



Open Library of Humanities

Lexical competition, sound change, and language dominance in Cantonese tone category recognition

Rachel Soo*, Department of Linguistics, The University of British Columbia, Vancouver, BC, Canada, rachel.soo@ubc.ca

Molly Babel, Department of Linguistics, The University of British Columbia, Vancouver, BC, Canada, molly.babel@ubc.ca

*Corresponding author.

Speech perception can be shaped by factors such as lexical competition, synchronic variation, and language dominance. Listeners can use lexical information to categorize sounds and recognize words, but systematic variation may act to neutralize lexical contrasts. Additionally, the detail in the phonological-lexical representations may vary along with linguistic experience. We examine the confluence of these factors in the current study, examining how lexical competition structures phonetic variation of merging vs. non-merging tone categories in Cantonese-English bilinguals varying in relative Cantonese vs. English dominance. Listeners categorized tokens from lexical tone continua generated from minimal pairs (Experiment 1: Word identification) and from word-nonword pairs (Experiment 2: Lexical decision). When there are lexical competitors at both endpoints of the continua, listeners maintained more discrete categorization functions for non-merging tones than merging tones. In the absence of lexical competition, there was no difference in response functions for merging and non-merging tones at the group level. English-dominant early Cantonese-English bilinguals consistently showed less sigmoidal response functions compared to Cantonese-dominant early bilinguals in both experiments suggesting that lexical representations may be stored with less precise tonal detail by English-dominant Cantonese speakers. Overall, these data suggest that the presence of a lexical competitor constrains the range of acceptable phonetic variation for non-merging tones, while listeners are more accepting of phonetic variation in the merging tone categories.



1. Introduction

The degree of categoricity in speech perception can be affected by several factors, such as presence of a lexical competitor (e.g., Ganong, 1980), participation in an ongoing sound change (e.g., Harrington, Kleber, & Reubold, 2008), and language dominance (e.g., Casillas & Simonet, 2016). To date most work has only addressed the influence of these factors in the segmental domain. In this study, we extend this work to a suprasegmental contrast, homing in on lexical tone. Cantonese presents an ideal linguistic landscape for this inquiry. Cantonese is a tone language and several of the tones are undergoing mergers, leading to increased ambiguity of phonological categories. The Cantonese lexicon also contains gaps where a given lexical item is missing a tonal lexical competitor, offering opportunities for lexical biases to emerge. While we focus on first language speakers of Cantonese, the Cantonese diaspora affords a global speech community with varying degrees of language dominance in Cantonese relative to another language – English, in our case. First language speakers of Cantonese who are English dominant, but still highly proficient in Cantonese, may encode lexical tone less precisely than Cantonese-dominant Cantonese-English bilinguals (Lam, 2018; Soo & Monahan, 2023). We build to our specific hypotheses in the paragraphs that follow, ultimately testing how Cantonese listeners resolve phonetic ambiguity in word-level perception given multiple factors: Lexical competition, synchronic variation for tone mergers-in-progress, and language dominance.

1.1. Lexical influence in phonetic encoding

Everyday speech perception requires listeners to process phonetically variable speech in acoustically variable conditions. Listeners may leverage lexical information to accomplish this task by mapping sounds to categories that create real words in their lexicon (Luthra, Guediche, Blumstein, & Myers, 2019; Pisoni & Tash, 1974; Wedel & Fatkullin, 2017). The theory that the lexicon plays a role in allowing listeners to process phonetically variable speech is well established (e.g., Connine & Clifton Jr, 1987; Marslen-Wilson, 1984; McClelland & Elman, 1986; Samuel, 1996, 1997, 2001). Early work by Ganong (1980) established that lexical context disambiguates ambiguous speech sounds. An ambiguous sound midway between a /t/ and /d/ was more likely to be identified as /t/ in a context like “?ask”, and /d/ in a context like “?ash”, because “task” is a real word (while *“dask” is not), and “dash” is a real word (while *“tash” is not). Similar Ganong-type lexical bias effects have been observed with different phoneme pairs (Connine, Titone, Deelman, & Blasko, 1997; Pitt, 1995), in different word positions (Pitt & Samuel, 1993, 1995), with words of varying lengths (Pitt & Samuel, 2006), and in tones (Fox & Unkefer, 1985; T. H. Yang, Jin, & Lu, 2019). Likewise, in phoneme restoration studies, for example, listeners report hearing a word as intact despite the presence of noise obscuring particular sounds in the word (Warren, 1970). This “restoration effect” is stronger in words than in nonwords, suggesting that phonetic encoding is supported by the lexicon (Samuel, 1981, 1996). In phoneme monitoring

tasks, target phonemes are identified faster in words than in nonwords (Rubin, Turvey, & Van Gelder, 1976). Furthermore, nonword processing has been shown to vary as a function of its similarity to real words (Connine et al., 1997; Wurm & Samuel, 1997), and lexical knowledge guides the retuning of phonetic category boundaries in perceptual learning paradigms (e.g., Norris, McQueen, & Cutler, 2003).

On a broad level, this body of literature shows that listeners are able to use their lexicon to guide the interpretation and learning of ambiguous speech sounds. Listeners also exhibit evidence of competition at the lexical level in their speech. For instance, listeners have knowledge of the specific lexical competitors for a word along a given acoustic dimension in production. Baese-Berk and Goldrick (2009) investigated how lexical competition affects stop voicing contrasts in a production study using words with a voiced stop competitor (e.g., “pox”, whose competitor is “box”) and words without a voiced stop competitor (e.g., “posh”, whose competitor is *’bosh”). Words with a voiced stop lexical competitor were produced with longer voice onset time (VOT) than those without, suggesting that listeners may be implicitly aware of the lexical competitors for a given word and exploit this in production for the purposes of contrast enhancement. These effects were replicated with word-initial alveolar /t/ and velar /k/ stops.¹ These effects are not limited to laboratory speech. Using the Buckeye Corpus, Wedel, Nelson, and Sharp (2018) observed that the existence of a minimal pair along specific phonetic cues (i.e., VOT for word-initial voiced and voiceless stops, and Euclidean formant distance for vowels) significantly predicted hyperarticulation of those cues, while a more general measure of neighbourhood density did not. As such, the realization of a particular contrast in production varies as a function of phonetically-granular measures of lexical competition along specific acoustic dimensions.

These findings are paralleled in tone perception. For example, Fox and Unkefer (1985) find support for the Ganong effect for Mandarin tones with Mandarin listeners, but not English listeners, who lack the requisite lexical knowledge. This Ganong effect was recently replicated by T. H. Yang et al. (2019). In Yang and colleagues’ study, Mandarin listeners categorized items from lexical tone continua synthesized in PSOLA (Charpentier & Stella, 1986) and TANDEM-STRAIGHT (Kawahara et al., 2008). They found that listeners were significantly biased towards the real word endpoints only for continua synthesized using TANDEM-STRAIGHT, and no

¹ Goldrick, Vaughn, and Murphy (2013) examined these effects on the English voiced stop series word-initially and word-finally, observing different results. In word-initial position, unlike in Baese-Berk and Goldrick (2009), there were no differences in the positive VOT between items with and without lexical competitors. In word-final position, there were clear differences in the vowel durations of voiced stops as a function of lexical competition, but not in the expected direction. The preceding vowels of words with a competitor (e.g., “bud”, whose competitor is “but”) were produced as shorter than those without minimal pairs (e.g., “thud”, whose competitor is *’thut”), acting to reduce the voicing contrast for stops in word-final position. Goldrick et al. (2013) speculate that these differences across phonological contrasts and word positions are a consequence of the lexicon interacting with phonetic and phonological restrictions that constrain the range of allowable variation.

significant effects were observed for the PSOLA continua. This observation of different results for the two methods is relevant because it points to the multidimensional nature of tone contrasts. When stimuli are created in PSOLA, the F0 of a token is directly targeted for manipulation, whereas TANDEM-STRAIGHT includes resynthesis and morphing of all acoustic cues on the speech selected for manipulation (e.g., F0, duration, voice quality, etc.). These algorithmic details between TANDEM-STRAIGHT and PSOLA are germane not only to the methodological choices we make in the current study on Cantonese, but also to the overall role in signal fidelity in spoken language processing (e.g., Schouten, Gerrits, & Van Hoesen, 2003).

1.2. Contrasts and competition in sound change

The Cantonese tone inventory is currently experiencing several tone mergers-in-progress. Before detailing the nature of the Cantonese tone space and the tone mergers-in-progress, we first outline the theoretical and empirical landscape with respect to contrast and competition in diachrony to complement the effects of contrast and competition synchronically described in Section 1.1.

Apart from contrast enhancement (Baese-Berk & Goldrick, 2009), which can be a synchronic adjustment, lexical competition may also play a role in contrast maintenance, as a diachronic pressure. Specifically, minimal pairs have been shown to inhibit diachronic mergers. Wedel, Kaplan, and Jackson (2013) carried out a corpus study on a series of phoneme mergers from a diverse set of languages. By calculating the functional load for the phonemes taking part in mergers across these languages, they found an inverse relationship between the probability of a phoneme merger and functional load. That is, the probability of a merger was less likely between phonemes that distinguish a greater number of words in the language (i.e., when there are more minimal pairs).

The notion that gradual phonetic change in a diachronic merger may be inhibited for the purpose of maintaining word recognition is also observed in cases of sound change where there is no long-term neutralization of contrasts. In push chains, for example, the gradual phonetic change of one category pushes it towards another category. In turn, that category moves in an effort to maintain a contrast. Hay, Pierrehumbert, Walker, and LaShell (2015) examined frequency effects on a push chain shift in New Zealand English involving DRESS, TRAP, and KIT words. Using corpus data that span 136 years, they established that low frequency words changed faster than high frequency words and reason that lower frequency items move first as a way to resist ambiguity and promote comprehension. This is supported by a computational model in Todd, Pierrehumbert, and Hay (2019), which demonstrated that high and low frequency words change at different rates with respect to one another, depending on how the particular change affects acoustic ambiguity and the phonetic realization of a phonological contrast.

The studies discussed thus far provide evidence for the effects of lexical competition in both production and perception. This invites discussion of the perception-production link and an acknowledgment that sound changes ultimately require the change to be actualized in both

modalities (Harrington et al., 2019). The link between production and perception is central to theoretical models of contrast maintenance, like Wedel and Fatkullin (2017). In that model, categories consist of a series of exemplars, and the contrast between two categories is maintained through a positive feedback loop between production and perception driven by category competition. Mergers-in-progress may, thus, be represented as cases where the distribution of two sound categories begin to encroach on one another and produce an area of overlap at category boundaries. In these cases, it is the presence of a lexical competitor that drives the distributions to compete for the incoming percept and ultimately assist in the retention of contrast by facilitating the maintenance of independent, non-overlapping portions at the opposing extreme ends of the distribution. In the absence of a lexical competitor however, listeners may be more forgiving of phonetic variation, mapping a wider range of variation to the word category and exhibiting a lexical bias in the face of ambiguity (Ganong, 1980). As such, the absence of a lexical competitor is predicted to produce less competition, and thus less contrast between phonological categories. This is expected to be the case for both mergers-in-progress and more robust phonological contrasts.

1.3. Cantonese tones

Cantonese is a Sino-Tibetan tone language spoken primarily in Hong Kong, Macau, and Guangzhou, with diaspora communities across the world. Hong Kong Cantonese (the focus of the current study) contains six lexical tones that occur on open syllables and on syllables with nasal codas (Bauer & Benedict, 1997). These six lexical tones consist of three level tones (high-level: T1, mid-level: T3, low-level: T6), two rising tones (high-rising: T2, mid-rising: T5), and one falling tone (low-falling: T4). The language also has three checked allotones of the three level tones (high-stopped T7, mid-stopped T8, low-stopped T9), which are realized on closed syllables ending in unreleased stops /p t k/ (Bauer & Benedict, 1997). We focus on the six phonemic lexical tones of Cantonese in the current study. This tone inventory has been provided in **Table 1** and the tone pairs of interest for the current study have been visually depicted in **Figure 1**.

