The influence of functional load on Cantonese lexical tone perception*

Roger Yu-Hsiang Lo
University of British Columbia

Abstract: This paper examines the role of acoustic distance and functional load in the perceived similarity of Cantonese lexical tones across three groups of participants with different linguistic backgrounds — homeland native Cantonese speakers, native English speakers who do not have knowledge about Cantonese, and heritage speakers of Cantonese. The perceived similarity was characterized using accuracy and reaction time in the current study. As expected, acoustic distance, estimated as the root-mean-square of the fundamental frequency difference between the two tones in question, is a significant predictor for the perceived similarity for all three groups. While functional load does emerge as a significant predictor for native and heritage speakers of Cantonese, it is also found to have effect on native English speakers. It is speculated that this unexpected effect of functional load on native English speakers might be attributed to the acoustic distance measure here not capturing all of the relevant psychoacoustic aspects of lexical tones. The results in the current study are also compatible with a model of language which treats phonological generalizations as over word-level units, and lend support to the importance and long-term benefits of early language experience, even if such experience is limited drastically beyond childhood.

Keywords: functional load, acoustic distance, speech perception, Cantonese, lexical tone, heritage language

1 Introduction

The perceived similarity of speech sounds is influenced by a number of factors. On the one hand, signal-based acoustic properties of the sounds in question are probably the most evident and fundamental to any model of perceptual similarity (e.g., Martin and Peperkamp 2017; Strange 2007). On the other hand, a few studies have pointed to more abstract factors relating to a language's phonological system and patterns of usage, such as the phonological status of a pair of sounds (e.g., Boomershine et al. 2008; Johnson and Babel 2010; Kazanina et al. 2006). The current study investigates how one such phonological aspect — functional load — of Cantonese lexical tones shapes their perceived similarity across three groups of speakers with different linguistic backgrounds — native Cantonese speakers, native English speakers, and early Cantonese-English bilinguals[1] who speak Cantonese as a heritage language (I refer to them as heritage speakers for simplicity henceforth). Given that acoustic similarity is fundamental to perceived similarity, by controlling for the acoustic properties

*I would like to acknowledge my colleague Zoe Lam for her assistance in helping me complete this project and thank Prof. Kathleen Currie Hall, Prof. Molly Babel, and Prof. Douglas Pulleyblank for their guidance.
Contact info: ylo@mail.ubc.ca

[1]Note that many of the participants in this group in the current study consider both Cantonese and English as their native languages, so in some sense they are also native speakers of Cantonese and English. The terms — native Cantonese, native English, and heritage speakers — used here and in the rest of the paper serve mainly as convenient identifiers for different linguistic groups. The native Cantonese and English speaker groups acquired the respective language as their first language and keep using it as a primary/dominant language. Heritage speakers, on the other hand, tend to learn Cantonese as their first language but with English eventually becoming their primary/dominant language. See the description in section 1.3 for more detailed information.

of Cantonese lexical tones, the study will focus on understanding the influence of the phonological aspects.

