Phonetic and lexical encoding of tone in Cantonese heritage speakers

Rachel Soo^{1,2,3} and Philip J. Monahan^{2,3,4}

1. Department of Linguistics, University of British Columbia, Canada
2. Department of Language Studies, University of Toronto Scarborough, Canada
3. Department of Linguistics, University of Toronto, Canada
4. Department of Psychology, University of Toronto Scarborough, Canada

Contact Information:

Rachel Soo
Department of Linguistics
University of British Columbia
Totem Field Studios
2613 West Mall
Vancouver, BC Canada V6T 1Z4
rachel.soo@ubc.ca

Running title: Heritage speaker translation priming

Abstract

Heritage speakers contend with at least two languages: the less dominant L1, that is, the heritage language, and the more dominant second language (L2). In some cases, their L1 and L2 bear striking phonological differences. In the current study, we investigate this in the context of Toronto-born Cantonese heritage speakers and their maintenance of Cantonese lexical tone, a linguistic feature that is absent from English, the more dominant L2. Across two experiments, Cantonese heritage speakers were tested on their phonetic/phonological and lexical encoding of tone in Cantonese. Experiment 1 was an AX discrimination task with varying inter-stimulus intervals (ISIs), which revealed that heritage speakers discriminated tone pairs with disparate pitch contours better than those with shared contours. Experiment 2 was a medium-term repetition priming experiment, designed to extend the findings of Experiment 1 by examining tone representations at the lexical level. We observed a positive correlation between English dominance and priming in tone minimal pairs that shared contours. Thus, while increased English dominance does not affect heritage speakers' phonological-level representations, tasks that require lexical access suggest that heritage Cantonese speakers may not robustly and fully distinctively encode Cantonese tone in lexical memory.

Keywords

Heritage speakers, Cantonese, Lexical Tone, Lexical Decision, Priming, Phonetic Encoding

1 Heritage speaker phonology

Sensitivity to phonetic contrasts is evident in the earliest stages of infancy (Aslin et al., 2002; Eimas, 1975; Eimas et al., 1971; Kuhl, 2004; Marean et al., 1992; Werker & Tees, 1984), and monolingual listeners largely use native language sound categories present at the end of their first year of life to encode lexical items. Those who learn a second language (L2) later in adulthood (i.e., late L2 learners) sometimes show L2 interference effects during first language (L1) processing (Ju & Luce, 2004; Spivey & Marian, 1999), but the L1 typically remains the dominant language (Birdsong, 2014).

The acquisition profile of second-generation heritage speakers, however, contrasts with monolinguals and late L2 learners (Kondo-Brown, 2006; Montrul, 2010, 2011a, 2011b; Polinsky, 2011; Polinsky & Kagan, 2007; Valdés, 2005). Second-generation heritage speakers (henceforth, heritage speakers) are raised in an environment where they are exposed to a locally ethnolinguistically minority language at birth (Benmamoun et al., 2013; Montrul, 2012). This “heritage language” is learned and utilized at home, but heritage speakers also acquire and ultimately transition to using the majority societal language more in daily life. As a result, the dominant language of adult heritage speakers is not the L1, but the L2. Early exposure alone to the heritage language confers lasting benefits. That is, even if the heritage language is not used regularly in adulthood, childhood exposure results in enhanced performance relative to late L2 learners in both perception and production (Au et al., 2002; Knightly et al., 2003; Oh et al., 2003). While such early exposure places heritage speakers at an advantage over late L2 learners, this acquisition trajectory, coupled with L2 dominance, also makes heritage speakers incomparable to native speakers who are more dominant in the L1 (Chang et al., 2011; Godson, 2004; Kan, 2020; Kang et al., 2016; Saadah, 2011). Although some speech sound categories may be shared by the L1 and L2 (e.g., Chang et al., 2011; Kang & Nagy, 2016; Nodari et al., 2019; Ronquest, 2012; Saadah, 2011), in some cases, there may be features of the heritage

language that the L2 simply does not possess. While establishing separate categories for sounds shared across languages may be difficult, there may also be difficulty maintaining phonetic and linguistic dimensions that are utilized in one language but not the other.

In the current paper, we investigate how heritage speakers maintain linguistic features that are routinely utilized in the heritage language for lexical contrast but play no role in the dominant language. Specifically, we test second-generation Toronto-born Cantonese heritage speakers, whose L2 is English, on their ability to maintain Cantonese lexical tone. Sensitivity to lexical tone is observed early in life (Mattock et al., 2008; Mattock & Burnham, 2006; Yeung et al., 2013), and while there is considerable individual variation, children learning a tone language typically acquire the tone system of the ambient language while they are still making segmental errors (Li & Thompson, 1977; Tse, 1978). If Cantonese heritage speakers have fully acquired their tone system early in life, then the question arises: To what extent can they maintain these contrasts in the face of their more dominant L2, English, where lexical tone is absent? With this in mind, we carried out two experiments. Experiment 1 was designed to study heritage speaker processing of tone at a phonetic and phonological level using an AX discrimination task with short and long inter-stimulus intervals (ISIs), respectively. We extend these findings in Experiment 2 by carrying out a medium-distance repetition priming task to examine the extent to which Cantonese heritage speakers encode tone at a lexical level.

1.1 Cantonese tone processing

Cantonese is a Sino-Tibetan language spoken in Hong Kong, Macau, and Guangzhou (Eberhard et al., 2020). It has six lexical tones that occur on open syllables with optional nasal codas (Bauer & Benedict, 1997; Matthews & Yip, 2013). See Table 1, which also includes information about items used in Experiment 1. These lexical tones are distinguished by pitch height, contour, and magnitude of change (Fok, 1974; Gandour, 1981; Khouw & Ciocca, 2007). Among these six tones, there are three level tones (T1, T3, T6), two rising tones (T2, T5) and one falling tone (T4). The level tones mainly differ in pitch height: T1 has the highest pitch,

followed by T3, followed by T6, which has the lowest pitch. The rising tones are characterized by a rising contour, with T2 ending at a higher pitch than T5. Finally, the falling tone, T4, is produced with a falling contour and typically contains considerable glottalization (Yu & Lam, 2014).¹

Table 1. Experiment 1 stimuli durations for each speaker (6 tokens × 2 speakers). Tone numbers are provided in addition to the numerical tone notation system devised by Chao (1930) in square brackets. For each word, a Chinese character, its Jyutping (Tang et al., 2002), and English gloss are included. Additionally, for each token, the arithmetic means (in ms), both by-talker and by-tone are provided.

	Contour	Word	Male duration (ms)	Female duration (ms)	Mean duration (ms)
T1 [55]	Level	衣 <i>ji1</i> “clothes”	583	680	632
T3 [33]		意 <i>ji3</i> “idea”	698	737	718
T6 [22]		二 <i>ji6</i> “two”	778	632	705
T2 [25]	Rising	椅 <i>ji2</i> “chair”	512	700	606
T5 [23]		耳 <i>ji5</i> “ear”	590	684	637
T4 [21]	Falling	疑 <i>ji4</i> “suspicious”	580	629	605
		Mean (ms)	624	677	650

Pitch height, contour shape, and direction all appear to be important cues to Cantonese tone (Gandour, 1981; Tse, 1978; Vance, 1976), and native Cantonese speakers weigh pitch height most heavily (Gandour, 1983). Empirical studies on heritage speaker lexical tone processing provide conflicting results, making it difficult to examine the relative weighting of these dimensions in an L2-dominant group. So (2000) tested heritage speakers on Cantonese tone discrimination in a six-alternative forced choice task using pictures. She observed overall higher accuracy rates for native speakers than for heritage speakers, but similar confusion patterns between T2-T5 and T4-T6. In a word identification paradigm, Lam (2018) also found that “homeland” speakers are better than heritage speakers at using tonal information alone to discriminate words. Conversely, in a discrimination paradigm, Soo and Monahan (2017) found

¹ Several lexical tones are undergoing tone mergers (Lam, 2018; Mok et al., 2013). We return to this issue in our discussion of the results of Experiment 1 and Experiment 2.

that heritage speakers performed as well as native speakers. Overall, however, both groups performed worse on pairs that shared contour type (e.g., rising-rising, T2-T5) compared to pairs that had disparate contour types (e.g., level-falling, T1-T4). Kan and Schmid (2019) likewise found that Cantonese heritage children (aged 5-11 years) performed worse on tone contrasts that share contour type compared to contrasts with disparate contour types in a discrimination task; however, in contrast to Soo and Monahan (2017), they found that heritage children performed worse than the control native Cantonese speaking children. A similar population of Cantonese heritage children also received lower native-likeness and comprehensibility ratings in their production (Kan, 2020). Each of these studies defined heritage speakers in a distinct manner. As such, observed differences in heritage speaker tone perception might stem from differences in how heritage speakers were defined.