Contour	Tone	Description	Example word
Level	1 [55]	High-level	衣 <i>ji1</i> 'clothes'
	3 [33]	Mid-level	意 <i>ji3</i> 'idea'
	6 [22]	Low-level	二 <i>ji6</i> 'two'
Rising	2 [25]	High-rising	椅 <i>ji2</i> 'chair'
	5 [23]	Mid-rising	耳 <i>ji5</i> 'ear'
Falling	4 [21]	Low-falling	疑 <i>ji4</i> 'suspicious'

Table 1: Cantonese phonemic tone inventory. Tone numerals following the Chao transcription system are given in square brackets to characterize the pitch contour (Chao, 1947). In the Jyutping transcription, tones are represented with numbers.

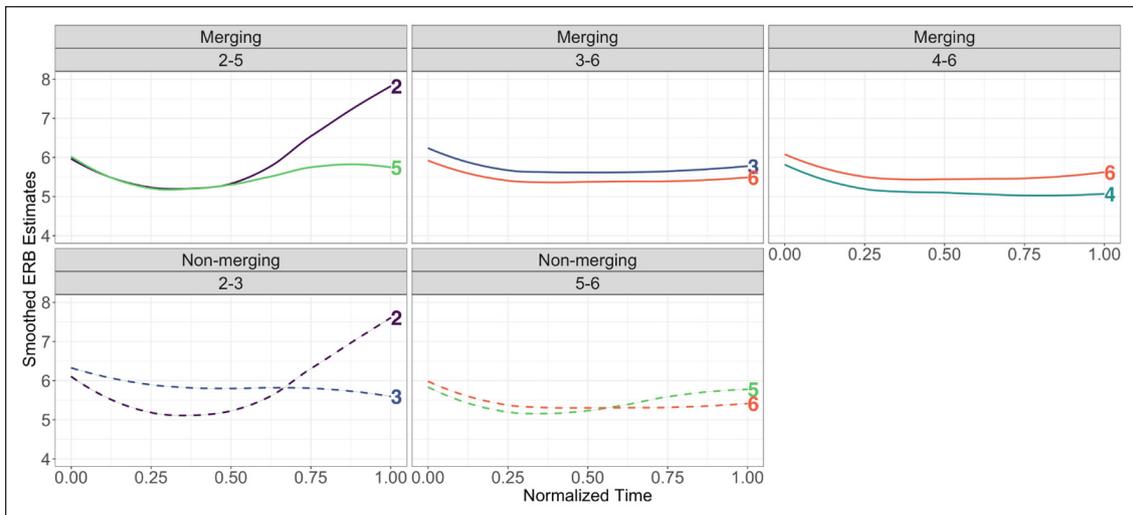


Figure 1: Smoothed ERB estimates for the items in each tone pair used in Experiments 1 and 2 plotted on a normalized time scale for ease of visualization. All stimuli were produced by a linguistically trained female native speaker of Cantonese originally from Hong Kong (35-years-old), who is also a Cantonese language teacher and self-reports not producing the tone mergers. Merging tone pairs (solid) are given in the top row, while non-merging tone pairs (dotted) are given in the bottom row. Each tone contour in a pair is directly labeled in the graph.

The literature documents ongoing mergers between T2-T5, T3-T6, T4-T6, and T3-T5, in both production and perception (Bauer, Kwan-Hin, & Pak-Man, 2003; Fung & Lee, 2019; Lam, 2018; Lee, Chan, Lam, Van Hasselt, & Tong, 2015; Mok, Zuo, & Wong, 2013; Tsui, 2012; Vance, 1976; Wong, 2008). In the current study, we focus on the mergers between T2-T5, T3-T6, and T4-T6, as these are the most well-documented in the literature.

In Cantonese, syllables are maximally (C)V(V)(N), plus tone. Since a larger proportion of all possible Cantonese monosyllables are real words compared to languages like English with more complex syllable templates (Matthews & Yip, 2013), nearly any word in Cantonese will face tonal lexical competition, though not from all possible tonal competitors. The perceived wordlikeness of Cantonese strings is positively correlated with measures of phonotactic probability (Kirby & Yu, 2007). As languages like Cantonese have comparatively more restricted syllable templates in addition to lexical tone, this substantially broadens the range of possible lexical items in the language. In other tone languages with a similarly restricted syllable template, like Mandarin, listeners more readily change nonwords into words by changing tones rather than vowels or consonants (Wiener & Turnbull, 2016), as changes in tone do little to narrow the range of lexical competitors compared to vowels or consonants (Cutler & Chen, 1997; Sereno & Lee, 2015; Ye & Connine, 1999). Wiener and Turnbull (2016) reason that this means tones have a lower priority in Mandarin, as tones matter less for lexical access than consonants and vowels. This mutability suggests that listeners essentially care less about the tonal identity compared to segments, and

will more readily change the tone. While there is additional behavioural evidence in support of tones carrying less information in Mandarin (Tong, Francis, & Gandour, 2008), quantification of the Mandarin lexicon indicates that vowels and tones contribute equally to functional load (Surendran & Levow, 2004). The role of lexical tone relative to segments in spoken word recognition is an active area of investigation (Malins & Joanisse, 2010; Q. Yang & Chen, 2022).

That said, the relative contribution of lexical competition in structuring tone merger sound changes in Cantonese is unclear. If, as with Mandarin, changes in tone do little to narrow the range of lexical competitors compared to vowels or consonants in Cantonese, the contribution of lexical competition may be different from that of previous studies investigating segmental sound changes.

We are interested in how lexical competition affects the mapping of phonetic variation to words in Cantonese tones that are merging or remain stable.

1.4. The bilingual space

Abroad and in the Cantonese-speaking homelands, Cantonese speakers are multilingual. In Hong Kong, few monolingual Cantonese talkers exist, given the socio-political landscape of Hong Kong over the past century. Many Cantonese talkers speak English, due to Hong Kong's history as a British colony, and Mandarin, for contemporary geo-political reasons. In the Canadian Cantonese diaspora, many individuals are not only bilingual in English (a majority societal language), they are also often more dominant in English. We focus on early Cantonese-English bilinguals in the current study because that is the population in our speech community. While these participants are unified by the fact that Cantonese is their first language, their linguistic profiles and usage patterns vary. As seen in Table 4 (see Appendix), the bilinguals in the current study differ in their degree of English dominance. This is relevant because lexical support has been shown to vary as a function of language proficiency (Samuel & Frost, 2015; Soo, Sidiqi, Shah, & Monahan, 2020). Since lexical support for phonetic encoding is only present insofar as listeners have developed and can access fully functional lexical representations, a phonologically well-specified lexicon is a clear necessity for successful speech perception in the language. This is formalized in the fuzzy lexicon hypothesis proposed for L2 speech processing (Gor, Cook, & Jackson, 2010), wherein the effect of lexical competition may be less robust if phonological-lexical representations themselves are stored with less phonological detail. This line of work also speaks to the interaction of lexical competition with category contrast, as studies have shown that confusable categories may produce lexical representations that lack phonological detail for L2 speakers (Darcy, Daidone, & Kojima, 2013).

First language Cantonese speakers who are English-dominant pattern equivalently to first language Cantonese speakers who are Cantonese-dominant speakers in AX discrimination tasks (Soo & Monahan, 2017, 2023, though cf. Kan & Schmid, 2019). However, there is some evidence that English-dominant Cantonese speakers encode lexical tone less precisely (Lam, 2018;

Soo & Monahan, 2023). In a comparison of “homeland” (e.g., Hong Kong born-and-raised) and “heritage” (e.g., Canada born-and-raised) speakers, Lam (2018) found that heritage listeners were more willing to ignore tone errors to maintain semantic coherence in sentences than homeland speakers. Soo and Monahan (2023) showed that Cantonese heritage listeners treat tone minimal pairs like identity pairs in a medium-term priming paradigm, whereas Hong Kong-based Cantonese listeners exhibit the expected lexical inhibition in that context. They interpret this as evidence that the heritage speakers encode the tone less precisely. It is with these results in hand that we consider how language dominance moderates listeners’ perception of tone ambiguity in the face of lexical competition. We focus on relative language dominance between Cantonese and English in lieu of categorically coding individuals’ language backgrounds (e.g., heritage versus homeland) because the nuance of a gradient measure more accurately captures the linguistic diversity in the Cantonese-speaking population.

Bilingualism also has relevance here for how phonetic variation may be realized (and shaped) in different bilingual communities. For example, Samuel and Larraza (2015) investigate a “mislabelling” of nonwords as words by early, highly proficient Spanish-Basque bilinguals, whose L1 is Basque. These speakers erroneously accept Basque nonwords that replace the “correct” apical affricate with an “incorrect” laminal affricate at fairly high rates despite the fact that Basque listeners can perceive the contrast. Samuel and Larraza (2015) speculate that this may be due to the fact that these speakers are regularly exposed to Spanish-accented Basque, where the laminal and apical affricate contrast is confounded. Samuel and Larraza (2015, p. 54) note, “...it would probably not even be correct to call it a problem, as it would be an appropriate adaptation to the nature of the input and its relationship to lexical entries.” To be clear, in these cases, items that would typically be called nonwords by L1 speaker/listeners are endorsed as words because of exposure to frequent “mispronunciations” of these items by L2 speakers. Thus, phonetic variation may be realized and accepted differently in specific bilingual communities. In the context of the Cantonese tone mergers, it is not habitual “mispronunciation” of tones that drives our predictions, but rather the variability induced by tone mergers and the role of lexical contrasts in maintaining boundaries.

1.5. Predictions for our study

The goal of this study is to test the role of lexical competition, sound change, and language dominance in Cantonese tone perception. To this end, we assess listener categorization of merging (T2-T5, T3-T6, T4-T6) and non-merging (T2-T3, T5-T6) tone pairs with and without lexical competitors in a within-subjects design across two experiments. Experiment 1 is a word identification task utilizing tone pairs with lexical competitors (i.e., tone minimal pairs) and Experiment 2 is a lexical decision task utilizing tone pairs *without* lexical competitors (i.e., word-nonword pairs).

We predict that the tone categories in non-merging tone pairs will be more perceptually distinct, showing more categoricity than merging tone pairs. Moreover, in line with the literature on sound change (Todd et al., 2019; Wedel, Jackson, & Kaplan, 2013), we would expect that lexical competition supports the maintenance of more distinct tone categories, manifested with more categorical response functions in items with lexical competitors than those without. Because items with lexical competitors (Experiment 1) and those without (Experiment 2) are presented in different experiments with different dependent measures, we cannot directly test this. But, an effect of lexical competition is predicted to be stronger in non-merging tone pairs than in merging tone pairs, since the distinction between words is already in the process of being neutralized in merging tone pairs. We test this in Experiment 1. In Experiment 2, we predict that in the absence of a lexical competitor, listeners will be overall more accepting of tone variation, endorsing items on the nonword-side of the continuum as words. Finally, across both experiments, we expect that the overall categoricity of the tone pairs and the effect of the tone merger status may vary as a function of the bilingual profile of our participants. It is likely that English-dominant early Cantonese-English bilinguals will demonstrate less categoricity overall, given prior work showing that phonological-lexical representations may be less well-specified in a bilingual's less dominant language. At the same time, given that Cantonese was acquired at a young age and not as a late second language, it may also be the case that their early experience with Cantonese has cemented their lexical knowledge such that there is no effect of language dominance on categorization. We know of no published evidence that speaks directly to Cantonese- and English-dominant early Cantonese-English bilinguals' participation in the tone mergers with respect to lexical competition, so we do not hypothesize about an interaction between merger status, lexical competition, and language dominance.

2. Methodology

2.1. Participants

Early Cantonese-English bilinguals who learned Cantonese before the age of five were recruited to take part in both Experiments 1 and 2 (within-subjects design). Due to research restrictions imposed by the COVID-19 pandemic, Experiments 1 and 2 were conducted in-person and online. In-person participants took part in Experiments 1 and 2 through E-Prime 2.0 (Psychology Software Tools, 2012) across two 45-minute sessions separated by approximately 1.5 weeks. Online participants took part in Experiments 1 and 2 through Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). To combat online fatigue, each experiment was separated into two parts for a total of four 15-minute online sessions, separated by approximately one week.

Forty-three individuals completed the study (30 in-person, 13 online). Seven individuals who did not learn Cantonese before the age of five were removed prior to analysis. One individual who reported learning Cantonese from age one, but self-reported Cantonese understanding ability as

0 was also removed. One participant who did not complete all four-parts of the online study was also removed. The remaining 34 early Cantonese-English bilinguals' demographic summaries are reported in Table 4 (see Appendix). All participants provided verbal informed consent and were compensated with partial course credit or remunerated \$30 CAD for those who participated via Prolific (Palan & Schitter, 2018) for their time. In the four-part online study, the \$30 CAD was distributed as follows – \$5 for each of the first three parts and \$15 for the fourth to incentivize completion of all four parts.