1.1 Phonology and perceived similarity

The influence of phonological systems on speech perception has been documented in various branches of linguistics. Studies on first and second language acquisition have shown that one's perceptual system is modulated by native language experience (e.g., Best et al. 1988; Kuhl et al. 1992; Polka and Werker 1994). In addition, the phonological status of a pair of sounds being contrastive or allophonic also contributes to its perceived similarity: contrastive pairs of sounds are in general perceived to be less similar or less confusable in comparison with allophonic pairs of sounds (e.g., Boomershine et al. 2008; Johnson and Babel 2010; Kazanina et al. 2006). For instance, Boomershine et al. (2008) found that, using an explicit similarity rating task on a scale of 1–5 and an AX discrimination task, the perceived similarity among [d], [ð], and [ɾ] by native Spanish and English speakers is determined by the sounds' phonological statuses in the respective language. In particular, being allophonic in Spanish but contrastive in English, [d] and [ð] are perceived as more similar by Spanish speakers than by English speakers. On the other hand, [d] and [ɾ] are judged as more similar by English speakers than by Spanish speakers owing to their being allophonic in English but contrastive in Spanish. Comparable perceived similarity patterns are also observed for contrasts at the suprasegmental level. For example, Huang and Johnson (2010) reported that, again using a similarity rating task and a discrimination task, the perceived similarities among Mandarin lexical tones are rated differently by native Mandarin speakers and native English speakers. In particular, Mandarin speakers perceive the tones that are neutralized by phonological tone sandhi rules in Mandarin as more similar than do English speakers. In another study, Sun and Huang (2012) investigated tone perception by native Taiwanese speakers and by native English speakers, using Taiwanese tone continua. Their results indicated that tone language speakers seem to perceive tones as phonemic categories while non-tone language speakers tend to perceive them as continuous. In short, the studies on speech perception generally agree on the influence of phonology on speech perception, both at the segmental and suprasegmental levels.

1.2 Functional load and perceived similarity

Statistical properties of phonological contrast have also been shown to play a role in perceived similarity. One such statistical property is commonly referred to as *functional load*. Informally, the functional load of a mechanism in a language can be thought of as the degree of reliance of the language on the mechanism to convey information (Hockett 1955; Surendran and Niyogi 2003, 2006). This measure is widely used to weigh the importance of binary oppositions, such as contrastive sounds. A pair of contrastive sounds that distinguishes many words is described as carrying high functional load, whereas one that only distinguishes few words has low functional load (see section 2.4.2 for more detailed descriptions on the quantification of functional load). Despite the fact that empirical studies are still limited, a few studies do provide suggestive or positive evidence regarding the role of functional load in perceived similarity. For example, Stevenson (2015) tested the combined effects of functional load and frequency on the perceived similarity among three vowel pairs in Laurentian French, and she found that the pair of sounds with the lowest functional load and the lowest frequency is rated as being most similar, compared with pairs with mid and highest values in the functional load and frequency measures. In another study, Hall and Hume (2015) directly

tested the impact of functional load on perceived similarity using perceptual similarity data based on vowel pairs in standard continental French collected from native French speakers. Their study found functional load to be a significant predictor of perceived similarity such that vowel pairs with lower functional load are correlated with a higher degree of perceived similarity. Using an ABX discrimination paradigm, Martin and Peperkamp (2017) also found that French words with a place mispronunciation (e.g., replacing /v/ with /ʒ/) are more difficult to recognize than the words with a voice mispronunciation (e.g., replacing /v/ with /f/). They attributed this observation to the fact that the place feature has a higher functional load than the voice feature, so that French speakers give more importance to place than to voicing cues during word recognition, making it harder to them to recognize a word with a mispronounced place feature than one with a mispronounced voice feature. Evidence from studies on phonological mergers also argues for the link between functional load and perceived similarity: if a pair of sounds that has a high functional load is perceived as less similar, then this pair of sounds should be less prone to merger on the perceptual basis. This argument is consistent with the finding that sounds pairs with a higher functional load tend to be more resistant to phonological merger (Bouchard-Côté et al. 2013; Tsui 2012; Wedel et al. 2013a,b). For instance, Wedel et al. (2013b) found that cross-linguistically vowels are less likely to merge when they have a higher functional load. Tsui (2012) also showed that two ongoing tonal mergers in Hong Kong Cantonese are correlated with the combined effects between high acoustic similarity and low functional load.

1.3 Current study

The concept of functional load can be generalized to contrast at the suprasegmental level, such as lexical tones. For instance, in Cantonese each syllable is associated with one of its lexical tones. The language has six lexical tones and three allotones (i.e., "checked/entering tones" in traditional Chinese phonology): Tone1 (high-level [55]), Tone2 (high-rising [25]), Tone3 (mid-level [33]), Tone4 (low-falling [21]), Tone5 (low-rising [23]), Tone6 (low-level [22]), Tone7 (high-stopped [5]), Tone8 (mid-stopped [3]), and Tone9 (low-stopped [2]). The six lexical tones, Tone1 to Tone6, appear in open syllables or syllables ending with [m, n, ŋ], while the three allotones, Tone7 to Tone9, only appear in syllables ending with unreleased stops [p̚, t̚, k̚]. Table 1 below summarizes the description above with examples, and Figure 1 shows the fundamental frequency (F0) traces of the six lexical tones.