1.2 Lexical-phonological representations in bilinguals

It has long been known that the L2 can influence L1 speech perception in bilinguals (Flege, 1987; Flege & Eefting, 1987; Major, 1992; Williams, 1979); however, later work realized a more nuanced set of findings when tasks require access to the lexicon. In a study by Pallier et al. (2001), highly proficient Catalan-Spanish bilinguals were tested on a medium-term repetition priming experiment designed to investigate how robustly Catalan-specific phonetic contrasts are lexically represented. It was previously established that this bilingual population was equally proficient in each language yet differed in their phonetic categorization of Catalan specific contrasts (Bosch et al., 2000; Pallier et al., 1997; Sebastián-Gallés & Soto-Faraco, 1999). In Pallier et al. (2001), half of the participants acquired Catalan first, and the other half acquired Spanish first; all participants acquired their L2 prior to 6-years old. Medium-term repetition priming requires participants to make lexical decisions on a series of isolated words and nonwords that are inherently paired as primes and targets across eight to 20 trials. Unlike immediate priming paradigms, medium- and long-term priming paradigms appear to tap into how phonological contrasts are lexically stored (Sumner & Samuel, 2009). Moreover, in

medium-term priming, nonwords and minimal pairs do not show priming effects. In Pallier et al. (2001), Spanish-first bilinguals showed medium-term priming for the [ɛ]-[e] Catalan minimal pairs (e.g., [nɛtə] “granddaughter”, [netə] “clean” (fem.)), as if the two vowels were members of the same category and consequently treated as identity pairs. Catalan-first bilinguals did not show medium-term priming for the minimal pairs, suggesting that these two vowel categories were contrastive in their lexicon. In sum, the findings of Pallier et al. (2001) indicated that even extremely proficient bilinguals show asymmetries in how phonological contrasts are maintained in the lexicon, and these representations are shaped by acquisition patterns early in life. As such, certain tasks that require lexical access can reveal subtle differences between highly proficient bilingual groups in their encoding of lexical-phonological representations, which may otherwise be difficult to distinguish.

1.3 Current study

Aside from Lam (2018), previous studies on Cantonese heritage speakers utilized tasks that tap phonetic discrimination and identification but not lexical access. While studies on heritage speaker tone perception generally converge on the notion that heritage speakers do not outperform native speakers, investigating lexical access allows us to begin to understand the robustness of heritage speakers’ tonal representations. More broadly, heritage Cantonese speakers represent an ideal test case to understand the maintenance of lexical tone following the acquisition of a more dominant L2, given the complex Cantonese lexical tone system. In the current study, we tested Cantonese heritage speakers and Cantonese native speakers on the phonetic/phonological discrimination and lexical encoding of Cantonese tone in two experiments.

In Experiment 1, we used an AX discrimination task, similar to Soo and Monahan (2017). Three groups of participants were tested: heritage Cantonese speakers, native Cantonese speakers, and native English speakers. The principal difference between the experimental design in Soo and Monahan (2017) and the current experiment was the addition of a 2500 ms

ISI, in addition to the 500 ms ISI. These ISIs were selected following Werker and Logan (1985), who used three distinct ISIs to induce three putative levels of speech processing: A 250 ms ISI appeared to induce acoustic-level processing, an 500 ms ISI appeared to induce a phonetic-level analysis of the signal, and a 1500 ms ISI appeared to force listeners to rely on phonemic representations to discriminate sound pairs. As such, in the current experiment, a shorter (500 ms) ISI and a longer (2500 ms) ISI were expected to force listeners to rely on phonetic or more abstract phonological representations, respectively (Werker & Logan, 1985; Yu et al., 2017). Following Soo and Monahan (2017), we predicted that both heritage and native Cantonese speakers would have higher discrimination sensitivity to pairs that had disparate contour types (e.g., rising versus level) than to pairs that shared contour types (e.g., both rising) in the short ISI condition. Moreover, we predicted that heritage speakers would have difficulty encoding tone at a deeper, phonological level of representation, and thus, poorer tone discrimination in the long ISI condition.

In Experiment 2, a medium-term repetition priming experiment similar to Pallier et al. (2001) was carried out. This task taps into lexical-phonological representations and reveals perceptual asymmetries even in extremely proficient bilingual participants. Like the Catalan-Spanish bilinguals in Pallier et al. (2001), these Cantonese heritage speakers acquired linguistic features of their L1 at an early age; however, in the case of lexical tone in Cantonese, they must additionally maintain sensitivity to a linguistic dimension (i.e., lexical tone) that is entirely absent from the L2. Thus, if heritage Cantonese speakers do not robustly encode lexical tone, potentially because their more dominant L2 (i.e., English) does not use lexical tone, we predicted increased repetition priming for tone minimal pairs as a function of increased L2 dominance. This is akin to the Spanish-first bilinguals who showed repetition priming for the [ɛ]-[e] pairs, treating them like identity pairs. For the Cantonese native speakers, only priming in the identity condition was expected.

2 Experiment 1: AX discrimination

2.1 Participants

Eighty native Cantonese, heritage Cantonese and native English speakers were recruited from the student population at the University of Toronto Scarborough. One participant was excluded from the analysis due to technical issues. This left 79 total participants. There were 26 native Cantonese speakers (19 female, mean age = 20.7 years, standard deviation (sd) = 1.9 years), 28 heritage Cantonese speakers (17 female, mean age = 19.4 years, sd = 1.4 years) and 25 native English speakers (18 female, mean age = 19.9 years, s = 2.6 years). Our sample sizes are comparable to previous studies investigating similar populations and questions (Kan, 2020; Lam, 2018; So, 2000; Yang, 2015).

All native Cantonese speakers learned Cantonese from birth, had parents who were native Cantonese speakers and self-rated Cantonese as their most proficient language in listening and speaking. Native Cantonese participants were raised in a location where Cantonese is the primary language (e.g., Hong Kong, Macau) and lived an average of 4.1 years in an English-speaking region of Canada (sd = 2.9 years; range: 5 months:10 years). Finally, native speakers self-reported using Cantonese 61.7% (sd = 16.0%) of the time in a typical day, compared to English, which they report using 25.6% (sd = 13.2%) in a typical day. All heritage Cantonese speakers had native Cantonese speaking parents, were formally educated in a language other than Cantonese (26 in English, 1 in French and 1 in Mandarin) and had lived in Canada for most of their life (mean = 18.5 years, sd = 2.91 years). Moreover, all heritage Cantonese speakers reported learning Cantonese from birth and had an average age of English acquisition of 2.79 years (sd = 2.44 years). In self-reported speaking and listening proficiency ratings, the heritage Cantonese speakers reported higher proficiency in English (speaking mean = 9.86/10, sd = 0.45; listening mean = 9.82/10, sd = 0.77) compared to Cantonese (speaking mean = 7.26/10, sd = 2.28; listening mean = 7.82/10, sd = 1.86). Finally, our heritage speakers

self-reported using English 72.7% (sd = 18.0%) of the time in a typical day, compared to Cantonese, which they report using 24.1% (sd = 16.8%). None of the native English participants reported familiarity with Cantonese or any other tone language.

Prior to the experiment, all participants also filled out a language background questionnaire. Heritage and native Cantonese speakers additionally completed the Bilingual Language Profile questionnaire (BLP; Gertken et al., 2014). The BLP assesses relative language dominance for a pair of languages, in this case, Cantonese, and English, and provides a quantitative dominance score on a scale of -218 to +218. In the current implementation of the BLP, a positive value indicates greater dominance in English relative to Cantonese, while a negative value indicates greater dominance in Cantonese relative to English. In particular, the questionnaire provides a holistic representation of language dominance by asking participants to provide self-reports of their language history (e.g., age of acquisition, education, family), language use (e.g., work, and familial settings), language proficiency (e.g., Likert scale ratings for reading, writing, speaking, and understanding), and language attitudes (e.g., cultural ties to linguistic identity). As is clear from the distribution of BLP dominance scores in Figure 1A, nearly all native Cantonese speakers in Experiment 1 are Cantonese-dominant, while all heritage Cantonese speaker are English-dominant. Thus, while native speakers were recruited in Toronto, what differentiated them from heritage speakers was their language dominance. Participants reported no known hearing, language, or neurological deficits. Instructions were provided to participants in either Cantonese for native Cantonese speakers or in English for heritage Cantonese and native English speakers. All participants provided written informed consent prior to the experiment and were remunerated for their time.

A Experiment 1: AX Discrimination

B Experiment 2: Medium-Term Repetition Priming

Figure 1. Distribution of BLP scores for the heritage and native Cantonese participants in (A) Experiment 1 and (B) Experiment 2. Each point represents an individual participant. Negative values reflect greater Cantonese dominance, while positive values reflect greater English dominance. Across both experiments, most native Cantonese participants scored as Cantonese dominant, whereas most heritage Cantonese participants scored as English dominant.

2.2 Stimuli

The items in the experiment were the syllable [ji] produced with all six lexical tones (see Table 1). This syllable was selected because a word is created when combined with each of the six tones. Moreover, it has been used in several previous studies of Cantonese tone perception with both native speakers (Ching, 1984; Fung & Lee, 2019; Jia et al., 2015; Mok et al., 2013; Tsang et al., 2011) and heritage speakers (Lam, 2018; Soo & Monahan, 2017). Both a male and a female native speaker of Cantonese recorded the stimuli. Each talker produced each syllable three times. The best token for each syllable for each talker was selected for inclusion in the experiment. We defined the best tokens as those that were free from artifacts or

mispronunciations. This resulted in 12 distinct syllables (6 tones $\times$ 2 genders $\times$ 1 token).

Recordings were made in a sound-attenuated cabin with a sampling frequency of 44.1 kHz and a 16-bit depth. All stimuli were digitally scaled to have an equal root mean square (RMS) intensity in Praat (Boersma & Weenink, 2020).

Figure 2. Pitch contours for each stimulus in Experiment 1. Tokens produced by the female talker are in the left panel, and tokens produced by the male talker are in the right panel. Tone category is coded by colour and labels to the right of each tone contour.