2.2. Procedure

The procedure for Experiments 1 and 2 were identical. In the online versions of Experiments 1 and 2, participants first took part in a headphone test to ensure that they were wearing adequate headphones (Woods, Siegel, Traer, & McDermott, 2017). Those who did not achieve 80% accuracy on the task were not permitted to proceed. In-person participants did not need to complete this headphone test as they completed the task with AKG K240 headphones in a quiet laboratory environment. Instead, they proceeded immediately to listening to a Cantonese version of “The North Wind and the Sun” story from the Aesop Language Bank produced by the same speaker who produced the materials to familiarize participants with the speaker’s tone range.² Online participants heard the story at the start of each of the four experimental sessions. After the passage, participants took part in the main task, which was a word identification task in Experiment 1, and a lexical decision task in Experiment 2. Following the main task, participants completed the Bilingual Language Profile Questionnaire (BLP; Gertken, Amengual, & Birdsong, 2014). The BLP questionnaire asks a series of questions about participants’ language history, background, and use. These responses are used to compute a quantified measure of language dominance called a “dominance score” on a scale of ± 218 , where positive scores indicate greater English dominance and negative scores indicate greater Cantonese dominance.³ As shown in Table 4 (see Appendix), dominance scores span a wide range, and this heterogeneity is typical of the Cantonese-English speech community in our subject population. Across the 34 participants in the study, 21 are English-dominant and 13 are Cantonese-dominant.

2.3. Materials

Items for Experiment 1 consisted of four tone minimal pairs selected for each merging tone pair (T2-T5, T3-T6 and T4-T6), and 11 tone minimal pairs for each non-merging tone pair (T2-T3 and

² While the six lexical tones are not equally represented in the story, crucially, the story included instances of T1 and T6, such that listeners could establish the pitch ceiling and pitch floor of the speaker. The full transcription of the story is available at <https://www.aesoplanguagelibrary.com/yue.html>.

³ We use the BLP because it provides a quantifiable global representation of dominance and because of the current lack of a metric that targets lexical proficiency.

T5-T6).⁴ This produced a total of 34 unique stimulus pairs in Experiment 1 (see Table 5 in the Appendix). Words in each stimulus pair were chosen to be familiar, and approximately matched for lexical frequency according to counts from the Hong Kong Cantonese Corpus (Luke & Wong, 2015).

Items in Experiment 2 consisted of eight word-nonword tone pairs selected for each merging pair (T2-T5, T3-T6 and T4-T6), and 22 such pairs for each non-merging tone pair (T2-T3 and T5-T6). As in T. H. Yang et al. (2019), these nonwords were accidental tonal gaps in the language. Since each tone in a stimulus pair had to be represented as a word and a nonword (e.g., 甜 *tim4* ‘sweet’ – **tim6*, where T6 is the nonword, and **dei4* – 地 *dei6* ‘ground’, where T4 is the nonword), this produced twice as many stimulus pairs for each tone pair compared to Experiment 1, in which half of the items in Experiment 2 were nonwords. This produced a total of 68 unique stimulus pairs in Experiment 2 (see Table 6 in the Appendix). Altogether, there were 102 unique stimulus pairs; tone minimal pairs were intended for Experiment 1 and word-nonword pairs were intended for Experiment 2. It is worth noting that we selected stimulus pairs with and without *specific tonal* lexical competitors for Experiments 1 and 2, respectively, as a means of more carefully examining the effect of lexical competition in merging and non-merging tone pairs. In doing so, we did not control for the existence of other tonal lexical competitors. For example, while 狼 *long4* ‘wolf’ was designed to bear 浪 *long6* ‘ocean wave’ as a lexical competitor for T4-T6, competitors for T1, T2, T3, or T5 for the syllable *long* were not controlled for. Overall, the presence of other tonal lexical competitors was unlikely to have played a role in the degree of acoustic-auditory distance, as evidenced by the non-significant effect of lexical competition on the acoustic-auditory distance (see Section 2.4).

Items were recorded in a sound-attenuated room using Audacity through a Samson C03U USB microphone at a sampling rate of 44.1 kHz and 24-bit depth. Following T.H. Yang et al. (2019), we used TANDEM-STRAIGHT (Kawahara et al., 2008) to synthesize 11-step continua.⁵ TANDEM-STRAIGHT considers all available acoustic cues on the entire tone-bearing syllable for resynthesis (e.g., F0, duration, voice quality, etc.). The algorithm decomposes signals into source and filter, and allows morphing between user-specified spectral and temporal anchors. Our anchors delineated the initial consonant and the rime as distinct intervals for spectral and temporal morphing.

2.4. Acoustic-auditory analysis

The purpose of Experiments 1 and 2 is to quantify the role that lexical competition plays in the perception of merging and non-merging tones. While several tone mergers have been documented

⁴ The imbalanced numbers are due to the fact this experiment served double duty: Testing the hypotheses described here about lexical competition and perceptual category structure in tones, and pre-testing for a future experiment on perceptual learning of tones in Cantonese. This is also why T2-T3 and T5-T6 were selected as the non-merging tone pairs of interest; each tone pair consists of a rising tone (T2, T5) and a level tone (T3, T6), making it possible to test for generalization of perceptual learning to tone pairs bearing similar contours.

⁵ All stimuli are available at <https://osf.io/bcmv4/>.

in the literature (see Section 1.3 and the General Discussion), not all speakers exhibit these tone mergers to the same degree, and some speakers maintain distinct categories. In order to conduct these experiments without confounds of, for example, the inherent ambiguity of merging tones or the reduction of contrast in the stimuli, we first quantify the acoustic-auditory distinctiveness of our stimuli.

While tones in Cantonese may be differentiated along a number of (psycho-)acoustic dimensions (see e.g., Chan, 1974; Gandour, 1981; Khouw & Ciocca, 2007), these analyses focus on F0 as the primary cue to Cantonese tone perception. F0 was estimated from the tone-bearing units in Voicesauce (Shue, Keating, Vicenik, & Yu, 2011) using the STRAIGHT algorithm (Kawahara, Masuda-Katsuse, & De Cheveigne, 1999), which generated F0 estimates every millisecond. These F0 estimates were transformed into equivalent rectangular bandwidths (ERBs, Moore & Glasberg, 1983). These ERB estimates were then smoothed using the *loess* function (Cleveland, Grosse, & Shyu, 1992) in R (R Core Team, 2022) with a span of 0.25 to allow for the maintenance of some degree of natural variation in the contour. **Figure 1** presents these smoothed ERB estimates on a normalized time scale for ease of visualization by tone pairs, with the three merging tone pairs on the top row and the two non-merging tone pairs on the bottom row.

To quantify acoustic-auditory distance between minimal pair items, an area between two curves algorithm was used (Jekel, Venter, Venter, Stander, & Haftka, 2019). This measure is computed by summing the area of the quadrilaterals that are constructed by joining consecutive pairs of points from the two signals. In the case that the signals do not have equal numbers of points, as is the case with tokens varying in duration, new points are created at the bisection of the segment with the largest Euclidean distance. Adding points in this manner does not change the area but allows for efficient computation. Computing similarity this way makes no assumption of the shape of the trajectories in terms of either local fluctuations or global patterns of rising or falling contours. A higher degree of similarity is thus quantified as a smaller area between trajectories; a curve would have no area between itself. In contrast to (Pearson R) correlation, computing the area between curves has no upper limit. **Figure 2** presents box plots of these data for each tone pair and separated by whether the items have lexical competitors or not.

To assess whether there was a difference in acoustic-auditory distance between merging and non-merging tone pairs, and between pairs with real word competitors (i.e., minimal pairs) and those without real word competitors (i.e., word-nonword pairs), the area between the smoothed tone curves was used as the dependent measure in a linear mixed effects regression model carried out using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). The model included fixed effects of Tone Type (Non-Merging, Merging) and Condition (Competitor, No competitor) with Non-merging and Competitor as reference levels. Tone Pair was a random effect with random slopes by Condition.⁶

⁶ Model Syntax: $\text{AreaBetween} \sim \text{ToneType} * \text{Condition} + (1 + \text{Condition} | \text{TonePair})$.

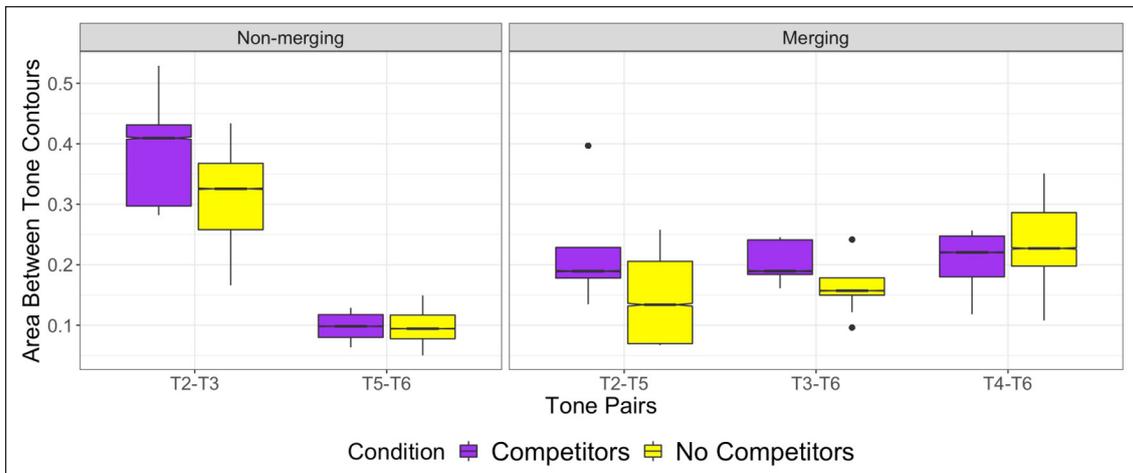


Figure 2: Acoustic-auditory distance between tones in non-merging (left) and merging pairs (right) across items with lexical competitors (purple) and without lexical competitors (yellow).

The model returned a significant intercept ($\beta = 0.4475$, $SE = 0.1712$, $t = 2.613$), but none of the effects, nor their interactions, were significant (Tone Type: $\beta = 0.073$, $SE = 0.268$, $t = 0.271$; Condition: $\beta = -0.089$, $SE = 0.069$, $t = -1.302$; Tone Type * Condition: $\beta = -0.00005$, $SE = 0.102$, $t = -0.001$). These results suggest that, overall, there were no significant differences between Merging and Non-Merging tone pairs in the magnitude of their acoustic-auditory distance, and that items with lexical competitors were not overall produced with greater acoustic distance than those without lexical competitors.

The tone pairs vary in their acoustic-auditory distinctiveness in a way that does not align with whether the tone pairs are merging or not. Note, for example, that the smallest acoustic-auditory distance is for the non-merging pair T5-T6 (see also Tsui, 2012). Having established that there are no confounds of the pure acoustic-auditory distinctiveness in our stimuli between merging and non-merging tones, and between pairs with and without lexical competitors, we move forward to interpret the results of Experiments 1 and 2 as a function of lexical competition and tone merger based listener biases, though we necessarily consider our results in the context of categorical perception and tone shape as well.