Table 1: Cantonese tone system. The classification here follows traditional Chinese dialectology, according to which the tones were derived from the four Middle Chinese tone classes, with each class splitting into two registers, upper and lower, depending on the voicing of the onset. Cantonese further split the upper checked tone class into two tones on the basis of vowel length.

Register	**Lexical tones**			**Checked tones**
Upper	Tone1 high-level [ji55] 'clothes'	Tone2 high-rising [ji25] 'chair'	Tone3 mid-level [ji33] 'idea'	Tone7 high-stopped [jɪk̚5] 'benefit' Tone8 mid-stopped [jak̚3] 'eat'
Lower	Tone4 low-falling [ji21] 'suspicious'	Tone5 low-rising [ji23] 'ear'	Tone6 low-level [ji22] 'two'	Tone9 low-stopped [jɪk̚2] 'wing'

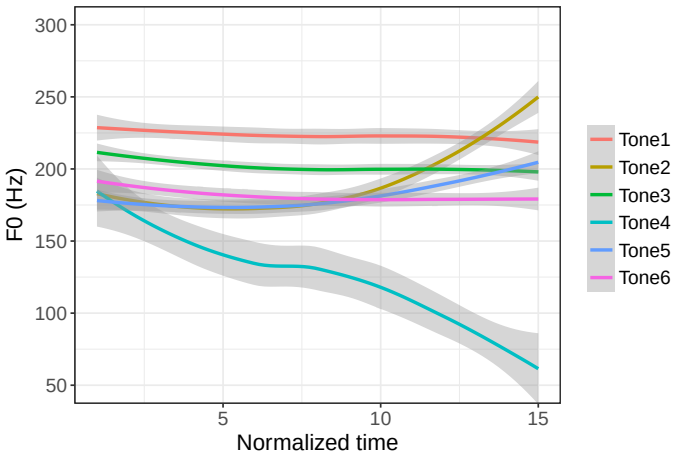

Figure 1: Illustration of F0 contours of the six lexical tones in Cantonese. The F0 data were from the stimuli used in the following experiment and were extracted using the autocorrelation algorithm implemented in Praat (Boersma and Weenink 2018) from the central 70% of monosyllables produced in isolation by a female native speaker of Hong Kong Cantonese.

The functional load of a tone pair can be estimated, for example, by quantifying the number of distinct lexical items that hinge upon only the tone pair in question. This generalization of functional load to lexical tones immediately raises the question of whether functional load of suprasegmental contrast also has impact on the perceived similarity of the contrast at the suprasegmental level. Based on Hall and Hume's (2015) finding that functional load is correlated with perceived similarity of contrastive sounds at the segmental level, it is reasonable to hypothesize that functional load should also influence perceived similarity of lexical tones, given that both lexical tones and phonemes serve essentially the same function of distinguishing lexical items. So far there are still no systematic studies tapping into this relationship between functional load and perceived similarity vis-à-vis lexical tones. By drawing on the rich tonal system of Cantonese, the current study aims to shed some light on this issue by explicitly testing the potential impact of functional load of Cantonese lexical tones on speakers' perception of the tones. In a perceptual experiment, perceptual similarity data of tone pairs were collected from native Cantonese speakers, native English speakers, and heritage speakers of Cantonese. The results were then modeled using linguistic groups and measures for acoustic distance and functional load as predictors.