Average stimuli durations are provided in Table 1. No duration adjustments were made to keep the syllables as natural as possible. Figure 2 presents the contours for each of the 12 tokens included in the current experiment. The average frequency for the female talker (mean = 194

Hz, $sd = 21$ Hz) was higher than that for the male speaker (mean = 112 Hz, $sd = 22$ Hz), as expected. The reason for including different talkers was twofold: 1) to increase task difficulty, especially for native speakers and 2) to prevent listeners from performing low-level acoustic matching between the two syllables within a trial. The pitch floor for our female talker (Tone 4, minimum pitch = 153 Hz) is approximately the pitch ceiling for our male talker (Tone 1, maximum pitch = 160 Hz), indicating that the tone spaces are quite physically distinct from one another.

2.3 Procedure

The experimental method was an AX discrimination task. Participants were seated in front of a computer monitor and wore Sennheiser HD 380 PRO headphones. The volume was adjusted to a comfortable level and remained constant across participants. The experiment was delivered using PsychoPy (Peirce et al., 2019). Each trial began with the presentation of a fixation point (“+”) which remained on the screen for 500 ms, after which the trial would begin. On each trial, participants listened to a pair of Cantonese [ji] syllables. All trials included one male-produced syllable and one female-produced syllable. The order of presentation of the male and female tokens was fully counterbalanced across trials, such that on half of the trials the female token was presented first and on the other half of trials, the male token was presented first.

Participants were asked to respond via keypress on a computer keyboard as to whether the two tokens on a given trial were the same (press “q”) or different (press “p”). There were six ‘same’ tone pairs (e.g., T1-T1) and 15 ‘different’ tone pairs (e.g., T1-T5). All 21 possible tone combinations were tested. Each ‘same’ pair was presented 20 times, while each ‘different’ pair was presented eight times. In total, there were 240 trials randomized in each block: 120 ‘same’ and 120 ‘different’ trials. Each participant completed two counterbalanced blocks, which were identical except for the ISI between the first and second members of a pair. As mentioned, previous reports have shown that the duration of the ISI affects the linguistic level at which participants are making judgments. Shorter intervals tap into phonetic-levels of representation,

while longer intervals tap into phonological-levels of representation (Werker & Logan, 1985).

Thus, in one block, the ISI was 500 ms (designed to assess phonetic-level processing), while in the other, the ISI was 2500 ms (designed to assess phonological-level processing). The order of blocks was counterbalanced across participants. The inter-trial interval (ITI) was 1000 ms.

Participants completed four practice trials at the beginning of each block to familiarize them with the task.

2.4 Results

All data aggregation and visualization were carried out using the packages {dplyr} (Wickham et al., 2020) and {ggplot2} (Wickham, 2016) in the statistical software R (R Core Team, 2019).

First, trials with extreme reaction times (i.e., > 10 seconds) were removed (less than 0.6% of the data), then, all trials with reaction times greater or less than 2.5 standard deviations from an individual's mean reaction time were excluded (3.7% of the data). Figure 3 presents the d' scores for all pairwise comparisons. Since previous studies have shown that shared tone contour types cause difficulty for Cantonese heritage speakers (Kan & Schmid, 2019; Lam, 2018; Soo & Monahan, 2017), tone pairs were collapsed into two categories for statistical analysis: those that shared pitch contour type ('shared', i.e., level: T1-T3, T1-T6, T3-T6; rising: T2-T5) and those that had disparate pitch contour types ('disparate', i.e., T1-T2, T1-T4, T1-T5, T2-T3, T2-T4, T2-T6, T3-T4, T3-T5, T3-T6, T4-T5, T4-T6, T5-T6). Figure 4 presents the d' scores and distributions for each participant group and tone contour type in the short and long ISI condition. Overall, pairs that shared contour types were more difficult to discriminate than those that had disparate contour types. This corroborates discrimination patterns observed in Soo and Monahan (2017), who tested similar items with only a 500 ms ISI.

Figure 3. Pairwise d' scores for each participant group for each tone pair. Numbers along x-axis refer to tone pairs (e.g., 12 = T1-T2). The dashed vertical line separates tone pairs with disparate contour types (e.g., T1-T4, left side) from those with shared contour types (e.g., T1-T3, right side). The dotted horizontal line is the intercept of the model in the Appendix (also the overall mean d' score). Error bars represent the standard error of the mean.

Figure 4. Distribution of responses for d' scores by whether the trial contained pairs that shared contour types (e.g., T1-T3) or had disparate contour types (e.g., T1-T4). Shading represents short (white) or long (grey) ISIs. Points are individual participant means. Black points are condition means. Error bars represent the standard error of the mean.

Subsequently, our results were submitted to a generalized linear mixed-effects model using the package `{lme4}` (Bates et al., 2015). The model included the simple-coded fixed effects of Contour (Disparate^{Ref}, Shared), Group (Native Cantonese^{Ref}, Heritage Cantonese, Native English), and ISI (Short ISI (500 ms)^{Ref}, Long ISI (2500 ms)) and their interactions. Simple coding allows us to interpret main effects as opposed to simple effects. The observation variable was d' score, which provides a measure of discrimination sensitivity (Macmillan & Creelman, 2004). Higher d' scores indicate better discrimination sensitivity, while a d' score of 0 represents chance-level discrimination sensitivity. The model's random effects structure included random by-participant slopes for Contour Type and Block, as well as random by-participant and by-item intercepts. This model was selected using stepwise model comparison based on an Akaike's Information Criteria (AIC), starting from a model with random by-participant slopes for Contour and Block and their interaction, as well as a random by-item slope for Block, and random by-participant and by-item intercepts. Degrees of Freedom were estimated with Satterthwaite's method as implemented in the package `{lmerTest}` (Kuznetsova

et al., 2017). The residuals were checked for homoscedasticity and approximately followed a normal distribution. For completeness, the output of the model with all pairwise comparisons is provided in Appendix A. Table 2 presents the output of the model.

Table 2. Model output for the linear mixed effects model. The dependent variable is d' score. Contour, ISI and Group were coded using simple coding. Significance codes: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

	β	SE	t	p	
(Intercept)	1.24	0.22	5.64	< 0.001	***
Contour	-1.56	0.43	-3.61	0.003	**
ISI	-0.11	0.05	-2.09	0.039	*
Group: HC	-0.12	0.13	-0.91	0.367	
Group: NE	-0.89	0.14	-6.51	< 0.001	***
Contour $\times$ ISI	-0.13	0.08	-1.66	0.097	.
Contour $\times$ Group: HC	-0.14	0.16	-0.87	0.386	
Contour $\times$ Group: NE	0.69	0.17	4.08	< 0.001	***
ISI $\times$ Group: HC	0.03	0.13	0.26	0.798	
ISI $\times$ Group: NE	< 0.001	0.13	0.00	0.998	
Contour $\times$ ISI $\times$ Group: HC	0.11	0.19	0.60	0.548	
Contour $\times$ ISI $\times$ Group: NE	0.15	0.19	0.76	0.447	

We observed a main effect of Contour ($\beta = -1.56$, $SE = 0.43$, $t = -3.61$, $p < 0.001$), with higher d' scores for pairs with disparate contour types (d' : mean = 2.04, sd = 1.35) than pairs with shared contour type (d' : mean = 0.47, sd = 1.22). There was also a main effect of ISI ($\beta = -0.11$, $SE = 0.05$, $t = -2.09$, $p = 0.039$), with higher d' scores (d' : mean = 1.66, sd = 1.49) for the shorter ISI than the longer ISI (d' : mean = 1.58, sd = 1.49), although the effect size was small (Cohen's $d = 0.099$). There was no difference between the heritage (d' : mean = 1.98, sd = 1.43) and native Cantonese speakers (d' : mean = 1.90, sd = 1.44); however, the native English speakers showed lower discrimination sensitivity (d' : mean = 0.93, sd = 1.36) than the native Cantonese speakers. To test the relative performance of the heritage Cantonese speakers compared to the native English speakers, we performed a post-hoc tests using the {phia}

package in R (Rosario-Martinez, 2015). The holm method for multiple comparisons was applied. Heritage Cantonese speakers showed better overall discrimination sensitivity to the tone pairs in the experiment compared to the native English speakers ($\chi^2(1) = 32.79, p < 0.001$). Finally, we observed a Contour $\times$ Group: NE interaction ($\beta = 0.69, SE = 0.17, t = 4.08, p < 0.001$). As above, a post-hoc test was carried out to assess the relative difference between groups for pairs with shared and disparate contours. For both shared and disparate pairs, Native English speakers performed significantly worse than native Cantonese (Disparate: $\chi^2(1) = 44.57, p < 0.001$; Shared: $\chi^2(1) = 17.22, p < 0.001$) and heritage Cantonese speakers (Disparate: $\chi^2(1) = 42.54, p < 0.001$; Shared: $\chi^2(1) = 7.53, p < 0.05$).

