table 5 ***Number of participants with scores (%) in different ranges***

Group	Contrast category	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-100
HS	Control	0	0	3	0	1	4	1	16	42
	Distinct	0	4	3	0	3	6	11	12	28
	Similar	1	7	0	9	19	16	7	2	6
HK	Control	0	0	0	0	0	0	1	4	59
	Distinct	0	0	0	0	0	2	1	17	44
	Similar	0	0	0	0	1	4	8	22	29

The results showed that there were participants from both language groups who reached a ceiling level of performance (100%), but the HSs had a larger range of scores (Figure 3). Many HSs scored above 70% on the Distinct and Control categories, and the average score was also reasonably high. However, for the similar pair the average score was lower, and many HSs scored below 70%. All HK participants scored at or above chance level in all contrast categories. Ten HSs (15% of group) scored below the HK range for the Distinct pair, and 17 HSs (25% of group) scored below the HK range for the Similar pair.

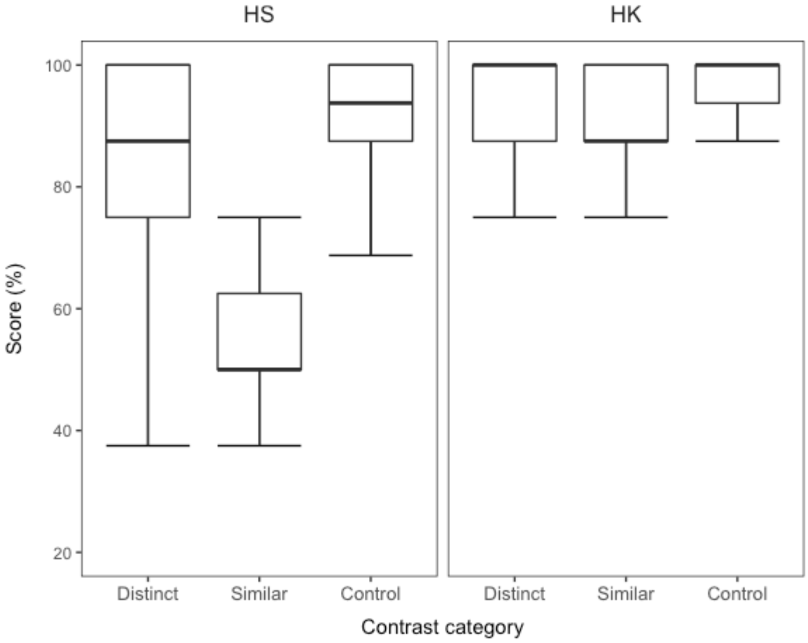


Figure 3 ***Distribution of scores by language group and contrast category. This figure was produced using one score for each speaker in each category***

Mixed effects logistic regressions were conducted using the 'lme4' package in R (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2016). The dependent variable was accuracy on each trial. Fixed-effects predictors with *p* values smaller than .05 were considered to be statistically significant. Only those predictors relevant to the analysis at hand were included, and non-significant fixed-effects predictors were retained in the final model. Fixed-effects predictors are listed in each section below.

3.1 Between-group comparison

The first regression compared the two language groups. As described in the methodology section, age-matched subgroups were used for this analysis. Fixed-effects predictors examined included Group (including the HS and HK levels) and contrast Category (including the Control, Distinct, and Similar levels). Both were given dummy coding, with reference levels HK and Control for Group and Contrast respectively. An interaction term between Group and Contrast was also included. Participants were entered as Subject, as a random factor.

The final model is shown in Table 6. The overall model fit was conditional $R^2 = .25$, marginal $R^2 = .40$ (calculated using the 'piecewiseSEM' package in R, Lefcheck, 2015). The HSs were less accurate compared to the HK participants in the Control category ($B = -1.20$, $SE = .29$, $p < .001$), and the HK group had lower scores in the Similar category compared to the Control ($B = -1.44$, $SE = .24$, $p < .001$), but there were no significant differences between scores in the Distinct and Control categories ($B = -.44$, $SE = .29$, $p = .13$). The significant terms also showed that the difference between the Similar and Control category scores was significantly larger in the HSs than in the HK group, but the difference between the Distinct and Control category scores was similar in the two groups.

Table 6 ***Model for between-group comparison***

	Estimate	Std. Error	*z* value	*p* value
(Intercept)	3.73	.24	15.66	< .001
Group: HS	-1.20	.29	-4.11	< .001
Category: Distinct	-.44	.29	-1.50	.13
Category: Similar	-1.44	.24	-5.90	< .001
Group: HS * Category: Distinct	-.37	.34	-1.07	.28
Group: HS * Category: Similar	-.07	.29	-2.63	.009

Post hoc Tukey pairwise comparisons with Holm-Bonferroni adjustments (calculated using the 'multcomp' package in R, Hothorn, Bretz, & Westfall, 2008) indicated that the HSs scored lower than the HK participants in both Distinct and Similar categories (*p*s < .001). The HSs scored lower in the Similar than the Distinct category ($p < .001$), as well as lower in both Distinct and Similar categories than in the Control category (*p*s < .001). The HK group also scored lower in the Similar category compared to the Distinct category ($p = .001$)

3.2 HSs' performance

To test our proposal that AOA and linguistic experience (cf. LILt) determine whether HL categories assimilate to the majority language, mixed effects logistic regression was conducted to examine factors affecting the HSs' performance ($n = 65$, two HSs were excluded because of missing data). The dependent variable was accuracy on trials for the Distinct and Similar categories combined; Control category trials were not included so as to focus on participants' ability to discriminate tones, and also because individual overall scores for Control category trials were used as a measure of participants' accuracy in completing the task. Subject was included as a random factor. Fixed-effects predictors included were:

- Age of testing (AOT), to test for HL development across the group
- Gender (dummy-coded as Male and Female, with Male as the reference level), to control for any differences between the two genders in completing the task
- Cantonese experience (percentage of Cantonese use with parents, treated as a continuous variable; see Section 2.1), as an indicator of linguistic experience
- Chinese literacy (dummy-coded as Literate and Not literate, with Literate as the reference level; see Section 2.1), as an indicator of linguistic experience
- Age of arrival (AOA), to test for the effect of having lived in a society with Cantonese as a majority language
- Task accuracy (individual score in the Control category), to control for participants' accuracy in completing the task

The final model is shown in Table 7. The overall fit was conditional $R^2 = .07$, marginal $R^2 = .08$.

Table 7 ***Fixed-effects predictors of the HSs' accuracy on Distinct and Similar tone contrasts***

	Estimate	Std. Error	*z* value	*p* value
(Intercept)	-2.17	.45	-4.84	< .001
Age of testing	.11	.04	3.13	.002
Gender: Female	-.20	.12	-1.66	.10
Cantonese experience	-.29	.29	-1.03	.31
Chinese literacy: Not literate	-.31	.13	-2.36	.02
Age of arrival	.06	.03	1.91	.06
Task accuracy	3.75	.35	10.65	< .001

The results indicated that older HSs performed better than younger HSs, and that HSs literate in Chinese performed better than HSs not literate in Chinese. There was also a significant effect of task accuracy ($p < .001$). No predictive effects were found for Cantonese experience, AOA, or Gender (*p*s > 0.05).

3.3 HK participants' performance

To examine factors affecting scores of the HK participants, the same analysis as for the HSs was carried out for the HK participants. Chinese literacy and AOA were excluded as fixed-effects predictors as there was next to no variance in these aspects. The final model is shown in Table 8. The overall fit was conditional $R^2 = .02$, marginal $R^2 = .02$. The results indicated that older children performed better than younger children ($p = .02$). There was also a significant effect of task accuracy ($p < .001$). There was no significant effect of Gender or Cantonese experience (ps $> .05$).

Table 8 ***Fixed-effects predictors of the HK group's accuracy on Distinct and Similar tone contrasts***

	Estimate	Std. Error	*z* value	*p* value
(Intercept)	-6.38	2.00	-3.18	.002
Age of testing	.13	.06	2.32	.02
Gender: Female	.15	.23	.67	.50
Cantonese experience	-.94	1.01	-.94	.35
Task accuracy	9.10	1.76	5.18	< .001

3.4 Frequency

The occurrence of syllables that are meaningful when carrying each tone of the target tone pairs was counted in CANCORP (adult and child utterances separately) and HKCC. Table 9 shows that syllables that are meaningful both when carrying both Tone 1 (high level) and when carrying Tone 4 (low falling) occur more frequently than syllables that are meaningful in both Tones 2 (mid rising) and 5 (low rising).

Table 9 ***Frequency of syllables in CANCORP and HKCC that are meaningful in both tones (percentage of all syllables in that particular dataset)***

Tones 1 & 4	Total	Target: Tone 1	Target: Tone 4
CANCORP (children)	29142 (11.45%)	21547 (8.46%)	7595 (2.98%)
CANCORP (adult)	65595 (14.73%)	47550 (10.68%)	18045 (4.05%)
HKCC	17111 (14.26%)	8744 (7.29%)	8367 (6.97%)
Tones 2 & 5	Total	Target: Tone 2	Target: Tone 5
CANCORP (children)	9089 (3.57%)	3701 (1.45%)	5388 (2.12%)
CANCORP (adult)	30047 (6.75%)	12478 (2.80%)	17569 (3.94%)
HKCC	7797 (6.50%)	2431 (2.03%)	5366 (4.47%)

4.0 Discussion

In this study, a discrimination task was conducted to examine the acquisition of heritage Cantonese tones. The first research question asked whether the PAM-S predicted HSs' accuracy in discriminating different pairs of tones. Good discrimination was expected for the Distinct pair of tones, undergoing Two-Category (TC) Assimilation, and poor discrimination was expected for the Similar pair, undergoing Similar Category (SC) Assimilation. These predictions were borne out by the HSs' lower scores for the Similar contrast compared to the Distinct contrast, showing that the PAM-S can be applied to the acquisition of heritage phonology, especially in determining which tonal contrasts are more likely to be acquired. However, the HSs also scored lower in the Distinct contrast compared to the Control stimuli, which did not meet the expectation that TC Assimilation would occur for both contrast categories and lead to similar performance. Possible reasons for this are discussed in Section 4.1.

The second research question asked how the occurrence of assimilation can be determined in HSs, and it was proposed that AOA and linguistic experience had an effect on forming HL intonational categories, as put forward in LILt. HSs with earlier AOA and/or less Cantonese experience would have lower scores on the discrimination task, whereas HSs with later AOA and/or more Cantonese experience would be more accurate in discriminating the target tones.

The results showed that in this study, the HSs with some level of Chinese literacy were more accurate in perceiving tones, which is consistent with previous evidence showing positive effects for having received formal education where the HL was the medium of instruction (Ahn et al., 2017, following Hakuta & D'Andrea, 1992). This indicates some benefits for HSs with more HL exposure in general. Therefore, the amount of HL learning and also of learning using the HL should be considered in future studies. A more sensitive measure is also needed to evaluate the effect of Chinese literacy; some HSs had been taught to read and write Chinese through Mandarin or a mix of Mandarin and Cantonese, and it is not yet certain in what ways exposure to Mandarin might affect Cantonese tone perception. For example, knowledge of one tone language may not be immediately beneficial for learning another tone language (e.g. So & Best, 2010).