3. Experiment 1: Word identification task

To examine how the presence of a lexical competitor structures the perception of merging and non-merging tone pairs, we carried out a word identification task. We predict that the tone categories in non-merging tone pairs will be more perceptually distinct, showing more categoricity than merging tone pairs. Since Cantonese literacy cannot be assumed in the English-dominant sample of our participant population, responses were elicited with pictures. An example screen shot of a trial is shown in **Figure 3**. Listeners were presented with a full

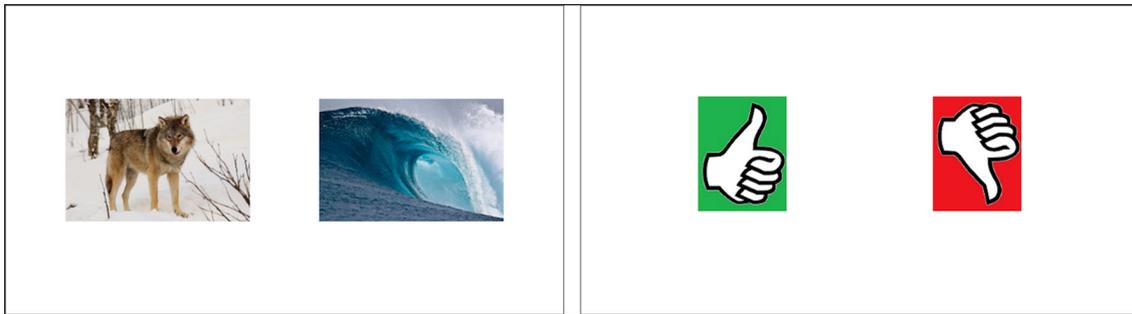


Figure 3: Left: Trial schematic of Experiment 1: Word identification task. Continua endpoints were real Cantonese words (e.g., 狼 *long4* ‘wolf’ – 浪 *long6* ‘ocean wave’) matched for frequency and pictured on either side of the screen. Right: Trial schematic of Experiment 2: Lexical decision task. Continua endpoints were either real Cantonese words or nonwords and pictured on either side of the screen with a thumbs up or down image, respectively. These left/right affiliation of images were counterbalanced across participants such that the presentation of images did not favour one side of the screen over the other.

randomization of the continuum steps from the merging (T2-T5, T3-T6, T4-T6) and non-merging (T2-T3, T5-T6) tone pairs, and asked to categorize each item using the pictures.⁷ Each item was repeated three times. Continuum steps 1 and 11 always correspond to the lower and higher numbered tones, respectively (e.g., continua for T4-T6 start at a resynthesis of a natural T4 production (step 1) and end at a resynthesis of a natural T6 production (step 11)). The tone at step 1 was always represented with an image on the left side of the screen (e.g., 狼 *long4* ‘wolf’). Step 11 was represented with an image on the right side of the screen (e.g., 浪 *long6* ‘ocean wave’). For the purposes of clarity, in the remainder of the paper, we describe the tonal categorization responses at continuum endpoints by referring to the proportion of responses to the left or right image. We do not refer to the specific tones corresponding to continuum endpoints as we ultimately collapse across several tone pairs for non-merging and merging tone pairs in subsequent analyses, making it difficult to refer to a specific tone at a particular continuum endpoint.

3.1. Analysis and results

Response latencies of less than 250 ms and those greater than three standard deviations from each individual subject’s mean were removed from the data set, removing approximately 4% of the responses. The remaining responses are visualized in **Figure 4**, where the proportion of responses for the image on the left is plotted as a function of continuum step for non-merging and

⁷ All of the pictures that were selected for the endpoints were included in a pretest. Six research assistants were asked to list the (English) words that were represented by each picture. The intended word was always included in the top three responses for each picture, suggesting that the selected pictures were representative of the intended word.

merging pairs. As mentioned, step 1 corresponds to the left image (e.g., 狼 *long4* ‘wolf’), while step 11 corresponds to the right image (e.g., 浪 *long6* ‘ocean wave’).

Data were analyzed with a Bayesian multilevel regression model using *brms* (Bürkner, 2017) in R using *cmdstanr* on the back end (Gabry & Češnovar, 2021; Stan Development Team, n.d.). Tone Type (Merging, Non-Merging; treatment coded with Non-Merging as the reference level), Continuum Step (z-transformed/standardized with respect to the original sequence/manipulation; e.g., Step 1 was consistently mapped to the same value, regardless of participants), and Language Dominance (z-transformed/standardized) were the population-level “effects”. The interaction between Step and Tone Type, and between Step and Dominance were included in the model.⁸ There were by-listener, by-item, and by-tone pair group-level effects (the “random intercepts”), with by-listener random slopes for Step, Tone Type, and their interaction; by-item random slopes for Step; by-tone pair random slopes for Step; and by-format (online versus in-person data collection) random slopes for Step, Tone Type, and their interaction. Item refers to the segmental string of the words and tone pair refers to the tones of the two items pitted against one another in a single trial. The model family was Bernoulli.⁹ The dependent variable was a binary response of whether the left image (1) or right image (0) was selected for each token.

Priors for all population-level effects were weakly informative priors of normal distributions with a mean of 0 and a standard deviation of 2 for the intercept and population-level parameters. The standard deviations for the group-level effects had a half normal distribution of mean 0 and standard deviation 1 as priors, and correlations used an LKJ prior of concentration 2. The model was fit using 4 Markov chains and 4000 samples each with 1000 warm-up samples per chain.

There were no divergent transitions and the \hat{R} values were all < 1.01 , suggesting well-mixed chains. Visual inspection of the graphical posterior predictive check indicated that the model fit the data well. To confirm the importance of the Step by Tone Type interaction, a second model was fit without the interaction in either the population or group-level effects with Subject and Format. The Bayesian leave-one-out estimate of expected log point-wise predictive density (ELPD-LOO) method was used for model comparison (Vehtari, Gelman, & Gabry, 2017). The

⁸ These particular two-way interactions were included in the statistical model because we had hypotheses about how Tone Type and Dominance will interact with Step – tones undergoing merger and English-dominant listeners are predicted to have shallower categorization functions. We did not have a hypothesis about the interaction between Tone Type and Dominance, so this interaction was not included in the model.

⁹ Final model syntax: Categorization \sim Step (centered, scaled) + Tone Type (Merging, Non-Merging; treatment-coded) + Dominance (centered, scaled) + Step:Tone Type + Step:Dominance + (1 + Step * Tone Type | Subject) + (1 + Step | Item) + (1 + Step | Tone Pair) + (1 + Step * Tone Type | Format (Online, in-person)), family = bernoulli(link = “logit”).

model with the Step by Tone Type interaction offered substantially better predictive accuracy (elpd-diff = -90.7, se-diff = 13.6). A third model with the three-way interaction of Step, Dominance, and Tone Type did not offer better predictive accuracy than the model that included the Step by Tone Type interaction (elpd-diff = -0.3, se-diff = 1.2). The model including the Step by Tone Type interaction, but not three-way interactions is reported below and in **Table 2**.

	$\hat{\beta}$	SE	95% CrI	PD
Intercept	0.97	0.72	[-0.57, 2.39]	0.92
Step	-1.81	0.63	[-2.83, -0.23]	0.98
Tone Type (Merging)	-0.87	0.87	[-2.55, 0.89]	0.85
Dominance	-0.06	0.06	[-0.17, 0.06]	0.82
Step:Tone Type (Merging)	1.23	0.69	[-0.36, 2.53]	0.955
Step:Dominance	0.22	0.08	[0.07, 0.37]	1.00

Table 2: Population-level or fixed-effect predictors for the Bayesian model for Experiment 1. The $\hat{\beta}$ estimate, standard error (SE), 95% Credible Interval (CrI), and Probability of Direction (PD) are reported.

The model results are interpreted as follows, adhering to the suggestions of Nicenboim and Vasishth (2016): When the 95% Credible Interval (CrI) for a given parameter excludes 0, this is considered strong evidence for an effect. The evidence for an effect is described as weak if the CrI includes 0, but the Probability of Direction (PD) is more than 0.95.

Figure 4 presents the empirical data showing the effect of Step and the interaction of Step by Tone Type in Experiment 1. There was strong evidence for an effect of Step (CrI: [-2.83, -0.23]; PD: 0.98). This indicates that listeners' categorization of the tokens changes in line with the continuum steps. There was also weak evidence for a Step by Tone Type interaction (CrI: [-0.36, 2.53]; PD: 0.95). For both merging and non-merging tone pairs, listeners showed sensitivity to tonal changes with each continuum step, but this sensitivity was greater for the non-merging pairs. This is shown by the fact that listeners were more extreme in the proportion of their responses at the endpoints, producing a more sigmoidal response function. Both of these descriptions suggest more categorical perception for non-merging pairs compared to merging pairs (Schertz & Clare, 2020).

Figure 4 also illustrates the interaction of Step by Dominance in Experiment 1, where Dominance, a continuous measure in the analysis, is split at 0 to visualize the distinction between Cantonese-dominant and English-dominant participants. We observed strong evidence for a Step by Dominance interaction (CrI: [0.07, 0.37]; PD: 1.00). These results show that the early Cantonese-English bilingual listeners who are more Cantonese-dominant have steeper categorization functions than listeners who are more English-dominant.

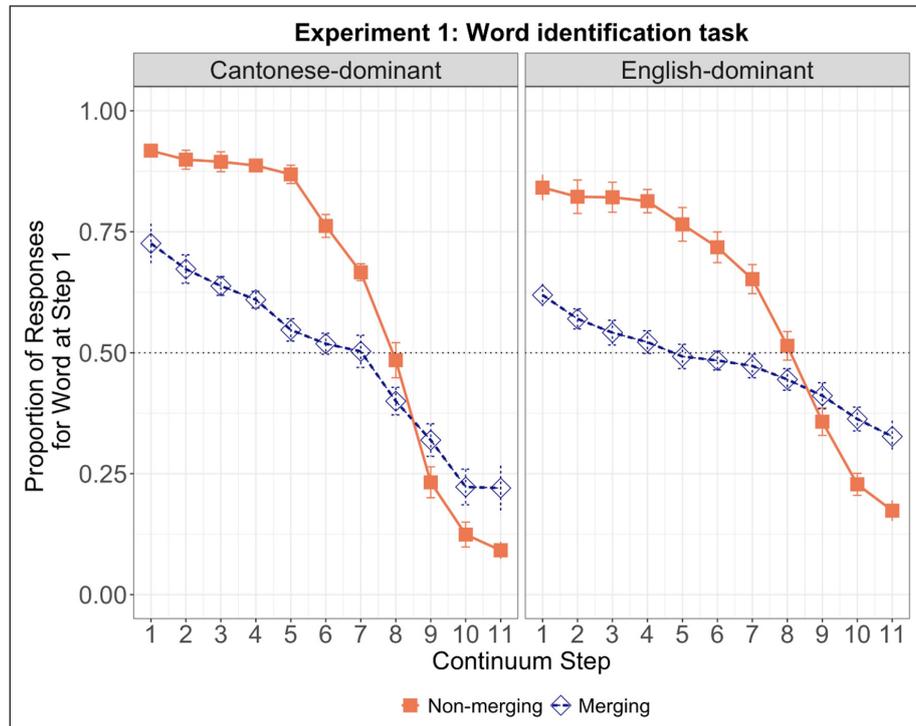


Figure 4: Proportion of trials categorized as the word at step 1 (i.e., the left image) for non-merging (orange solid squares) and merging (navy dashed diamonds) tone pairs by Cantonese-dominant (left) and English-dominant (right) participants in Experiment 1: Word identification task.

3.2. Interim discussion

In Experiment 1, where listeners were presented with tone continua where each endpoint is a real word, there is strong evidence that their responses vary according to the continuum step. Listeners are more likely to categorize each end of the continua as the appropriate word. The Bayesian analysis also indicates that there is weak evidence in support of our hypothesis that, crucially, Step is mediated by whether the tone pair is undergoing a merger or not. Listeners perceived non-merging tone pairs more categorically than merging tone pairs.

While there was strong evidence for an effect of Step, listeners' responses at continua endpoints were not at extreme probabilities. That is, the mean response for steps 1 and 11 were not 100% and 0%, respectively, for either of the Tone Types. It is not particularly surprising that listeners are not at floor and ceiling at these endpoints. While this was not a discrimination task, tone perception has been documented to be less categorical than segmental perception, particularly outside of sentential contexts that allow listeners to do more local normalization (Francis, Ciocca, & Ng, 2003; Sun & Huang, 2012). Listeners' acceptability of tone variation at the endpoints should not, therefore, be considered alarming.

We also observed strong evidence for an interaction between Step and Language Dominance. While all participants are early Cantonese-English bilinguals, thirteen are Cantonese-dominant

and 21 are English-dominant. The Cantonese-dominant listeners had a steeper categorization function than the English-dominant listeners across all tone types. That English-dominant early Cantonese-English bilinguals would be less categorical in their word categorization behavior across these tone continua is expected. As mentioned, a body of work has shown that phonological-lexical representations may be less well-specified for less proficient bilinguals. In so far as proficiency in Cantonese is correlated with dominance in Cantonese, Cantonese-dominant bilinguals may demonstrate more categoricity in their response functions as their Cantonese phonological-lexical representations are more well-specified. These results thus extend our understanding of language proficiency in the context of language dominance as a bilingual factor.