2.5 Interim discussion

Overall, listeners had difficulty with the task. Across all three groups and ISI conditions, participants showed poorer discrimination sensitivity when pairs shared contour type (e.g., T1-T3) compared to when they possessed disparate contour types (e.g., T1-T4). This was true even for the native Cantonese speakers, who typically show high discrimination sensitivity in previous reports. For instance, Fung and Lee (2019) observed ceiling-level performance across most comparisons. The native Cantonese speakers in the current study had d' scores of less than 3.0, and for pairs that shared contour type, their mean d' was less than 1.0, indicating relatively poor discrimination sensitivity. It is worth noting that Fung and Lee (2019) had the same talker produce both syllables in a pair, making their stimuli much more acoustically similar than the tokens used here. The large physical differences in pitch between our two talkers (see Figure 2) required listeners to normalize across these disparate pitch ranges, which likely increased task difficulty. In particular, this experimental design might have encouraged a phonological-level analysis of the stimuli, even in the short ISI block. The inclusion of two talkers in each AX trial pair with phonetically distinct tone ranges (see Figure 1) potentially forced participants to compare each member of a trial-pair at a phonological level irrespective of the ISI. If true, this would eliminate the hypothesis that the two ISIs encouraged an analysis of the

stimuli at distinct levels of representation, that is, the short ISI block encourages a phonetic-level analysis, while the long ISI block encourages a phonological-level analysis. We leave further exploration of this possibility for future research. Furthermore, there was a main effect of ISI, although the effect size was small and similar across participant groups (native Cantonese $\Delta d' = 0.069$; heritage Cantonese $\Delta d' = 0.063$; native English $\Delta d' = 0.10$); the largest difference was for the native English speakers. The fact that this difference of a similar size was evident across all three groups suggests that these effects are not likely due to phonological encoding of tone, as the English speakers lack lexical tone in their phonology. Instead, these small differences may be attributed to the added cost of retaining pitch information in auditory memory.

Finally, as expected, the native English speakers, who had no familiarity with Cantonese and would need to rely on acoustic representations alone, showed lower discrimination sensitivity compared to the native and heritage Cantonese speakers, while there was no difference between the two Cantonese groups. The lack of a difference between native and heritage Cantonese speakers suggests that they do not encode tone differently in a task that requires auditory comparison at a phonetic, and putatively phonological, representation level. While each of the items in Experiment 1 was an existing Cantonese word, the nature of the task forces participants to focus on the phonetic characteristics of the stimuli, and in particular, their pitch contour. Thus, it is unclear whether Cantonese heritage speakers encode tone in a manner similar to native Cantonese speakers when they are required to access the lexicon and store words in auditory memory. To test this, in Experiment 2, we performed a medium-term auditory repetition priming experiment. Neither minimal pairs nor nonwords show priming in this task, which has been shown to distinguish between highly proficient bilingual speakers in terms of lexical and phonological processing (Pallier et al., 2001). In particular, Experiment 2 required participants to make continuous lexical decision judgments to individual words and nonwords that were occasionally followed by an identical repetition or an item that differed in lexical tone (i.e., identity pairs and tone minimal pairs). Recall that in Experiment 1, discrimination sensitivity

was significantly lower for pairs with shared contour types than for pairs with disparate contour types across all participant groups and ISIs (see Figure 4). If tone is not as well encoded in heritage Cantonese speakers, we predict that they will show repetition priming effects in response to tone minimal pairs, especially when they have a shared contour type. This is akin to the findings in Pallier et al. (2001), where even extremely proficient Spanish-Catalan bilinguals whose L1 was Spanish showed repetition priming to [ɛ]-[e] minimal pairs, suggesting that they heard the two words as the same. Concretely, in our experiment, this should be manifested in the form of increased repetition priming for tone minimal pairs as a function of increased English dominance.

3 Experiment 2: Medium-term auditory repetition priming

3.1 Participants

Thirty-three heritage Cantonese speakers were recruited from the University of Toronto Scarborough. Thirty-five native Cantonese speakers were recruited from the University of Toronto Scarborough ($n = 14$) and from Hong Kong University ($n = 21$). Two heritage Cantonese participants were excluded as they did not acquire Cantonese as a first language, and two additional heritage speakers were excluded as they did not complete the BLP. This left 29 heritage Cantonese speakers (21 female, mean age = 20.8 years, $sd = 2.2$ years) and 35 native Cantonese speakers (mean age = 22.7 years, $sd = 4.9$ years) for remaining analyses.

All native Cantonese speakers learned Cantonese from birth and had at least one Cantonese-speaking parent. All native Cantonese participants were raised in a location where Cantonese is the primary language (e.g., Hong Kong, Guangdong), and participants recruited in Toronto lived an average of 5.3 years in an English-speaking region of Canada ($sd = 6.3$ years; range: 3 months:10 years). The average English age of acquisition for all native speakers was 4.7 years ($sd = 2.9$ years). In self-reported speaking and listening proficiency ratings, all native Cantonese speakers reported higher proficiency in Cantonese (speaking mean = 9.67/10, $sd =$

0.99; listening mean = 9.67/10, sd = 1.05) compared to English (speaking mean = 6.63/10, sd = 2.04; listening mean = 6.97/10, sd = 1.95). Finally, native speakers self-reported using Cantonese 64.3% (sd = 21.9%) of the time in a typical day, compared to English, which they report using 21.7% (sd = 15.0%). All heritage Cantonese speakers learned Cantonese from birth and had at least one Cantonese speaking parent. Their average English age of acquisition was 2.2 years (sd = 2.1 years). In self-reported speaking and listening proficiency ratings, heritage Cantonese speakers reported higher proficiency in English (speaking mean = 9.63/10, sd = 1.03; listening mean = 9.57/10, sd = 1.19) compared to Cantonese (speaking mean = 6.76/10, sd = 2.63; listening: mean = 6.72/10, sd = 2.53). Finally, our heritage speakers self-reported using English 63.9% (sd = 23.7%) of the time in a typical day, compared to Cantonese, which they report using 30.6% (sd = 22.5%). As in Experiment 1, prior to the main task, participants filled out a language background questionnaire and completed the BLP (Gertken et al., 2014). The distribution of BLP dominance scores in Figure 1B show that most of the native Cantonese speakers (both those recruited from Toronto and from Hong Kong) in our experiment were more dominant in Cantonese, while all heritage Cantonese speakers were more dominant in English. Participants reported no known hearing, language, or neurological deficits. All participants provided written informed consent prior to the experiment and were remunerated for their time.

3.2 Stimuli

Forty-eight Cantonese word minimal pairs were selected for the study. All items were monosyllabic with an onset consonant, a vowel, and optionally a nasal coda. Of these 48 pairs, 16 were tone minimal pairs (see Table 3). Considering the results from Experiment 1, these tone minimal pairs were created to either share tone contour type (i.e., two minimal pairs for T1-T3, three minimal pairs for T3-T6, three minimal pairs for T2-T5) or have disparate tone contour types (i.e., three minimal pairs for T3-T4, three minimal pairs for T4-T5, two minimal pairs for T5-T6). In addition to the tone minimal pairs, 16 vowel minimal pairs and 16 consonant minimal

pairs were created, which represented filler trials. For each tone, consonant, and vowel minimal pair, an identity pair was created by duplicating the first member of each pair. The minimal pairs and their corresponding identity pairs were counterbalanced across two separate lists.

Frequency counts were acquired from the Hong Kong Cantonese Corpus (Luke and Wong, 2015) using PyCantonese (Lee, 2015). An omnibus ANOVA with the factors Order (First, Second) and Cue (ID, Tone, Vowel, Consonant) revealed no significant main effects or interactions (all $ps > .05$).²

Forty-eight additional monosyllabic nonword pairs were created that likewise differed only in onset consonant, vowel or tone. Nonwords were phonotactically legal Cantonese syllables representing systematic and accidental gaps in the language. All stimuli were produced by a female native speaker of Cantonese who did not produce the stimuli in Experiment 1. Recordings were made in a sound-attenuated cabin with a sampling frequency of 44.1 kHz and a 16-bit depth. All stimuli were digitally scaled to have an equal root mean square (RMS) intensity in Praat (Boersma & Weenink, 2020).

Table 3. Experiment 2 example stimuli. The Jyutping (Tang et al., 2002) is provided for each item, as well as the Chinese character and English gloss for the real words. Pairs were either identical or differed only in tone. For each pair type, there were also nonword pair counterparts.

	Pair type	Prime	Target
Real words	Identity	浪 <i>long2</i> “rinse”	浪 <i>long2</i> “rinse”
	Tone: Disparate	褲 <i>fu3</i> “pants”	扶 <i>fu4</i> “to hold onto”
	Tone: Shared	叫 <i>giu3</i> “to call out”	撬 <i>giu6</i> “to pry open”
Nonwords	Identity	<i>hi3</i>	<i>hi3</i>
	Tone: Disparate	<i>pau4</i>	<i>pau5</i>
	Tone: Shared	<i>su2</i>	<i>su5</i>

3.3 Procedure

² Frequency counts were collected from a native speaker corpus, and the ANOVA served to ensure that the selected stimuli were matched within each member of a Pair and across each Cue. We cannot ensure that the items in Experiment 2 are as familiar to heritage speakers as they are to native speakers; however, the accuracy data, broken down by pair type in Table 4, indicates that both speaker groups had high accuracy rates (all > 80%) in the lexical decision task.

Participants were seated in front of a computer monitor and wore Sennheiser HD 380 PRO headphones to complete the task. The volume was adjusted to a comfortable level and remained constant across participants. The experiment was delivered using PsychoPy (Peirce et al., 2019). In the medium-term auditory repetition priming task, participants made lexical decision responses to the individual items in identity and minimal pairs. Specifically, on each trial, participants were presented with a fixation point (“+”) for 500 ms, followed by an auditory token (e.g., 叫 *giu3* “to call out”). On each trial, participants were instructed to respond as quickly and as accurately as possible via keyboard press as to whether the stimulus was a real or nonword of Cantonese. They pressed “f” if they thought the stimulus was a real word of Cantonese and “j” if they thought the stimulus was a nonword. Following Pallier et al. (2001), the corresponding member of the pair was presented separately eight to 20 trials later (e.g., 撬 *giu6* “to pry open”). The ITI was 1000 ms. Participants completed four practice trials at the beginning of the experiment to familiarize them with the task. Experimental trials were presented in a pseudo-randomized order that was fixed across participants to ensure the eight to 20 trial-spacing.