On the other hand, Cantonese experience and age of arrival (AOA) played no role in HS scores on tone discrimination, i.e. contrary to the predictions, there was no benefit in more contact with Cantonese at home or living in a Cantonese-majority environment for longer, as far as perceiving the target tone contrasts was concerned. These findings do not agree with other research demonstrating the effects of amount of exposure and AOA on HL abilities (e.g. Ahn et al., 2017; Law & So, 2006; Montrul, 2008; So, 2000; Unsworth, 2013).

While AOA was used in this study as an indicator for how much Cantonese exposure the HSs had received prior to immigration, in other studies it has been used as a proxy for neural plasticity (e.g. Flege, Munro, & MacKay, 1995; Munro, Flege, & MacKay, 1996). Late AOAs indicate arrival at an age when neural plasticity is lower, and if speakers with late AOAs start acquiring the L2 (or are exposed to it as a majority language) at an older age, then the less likely they are to become proficient in the L2 (either because the L1 is more developed or speakers are less able to learn languages when older, see for example Pallier, 2007). It can be hypothesised that the HL will be less likely to assimilate to the L2 when the L2 intonational categories are weaker, and therefore late AOAs lead to more target-like HL categories. The LILt suggests that AOA predicts L2 learning, but does not specify why. With the preceding reasoning, PAM-S and LILt can be integrated to predict HL development or maintenance. Of course, AOA was not a significant predictor in this study, so whether AOA is interpreted is exposure and/or neural plasticity, its effects may not be apparent in childhood; the HSs were relatively young and none of them arrived past the turning point of age 12 (Ahn et al., 2017), so there may not have been enough variance in the AOAs for a significant effect to be found.

As for Cantonese exposure, it is possible that the expected variation among the HSs may not have been adequately reflected in the indicators used. The present study only used a rough measurement of the current proportion of Cantonese use and only use with parents, so detailed measurements of Cantonese and English input and output throughout HSs' lives might reveal more subtle relations between language experience and acquisition. The HSs could have received less input from their parents in absolute terms, since many parents were employed in service or catering positions and probably worked in the evenings. In addition, a larger difference between the two groups may be found if language use outside the home was also taken into account: with English spoken as the majority language of the United States, Cantonese use for the HSs is restricted to the home and the Cantonese-speaking community, while it is the majority language in Hong Kong. Therefore, the HSs may not have been exposed to a sufficient amount of Cantonese to acquire all its tonal contrasts (cf. Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012).

Interestingly, the LBQ revealed no difference between the two groups in terms of the proportion of Cantonese used by parents at home (i.e. input), but only in what the participants used with their parents (i.e. output). This is surprising because the perception task in this study targeted passive knowledge, so experience in perception (that is, receiving more input) might be expected to be more relevant than experience in production. The result here could reflect the stronger influence of English in the United States, such that HSs used less Cantonese at home even while their parents used the same proportion of Cantonese as parents in Hong Kong. Future research should consider whether both input and output are equally relevant in perception or production tests. In short, the limitations of the language background measurements may explain why no effects were found for Cantonese experience. Using the full, original version of the BiLEC (Unsworth, 2013), or other detailed questionnaires for HSs (e.g. Lee-Ellis, 2012, also used in Ahn et al., 2017), might have resulted in findings more comparable to previous research.

It remains that the variance in the HSs' scores is largely unexplained. Age of testing (AOT) and task accuracy had a predictive effect on participants' scores, so the cognitive load of the task, specifically on memory span and processing, could have disfavoured younger children (Gathercole, 1998), while older children had better concentration and were less likely to make mistakes due to fatigue. The predictive effects of AOT for both groups also suggest a limitation of the original form of PAM-S when applied to young HSs, in that poor discrimination could on the surface be attributed to both (e.g. SC) assimilation and tone acquisition that was still ongoing; even children in Cantonese-speaking environments do not achieve adult-like application of tonal knowledge until age 9 or 10 (Ching, 1984; Ciocca & Lui, 2003), and in general level tones are acquired before contour tones (e.g. Li & Thompson, 1977; So & Dodd, 1995; J. K.-P. Tse, 1973). However, AOT and task accuracy explained only some of the variance, so future work should look to other factors that were not examined in this study, including attitude towards the Cantonese language (Shin, 2010), attitude towards the testing session (Nagy, 2015), language aptitude (Bylund, Abrahamsson, & Hyltenstam, 2012), as well as predictors of individual perception ability (Jeon, 2001), such as linguistic and non-linguistic pitch processing abilities (Bowles et al., 2016).

4.1 Group differences

The third research question addressed the predictions of PAM-S for between-group differences. It was predicted that the HSs would perform less accurately than the HK group for the Similar pair of tones because of SC assimilation, leading to an overall lower score. In fact, the HSs had lower scores in all three contrast conditions. One explanation could be that HSs pay less attention to pitch differences as phonemic cues. Speakers attend to the pitch cues that are relevant to their own language(s) (e.g Wayland & Guion, 2004), so as HSs became more proficient in English, perhaps they also became less sensitive to the onset pitch or contour of each syllable, as these cues are less relevant in English. As a result, the HSs attended less to these cues even when listening to Cantonese. The global intonation patterns and prosodic stress of English sentences, which rely on pitch height, are also in direct competition with Cantonese tones. Even though there is a large F0 difference between Tones 1 (high level) and 4 (low falling), the contours of the two tones are similar, so the HSs may have perceived them to be more similar than expected. This could explain why the HSs found even an acoustically salient tonal contrast (between Tones 1 and 4) to be more difficult to perceive than a segmental contrast, while the HK group obtained similar scores in the Distinct and Control conditions. (In no way does this mean that HSs cannot distinguish tones, only that they were less accurate.)