4. Experiment 2: Lexical decision task

In Experiment 2, we test how the absence of a lexical competitor affects tonal category boundaries, as the absence of an abutting lexical item may lead listeners to be more accepting of phonetic variation and a broader range of pronunciation variants. We expect that the absence of a tonal lexical competitor will increase the acceptability of a wider range of pronunciation variants for both merging and non-merging tone pairs, eliciting word endorsements from theoretical nonwords.

Listeners were presented with a full randomization of the continuum steps from the merging and non-merging tone pairs, and asked to respond as to whether they heard a real or nonword, by selecting the thumbs up or thumbs down image, respectively, since the word endpoint of each continua was not always imageable. An example screen shot of a trial is shown in **Figure 3**. As in Experiment 1, step 1 always corresponded to the first, lower numbered tone in each tone pair, while step 11 always corresponded to the second, higher numbered tone in each tone pair. However, as mentioned, since word-nonword tone pairs were used in Experiment 2, and each endpoint for a given tone pair was represented with a nonword (e.g., 甜 *tim4* ‘sweet’ – **tim6*, where T6 is the nonword, and **dei4* – 地 *dei6* ‘ground’, where T4 is the nonword), the lower numbered tone in each tone pair corresponded to a word for only half the items. The word/nonword response options (i.e., the thumbs up and thumbs down images) were counterbalanced across participants such that the presentation of images did not favour one side of the screen over the other. For consistency in subsequent analyses, we flip the continuum steps to position the real word endpoint for each tone pair at continuum step 1, and quantify the proportion of *word endorsements* at step 1. Items were not repeated so as to keep the experiment at a reasonable length. The order of the images was counter-balanced across listeners.

4.1. Analysis and results

Response latencies of less than 250 ms and those greater than three standard deviations from subject’s individual means were removed from the data set, removing approximately 4% of the responses. As mentioned, continuum step was flipped such that the real word endpoints of the continua were consistently positioned at continuum step 1. The data were analyzed with a

Bayesian multilevel regression model using *brms* (Bürkner, 2017) in R using *cmdstanr* on the back end (Gabry & Češnovar, 2021; Stan Development Team, n.d.).

Tone Type (Merging, Non-Merging; treatment coded with Non-Merging as the reference level), Continuum Step (z-transformed/standardized with respect to the original sequence/manipulation; e.g., Step 1 was consistently mapped to the same value, regardless of participants), and Language Dominance (z-transformed/standardized) were the population-level “effects”. The interaction between Step and Tone Type and that between Step and Dominance were included in the model. There were by-listener, by-item, and by-tone pair group-level effects (the random intercepts), with by-listener random slopes for Step, Tone Type, and their interaction; by-item random slopes for Step; by-tone pair random slopes for Step; and by-format (online versus in-person data collection) random slopes for Step, Tone Type, and their interaction. The model family was Bernoulli.¹⁰ The dependent variable was the binary response of word (1) or nonword (0).

Priors for all population-level effects were weakly informative priors of normal distributions with a mean of 0 and a standard deviation of 2 for the intercept and population-level parameters. The standard deviations for the group-level effects had a half normal distribution of mean 0 and standard deviation 1 as priors, and correlations used an LKJ prior of concentration 2. The model was fit using 4 Markov chains and 4000 samples each with 1000 warm-up samples per chain.

As with Experiment 1, to confirm the importance of the Step by Tone Type interaction, a second model was fit without the interaction in either the population or group-level effects with subject and format. The ELPD-LOO method was used for model comparison (Vehtari et al., 2017). The model with the Step by Tone Type interaction offered substantially better predictive accuracy (elpd-diff = -16.8, se-diff = 5.9). A third model with the three-way interaction of Step, Dominance, and Tone Type did not offer better predictive accuracy than the model that included the Step by Tone Type interaction (elpd-diff = -0.7, se-diff = 1.2). The model including the Step by Tone Type interaction is reported below and in **Table 3**. There were no divergent transitions and the \hat{R} values were all < 1.01, suggesting well-mixed chains. Visual inspection of the graphical posterior predictive check indicated that the model fit the data well.

Figure 5 illustrates the effect of Step and the interaction of Step by Tone Type in Experiment 2. There was no evidence for an effect of Step (CrI: [-1.71, 0.45]; PD: 0.92]), nor was there evidence for an interaction of Step and Tone Type (CrI: [-0.67, 1.28]; PD: 0.85), as indicated by the fact that the CrI includes 0 and the probability of direction is less than 0.95 in both cases. These results suggests that listeners are accepting of phonetic variation when there is no lexical competitor irrespective of the tone type.

¹⁰ Final model syntax: Lexical Decision ~ Step (centered, scaled) + Tone Type (Merging, Non-Merging; treatment-coded) + Dominance (centered, scaled) + Step:Tone Type + Step:dominance + (1 + Step * Tone Type | Subject) + (1 + Step | Item) + (1 + Step | Tone Pair) + (1 + Step * Tone Type | Format (Online, in-person)), family = bernoulli(link = “logit”).

	$\hat{\beta}$	SE	95% CrI	PD
Intercept	0.50	0.51	[-0.57, 1.46]	0.85
Step	-0.64	0.43	[-1.39, 0.35]	0.92
Tone Type (Merging)	0.48	0.51	[-0.60, 1.45]	0.84
Dominance	-0.02	0.12	[-0.25, 0.23]	0.55
Step:Tone Type (Merging)	0.32	0.40	[-0.59, 1.09]	0.84
Step:Dominance	0.14	0.06	[0.03, 0.25]	1.0

Table 3: Population-level or fixed-effect predictors for the Bayesian model for Experiment 2. The $\hat{\beta}$ estimate, Standard Error (SE), 95% Credible Interval (CrI), and Probability of Direction (PD) are reported.

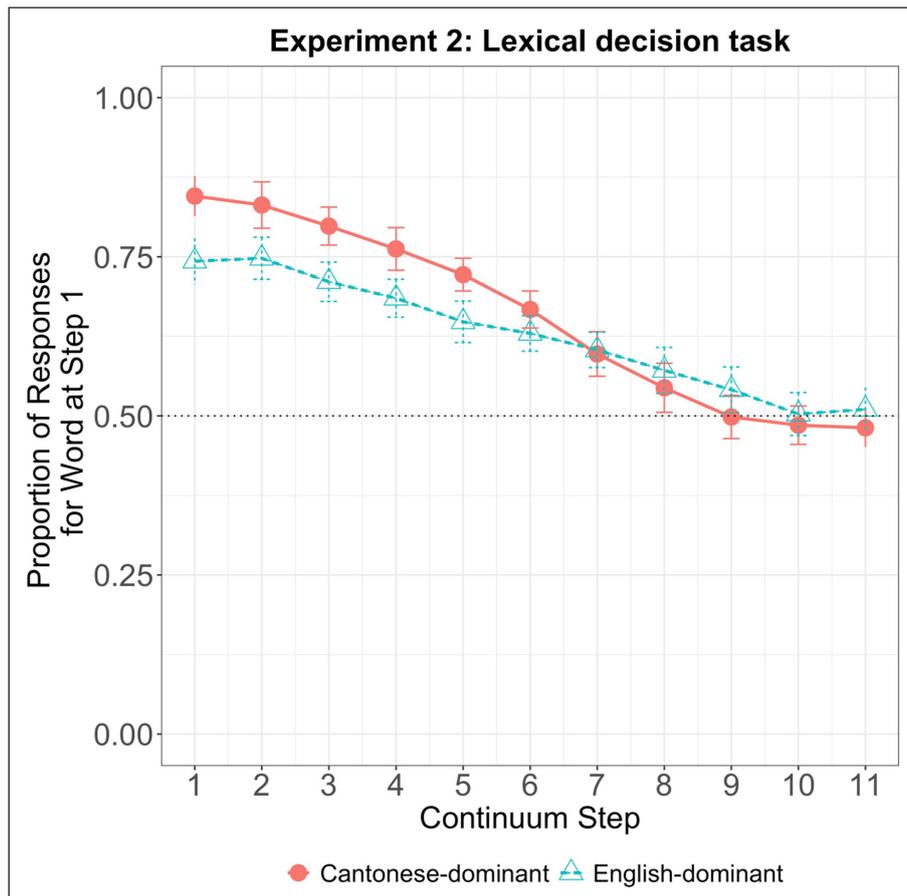


Figure 5: Proportion of trials with word responses for Cantonese-dominant (pink solid circles) and English-dominant (blue dashed triangles) participants in Experiment 2. The steps have been arranged so that the word endpoint is at step 1 and the nonword endpoint is at step 11.

Figure 5 illustrates the effect of Step and the interaction of Step by Dominance in Experiment 2. Again, this visualization splits the continuous Dominance measure at 0 to present the group-level results for the Cantonese-dominant and English-dominant participants. We observed strong evidence for a Step by Dominance interaction (CrI: [0.03, 0.24]; PD: 1.00) as the CrI excludes 0 and the PD is high at 1.00. These results show that the response functions for both Cantonese-dominant and English-dominant listeners are relatively flat, with the English-dominant listeners' functions slightly more so.

The analysis provides no evidence for the other factors, including, somewhat surprisingly the effect of Step. The Credible Intervals are all quite large, encompassing 0. To investigate this variation, we present **Figure 6**, which presents the by-participant responses by Step, colour coded by a categorical treatment of language dominance. These data show that participant responses nearly run the gamut. A handful of individuals look like S1, hovering around the 0.5 mark, being equally likely at any step to call the item a word or non-word. More individuals look like S12, showing a declination of a probability of word responses as the step approaches the non-word side of the continuum, but never dropping much, if at all, below chance (0.5) levels of word endorsement. Some participants, like s53, are well above 0.5 across the continua, being more likely to identify all items as words.

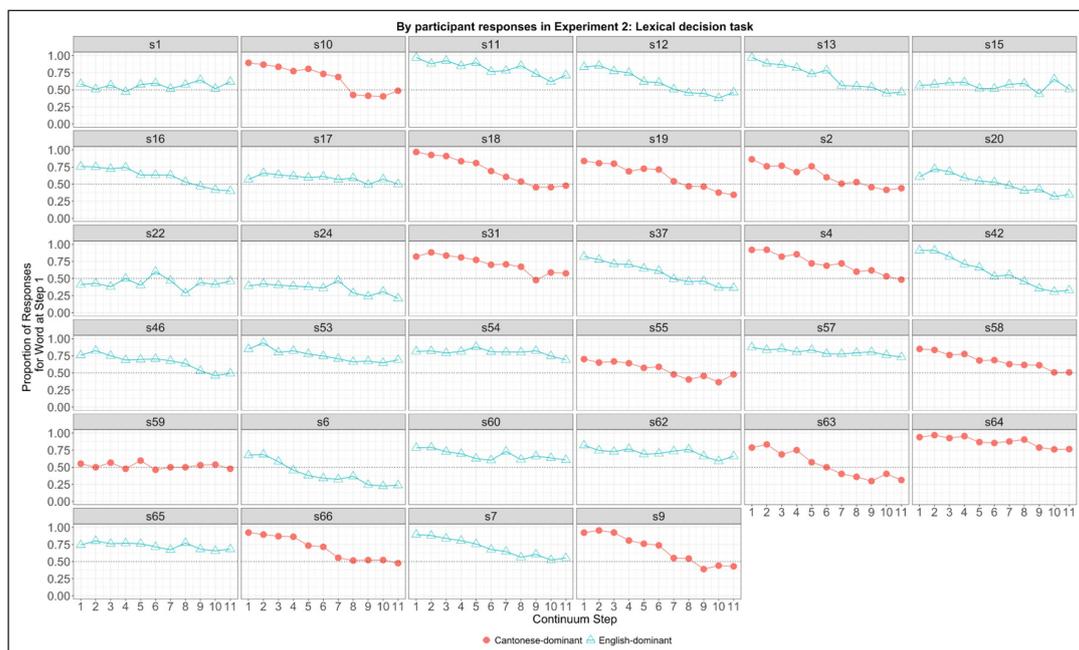


Figure 6: By-participant proportion of trials with word responses by continuum step for Experiment 2. Each panel represents a participant and their respective language dominance (pink solid circles: Cantonese-dominant participants, blue dashed triangles: English-dominant participants). The steps have been arranged so that the word endpoint is at step 1 and the nonword endpoint is at step 11.