3.4 Results

All data aggregation and visualization were carried out using the packages {dplyr} (Wickham et al., 2020) and {ggplot2} (Wickham, 2016) in the statistical software R (R Core Team, 2019). Trials with extreme reaction times (i.e., > 5 seconds) were removed from the analysis (2.9% of the data). Then, all trials with reaction times greater or less than 2.5 standard deviations of an individual’s mean reaction time were excluded. If either member of a pair was considered an outlier based on these criteria, the corresponding member was removed (6.4% of the data). Following Pallier et al. (2001), priming magnitudes were calculated by subtracting the reaction time to the second item in a pair from the reaction time to the first item in a pair. This

was conducted on accurate tone minimal pairs that i) shared contour type, ii) differed in contour type, and iii) their corresponding identity pairs. The mean priming magnitudes and mean accuracy data for each group and each pair type is summarized in Table 4.

Table 4. Mean priming magnitudes (and mean accuracy in parentheses) for heritage and native speakers for tone identity pairs and tone minimal pairs.

Group	Identity	Tone: Disparate	Tone: Shared
Heritage	+52 ms (84.0%)	-50 ms (81.4%)	+20 ms (80.7%)
Native	+72 ms (85.8%)	+22 ms (87.9%)	-61 ms (85.7%)

Following this, we submitted the log transformed reaction times to a linear mixed effects model (Baayen et al., 2008) using the package {lme4} (Bates et al., 2015). The package {lmerTest} in R (Kuznetsova et al., 2017) was used to estimate degrees of freedom and compute probability values. The model included simple coded fixed effects of BLP Dominance scores as a continuous variable, Pair Type (Identity^{Ref}, Tone: Disparate, Tone: Shared), Order (First^{Ref}, Second) and their interactions. The random effects structure of the model included random by-participant and by-item intercepts, and random by-participant slopes for Order. Degrees of freedom were estimated using Satterthwaite's method. The full output of the linear model is provided in Table 5. We observed a main effect of BLP Dominance scores ($\beta = < 0.01$, $SE = < 0.01$, $t = 2.59$, $p = 0.012$), an interaction between BLP Dominance $\times$ Tone: Shared ($\beta = - < 0.01$, $SE = < 0.01$, $t = -2.43$, $p = 0.015$) and a three-way interaction between BLP Dominance $\times$ Tone: Shared $\times$ Order ($\beta = - < 0.01$, $SE = < 0.01$, $t = -2.73$, $p = 0.006$). We carried out post-hoc tests using the {phia} package in R (Rosario-Martinez, 2015) and the holm method for multiple comparisons. Pairwise comparisons between the First and Second member of a pair provide an approximation of the priming magnitude differences for each Pair type and revealed a significant effect for Tone: Shared pairs ($\chi^2(1) = 6.18$, $p = 0.038$). These post-hoc results are visualized in Figure 5, where the priming magnitudes for the identity and tone minimal pairs are correlated with the BLP dominance scores (Gertken et al., 2014; see Figure 5). This visualization shows that the priming magnitudes differed as a function of participants' dominance in English (as

indexed by the BLP dominance scores) for tone minimal pairs that shared contours. To quantify the strength and direction of the relationship between priming magnitude and dominance scores, we calculated Kendall's rank correlation tau (Kendall, 1955) for each Pair type. The correlation between the priming magnitudes and the dominance scores for identity pairs was not significant ($z = -1.43$, $p = 0.15$, $\tau = -.05$), nor was it significant for tone pairs of disparate contour types was not significant ($z = 0.60$, $p = .54$, $\tau = -0.03$); however, there was a significant positive correlation for tone pairs that shared contour type ($z = 2.84$, $p < 0.01$, $\tau = 0.13$), suggesting that the priming for these tone minimal pairs was stronger for participants who were more dominant in English. Testing the nonword pairs, there were no significant correlations between BLP dominance scores and priming magnitude for any of the conditions (all τ s $< |.04|$, p s > 0.05).

Table 5. Full output of the linear model from the data in Experiment 2. Categorical fixed effects were simple coded, allowing us to interpret main effects. The model included fixed effects of BLP Dominance scores as a continuous variable, Pair Type (Identity^{Ref}, Tone: Disparate, Tone: Shared), Order (First^{Ref}, Repetition) and their interactions. The random effects structure of the model included random by-participant and by-item intercepts, as well as random by-participant slopes for Pair Type. The observation variable was log transformed reaction time. Significance codes: *** < 0.001 , ** < 0.01 , * < 0.05 , . < 0.1 .

	β	SE	t	p	
(Intercept)	0.24	0.03	7.67	< 0.001	***
BLP Dominance	< 0.01	< 0.01	2.59	0.012	*
Tone: Disparate	0.04	0.02	1.63	0.109	
Tone: Shared	0.03	0.02	1.39	0.169	
Order	-0.023	0.03	-0.90	0.373	
BLP Dominance x Tone: Disparate	< -0.01	< 0.01	-0.35	0.723	
BLP Dominance x Tone: Shared	< -0.01	< 0.01	-2.42	0.015	*
BLP Dominance x Order	< -0.01	< 0.01	-0.78	0.433	
Tone: Disparate x Order	0.06	0.05	1.38	0.171	
Tone: Shared x Order	0.04	0.05	0.90	0.370	
BLP Dominance x Tone: Disparate x Order	< -0.01	< 0.01	-0.32	0.749	
BLP Dominance x Tone: Shared x Order	< -0.01	< 0.01	-2.73	0.006	**

Figure 5. Correlation of priming magnitudes (ms) with BLP dominance scores for real word identity pairs and tone minimal pairs with Shared (T1-T3, T2-T5, T3-T6) and Disparate (T3-T4, T4-T5, T5-T6) tone contour type contrasts. Priming magnitudes were calculated by subtracting the reaction time to the second item of a pair from the reaction time to the first item of a pair. Regression line is line of best fit using a linear model. Points represent individual participants, and the shaded area around the regression line is the standard error. **The dotted $y = 0$ intercept line represents no difference in the reaction time between the first and second presentation of a pair (i.e., no medium-term repetition priming).**

3.5 Interim discussion

Experiment 1 showed that heritage and native Cantonese speakers discriminated tone contrasts similarly across both short and long ISIs in an AX discrimination task. Overall, both groups appear to encode tone similarly at a phonetic, and putatively phonological, level of representation, but this leaves open the question of how tone is encoded in the lexicon. We examined this in Experiment 2 with a medium-term auditory repetition priming task. This task required participants to store lexical items in long-term memory and has been shown to elucidate lexical and phonological differences between highly proficient bilingual speakers (Pallier et al., 2001). In this task, minimal pairs were not expected to show priming across medium distances. Thus, observing minimal pair priming would indicate that pairs are being processed like identity pairs and that the cue in question is weakly encoded. In this vein, we

predicted that tone would not be robustly encoded at a lexical level for English-dominant individuals, and that this would be manifested in the form of repetition priming for tone minimal pairs.

While we observed clear identity priming in both the heritage and native speakers, there was greater identity priming for native speakers. Furthermore, as predicted, positive priming magnitudes were also observed for the heritage speakers in the tone minimal pairs that shared tone contour type. This was not the case for native speakers, for whom the priming magnitudes in such pairs were negative, indicative of inhibition. That is, presentation of the first item did not facilitate, and instead slowed recognition of the second item of the pair. Inhibition was also observed for the heritage speakers in the tone minimal pairs of disparate contour types. Though, interestingly, for the native speakers, these were the tone minimal pairs that demonstrated positive priming magnitudes. This finding was echoed in the correlation between dominance scores and priming magnitudes. We observed a significant positive correlation between dominance scores and tone minimal pair priming suggesting that these tones are not only weakly encoded at a phonetic level, but also a lexical level for English-dominant speakers. Specifically, the positive correlation between dominance scores and priming magnitudes was only present for tone minimal pairs representing shared contour types (i.e., T1-T3, T2-T5, T3-T6).

Recall that in the Experiment 1, both groups also showed low discrimination sensitivity to these tone pairs. It is worth noting that two of the three tone pairs that share contour type are reported to be merging in the literature: T2-T5 and T3-T6 (Mok et al., 2013). The T2-T5 merger is well-established in both production (Bauer & Benedict, 1997; Kej et al., 2002; Yiu, 2009) and perception (Lee et al., 2015; Mok et al., 2013; Yiu, 2009), while the status of the T3-T6 merger is less certain, with conflicting reports about its status in production versus perception (Fung & Lee, 2019; Mok et al., 2013; Peng & Wang, 2005). While our results may provide evidence for these mergers in both perception and recognition, it is unlikely that our results are due to the

mergers alone, as this explanation would leave the data for T1-T3 (not reported to be merging) unaccounted for. Thus, it is unclear if these results are due to the mergers or since both members of each pair share a tone contour type.