If this is the case, then the assimilation type of Tones 1 (high level) and 4 (low falling) to the English intonational categories may be one other than the TC assimilation adopted in this study. For example, if they underwent Category Goodness (CG) assimilation where the two tones assimilated unequally to one English intonational category, fair to good perception would be predicted at an accuracy between the levels for TC and SC assimilation (Best, 1993, 1995), which would be supported by the results of this study. Hallé et al. (2004) studied French speakers with no prior exposure to tone languages listening to Mandarin, and suggested that tonal categories were not categorised by speakers of non-tone languages and were perceived as uncategorised or non-speech intonational categories. Since the HSs in the present study were exposed to tones, Hallé et al.'s argument is not directly applicable, but the analysis of tonal categories as non-phonemic is a direction worth considering. Classification tasks targeting cross-language correspondence, as conducted in other studies based on the PAM(-S) (e.g. So &

Best, 2011), could be used to further investigate the relationship between Cantonese and English prosodic categories.

As for the group difference in the control trials, the HSs' potentially lower Cantonese proficiency could have led to an overall disadvantage. Since the ABX task has memory demands and the stimuli were Cantonese, the HSs could have been less able to fully utilise their working memory capacity compared to if they had been more proficient (Gass & Lee, 2001). They could have also been not used to speaking Cantonese outside of their home, especially in a classroom where normally English is used during school hours. Even though the task was presented as a game and various strategies were used to make the participants feel more comfortable before the session (e.g. playing games with them in the testing classroom, letting them interact with the test equipment), the participants might still have felt nervous or out of their comfort zone, and hence performed less well.

Participants aged 5–12 were selected for this study in order to compare how the two groups acquired tone discrimination. Age of testing (AOT) predicted discrimination accuracy for both groups, indicating that even with maximal Cantonese exposure (for the HK group speaking mainly Cantonese both at home and in society), tone development continues in later childhood (e.g. Ching, 1984; Ciocca & Lui, 2003). Therefore, the lower scores of both groups in the Similar category compared to the Distinct category could be because the participants were still acquiring tones. However, the age range of the participants could also have affected the comparison between the two groups: if the HSs diverged from the HK speakers at a young age, the between-group differences could be compounded as children grew older. Accordingly, there would be more apparent differences between the two groups in features that are acquired later. Indeed, there was bigger difference in scores between the two groups in the Similar category, targeting the contour tones which are acquired later (Li & Thompson, 1977; So & Dodd, 1995), compared to in the Distinct category (Section 3.1).

4.2 Quality of input

Differences in quality between the input available to the two groups is another reason for the divergence in the acquisition of Cantonese tones. Previous studies refer to the benefits of a diverse source of HL (e.g. Pascual y Cabo & Rothman, 2012; Rinke & Flores, 2014); such variation is absent for the HSs in the present study, since a large proportion of Cantonese input comes from participants' parents, and there is only a limited number of other Cantonese speakers or range of media in the United States providing Cantonese input.

Quality in terms of similarity with the homeland variety can also be considered. If the tones in the input available to the HSs differed from what was available to the HK participants, then the tonal system acquired by the two groups of participants would naturally be different. Sound changes can occur after even a short period of immigration, most commonly due to influence from the new environmental language (e.g. Chang et al., 2011; Tse, 2016), and adult immigrants have been shown to neutralise L1 phonological contrasts (de Leeuw, Tusha, & Schmid, 2018). Therefore, the HSs' parents or other Cantonese speakers in the United States may have undergone such change, and provided input to the HSs that differed from the input provided to the HK group. Previous studies suggest that phonological contrasts are maintained in the speech of HSs of Chinese (Cantonese, Mandarin) even into the next generation (Chang et al., 2011; Tse, 2016), but the situation may differ for patterns that are already changing in the homeland variety.

There is on-going/complete merging of Tones 2 (mid rising) and 5 (low rising) in various varieties of Cantonese (e.g. Soo & Monohan, 2017; Zhang, 2018), which results in input with a less evident Tone 2–5 contrast or that lacks the contrast altogether. Merging can explain why both groups were less accurate in discriminating the Similar tones than the Distinct tones, a finding consistent with previous studies showing that similar tones are more difficult to distinguish (e.g. Ching, 1984; Ciocca & Lui, 2003; So & Best, 2010). The lack of the Tone 2–5 contrast in the HSs' input (Section 4.2) could also explain why Cantonese experience had no effect on accuracy among the HSs. The different extent of merging in the Cantonese spoken in different regions could have contributed to the HSs' lower scores in the Similar category compared to the HK group: if the merging is more advanced in the mainland than in Hong Kong (Zhang, 2018), then naturally the HSs, whose parents came mostly from the mainland, would be less likely to acquire the Tone 2–5 contrast compared to the HK group, whose parents were more often from Hong Kong. Differences among the mainland varieties would also help explain the larger range of scores in the HSs, but the extent of merging in the parents' Cantonese could not be determined here: no speech data was collected from the parents, and most of the participants' parents did not give a more specific birthplace than 'mainland China' or 'Guangdong'. None of the parents who specified a town or city in the mainland were born in one where recent tone merging has been documented.