4.2. Interim discussion

The results of Experiment 2 are simple. While the empirical responses show a shallow decline in word endorsement rates across the continua, the Bayesian results provide no evidence that listeners' identification of the tokens as words or nonwords along the tone continua varied by step. While the empirical results in **Figure 5** suggest that the majority of responses are above 0.5, suggesting that listeners provide more word endorsements until step 9 of the continuum, the results of the Bayesian model provide no evidence of this. The individual response functions shown in **Figure 6** reveal considerable listener variation. For example, participants S1 and S59 hover at chance across the continuum, while some listeners show more s-like patterns that cross the 0.5 line or stay above it across the continuum steps. A Step by Dominance interaction provides strong evidence, however, that Cantonese-dominant listeners have a steeper categorization function than English-dominant listeners. In the context of bilingualism, these results suggest that language dominance affects the encoding of phonetic detail, with greater dominance in Cantonese resulting in having better-specified phonological-lexical representations that allow for drawing more of a distinction between words and nonwords that differ in tone in Cantonese.

5. General discussion

The current study investigated the effect of lexical competition on the perception of Cantonese tone categories in Cantonese-English bilinguals, who varied in Cantonese dominance. Since lexical competition has been shown to play a role in contrast maintenance for diachronic sound changes, we examined a selection of both merging tone pairs (T2-T5, T3-T6, T4-T6) and non-merging tone pairs (T2-T3, T5-T6) in the language. To do so, we created tone minimal pairs (e.g., 狼 *long4* 'wolf' – 浪 *long6* 'ocean wave') and word-nonword tone pairs (e.g., 甜 *tim4* 'sweet' – **tim6* and **dei4* – 地 *dei6* 'ground') for the set of merging and non-merging tone pairs. Eleven-step continua were synthesized for each of the minimal pairs and word-nonword pairs in TANDEM-STRAIGHT (Kawahara et al., 2008). As minimal pairs bore real words at both endpoints, they were tested in a word identification task utilizing pictures (Experiment 1) to examine how the presence of a lexical competitor may affect listener categorization functions. Likewise, since word-nonword pairs only had a real word on one endpoint (while the other endpoint was a nonword), they were included in a lexical decision task (Experiment 2) to examine how the absence of a lexical competitor would affect listener categorization functions.

As predicted, merging tones exhibited less categorical response functions than non-merging tones in Experiment 1, where both continuum endpoints were words. In merging tone pairs, listeners were more accepting of pronunciation variation across the continuum, resulting in the endpoints being categorized as either lexical item. This suggests that words containing merging tones are perceived less categorically than those with non-merging tones. Importantly, the evidence in support of this was extant, but weak in our Bayesian analysis. The subtlety afforded

by a Bayesian analysis and interpretation is insightful here because we expect that such effects would be subtle, as these mergers are still progressing and listeners are still able to map the pronunciations to the intended lexical item.

In Experiment 2, where listeners engaged in a lexical decision task with one endpoint being a real word and the other being a nonword, there was no overall evidence for the merger status of the tone pair playing a role. The credible intervals were wide due to individual variation in Experiment 2. While many listeners were highly accepting of variation across the continua, mapping the pronunciation variation to the real word endpoint, some listeners' response functions hovered around 50% across the continua.

The degree to which listeners show robust categorical perception for tones has been shown to vary according to tone shape. Francis et al. (2003) find that, generally, Cantonese tone contrasts that involve contours are perceived more categorically than those that involve level tones. These findings that contour tone contrasts are perceived more categorically than level tones has been found with other tonal languages as well (e.g., Sun & Huang, 2012, Taiwanese). Additionally, there are individual differences in listeners' sensitivity to the merging tone contrasts in both behavioural and neurolinguistic (e.g., ERP) data (Ou & Law, 2017; Ou, Law, & Fung, 2015). There is also evidence for gradience in the encoding of different tone pairs. Maggu, Liu, Antoniou, and Wong (2016) examined brain stem, cortical, and behavioural responses to four different Cantonese tone pairs ranging from (and we use their terms here) "unmerged" (T1-T2) to "quasi-merger" (T3-T6) to "near-merger" (T4-T6) to "fully merged" (T2-T5). Across the incline of tone mergedness, they find gradience in perceptual sensitivity at all levels.

Yet, the Cantonese tone mergers affect tone pairs of all shapes (e.g., T2-T5, both contour; T3-T6, both level; and T4-T6, contour and level). A given tone pair's participation in a merger is, therefore, not entirely accounted for by the relationship between perceptual distinctiveness and tone shape (Francis et al., 2003; Sun & Huang, 2012). Given that we do not pair a categorization task with a discrimination task, we cannot speak directly to *categorical perception*. However, across both experiments, we see empirically that neither Cantonese- nor English-dominant listeners consistently respond with a particular word (or nonword) label on either end of the continuum. This suggests that, at least outside of a sentential context and syntactic, semantic, and pragmatic clues, listeners do not have particularly robust categorical responses at the end points. While direct comparison across our two experiments is not possible because of the lack of a shared dependent variable, the results are reminiscent of T. H. Yang et al. (2019) and Fox and Unkefer (1985), both of which observed more categorical perception of (non-merging) tones by Mandarin listeners when a word was present at both ends of a tonal continuum. Again, while we cannot make direct comparisons across our experiments, the current results are not inconsistent with the broader claim that lexical competitors restrict the range of acceptable phonetic variation (Baese-Berk & Goldrick, 2009; Goldrick et al., 2013), extending

such conclusions to the realm of suprasegmentals. Lexical competition may support the more categorical mapping of phonetic variation to words with tones, but more research is needed to corroborate this interpretation.

The results of both analyses revealed strong evidence that Cantonese-dominant Cantonese-English bilinguals have steeper categorization functions than English-dominant Cantonese-English bilinguals independent of the tone mergers.¹¹ This may suggest that English-dominant Cantonese-English bilinguals may have less phonological detail in their Cantonese phonological-lexical representations, prompting more variable labeling of words. This would be consistent with the fuzzy lexicon hypothesis proposed for L2 speech processing (Gor et al., 2010). Several studies have shown that categories that are confusable may produce lexical representations that lack phonological detail for L2 speakers (Darcy et al., 2013). While much of this literature situates confusable categories in the context of L2 learners who struggle to differentiate new L2 categories from similar L1 categories (Cook, Pandža, Lancaster, & Gor, 2016; Gor et al., 2010), the case of fuzzy tonal representations for the English-dominant Cantonese-English bilinguals represents a unique departure from this body of work in two ways. Firstly, these individuals are early bilinguals who are more dominant in their L2, English (see Table 4 in the Appendix). Secondly, the confusability of the sound categories in question (i.e., lexical tones) cannot be due to similarity equivalence with a category in the L2 (English). The L1 tonal categories bear no direct analogue in the L2. These results are not the first to suggest that the tone representations of some early Cantonese-English bilinguals may be fuzzy, though previous work has not used this term. Lam (2018) compared the role of tone in word identification with Cantonese-dominant and English-dominant early Cantonese-English bilinguals. Two findings from Lam's work speak to the fuzziness of tonal representations in the English-dominant group. English-dominant listeners were less accurate at identifying low-pass filtered words from their retained tonal information alone, and in the context of sentences, English-dominant listeners relied more on semantic context than tone, whereas Cantonese-dominant listeners attended to the tonal information at the expense of semantic predictability. More recently, Soo and Monahan (2023) provide evidence from a medium-term priming task that "heritage" Cantonese speakers treat tone minimal pairs like identity pairs, suggesting that tone may be encoded less precisely for English-dominant Cantonese-English bilinguals. All together, this suggests that lexical representations with less precise phonological detail may be a feature of a less-dominant language, not simply a late acquired second language (see also Soo & Monahan, 2022).

These results provide additional evidence for how these mergers affect phonetic and phonological category structure in Cantonese. Discrimination and categorization tasks with

¹¹ Model comparison did not warrant the inclusion of a three-way interaction with Continuum Step, Tone Type, and Language Dominance, which we interpret as dominance globally affecting categorization independent of the tone mergers.

naturally-produced Cantonese tones (not synthesized continua, as was used in this experiment) consistently show that the listeners are less accurate with merging tones (Fung & Lee, 2019; Lam, 2018; Mok et al., 2013; Soo & Monahan, 2017). In an AX discrimination task testing all possible tonal combinations, Fung and Lee (2019) observed the lowest accuracy rates for the merging pairs T2-T5 and T4-T6 by Hong Kong Cantonese speakers. Mok et al. (2013) also observed that participants who are merged in production were likewise slower at discriminating merging tone pairs in an AX task. These results were paralleled in a heritage speaker population by Soo and Monahan (2017) who observed low sensitivity (d') scores for T2-T5 as well as T3-T6 in an AX task. In a word-identification paradigm, Lam (2018) also observed overall higher confusion rates between tones in pairs T2-T5, T3-T6, and T4-T6 in heritage (raised in Canada) and homeland (raised in Hong Kong) listeners. The existing body of work on tone mergers is united by its use of naturally produced stimuli. By probing listener categorization across synthesized continua, our results provide a more comprehensive picture of the nature of the category boundaries of merging and non-merging Cantonese tones, and suggest that merging tones have less discrete category boundaries compared to non-merging tones when they have lexical competitors.

Overall, what do these results suggest about the nature of tone categories? Following Wedel and Fatkullin (2017) and Yu (2007), we propose that phonological categories are generalizations over distributions of encoded items (Pierrehumbert, 2001; Yu, 2007), which, if merging, bear potentially substantial areas of overlap with other tonal distributions at category boundaries. Under this assumption, speech recognition involves the mapping of percepts to categories which occupy a similar perceptual space, and competition between categories acts to maintain category contrasts (Wedel & Fatkullin, 2017). In a one-dimensional space, the distribution of these categories are subject to “entrenchment”, acting to tighten category distribution and “noise”, acting to broaden the category distribution. The balance between these two forces can be disrupted in cases where two categories approach one another along a shared acoustic dimension. As the region of overlap at the category boundary begins to increase, usage frequency comes into play in promoting greater activation of one of these categories (Pierrehumbert, 2001). Since categories compete for new percepts, the category with increased activation will be assigned more percepts, gradually increasing the area of overlap between the two categories even more (and eventually, theoretically, eradicating the distinction between the categories entirely in favour of one category). In this sense, the boundary between merging tone categories may be less crisp than that of non-merging tones, as observed in Experiment 1, given that the distribution of exemplars for merging category distributions overlap at category boundaries (e.g., Fung & Lee, 2019; Mok et al., 2013). The nature of the category overlap need not be symmetric (e.g., Harrington et al., 2019) and, indeed, they likely are not in Cantonese tones. In a study of the Cantonese tone mergers, Mok et al. (2013) describes T5 as more variable than T2, T6 as more variable than T4, and T6 and T3 as equally variable. These findings appear to align with tonal

type frequency; the more variable tone in a given merging tone pair is also the one with lower type frequency. Yet, the perceptual mislabelling of these tones is rather variable within and across speakers; for instance, sometimes T2 is labeled as T5 more often despite T5 being more variable. This suggests that the nature of the mergers, in terms of their distributional changes, is subject to considerable variation across the population of Cantonese speakers. At the same time, listeners are still able to perceive the difference between merging tones overall, as the extreme ends of these distributions do not overlap, at least for some speakers, including the one used in the current study.

Wedel and Fatkullin (2017) also provide an account for the contribution of lexical factors. When two categories begin to approach one another, noise in the opposing extreme ends of each distribution can pull the category distributions away from the regions of overlap at category boundaries. This “variant trading” behaviour, which would normally act to sharpen category boundaries, is inhibited when ambiguous incoming percepts are disambiguated by the lack of a lexical competitor. In other words, the percept can easily be categorized using lexical knowledge about whether the resulting token would produce a real word in the language (e.g., Ganong, 1980; Norris et al., 2003). The *presence* of a lexical competitor thus, promotes variant trading and category contrast, while the *absence* of a lexical competitor does not. This is a potential explanation for our finding that listeners’ categorization at each continuum step varies as a function of tone type in Experiment 1 where there were lexical competitors, but not in Experiment 2 where there were no lexical competitors.