That tone encoding is potentially modulated by relative language dominance hints at the heterogeneity of heritage speaker populations. While all heritage speakers in the current experiment learned Cantonese from birth, and self-reported being more proficient in English compared to Cantonese (see section 3.1), this analysis reveals that more fine-grained differences both within and between speaker groups is meaningful for the lexical encoding of tone. Specifically, the BLP is a composite score that takes into consideration several factors, for example, age of acquisition, time spent in English and Cantonese environments, daily usage, group identity, language attitudes, etc. As such, finding that language dominance correlates with differences in priming magnitudes suggests that a heritage speaker's proficiency in the heritage language necessarily requires the consideration of a myriad of factors. Indeed, despite the fact that we utilized what Polinsky and Kagan (2007) refer to as a "narrow" definition of a heritage speaker, consideration of the heritage speaker's family, quality of the input, and a number of other factors are important in understanding how heritage speakers process and represent speech (see Benmamoun et al., 2013 for a review).

4 General discussion

The linguistic profile of heritage speakers sets them apart as bilinguals who are more dominant in their L2 (cf. late L2 learners), and less dominant in their L1 (cf. monolingual speakers). For heritage speakers, maintaining fluency in the L1 may be difficult in instances where the L1 bears contrastive dimensions not utilized in the L2. In the current paper, we examined the extent to which heritage speakers maintain such L1 contrasts that are not paralleled in the L2 by testing heritage speakers of Cantonese on their discrimination and lexical encoding of tone. To investigate how heritage speakers of Cantonese discriminate lexical tones in Cantonese, we

carried out an AX discrimination task. The task included two ISIs, which were intended to assess distinct levels of linguistic representation (Werker & Logan, 1985; Yu et al., 2017). The shorter ISI (i.e., 500 ms) was intended to probe phonetic levels of representation, while the longer ISI (i.e., 2500 ms) was intended to probe more abstract, phonological levels of representation. Native English speakers with no exposure to a tone language were also included as a control group since they presumably lack phonetic and phonological tone representations. Overall, native English participants were least sensitive to lexical tone differences, and they performed significantly worse than heritage Cantonese speakers and native Cantonese speakers. This poor performance is consistent with previous reports that non-tone language speaking participants struggle to discriminate lexical tone (Burnham et al., 2015; Chang et al., 2017; Lee et al., 1996; Mok & Zuo, 2012; Peng et al., 2010; Qin & Jongman, 2016; Qin & Mok, 2011; Sun & Huang, 2012; Wang et al., 1999; Xu et al., 2006).

Between the heritage and native Cantonese groups, there was no difference. Overall, tone pairs that shared contour type (e.g., T2-T5, both rising tones) were more difficult to discriminate than those that had disparate contour types (e.g., T1-T2, a level-rising tone pair), consistent with previous findings (Kan & Schmid, 2019; Lam, 2018; Soo & Monahan, 2017). In a task that required participants to rely on phonetic or phonological tone representations, no differences emerged between heritage and native Cantonese speakers. Thus, early acquisition of tone may support robust tone discrimination later in life, even when participants are more dominant in an L2 that does not utilize lexical tone.

In addition, while there was a statistically significant effect of ISI, it was small in effect size. These results are in line with previous reports that find that ISI plays a small role in tone discrimination sensitivity irrespective of participants' language background. Wayland and Guion (2004) observed no effect of ISI in discriminating the Thai mid-tone versus low-tone contrast by native Thai, native Mandarin, and native English listeners. Additionally, Lee et al. (1996) found no effect of ISI in native Cantonese, native Mandarin, or native English in discriminating

Cantonese lexical tone. Participants' sensitivity declined only when they were asked to perform a task (i.e., backward counting) during the ISI. Finally, Wayland and Li (2008) showed that before training, native Mandarin speakers better discriminated the Thai mid-tone versus low-tone contrast at the longer ISI compared to the shorter ISI. After perceptual training, however, native Mandarin speakers showed no effect of ISI. No ISI effect was observed in native English either before or after training.

Our results corroborate previous findings that the duration of ISI has little effect on tone discrimination sensitivity irrespective of tone language background. This may be a result of task difficulty. In fact, our participants were not as proficient in the task as previous reports (Fung & Lee, 2019). We suspect that this is since members of a pair were always produced by different talkers with considerable differences between their tone spaces (see Figure 2), whereas previous reports used a single talker. That is, having to compare across two very physically distinct tone spaces likely resulted in poorer than anticipated performance, even for the native Cantonese speakers. Chang et al. (2017) showed that different tasks require different speaker normalization demands and in turn, produce different perceptual results. By using a high-variability experiment, in which auditory stimuli were produced by 24 speakers, they observed that Cantonese and Mandarin listeners were better able to categorize all Cantonese tone and gender combinations above chance. These findings are in line with studies of talker variability, where perceptual adaptation to accented speech is improved when listeners are exposed to different talker voices (Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Palmeri et al., 1993). While these previous reports suggest that additional variation is beneficial to the listener, it might be the case that in discrimination tasks that require the direct comparison between two auditory stimuli, variation can hinder performance. Alternatively, it may have been the case that the inclusion of two talkers in the AX task forced participants to make discriminations at a phonological level irrespective of the ISI (see Section 2.5).

The AX discrimination task did not identify a difference between our Cantonese groups

in terms of phonetic encoding of lexical tone. Because all items in the experiment were Cantonese words, successful performance did not rely on access to the lexicon; in the AX discrimination task, listeners made judgements on the perceptual and/or representational properties of the stimuli.

Subsequently, we asked whether differences would emerge when the task requires access to the lexicon. That is, how well do heritage Cantonese speakers encode tone in the lexicon? To address this question, we conducted a medium-term auditory repetition priming experiment, where participants made a lexical decision to Cantonese words and nonwords. These items were either followed by a member that was identical or differed in tone eight to 20 trials later. Pallier et al. (2001) showed that only identity pairs prime across medium distances, while minimal pairs do not (cf. immediate auditory priming for minimal pairs: Radeau et al., 1995; Slowiaczek et al., 1987; Slowiaczek & Pisoni, 1986; Sumner & Samuel, 2009). Thus, if Cantonese tone representations are not robustly encoded in the lexicon due to English dominance, we expected to see increased tone minimal pair priming as a function of increased English dominance. This is because speakers would potentially hear a pair of words that only differ in tone as an identity pair.

We observed clear identity priming (see Table 4), which validates the efficacy of a task and demonstrates that speakers can be primed over medium-term distances (Church & Schacter, 1994; Goldinger, 1996; Luce & Lyons, 1998). For tone minimal pairs that shared contour types, we observed that heritage Cantonese speakers quantitatively showed more priming than native Cantonese speakers (see Table 4). To further examine the effects of language dominance and priming, we correlated BLP dominance scores with priming magnitudes and found a positive correlation only when the two members of a tone pair shared contour type. These were the same pairs that were more difficult in Experiment 1. That is, the more English dominant the participant was (i.e., the higher their BLP score), the more likely they were to exhibit priming in tone minimal pair cases (see Figure 5). As such, they were more likely

to hear two items that differed only in tone but shared contour type as identical. This suggests that later acquisition and dominance of a non-tone language, in this case, English, impacts heritage speakers' ability to encode tone robustly at a lexical level and these effects are evident in tasks that rely on access to the lexicon.

In this respect, the heritage speaker results from Experiment 2 are consistent with previous studies testing both early and late L2 learners. For example, Zou (2017) tested late Dutch L2 learners of Mandarin—along with native Mandarin speakers—in both a discrimination and a lexical decision task. In the AXB discrimination task, late L2 learners showed phonological discrimination of Mandarin tone contrasts. In a lexical decision task, however, late L2 learners displayed relatively poorer performance compared to native Mandarin speaker counterparts. Similarly, Díaz et al. (2012) reported that late Dutch-English bilinguals performed better in phonetic categorization tasks than in lexical decision tasks involving English vowel contrasts. Finally, across several tasks, Sebastián-Gallés and Baus (2005) reported that early Spanish-Catalan bilinguals, who learned Spanish first, performed better on phonetic categorization tasks with Catalan stimuli and considerably worse on lexical decision tasks with Catalan items relative to Catalan-first, early Catalan-Spanish bilinguals. In short, for both early L2 learners (Pallier et al., 2001; Sebastián-Gallés & Baus, 2005) and late L2 learners (Díaz et al., 2012; Zou, 2017), the later acquisition of the L2 potentially prevents listeners from forming native-like L2 phonetic or phonological categories; thus, differences between early/late L2 learners and native speaker controls appear to arise in tasks that tap into lexical representation. In the current experiment, the heritage speaker's eventual dominance in an L2 that does not possess the phonetic dimension at all (i.e., lexical tone) may interfere with the maintenance of L1 categories and these differences emerge principally in tasks that tap into lexical representations. At the same time, language dominance incorporates multiple bilingual variables (Chang & Yao, 2022; Gathercole, 2016; Luk & Bialystok, 2013), and as such, it is difficult to identify which variable (e.g., reduced L1 usage) is principally responsible for the current set of results.