Granted, regional variation is not solely responsible for the differences between the two groups. For example, there is no evidence of merging between Tones 1 (high level) and 4 (low falling) that could explain the HSs' lower scores with these two tones compared to the HK group. It is also notable that despite the results of Soo and Monohan (2017) showing complete merging in the perception of Tones 2 (mid rising) and 5 (low rising) among adult Hong Kong participants, the scores of the young HK children in this study were reasonably high for the Similar tone pair. A foreseeable difficulty in further applying PAM-S to Cantonese is that Tones 2 (mid rising) and 5 (low rising) are not the only tones showing relatively rapid merging even in homeland varieties (e.g. Ou, 2012), so any two categories that are hypothesised to show SC Assimilation might also be susceptible to merging, because these two categories were acoustically similar in some ways to begin with.

4.3 Frequency

An alternative explanation for the HSs' different abilities with regards to the two pairs of tones is based on the frequency of occurrence of tones. Previous research found that there was no relation between how frequently tone pairs occurred and how well they were discriminated (Ciocca & Lui, 2003). However, with different tone pairs considered and a different method of calculation used, support was found for a frequency-based explanation for poorer discrimination. It was shown that there syllables that are meaningful in both Tones 1 (high level) and 4 (low falling) occurred more frequently than syllables that are meaningful in both Tones 2 (mid rising) and 5 (low rising). Therefore, hypothetical interlocutors would need to discriminate between Tones 1 and 4 more frequently than between Tones 2 and 5, which might explain why young children acquire the Distinct contrast earlier than the Similar one.

Another manifestation of frequency effects is that participants may perform better if the stimuli are frequent words and familiar to them. Although only frequent words were used in the stimuli, there was no guarantee that all the participants, especially the younger ones, knew all the words. In future studies, each participant's familiarity with words used in the stimuli should be checked before the discrimination task.

5.0 Conclusion

The performance of the HSs in this study and the comparison with the HK participants raises interesting theoretical questions concerning the status of HLs and the nature of phonological knowledge. Since the HSs enjoyed early and a relatively high amount of exposure to Cantonese at home and in the local community, they may be expected to acquire Cantonese phonology successfully as an L1 speaker. However, not all of the HSs could discriminate the two target pairs of tones like their peers living in a majority language environment. The PAM-S framework was combined with LILt to explore the HSs' abilities to perceive tonal contrasts, and some (but not all) predictions were borne out in the results. While the net pattern observed in the present study was one of development, it cannot be determined whether HSs would eventually 'catch up' with the HK participants based on the available evidence. However, the results here are crucial in any attempt to construct the developmental trajectory of HSs' phonological abilities (e.g. Ahn et al., 2017; Polinsky, 2011). The high accuracy in some of the HSs suggest that resistance to assimilation is not necessarily futile, but there are also some unanswered questions, such as why, despite the functional importance of tones in Cantonese, there was not more preservation of the tonal contrasts.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors are deeply grateful to the editors and reviewers for their constructive advice and suggestions for improving this article.

References

Ahn, S., Chang, C. B., DeKeyser, R., & Lee-Ellis, S. (2017). Age effects in first language attrition: Speech perception by Korean-English bilinguals. *Language Learning*, *67*(3), 694–733. http://doi.org/10.1111/lang.12252

Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13*(3), 238–243. http://doi.org/10.1111/1467-9280.00444

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Bauer, R. S. (1998). Hong Kong Cantonese tone contours. In S. Matthews (Ed.), *Studies in Cantonese Linguistics* (pp. 1–33). Hong Kong: Linguistic Society of Hong Kong.

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

Bauer, R. S., Cheung, K., & Cheung, P. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change*, *15*(2), 211–225. http://doi.org/10.1017/S0954394503152039

Best, C. T. (1993). Learning to perceive the sound pattern of English. *Haskins Laboratory Status Report on Speech Research*, *SR-114*, 31–80.

Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 171–204). Baltimore: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775–794. http://doi.org/10.1121/1.1332378

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE*, *7*(2), e30843. http://doi.org/10.1371/journal.pone.0030843

Boomershine, A. (2013). The perception of English vowels by monolingual, bilingual, and heritage speakers of Spanish and English. *Proceedings of the 15th Hispanic Linguistics Symposium*, 103–118.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*, 189–221. http://doi.org/10.1080/09541446.2000.10590222

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808. http://doi.org/10.1111/lang.12159

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper nativelikeness in a second language? A study of ultimate attainment in early bilinguals. *Studies in Second Language Acquisition*, *34*(2), 215–241. https://doi.org/10.1017/S0272263112000034

Cantonese Pronunciation Electronic Dictionary Team (Research Institute for the Humanities, Chinese University of Hong Kong) (1999). A Chinese talking syllabary of the Cantonese dialect: An electronic depository. Retrieved from http://humanum.arts.cuhk.edu.hk/Lexis/Canton2/

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791–809.