The reasons for the inclusion of heritage speakers of Cantonese as participants are pragmatic as well as linguistic. The children born in the 1990s to first-generation Cantonese-speaking immigrant families from Hong Kong to Vancouver are now attending universities, forming a timely and available pool of potential participants for linguistic experiments. Linguistically, heritage speakers of Cantonese represent a unique bilingual population. Here I adopt Benmamoun et al.'s (2013) definition of heritage speakers as early bilinguals who grew up hearing and speaking the heritage language, alongside the majority language in early childhood, but for whom the majority language eventually became the primary language around the onset of schooling. Consequently, by early childhood, heritage speakers can already be strongly dominant in the majority language, with the heritage language being only secondary. Heritage speakers both resemble and differ from monolingual native speakers. The two groups are similar in terms of their early exposure to the target language (i.e., within the critical period). However, the two groups diverge in the quantity and va-

riety of linguistic input received and the social domain of language use. Reflecting this difference, a common finding emerging from the literature (see Oh et al. 2010 for Korean, Saadah 2011 for Arabic, Chang et al. 2011 for Mandarin, Levine 2015 for Yiddish) is that heritage speakers do not pattern with monolingual native speakers in various modules of grammar. For this reason, heritage speakers are assumed to be situated in the middle of a continuum that represents native competence, with first language (L1) native speakers and late second language (L2) speakers capping the two ends. With their rather distinct linguistic background and behavior, heritage speakers would be an ideal source of information for some questions concerning multilingual competence.

Given that functional load can be used to characterize phonological contrastiveness among lexical tones, and that previous studies have demonstrated the correlation between tonal phonology and perceived similarity, it is predicted that the functional load for lexical tones should also be correlated with tone perceived similarity. Specifically, after acoustic distance is taken into account, functional load is expected to be a significant predictor for the tone perceptual patterns in native Cantonese speakers. The prediction for heritage speakers is less straightforward as heritage speakers might have a Cantonese lexicon that is rather different from that of native Cantonese native speakers due to the more restricted domains of language use for heritage speakers. The same functional load estimate from the lexicon of native Cantonese speakers may not be representative of the lexicon of heritage speakers. At this moment, a sufficiently-large corpus of heritage Cantonese is still absent, so a more accurate functional load estimate for heritage speakers is infeasible. A plausible scenario, however, is that the same functional load estimate might still be predictive for heritage speakers, but to a lesser extent compared with native Cantonese speakers. As for native English speakers that do not have any knowledge about Cantonese, functional load is not supposed to contribute to the tone perceived similarity for the speakers. To foreshadow the results, the predictions for native and heritage Cantonese speakers are indeed borne out, yet unexpectedly functional load also appears to be predicative of tone perceived similarity for native English speakers.

The organization of the remainder of the paper is as follows. In section 2, we provide details on the perceptual experiment designed to collect the similarity data and the definitions and quantification of the predictors used to model the data. In section 3, we discuss the findings, and finally section 4 concludes the paper.

2 Experiment

2.1 Stimuli

Following the line of similar research in this area (e.g., Boomershine et al. 2008; Hall and Hume 2015; Johnson and Babel 2010; Martin and Peperkamp 2017), the stimuli used in the perception experiment are all non-words. Non-words also have the additional advantage over real words in avoiding the introduction of some confounding factors, such as word frequency associated with real words. In Cantonese, real words formed from combining the same syllable with different lexical tones usually have different frequencies. By using non-words as stimuli, we therefore eliminate the possibility that the observed perceived similarity is induced by word frequency. The non-words used consist of six sets of monosyllables — *[faʊ], *[jɔn], *[kɛŋ], *[mu], *[tɛ], *[tsɛm] — that follow Cantonese phonotactic rules (Kirby and Yu 2007), and each set is composed of the same monosyllable produced with the six lexical tones in Cantonese. In total, there are hence 6 monosyllables $\times$ 6 lexical tones $=$ 36 basic stimulus types.