Without a strong theory of the task for medium-term repetition priming, it is difficult to identify where heritage speakers experience difficulty. Several word recognition models posit that a set of lexical candidates phonetically consistent with the speech signal are activated until the appropriate candidate is selected (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994; Weber & Scharenborg, 2012). Priming occurs when a previously processed stimulus facilitates recognition of subsequent perceptual objects (Neely, 1977; Schvaneveldt & Meyer, 1973; Tulving & Schacter, 1990). In the case of word recognition, the selected lexical candidate has the potential to facilitate understanding of other perceptual objects if they are identical or overlap in form or meaning. In attempting to account for token-specific episodic effects in priming, Goldinger (1996) suggested that a “record” of a spoken word by an unfamiliar voice is created. This record is used when the same token is heard again at a later point in time—sometimes weeks later. The reuse of this record is responsible for the repetition effect. Extrapolating to the current experiment, listeners activate a set of candidates that are phonologically consistent with the input. Then, the candidate that best matches the input is selected and a lexical decision is made. A record is created, and this record can facilitate processing subsequent lexical items. It appears, however, that facilitation over longer distances only occurs when there is complete overlap between the prime and target (Pallier et al., 2001) or when both pronunciation variants are lexically stored (Sumner & Samuel, 2009).

In the current experiment, native Cantonese speakers do not experience medium-term repetition priming for minimal pairs because these entries are phonologically distinct and link to separate lexical entries. In the heritage Cantonese speakers, however, it is possible that tone representations are not robust in the lexicon, being stored with less precision and detail. That is, the later acquisition and dominance of an L2, which does not utilize tone, may affect how tone is stored in memory. This might allow a word with T3 (e.g., 叫 *giu3* “to call out”) to create a record for lexical items (that share segments) with both T3 (i.e., 叫 *giu3* “to call out”) and T6 (i.e., 攞

giu6 “to pry open”). So, while the tones are proficiently discriminated in AX tasks that require only phonetic or phonological analysis of the signal, the memory representation for items with shared tone contour types may not be wholly unique due to the memory demands of lexical storage (see Figure 5). Moreover, note that the effects observed in our experiment are limited only to tones that share contour type and not those with disparate contour types and overall, these hypotheses require empirical investigation.

5 Conclusion

The goal of the current paper was to examine how heritage Cantonese speakers discriminate and lexically encode tone in Cantonese. The results of the AX discrimination task showed no differences between Cantonese heritage speakers and Cantonese native speakers, while both performed better than English native speaker controls. Overall, tone pairs with shared tone contours were discriminated more poorly than tone pairs with disparate contours. To assess whether heritage speakers robustly encode tone in their lexicon, a medium-term auditory repetition priming task was conducted. We observed more identity priming in the Cantonese native speakers than in the heritage speakers. We also found that priming magnitude was positively correlated with English dominance for hard to discriminate tone pairs: The more English dominant the participant was, the more likely they were to experience medium-term priming for tone pairs that share contour type. This would suggest that heritage speakers were potentially encoding minimal pair lexical items that share tone contour type, and differ only in tone, as identity pairs. That is, they were not storing Cantonese lexical items with as much tonal detail. Thus, while the later dominance in English did not impact heritage speaker performance on tasks that rely on phonetic or phonological representations of tone, dominance-based effects on the lexical encoding of tone emerge in tasks that require listeners to access the lexicon.

Acknowledgements

We thank members of the Computation and Psycholinguistics Laboratory at the University of Toronto Scarborough for their assistance. We also thank Stephen Matthews and Diana Archangeli for facilitating data collection at Hong Kong University. This work was supported in part by the Social Sciences and Humanities Research Council (SSHRC) of Canada: CGS-M and CGS-D grants awarded to RS and an IDG grant awarded to PJM (Grant No. IDG430-15-00647).

Competing interests

The authors have no competing interests to declare.

References

- Aslin, R. N., Werker, J. F., & Morgan, J. L. (2002). Innate phonetic boundaries revisited (L). *The Journal of the Acoustical Society of America*, 112(4), 1257–1260.
<https://doi.org/10.1121/1.1501904>
- Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, 13(3), 238–243.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>
- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Walter de Gruyter.
- Benmamoun, E., Montrul, S. A., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3–4), 129–181.
<https://doi.org/10.1515/tl-2013-0009>
- Birdsong, D. (2014). Dominance and Age in Bilingualism. *Applied Linguistics*, 35(4), 374–392.
<https://doi.org/10.1093/applin/amu031>
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (6.6.16) [Computer software]. retrieved 6 June 2020 from <http://www.praat.org/>
- Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12(2), 189–221. <https://doi.org/10.1080/09541446.2000.10590222>
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. <https://doi.org/10.1016/j.cognition.2007.04.005>
- Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. X., Schwarz, I.-C., & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, 36(6), 1459–1491. <https://doi.org/10.1017/S0142716414000496>
- Chang, C. B., & Yao, Y. (2022). An individual-differences perspective on variation in heritage Mandarin speakers. In R. Rao (Ed.), *The Phonetics and Phonology of Heritage Languages*. Cambridge University Press.
- 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. <https://doi.org/10.1121/1.3569736>
- Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, 141(2), EL120–EL126.
- Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, 45, 24–27.
- Ching, T. Y. C. (1984). Lexical tone pattern learning in Cantonese children. *Language Learning and Communication*, 3(3), 243–414.
- Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 521.
- Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, 47(3), 207–238.
- Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, 22(6), 680–689.
<https://doi.org/10.1016/j.lindif.2012.05.005>

- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2020). *Ethnologue: Languages of the World* (Twenty-third edition). SIL International. <http://www.ethnologue.com>
- Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, 18(5), 341–347. <https://doi.org/10.3758/BF03211210>
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, New Series*, 171(3968), 303–306.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65. [https://doi.org/10.1016/S0095-4470\(19\)30537-6](https://doi.org/10.1016/S0095-4470(19)30537-6)
- Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of english. *Speech Communication*, 6(3), 185–202. [https://doi.org/10.1016/0167-6393\(87\)90025-2](https://doi.org/10.1016/0167-6393(87)90025-2)
- Fok, C. Y.-Y. (1974). *A perceptual study of tones in Cantonese*. University of Hong Kong, Centre of Asian studies.
- Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal of the Acoustical Society of America*, 146(5), EL424–EL430. <https://doi.org/10.1121/1.5133661>
- Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, 20–36.
- Gandour, J. (1983). Tone Perception in Far Eastern Languages. *Journal of Phonetics*, 11(2), 149–175.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating Form and Meaning: A Distributed Model of Speech Perception. *Language and Cognitive Processes*, 12(5–6), 613–656. <https://doi.org/10.1080/016909697386646>
- Gathercole, V. C. M. (2016). Factors moderating proficiency in bilingual speakers. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism Across the Lifespan* (pp. 123–140). American Psychological Association. <https://doi.org/10.1037/14939-008>
- Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 208–225). Multilingual Matters.
- Godson, L. (2004). Vowel Production in the Speech of Western Armenian Heritage Speakers. *Heritage Language Journal*, 2(1), 44–69.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183. <https://doi.org/10.1037/0278-7393.22.5.1166>
- Jia, S., Tsang, Y.-K., Huang, J., & Chen, H.-C. (2015). Processing Cantonese lexical tones: Evidence from oddball paradigms. *Neuroscience*, 305, 351–360. <https://doi.org/10.1016/j.neuroscience.2015.08.009>
- Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15(5), 314–318.
- Kan, R. T. (2020). Phonological Production in Young Speakers of Cantonese as a Heritage Language. *Language and Speech*, 0023830920910460.
- Kan, R. T., & Schmid, M. S. (2019). Development of tonal discrimination in young heritage speakers of Cantonese. *Journal of Phonetics*, 73, 40–54. <https://doi.org/10.1016/j.wocn.2018.12.004>
- Kang, Y., George, S., & Soo, R. (2016). Cross-language Influence in the Stop Voicing Contrast in Heritage Tagalog. *Heritage Language Journal*, 13(2), 184–218.
- Kang, Y., & Nagy, N. (2016). VOT merger in Heritage Korean in Toronto. *Language Variation and Change*, 28(2), 249–272. <https://doi.org/10.1017/S095439451600003X>
- Kej, J., Smyth, V., So, L. K. H., Lau, C. C., & Capell, K. (2002). Assessing the accuracy of

- production of Cantonese lexical tones: A comparison between perceptual judgement and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing*, 7(1), 25–38. <https://doi.org/10.1179/136132802805576535>
- Kendall, M. G. (1955). *Rank correlation methods*. Hafner Publishing Co.
- Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, 35(1), 104–117. <https://doi.org/10.1016/j.wocn.2005.10.003>
- Knightly, L. M., Jun, S.-A., Oh, J. S., & Au, T. K. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America*, 114(1), 465–474.
- Kondo-Brown, K. (2006). *Heritage language development: Focus on East Asian immigrants* (Vol. 32). John Benjamins Publishing.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843. <https://doi.org/10.1038/nrn1533>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(1), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lam, W. M. (2018). *Perception of lexical tones by homeland and heritage speakers of Cantonese* [Doctoral Dissertation]. University of British Columbia.
- Lee, J. (2015). *PyCantonese: Cantonese linguistic research in the age of big data* [Talk]. <https://pycantonese.org/>
- Lee, K. Y. S., Chan, K. T. Y., Lam, J. H. S., van Hasselt, C. A., & Tong, M. C. F. (2015). Lexical tone perception in native speakers of Cantonese. *International Journal of Speech-Language Pathology*, 17(1), 53–62. <https://doi.org/10.3109/17549507.2014.898096>
- Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527–542. <https://doi.org/10.1007/BF01758181>
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185–199. <https://doi.org/10.1017/S0305000900001598>
- Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, 26(4), 708–715.
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*. <https://www.tandfonline.com/doi/full/10.1080/20445911.2013.795574>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates, Inc.
- Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, 76(2), 190–208.
- Marean, G. C., Werner, L. A., & Kuhl, P. K. (1992). Vowel categorization by very young infants. *Developmental Psychology*, 28(3), 396.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102. [https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9)
- Matthews, S., & Yip, V. (2013). *Cantonese: A comprehensive grammar*. Routledge.
- Mattock, K., & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, 10(3), 241–265. https://doi.org/10.1207/s15327078in1003_3
- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3), 1367–1381. <https://doi.org/10.1016/j.cognition.2007.07.002>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- Mok, P. P. K., & Zuo, D. (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *The Journal of the Acoustical Society of America*,