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *Journal of the Acoustical Society of America*, *138*(6), 3703–3716. http://doi.org/10.1121/1.4937612

Chang, C. B., & Yao, Y. (2016). Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal*, *13*(2), 134–160.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, *129*(6), 3964–3980. http://doi.org/10.1121/1.3569736

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

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

Choi, J., Broersma, M., & Cutler, A. (2017). Early phonology revealed by international adoptees' birth language retention. *PNAS*, *114*(28), 7307–7312. http://doi.org/10.1073/pnas.1706405114

Ciocca, V., & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders*, *1*(2), 141–147. http://doi.org/10.1080/1476967031000090971

de Leeuw, E., Tusha, A., & Schmid, M. S. (2018). Individual phonological attrition in Albanian-English late bilinguals. *Bilingualism: Language and Cognition*, *21*(2), 278–295. http://doi.org/10.1017/S1366728917000025

Fishman, J. (2001). 300-plus years of heritage language education in the United States. In J. K. Ranard, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 81–89). Washington, DC: CAL, ERIC.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–277). Maryland: York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & E. Hualde (Eds.), *Laboratory phonology* (pp. 353–382). Berlin: Mouton de Gruyter.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-B

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 476–491. http://doi.org/10.1016/S0167-6393(02)00128-0

Flores, C., & Rato, A. (2016). Global accent in the Portuguese speech of heritage returnees. *Heritage Language Journal*, *13*(2), 161–183. http://doi.org/10.1177/1367006915589424

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

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294. http://doi.org/10.1016/j.wocn.2007.06.005

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

Gass, S., & Lee, J. (2011). Working memory capacity, inhibitory control, and proficiency in a second language. In M. S. Schmid & W. Lowie (Eds.), *Modeling bilingualism: From structure to chaos* (pp. 59–84). Amsterdam: John Benjamins.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, *39*(1), 3–27. http://doi.org/10.1017/S0021963097001753

Godson, L. (2004). Vowel production in the speech of Western Armenian heritage speakers. *Heritage Language Journal*, *2*(1), 44–69. http://doi.org/10.1121/1.3569736

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. http://doi.org/10.1016/S0095-4470(03)00016-0

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. http://doi.org/10.1016/j.wocn.2011.11.001

Hakuta, K., & D'Andrea, D. (1992). Some properties of bilingual maintenance and loss in Mexican background high-school students. *Applied Linguistics*, *13*, 72–99. http://doi.org/10.1093/applin/13.1.72

HKSAR Education Bureau (2008). *Hong Kong Chinese lexical lists for primary learning*. Retrieved from http://www.edbchinese.hk/lexlist_ch/

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363. http://doi.org/10.1002/bimj.200810425

Jeon, M. (2001). Avoiding FOBs: An account of a journey. *Working Papers in Educational Linguistics*, *17*(1), 83–106.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207. http://doi.org/10.1006/cogp.1999.0716

Kei, J., Smyth, V., So, L., Lau, C. C., & Capell, K. (2002). Assessing the accuracy of production of Cantonese lexical tones: A comparison between perceptual judgment and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing*, *7*, 25–38. http://doi.org/10.1179/136132802805576535

Kim, J. Y. (2016). *The perception and production of prominence in Spanish by heritage speakers and L2 learners.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Knightly, L., Jun, S., Oh, J., & Au, T. (2003). Production benefits of childhood overhearing. *Journal of the Acoustical Society of America*, *114*, 465–474. http://doi.org/10.1121/1.1577560

Kuhl, P. K. (1985). Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion* (pp. 231–262). Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*, 812–822.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843. http://doi.org/10.1038/nrn1533

Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., Lein, T., & van de Weijer, J. (2014). Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal*, *11*(2), 123–150. http://doi.org/10.1017/CBO9781107415324.004

Law, C. (2006). *Tonal characteristics of early English-Cantonese bilinguals.* MA dissertation, University of Hong Kong. Retrieved from http://hdl.handle.net/10722/51744

Law, N. C. W., & So, L. K. H. (2006). The relationship of phonological development and language dominance in bilingual Cantonese-Putonghua children. *International Journal of Bilingualism*, *10*(4), 405–428. http://doi.org/10.1177/13670069060100040201

Lee, J. L. (2015). *PyCantonese: Cantonese linguistic research in the age of big data*. Talk at the Childhood Bilingualism Research Centre, Chinese University of Hong Kong, September 15.

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*, 53–62. http://doi.org/10.3109/17549507.2014.898096

Lee, T. H. T., Wong, C. H., Leung, S., Man, P., Cheung, A., Szeto, K., & Wong, C. S. P. (1996). *The development of grammatical competence in Cantonese-speaking children: Report of RGC earmarked grant 1991-94*.

Lee-Ellis, S. (2012). *Looking into bilingualism through the heritage speaker's mind*. Unpublished doctoral dissertation, University of Maryland, College Park, MD.

Lefcheck, J. S. (2015). piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573–579. http://doi.org/10.1111/2041-210X.12512

Leung, M.-T., Law, S.-P., & Fung, S.-Y. (2004). Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(3), 500–5. http://doi.org/10.3758/BF03195596

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185–199. http://doi.org/10.1017/S0305000900001598

Li, X., To, C. K. S., & Ng, M. W. (2017). Effects of L1 tone on perception of L2 tone - A study of Mandarin tone learning by native Cantonese children. *Bilingualism: Language and Cognition*, *20*(3), 549–560. http://doi.org/10.1017/S1366728916000195

Luke, K. K., & Wong, M. L.-Y. (2015). The Hong Kong Cantonese Corpus: Design and uses. *Journal of Chinese Linguistics*, *25*, 309–330.

Lukyanchenko, A., & Gor, K. (2011). Perceptual correlates of phonological representations in heritage speakers and L2 learners. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp. 414–426). Somerville, MA: Cascadilla Press.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *Journal of the Acoustical Society of America*, *120*(6), 3978–3987. http://doi.org/10.1121/1.2363927

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. http://doi.org/10.3758/s13428-011-0168-7

Matthews, S., & Yip, V. (2001). *Cantonese: A comprehensive grammar* (2nd ed.). London: Routledge.