Finally, on a broader level, this study provides evidence for the notion that gradual phonetic change may be affected in cases where maintaining recognition in competing categories is paramount. Words with and without lexical competitors are parsed differently with respect to phonetic variation. The category threshold for tones is subject to the lexical status of the word in which it is part. Our research question centered on the role of mergers in the perception of phonetic variation, which meant we grossly grouped tones into their respective merging and non-merging categories, despite each tone merger being at a different stage. Future work should explore the tone pair specific behavior, which may not only be conditioned by its merger status, but also its functional load (Tsui, 2012).

6. Conclusion

We carried out two experiments to examine the effect of lexical competition on the categorization of merging and non-merging tone pairs in Cantonese. Testing early Cantonese-English bilinguals, Experiment 1 showed that category boundaries between non-merging tone categories are more categorical, sharpened by the presence of a lexical competitor in the abutting tone category, while merging tone boundaries were less categorical in the face of lexical competitors. Experiment 2

demonstrated some listeners were accepting of pronunciation variation across the board for merging and non-merging tones when there is no abutting lexical competitor, but there were no group-level differences based on continuum step or a tone pair's participation in a merger. In both experiments, Cantonese-dominant Cantonese-English bilinguals exhibited greater differences across continua than English-dominant Cantonese-English bilinguals.

These data can be explained if we conceive of a tone category as a distribution of perceptual exemplars. Since tone categories that are merging bear areas of potential overlap with one another, the boundary between these categories will be less discrete while the distribution of the categories themselves encapsulates a wider range of phonetic variation as good exemplars of the corresponding merging tone category. Furthermore, we observed that language dominance may play a role in the specificity of phonological-lexical representations, as category boundaries were more discrete for Cantonese-dominant bilingual listeners overall. These data provide insight into the lexicon's contribution to the ongoing tone mergers in Cantonese, demonstrate the effects of lexical competition in structuring phonetic variation for suprasegmental sound categories, and contribute to our understanding of bilingual variation in phonological and lexical representations.

A. Appendix

Subject	Dominance	Cantonese AoA	Cantonese Rating	English AoA	English Rating
s1	85	Birth	5	5	6
s2	-50	Birth	6	Birth	4
s4	-6	Birth	6	1	6
s6	79	Birth	5	4	6
s7	79	Birth	4	1	6
s9	-46	Birth	6	6	6
s10	-59	Birth	6	4	5
s11	34	2	5	8	5
s12	121	Birth	4	Birth	6
s13	56	Birth	5	3	6
s15	127	Birth	4	Birth	6
s16	97	Birth	3	3	5
s17	139	Birth	3	8	6
s18	-26	Birth	6	2	5
s19	-13	Birth	6	2	6
s20	52	Birth	5	Birth	6
s22	60	Birth	4	6	5
s24	25	Birth	6	3	6
s31	-28	Birth	5	9	5
s37	55	Birth	5	3	6
s42	108	Birth	5	5	6
s46	51	4	5	7	5
s53	63	Birth	4	2	6
s54	139	Birth	6	0	6
s55	-1	Birth	5	0	6
s57	80	Birth	4	3	6
s58	-99	1	6	8	4
s59	-0.36	Birth	6	6	6
s60	55	3	5	5	6
s62	102	Birth	5	5	6
s63	-15	Birth	6	0	6
s64	-0.08	3	6	10	5
s65	123	Birth	2	3	6
s66	-19	Birth	6	1	5

Table 4: Participant information about Cantonese and English Age of Acquisition (AoA), Cantonese and English self-ratings for understanding (out of 7), and BLP dominance scores rounded to the nearest whole number.

Tone Pair	Jyutping	Character	English gloss	Jyutping	Character	English gloss
T2-T3	fan2	粉	powder	fan3	瞓	sleep
T2-T3	gwai2	鬼	ghost	gwai3	貴	expensive
T2-T3	geng2	頸	neck	geng3	鏡	mirror
T2-T3	deng2	頂	roof	deng3	掙	throw
T2-T3	fu2	虎	tiger	fu3	褲	pants
T2-T3	gwan2	滾	boil	gwan3	棍	rod
T2-T3	tong2	糖	candy	tong3	熨	iron
T2-T3	zin2	剪	to cut	zin3	箭	arrow
T2-T3	zaau2	爪	a claw	zaau3	罩	mouth mask
T2-T3	sai2	洗	wash	sai3	細	small
T2-T3	paa2	扒	steak	paa3	怕	fearful
T5-T6	hau5	厚	thick	hau6	後	behind
T5-T6	jim5	染	dye	jim6	驗	written test
T5-T6	jyun5	軟	soft	jyun6	願	wish
T5-T6	lai5	禮	a gift	lai6	荔	lychee
T5-T6	laan5	懶	lazy	laan6	爛	rotten
T5-T6	lei5	浬	sail	lei6	脷	tongue
T5-T6	lou5	老	old	lou6	路	road
T5-T6	maa5	馬	horse	maa6	罵	scold
T5-T6	maan5	晚	night time	maan6	慢	slow
T5-T6	mun5	滿	full	mun6	悶	bored
T5-T6	ngo5	我	me	ngo6	餓	hungry
T2-T5	si2	屎	poop	si5	市	city
T2-T5	jyu2	瘀	a bruise	jyu5	雨	rain
T2-T5	sin2	癬	ringworm	sin5	鱧	eel
T2-T5	tou2	禱	prayer	tou5	肚	stomach
T3-T6	beng3	柄	a handle	beng6	病	sick
T3-T6	dung3	凍	cold	dung6	動	exercise
T3-T6	giu3	叫	call out	giu6	撬	pry open
T3-T6	zoeng3	醬	sauce	zoeng6	象	elephant
T4-T6	long4	狼	wolf	long6	浪	ocean wave
T4-T6	jau4	游	swim	jau6	右	right (direction)
T4-T6	se4	蛇	snake	se6	射	shoot
T4-T6	mou4	毛	fur	mou6	霧	fog

Table 5: Minimal pairs used in Experiment 1: Word identification task.

Tone Pair	Jyutping	Character	English gloss	Nonword Match in Jyutping
T2-T3	pui3	沛	abundant	pui2
T2-T3	ping3	聘	invite for service	ping2
T2-T3	paai3	派	distribute	paai2
T2-T3	leng3	靚	beautiful	leng2
T2-T3	kau3	釦	buckle	kau2
T2-T3	kaau3	靠	lean on for a favour	kaau2
T2-T3	hing3	熨	to warm	hing2
T2-T3	gwaan3	慣	accustomed to	gwaan2
T2-T3	goe3	鋸	to saw	goe2
T2-T3	faai3	快	fast	faai2
T2-T3	caa3	岔	crossroads	caa2
T2-T3	wun2	碗	bowl	wun3
T2-T3	mung2	懵	ignorant	mung3
T2-T3	mo2	摸	touch	mo3
T2-T3	mang2	攞	to be irritated	mang3
T2-T3	loeng2	兩	a unit of weight	loeng3
T2-T3	lem2	舐	lick	lem3
T2-T3	hoi2	海	ocean	hoi3
T2-T3	dai2	抵	worth the money	dai3
T2-T3	daa2	打	to fight	daa3
T2-T3	cin2	淺	shallow	cin3
T2-T3	ceng2	請	invite	ceng3
T5-T6	waai6	壞	spoilt	waai5
T5-T6	meng6	命	one's life	meng5
T5-T6	gwai6	跪	kneel	gwai5
T5-T6	gui6	劬	tired	gui5
T5-T6	gau6	舊	worn, old	gau5
T5-T6	ding6	定	or	ding5
T5-T6	daam6	啖	a bite, mouthful	daam5
T5-T6	daai6	大	big	daai5
T5-T6	bou6	步	a step	bou5
T5-T6	bei6	備	ready	bei5
T5-T6	baan6	扮	dress up as someone else	baan5

(Contd.)

T5-T6	tyun5	斷	break	tyun6
T5-T6	pui5	倍	double	pui6
T5-T6	pou5	抱	hug	pou6
T5-T6	pei5	被	blanket	pei6
T5-T6	leng5	領	shirt collar	leng6
T5-T6	keoi5	佢	him, her, she, he	keoi6
T5-T6	kei5	倚	to stand	kei6
T5-T6	kan5	近	near, close to	kan6
T5-T6	cou5	儲	save up	cou6
T5-T6	co5	坐	sit	co6
T5-T6	ci5	似	similar to	ci6
T2-T5	wing5	永	forever	wing2
T2-T5	mei5	美	pretty	mei2
T2-T5	laang5	冷	chilly	laang2
T2-T5	je5	嘢	thing	je2
T2-T5	zai2	仔	son	zai5
T2-T5	sau2	手	hand	sau5
T2-T5	fo2	火	fire	fo5
T2-T5	beng2	餅	cookie	beng5
T3-T6	maai6	賣	sell	maai3
T3-T6	lyun6	亂	messy, disorganized	lyun3
T3-T6	lin6	練	practice	lin3
T3-T6	doi6	代	era	doi3
T3-T6	taam3	探	visit	taam6
T3-T6	hoeng3	向	towards	hoeng6
T3-T6	gwaa3	掛	hang up	gwaa6
T3-T6	gaai3	界	boundary	gaai6
T4-T6	daan6	彈	elasticity	daan4
T4-T6	jaa6	廿	twenty	jaa4
T4-T6	gam6	揸	to press	gam4
T4-T6	dei6	地	ground	dei4
T4-T6	tim4	甜	sweet	tim6
T4-T6	tau4	頭	head	tau6
T4-T6	kung4	窮	destitute, poor	kung6
T4-T6	haang4	行	walk	haang6

Table 6: Word-nonword pairs used in Experiment 2: Lexical decision task.

Supplementary files

The stimuli, and the complete data set and the code for the main analysis are available on <https://osf.io/bcmv4/>.

Ethics and consent

All of the experiments reported in this study were approved by the Behavioral Research Ethics Board at the University of British Columbia.

Acknowledgements

We thank Martin Oberg for programming assistance on this project and we thank Roger Lo for his guidance on our Bayesian analysis. We also thank Zoe Lam for recording stimuli, and Khia Johnson, Kathleen Currie Hall and Amanda Cardoso for comments on the manuscript. Lastly, our appreciation goes to Angelina Lloy, Fion Fung, Christina Sen, Ariana Zattera, Ellie Yoon, Vince Dinoto, and other members of the Speech in Context Laboratory for their assistance on various stages of this project.

Funding information

This work has been supported by an NSERC Discovery grant (F17-05152) to MB.

Competing interests

The authors have no competing interests to declare.

Author contributions

RS contributed to conceptualization, data curation, investigation, methodology, formal analysis, visualization, writing-original draft, and writing-reviewing and editing. MB contributed to conceptualization, data curation, investigation, methodology, formal analysis, visualization, writing-original draft, and writing-reviewing and editing.