- 132(4), 2711–2720.
- 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. <https://doi.org/10.1017/S0954394513000161>
- Montrul, S. A. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30, 3–23.
- Montrul, S. A. (2011a). Multiple interfaces and incomplete acquisition. *Lingua*, 121(4), 591–604.
- Montrul, S. A. (2011b). The Linguistic Competence of Heritage Speakers. *Studies in Second Language Acquisition*, 33(2), 155–161. <https://doi.org/10.1017/S0272263110000719>
- Montrul, S. A. (2012). Is the heritage language like a second language? *EUROSLA Yearbook*, 12(1), 1–29. <https://doi.org/10.1075/eurosla.12.03mon>
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106(3), 226.
- Nodari, R., Celata, C., & Nagy, N. (2019). Socio-indexical phonetic features in the heritage language context: Voiceless stop aspiration in the Calabrian community in Toronto. *Journal of Phonetics*, 73, 91–112.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234. [https://doi.org/10.1016/0010-0277\(94\)90043-4](https://doi.org/10.1016/0010-0277(94)90043-4)
- Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K. (2003). Holding on to childhood language memory. *Cognition*, 86(3), B53–B64. [https://doi.org/10.1016/S0010-0277\(02\)00175-0](https://doi.org/10.1016/S0010-0277(02)00175-0)
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64(3), B9–B17. [https://doi.org/10.1016/S0010-0277\(97\)00030-9](https://doi.org/10.1016/S0010-0277(97)00030-9)
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. *Psychological Science*, 12(6), 445–449. <https://doi.org/10.1111/1467-9280.00383>
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic Encoding of Voice Attributes and Recognition Memory for Spoken Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 309–328.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peng, G., & Wang, W. S.-Y. (2005). Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication*, 45(1), 49–62. <https://doi.org/10.1016/j.specom.2004.09.004>
- Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 38(4), 616–624. <https://doi.org/10.1016/j.wocn.2010.09.003>
- Polinsky, M. (2011). Reanalysis in adult heritage language: New evidence in support of attrition. *Studies in Second Language Acquisition*, 33(2), 305–328.
- Polinsky, M., & Kagan, O. (2007). Heritage languages: In the “wild” and in the classroom. *Language and Linguistics Compass*, 1(5), 368–395. <https://doi.org/10.1111/j.1749-818X.2007.00022.x>
- Qin, Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech*, 59(3), 318–338.
- Qin, Z., & Mok, P. P. K. (2011). Perception of Cantonese Tones by Mandarin, English and French Speakers. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1654–1657. [http://www.phonetics.ucla.edu/voiceproject/Publications/Shue-et-al 2011 ICPHS.pdf](http://www.phonetics.ucla.edu/voiceproject/Publications/Shue-et-al%202011%20ICPhS.pdf)
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

- Radeau, M., Morals, J., & Segui, J. (1995). Phonological Priming Between Monosyllabic Spoken Words. *Journal of Experimental Psychology. Human Perception and Performance*, 21(6), 1297–1311.
- Ronquest, R. (2012). *An acoustic analysis of heritage Spanish vowels* [Doctoral Dissertation]. Indiana University.
- Rosario-Martinez, H. D. (2015). *phia: Post-Hoc Interaction Analysis*. <https://CRAN.R-project.org/package=phia>
- Saadah, E. (2011). *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic* [PhD Thesis]. University of Illinois at Urbana-Champaign.
- Schvaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and comparison processes in semantic memory. *Attention and Performance IV*, 395–409.
- Sebastián-Gallés, N., & Baus, C. (2005). On the Relationship Between Perception and Production in L2 Categories. In A. Cutler (Ed.), *On the Relationship Between Perception and Production in L2 Categories* (pp. 279–292). Routledge. <https://doi.org/10.4324/9781315084503-20>
- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72(2), 111–123. [https://doi.org/10.1016/S0010-0277\(99\)00024-4](https://doi.org/10.1016/S0010-0277(99)00024-4)
- Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, David B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1), 64–75.
- Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, 14(3), 230–237. <https://doi.org/10.3758/BF03197698>
- So, K. L. C. (2000). *Tonal production and perception patterns of Canadian raised Cantonese speakers* [Doctoral Dissertation]. University of Victoria.
- 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. In *Proceedings of the 43rd Annual Meeting of the Berkeley Linguistics Society: Vol. II* (pp. 47–54). Berkeley Linguistic Society.
- Spivey, M. J., & Marian, V. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science*, 10(3), 281–284. <https://doi.org/10.1111/1467-9280.00151>
- Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4), 487–501. <https://doi.org/10.1016/j.jml.2009.01.001>
- Sun, K.-C., & Huang, T. (2012). A cross-linguistic study of Taiwanese tone perception by Taiwanese and English listeners. *Journal of East Asian Linguistics*, 21(3), 305–327.
- Tang, S.-W., Kwok, F., Lee, T. H.-T., Lun, C., Luke, K. K., Tung, P., & Cheung, K. H. (2002). *Guide to LSHK Cantonese Romanization of Chinese Characters*. Linguistic Society of Hong Kong.
- Tsang, Y.-K., Jia, S., Huang, J., & Chen, H.-C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters*, 487(3), 268–272. <https://doi.org/10.1016/j.neulet.2010.10.035>
- Tse, J. K.-P. (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language*, 5(2), 191–204. <https://doi.org/10.1017/S0305000900007418>
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247(4940), 301–306.
- Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal*, 89(3), 410–426.
- Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese.

- Phonetica*, 33(5), 368–392.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649–3658.
- 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.
- Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250–267. <https://doi.org/10.1016/j.wocn.2007.06.004>
- Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *WIREs Cognitive Science*, 3(3), 387–401. <https://doi.org/10.1002/wcs.1178>
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37(1), 35–44. <https://doi.org/10.3758/BF03207136>
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, 75(6), 1866–1878. <https://doi.org/10.1121/1.390988>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.
- Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics*, 26(2), 95–104.
- Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2), 1063–1074. <https://doi.org/10.1121/1.2213572>
- Yang, B. (2015). *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*. Springer Berlin Heidelberg. <https://search.library.utoronto.ca/details?10026620>
- 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. <https://doi.org/10.1016/j.jml.2012.09.004>
- Yiu, C. Y. (2009). A Preliminary Study on the Change of Rising Tones in Hong Kong Cantonese: An Experimental Study. *Language and Linguistics*, 10(2), 269–291.
- Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, 136(3), 1320–1333. <https://doi.org/10.1121/1.4887462>
- Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017). Neurophysiological and Behavioral Responses of Mandarin Lexical Tone Processing. *Frontiers in Neuroscience*, 11. <https://doi.org/10.3389/fnins.2017.00095>
- Zou, T. (2017). *Production and perception of tones by Dutch learners of Mandarin* [PhD Thesis, Leiden University]. <https://hdl.handle.net/1887/52977>

Appendix A

	β	SE	df	t	p	
(Intercept)	1.61	0.07	63.12	22.2	< 0.001	***
T1-T2	-0.13	0.08	83.98	-1.65	0.103	
T1-T3	-0.84	0.09	79.8	-9.47	< 0.001	***
T1-T4	1.45	0.08	81.67	17.35	< 0.001	***
T1-T5	0.11	0.07	85.47	1.52	0.1326	
T1-T6	-0.09	0.1	79.46	-0.93	0.3557	
T2-T3	-0.21	0.07	85.53	-2.92	0.0045	**
T2-T4	0.35	0.06	115.09	6.07	< 0.001	***
T2-T5	-2.05	0.1	78.04	-19.74	< 0.001	***
T2-T6	-0.21	0.07	93.86	-3.09	0.0026	**
T3-T4	1.36	0.09	80.57	15.06	< 0.001	***
T3-T5	-0.11	0.07	86.7	-1.47	0.1442	
T3-T6	-1.62	0.09	81.38	-18.01	< 0.001	***
T4-T5	0.64	0.07	90.73	9.5	< 0.001	***
T5-T6	-0.07	0.08	81.96	-0.94	0.3492	
ISI	-0.08	0.05	78.32	-1.62	0.1093	
NC-HC	0.02	0.08	106.39	0.26	0.7942	
NC-NE	-0.28	0.09	106.39	-3.3	0.0013	**

Table A1. Full output of the linear model from the data in Experiment 1. Simple coding was used for the variables Group (Native Cantonese^{Ref}, Heritage Cantonese, Native English) and ISI (Short ISI^{Ref} (500 ms), Long ISI (2500 ms)). For Tone Pair, sum coding was used, so that every level is compared to the grand mean. Following Fung and Lee (2019), we selected the variable with the highest accuracy (T4-T6) as the level coded with -1 in the coding scheme. This omits this comparison from the model output. The fixed effects included Tone Pair, Group and ISI. Random effects included random by-participant intercepts and random by-participant slopes for Tone Pair and ISI.