Mennen, I. (2015). Beyond segments: Towards an L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 171–188). Berlin, Heidelberg: Springer-Verlag. http://doi.org/10.1007/978-3-662-45168-7

Mok, P. P. K., & Wong, P. W. (2010). Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. *Proceedings of Speech Prosody*, *100916*, 1–4.

Mok, P. P. K., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370. http://doi.org/10.1017/S0954394513000161

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins. http://doi.org/10.1075/sibil.39

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781139030502.001

Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(3), 313-334. https://doi.org/10.1017/S0142716400007967

Nagy, N. (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua*, *164*, 309–327. http://doi.org/10.1016/J.LINGUA.2014.04.012

Oh, J. S., Au, T. K.-F., & Jun, S. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language*, *37*, 1123–1132. http://doi.org/10.1017/S0305000909990286

Oh, J. S., Jun, S.-A., Knightly, L., & Au, T. (2003). Holding on to childhood language memory. *Cognition*, *86*(3), B53–B64. http://doi.org/10.1016/S0010-0277(02)00175-0

Ou, J. (2012). *Tone merger in Guangzhou Cantonese.* MPhil dissertation, Hong Kong Polytechnic University.

Pallier, C. (2007). Critical periods in language acquisition and language attrition. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 155-168). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.33

Pallier, C., Dehaene, S., Poline, J.-B., Argenti, A.-M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity: Can a second language replace the first ? *Cerebral Cortex*, *13*, 155–161. http://doi.org/10.1093/cercor/13.2.155

Pascual y Cabo, D., & Rothman, J. (2012). The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics*, *33*(4), 450–455. http://doi.org/10.1093/applin/ams037

Polinsky, M. (2011). Reanalysis in adult heritage language: A case for attrition. *Studies in Second Language Acquisition*, *33*, 305–328. http://doi.org/10.1017/S027226311000077X

Polinsky, M., & Kagan, O. (2007). Heritage languages: In the "wild" and in the classroom. *Language and Linguistics Compass*, *1*(5), 368–395. http://doi.org/10.1111/j.1749-818X.2007.00022.x

R Core Team. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rao, R. (2015). Manifestations of /bdg/ in heritage speakers of Spanish. *Heritage Language Journal*, *12*(1), 48–74. http://doi.org/10.1093/applin/amr040

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, and Psychophysics*, *77*(2), 571–591. http://doi.org/10.3758/s13414-014-0791-3

Rinke, E., & Flores, C. (2014). Morphosyntactic knowledge of clitics by Portuguese heritage bilinguals. *Bilingualism: Language and Cognition*, *17*(4), 681–699. http://doi.org/10.1017/S136672891300076X

Ronquest, R. E. (2013). An acoustic examination of unstressed vowel reduction in heritage Spanish. In C. Howe, S. E. Blackwell, & M. Quesada (Eds.), *Selected Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 175–171). Somerville, MA: Cascadilla Press.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, *13*(2), 155–163. http://doi.org/10.1177/1367006909339814

Shi, R., Gao, J., Achim, A., & Li, A. (2017). Perception and representation of lexical tones in native Mandarin-learning infants and toddlers. *Frontiers in Psychology*, *8*, 1117. http://doi.org/10.3389/fpsyg.2017.01117

Shin, S. J. (2010). "What about me? I'm not like Chinese but I'm not like American": Heritage-Language learning and identity of mixed-heritage adults. *Journal of Language, Identity & Education*, *9*(3), 203–219. http://doi.org/10.1080/15348458.2010.486277

Singh, L., Hui, T., Chan, C., & Golinkoff, R. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science*, *17*, 94–109. http://doi.org/10.1111/desc.12097

So, C. K. (2012). Cross-language categorization of monosyllabic foreign tones: Effects of phonological and phonetic properties of native language. *Monosyllables: From Phonology to Typology*, 55–69.

So, C. K., & Best, C. T. (2008). Do English speakers assimilate Mandarin tones to English prosodic categories? *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2*(12), 1120.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. http://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2011). Categorizing mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznan Studies in Contemporary Linguistics*, *47*(1), 133–145. http://doi.org/10.2478/psicl-2011-0011

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. http://doi.org/10.1017/S0272263114000047

So, K. L. C. (2000). *Tonal production and perception of Canadian raised Cantonese speakers.* MA dissertation, Simon Fraser University.

So, L. K. H. (1996). Tonal changes in Hong Kong Cantonese. *Current Issues in Language and Society*, *3*(2), 186–189. http://doi.org/10.1080/13520529609615467

So, L. K. H., & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*(3), 473–95. http://doi.org/10.1017/S0305000900009922

Soo, R., & Monahan, P. J. (2017). Language exposure modulates the role of tone in perception and long-term memory: Evidence from Cantonese native and heritage speakers. *Proceedings of the 43rd Annual Meeting of the Berkeley Linguistics Society*, *2*, 47–54.

Stoehr, A., Benders, T., Van Hell, J. G., & Fikkert, P. (2017). Heritage language exposure impacts voice onset time of Dutch-German simultaneous bilingual preschoolers. *Bilingualism: Language and Cognition*. http://doi.org/10.1017/S1366728917000116

Tsao, F.-M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology*, *50*, 111–124. http://doi.org/10.3389/fpsyg.2017.00558

Tse, A. C.-Y. (1991). *The acquisition process of Cantonese phonology: A case study.* MPhil dissertation, The University of Hong Kong.

Tse, H. (2016). Variation and change in Toronto Heritage Cantonese: An analysis of two monophthongs across two generations. *Asia Pacific Language Variation*, *2*(2), 124–156. http://doi.org/10.1075/aplv.2.2.02tse

Tse, J. K.-P. (1973). *The upper even tone in Cantonese: An instrumental investigation.* MA dissertation, National Taiwan Normal University.