References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. DOI: <https://doi.org/10.3758/s13428-019-01237-x>
- Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527–554. DOI: <https://doi.org/10.1080/01690960802299378>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>

- Bauer, R. S., & Benedict, P. K. (1997). *Modern cantonese phonology* (Vol. 102.;102;). New York; Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110823707>
- Bauer, R. S., Kwan-Hin, C., & Pak-Man, C. (2003). Variation and merger of the rising tones in hong kong cantonese. *Language Variation and Change*, 15(2), 211. DOI: <https://doi.org/10.1017/S0954394503152039>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. DOI: <https://doi.org/10.18637/jss.v080.i01>
- Casillas, J. V., & Simonet, M. (2016). Production and perception of the english/æ/–/A/ contrast in switched-dominance speakers. *Second Language Research*, 32(2), 171–195. DOI: <https://doi.org/10.1177/0267658315608912>
- Chan, Y.-Y. F. (1974). *A perceptual study of tones in cantonese*. Centre of Asian Studies, University of Hong Kong.
- Chao, Y. R. (1947). *Cantonese primer: Cambridge, mass., pub. for the harvard-yenching institute [by harvard university press, 1947]. character text*. University Microfilms.
- Charpentier, F., & Stella, M. (1986). Diphone synthesis using an overlap-add technique for speech waveforms concatenation. In *Icassp'86. ieee international conference on acoustics, speech, and signal processing*, 11, 2015–2018. DOI: <https://doi.org/10.1109/ICASSP.1986.1168657>
- Cleveland, W., Grosse, E., & Shyu, W. (1992). Local regression models. chapter 8 in statistical models in s (jm chambers and tj hastie eds.), 608 p. *Wadsworth & Brooks/Cole, Pacific Grove, CA*.
- Connine, C. M., & Clifton, C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human perception and performance*, 13(2), 291. DOI: <https://doi.org/10.1037/0096-1523.13.2.291>
- Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37(4), 463–480. DOI: <https://doi.org/10.1006/jmla.1997.2535>
- Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology*, 7, 1345. DOI: <https://doi.org/10.3389/fpsyg.2016.01345>
- Cutler, A., & Chen, H.-C. (1997). Lexical tone in cantonese spoken-word processing. *Perception & Psychophysics*, 59(2), 165–179. DOI: <https://doi.org/10.3758/BF03211886>
- Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, 8(3), 372–420. DOI: <https://doi.org/10.1075/ml.8.3.06dar>
- Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, 69–90.
- Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics*, 65(7), 1029–1044. DOI: <https://doi.org/10.3758/BF03194832>
- Fung, R. S., & Lee, C. K. (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. DOI: <https://doi.org/10.1121/1.5133661>

- Gabry, J., & Češnovar, R. (2021). cmdstanr: R interface to 'cmdstan' [Computer software manual]. (<https://mc-stan.org/cmdstanr>, <https://discourse.mc-stan.org>)
- Gandour, J. (1981). Perceptual dimensions of tone: Evidence from cantonese. *Journal of Chinese Linguistics*, 20–36.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110. DOI: <https://doi.org/10.1037/0096-1523.6.1.110>
- Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. *Measuring L2 Proficiency: Perspectives from SLA*, 208–225. DOI: <https://doi.org/10.21832/9781783092291-014>
- Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America*, 134(2), EL172–EL177. DOI: <https://doi.org/10.1121/1.4812821>
- Gor, K., Cook, S., & Jackson, S. (2010). Lexical access in highly proficient late l2 learners: Evidence from semantic and phonological auditory priming. In *Second language research forum (slrf)*, university of maryland.
- Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/- fronting, and sound change in standard southern british: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, 123(5), 2825–2835. DOI: <https://doi.org/10.1121/1.2897042>
- Harrington, J., Kleber, F., Reubold, U., Schiel, F., Stevens, M., & Assmann, P. (2019). The phonetic basis of the origin and spread of sound change. *Routledge Handbook of Phonetics*, 401–426. DOI: <https://doi.org/10.4324/9780429056253-15>
- Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, 139, 83–91. DOI: <https://doi.org/10.1016/j.cognition.2015.02.012>
- Jekel, C. F., Venter, G., Venter, M. P., Stander, N., & Haftka, R. T. (2019, may). Similarity measures for identifying material parameters from hysteresis loops using inverse analysis. *International Journal of Material Forming*. DOI: <https://doi.org/10.1007/s12289-018-1421-8>
- Kan, R. T., & Schmid, M. S. (2019). Development of tonal discrimination in young heritage speakers of cantonese. *Journal of Phonetics*, 73, 40–54. DOI: <https://doi.org/10.1016/j.wocn.2018.12.004>
- Kawahara, H., Masuda-Katsuse, I., & DeCheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous frequency-based f0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 27(3–4), 187–207. DOI: [https://doi.org/10.1016/S0167-6393\(98\)00085-5](https://doi.org/10.1016/S0167-6393(98)00085-5)
- Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 3933–3936). DOI: <https://doi.org/10.1109/ICASSP.2008.4518514>

- Khouw, E., & Ciocca, V. (2007). Perceptual correlates of cantonese tones. *Journal of Phonetics*, 35(1), 104–117. DOI: <https://doi.org/10.1016/j.wocn.2005.10.003>
- Kirby, J. P., & Yu, A. C. (2007). Lexical and phonotactic effects on wordlikeness judgments in cantonese. In *Proceedings of the international congress of the phonetic sciences xvi* (Vol. 13891392).
- Lam, W. M. (2018). *Perception of lexical tones by homeland and heritage speakers of Cantonese* (Unpublished doctoral dissertation). University of British Columbia.
- Lee, K. Y., Chan, K. T., Lam, J. H., Van Hasselt, C., & Tong, M. C. (2015). Lexical tone perception in native speakers of cantonese. *International Journal of Speech-Language Pathology*, 17(1), 53–62. DOI: <https://doi.org/10.3109/17549507.2014.898096>
- Luke, K.-K., & Wong, M. L. (2015). The hong kong cantonese corpus: design and uses. *Journal of Chinese Linguistics*, 25(2015), 309–330.
- Luthra, S., Guediche, S., Blumstein, S. E., & Myers, E. B. (2019). Neural substrates of subphonemic variation and lexical competition in spoken word recognition. *Language, Cognition and Neuroscience*, 34(2), 151–169. DOI: <https://doi.org/10.1080/23273798.2018.1531140>
- Maggu, A. R., Liu, F., Antoniou, M., & Wong, P. (2016). Neural correlates of indicators of sound change in Cantonese: Evidence from cortical and subcortical processes. *Frontiers in Human Neuroscience*, 652. DOI: <https://doi.org/10.3389/fnhum.2016.00652>
- Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, 62(4), 407–420. DOI: <https://doi.org/10.1016/j.jml.2010.02.004>
- Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. *Attention and Performance: Control of Language Processes*, 125–150.
- Matthews, S., & Yip, V. (2013). *Cantonese: A comprehensive grammar*. Routledge. DOI: <https://doi.org/10.4324/9780203835012>
- McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, 18(1), 1–86. DOI: [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- Mok, P. P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: Tone merging in hong kong cantonese. *Language Variation and Change*, 25(3), 341. DOI: <https://doi.org/10.1017/S0954394513000161>
- Moore, B. C., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3), 750–753. DOI: <https://doi.org/10.1121/1.389861>
- Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—part ii. *Language and Linguistics Compass*, 10(11), 591–613. DOI: <https://doi.org/10.1111/lnc3.12207>
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. DOI: [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Ou, J., & Law, S.-P. (2017). Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. *Attention, Perception, & Psychophysics*, 79(3), 945–963. DOI: <https://doi.org/10.3758/s13414-017-1283-z>

- Ou, J., Law, S.-P., & Fung, R. (2015). Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review*, 22(6), 1725–1732. DOI: <https://doi.org/10.3758/s13423-015-0839-y>
- Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. DOI: <https://doi.org/10.1016/j.jbef.2017.12.004>
- Pierrehumbert, J. (2001). Lenition and contrast. *Frequency and the Emergence of Linguistic Structure*, 45, 137. DOI: <https://doi.org/10.1075/tsl.45.08pie>
- Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285–290. DOI: <https://doi.org/10.3758/BF03213946>
- Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 1037. DOI: <https://doi.org/10.1037/0278-7393.21.4.1037>
- Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19(4), 699. DOI: <https://doi.org/10.1037/0096-1523.19.4.699>
- Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, 29(2), 149–188. DOI: <https://doi.org/10.1006/cogp.1995.1014>
- Pitt, M. A., & Samuel, A. G. (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1120. DOI: <https://doi.org/10.1037/0096-1523.32.5.1120>
- Psychology Software Tools. (2012). *E-prime 2 (Version 2.0)*. Retrieved from <https://www.pstnet.com/eprime.cfm>
- R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org>
- Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 19(5), 394–398. DOI: <https://doi.org/10.3758/BF03199398>
- Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474. DOI: <https://doi.org/10.1037/0096-3445.110.4.474>
- Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125(1), 28. DOI: <https://doi.org/10.1037/0096-3445.125.1.28>
- Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, 32(2), 97–127. DOI: <https://doi.org/10.1006/cogp.1997.0646>
- Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12(4), 348–351. DOI: <https://doi.org/10.1111/1467-9280.00364>
- Samuel, A. G., & Frost, R. (2015). Lexical support for phonetic perception during nonnative spoken word recognition. *Psychonomic Bulletin & Review*, 22(6), 1746–1752. DOI: <https://doi.org/10.3758/s13423-015-0847-y>

- Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language*, *81*, 51–71. DOI: <https://doi.org/10.1016/j.jml.2015.01.003>
- Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2), e1521. DOI: <https://doi.org/10.1002/wcs.1521>
- Schouten, B., Gerrits, E., & Van Hoesen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. DOI: [https://doi.org/10.1016/S0167-6393\(02\)00094-8](https://doi.org/10.1016/S0167-6393(02)00094-8)
- Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in mandarin spoken word processing. *Language and Speech*, *58*(2), 131–151. DOI: <https://doi.org/10.1177/0023830914522956>
- Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). Voicesauce: A program for voice analysis. *Proceedings of the ICPHS XVII*, 1846–1849.
- 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 *Berkeley linguistics society* (p. 47).
- Soo, R., & Monahan, P. J. (2022). Express: Language dominance and order of acquisition affect auditory translation priming in heritage speakers. *Quarterly Journal of Experimental Psychology*, *17470218221091753*. DOI: <https://doi.org/10.31219/osf.io/54j9h>
- Soo, R., & Monahan, P. J. (2023). Phonetic and lexical encoding of tone in cantonese heritage speakers. *Language and Speech*, *00238309221122090*. Retrieved from (PMID: 36172645). DOI: <https://doi.org/10.1177/00238309221122090>
- Soo, R., Sidiqi, A., Shah, M., & Monahan, P. J. (2020). Lexical bias in second language perception: Word position, age of arrival, and native language phonology. *The Journal of the Acoustical Society of America*, *148*(4), EL326–EL332. DOI: <https://doi.org/10.1121/10.0002116>
- Stan Development Team. (n.d.). *Stan modeling language users guide and reference manual*.
- 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. DOI: <https://doi.org/10.1007/s10831-012-9092-9>
- Surendran, D., & Levow, G.-A. (2004). The functional load of tone in mandarin is as high as that of vowels. In *Speech prosody 2004, international conference* (pp. 99–102). Retrieved from <https://www.isca-speech.org/archiveopen/sp2004/sp04099.html>
- Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, *185*, 1–20. DOI: <https://doi.org/10.1016/j.cognition.2019.01.004>
- Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in mandarin chinese. *Language and Cognitive Processes*, *23*(5), 689–708. DOI: <https://doi.org/10.1080/01690960701728261>
- Tsui, T.-H. (2012). Tonal variation in hong kong cantonese: Acoustic distance and functional load. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *48*, 579–588.
- Vance, T. J. (1976). An experimental investigation of tone and intonation in cantonese. *Phonetica*, *33*(5), 368–392. DOI: <https://doi.org/10.1159/000259793>

- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5), 1413–1432. DOI: <https://doi.org/10.1007/s11222-016-9696-4>
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917), 392–393. DOI: <https://doi.org/10.1126/science.167.3917.392>
- Wedel, A., & Fatkullin, I. (2017). Category competition as a driver of category contrast. *Journal of Language Evolution*, 2(1), 77–93. DOI: <https://doi.org/10.1093/jole/lzx009>
- Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*, 56(3), 395–417. DOI: <https://doi.org/10.1177/0023830913489096>
- Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2), 179–186. DOI: <https://doi.org/10.1016/j.cognition.2013.03.002>
- Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, 100, 61–88. DOI: <https://doi.org/10.1016/j.jml.2018.01.001>
- Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in mandarin chinese. *Language and Speech*, 59(1), 59–82. DOI: <https://doi.org/10.1177/0023830915578000>
- Wong, T.-S. (2008). The beginning of merging of the tonal categories b2 and c1 in hong kong cantonese. *Journal of Chinese Linguistics*, 36(1), 155–174.
- Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79(7), 2064–2072. DOI: <https://doi.org/10.3758/s13414-017-1361-2>
- Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language*, 36(2), 165–187. DOI: <https://doi.org/10.1006/jmla.1996.2482>
- Yang, Q., & Chen, Y. (2022). Phonological competition in mandarin spoken word recognition. *Language, Cognition and Neuroscience*, 1–24. DOI: <https://doi.org/10.1080/23273798.2021.2024862>
- Yang, T. H., Jin, S. J., & Lu, Y. A. (2019). The effect of mandarin accidental gaps on perceptual categorization. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 2022–2026*.
- Ye, Y., & Connine, C. M. (1999). Processing spoken chinese: The role of tone information. *Language and Cognitive Processes*, 14(5–6), 609–630. DOI: <https://doi.org/10.1080/016909699386202>
- Yu, A. C. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, 187–214. DOI: <https://doi.org/10.1017/S0952675707001157>