Tse, J. K.-P. (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language*, *5*(2), 191–204. http://doi.org/10.1017/S0305000900007418

Tse, S.-M. (1982). *The acquisition of Cantonese phonology.* Unpublished doctoral dissertation, The University of British Columbia.

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*, 86–110. http://doi.org/10.1017/S1366728912000284

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. Kreeft Peyton, D. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37–77). Washington, DC: Center for Applied Linguistics and Delta Systems.

Ventureyra, V. A. G., Pallier, C., & Yoo, H. (2003). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, *17*, 79–91. http://doi.org/10.1016/S0911-6044(03)00053-8

Ventureyra, V. A. G., & Pallier, C. (2004). In search of the lost language: The case of adopted Koreans in France. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 207–221). Amsterdam: John Benjamins.

Wang, Y., Behne, D. M., Jongman, A., & Soreno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tones. *Applied Linguistics*, *25*, 449–466. http://doi.org/10.1017.S0142716404001213

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712. http://doi.org/10.1111/j.1467-9922.2004.00283.x

Werker, J. F., Byers-Heinlein, K., & Fennell, C. T. (2009). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society B*, *364*, 3649–3663. http://doi.org/10.1016/S0010-0277(97)00040-1

Wong, P. (2012). Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and in the United States: Interjudge reliability and perceptual results. *Journal of Speech, Language, and Hearing Research*, *55*(October), 1423–1438. http://doi.org/10.1044/1092-4388(2012/11-0273)

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

Wong, P., 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. http://doi.org/10.3389/fpsyg.2017.01450

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2000). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1065–1079. http://doi.org/10.1044/1092-4388(2005/074)

Wong, P., & Strange, W. (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PLoS ONE*, *12*(8), e0182337. http://doi.org/10.1371/journal.pone.0182337

Wu, W. (2006). *A comparative analysis of the phonetics of Hong Kong Cantonese and Guangzhou Cantonese*. University of Hong Kong.

Xu, B. R., & Mok, P. P. K. (2011). Final rising and global rising in Cantonese intonation. *Proceedings of the 17th International Congress of Phonetic Sciences ICPhS 2011*, (August), 2173–2176. http://doi.org/10.1111/j.1467-9922.2004.00283.x

Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners.* Berlin: Springer Verlag. http://doi.org/10.1007/978-3-662-44645-4

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*(2), 123–139. http://doi.org/10.1016/j.jml.2012.09.004

Zhang, C. C., Peng, G., A., & Wang, W. S. Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *Journal of the Acoustical Society of America*, *132*, 1088–1099.

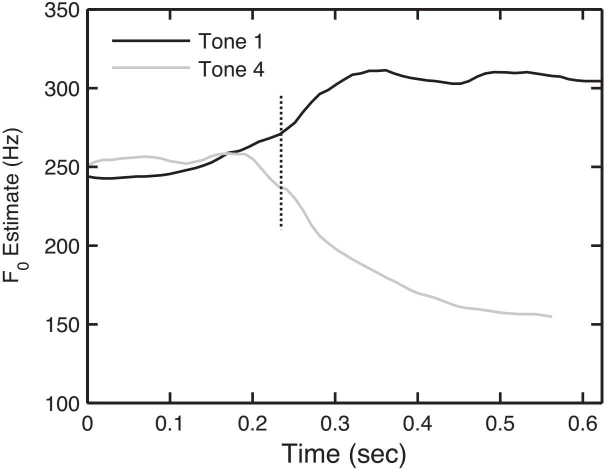
Zhang, J. (2018). *Tone merging in the Cantonese spoken in Hong Kong, Macau and Zhuhai*. Paper presented at Variation and Change in Chinese, La Trobe University, Melbourne.

Zhou, W. (2015). *Assessing birth language memory in young adoptees.* Unpublished doctoral dissertation, Radboud University Nijmegen.

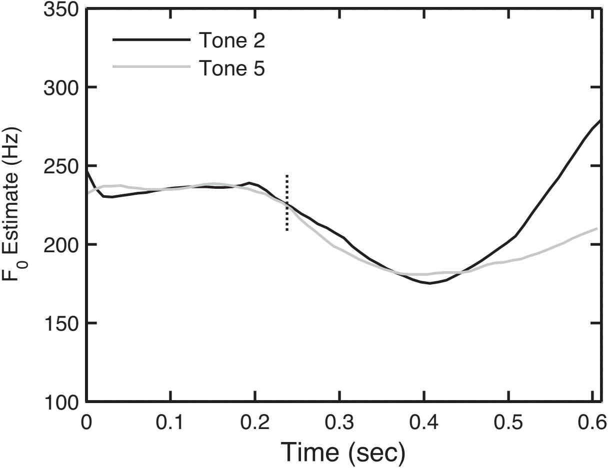
Appendix

Table A.1 ***Distribution of the age of testing of participants in each language group (n)***

Age of testing	HS	HK
5-6	1	1
6-7	8	5
7-8	17	9
8-9	9	13
9-10	16	10
10-11	14	11
11-12	2	12
12-13	0	3



350
300
250
200
150
100
0
0.1
0.2
0.3
0.4
0.5
0.6
Tone 1
Tone 4
F0 Estimate (Hz)
Time (sec)


Tone 2
Tone 5
350
300
250
200
150
100
0
0.1
0.2
0.3
0.4
0.5
0.6
F_0 Estimate (Hz)
Time (sec)