The stimuli for the perception experiment were recorded by a 34-year-old female native speaker of Hong Kong Cantonese. The speaker is from Hong Kong, is fluent in English, and also has linguistic training. She was asked to produce the stimuli as clearly and naturally as possible. The recording was carried out in a sound-attenuated booth at the University of British Columbia, using a head-mounted microphone. Stimuli were recorded using Audacity (https://www.audacityteam.org/), and the sampling frequency was 44100 Hz.

All the stimuli were listed out on a piece of paper, with one stimulus per line, and appeared as non-words in isolation. The stimuli were presented in romanization (in the same form as appeared in Appendix A) since they are all non-words which have no corresponding Chinese characters. Stimuli were blocked by syllable, and, for each of the six monosyllables, the tokens appeared in the order from Tone1 to Tone6. The speaker were asked to produce each stimulus three times at her own pace, and three recording sessions were carried out, for a total of 36 basic stimulus types $\times$ 3 repetitions $\times$ 3 sessions $=$ 324 recordings. Of the nine repetitions for each stimulus, two versions of each were chosen manually, taking into account clarity of production and similarity of duration across tokens in the same monosyllable set. The duration of tokens across different monosyllable sets was not controled, as different types of syllables have different intrinsic durations. In addition, control over duration across different monosyllable sets was not necessary for the current AX experimental paradigm, as participants would only hear two stimuli from the same monosyllable set in each trial.

To avoid ceiling performance for native Cantonese speakers, each stimulus was further masked by speech-shaped noise, which was generated based on the long-term average spectra (LTAS) of all the chosen stimuli (but see section 4 for discussion on the issue of using noise-masking). The signal-to-noise ratio (SNR) was set to -5 dB, and the final intensity of masked stimuli was matched to the intensity of original stimuli. The degradation of acoustic signal should increase the difficulty of the discrimination task.

2.2 Participants

One group of native Cantonese speakers, one group of native English speakers, and one group of heritage Cantonese speakers participated in the experiment. The native Cantonese speakers ($N =$ 15, 4 males, 11 females) were students or friends of students at the University of British Columbia, and were from a number of Cantonese-speaking regions in China, such as Hong Kong, Guangdong, and Macao. All of them listed Cantonese as their first language, and, except for two speakers, the rest of the speakers relocated to Vancouver after age 18 (the two speakers moved to Vancouver at 14 and 16 respectively). There was no restriction with regard to the maximum number of years they could have spent in Vancouver to be included in this group. They either volunteered or received partial course credit for their participation in the experiment. The native English speakers ($N = 23$, 7 males, 16 females) were undergraduate students at the University of British Columbia enrolled in linguistics courses who received partial course credit for their participation in the experiment. They were screened in a post-test questionnaire, and only participants that had no Cantonese or other tone/pitch-accent languages speaking experience were included in the experiment although some did have knowledge of another foreign language (e.g., French and Spanish). The heritage Cantonese speakers ($N = 23$, 3 males, 20 females) were likewise undergraduates at the University of British Columbia and received partial course credit for their participation in the experiment. The inclusion criterion for this group is their reporting that Cantonese is their first language, but that they were either born and raised in Vancouver or moved to the city before age 5. Therefore, they

all received their education in English. Typically, they reported their Cantonese comprehending and speaking ability to be fair or good but their reading and writing ability to be poor, as summarized in Table 2. None of the speakers from the three groups reported any history of speech or hearing disorders or disabilities.

Table 2: Mean self-ratings for Cantonese and English proficiency across the three participant groups.

Group	Cantonese				English			
	Understanding	Speaking	Reading	Writing	Understanding	Speaking	Reading	Writing
Cantonese	3.00*	3.00	2.92	2.85	2.46	2.23	2.38	2.38
Heritage	2.52	2.17	1.22	1.22	3.00	3.00	3.00	3.00
English	0.00	0.00	0.00	0.00	3.00	3.00	3.00	3.00

* The ratings range from 1 to 4, with 0 = not at all, 1 = poorly, 2 = fairly well, and 3 = fluently. The ratings for English from the group of native Cantonese speakers were averaged over 14 participants, instead of 15, as one participant did not provide the ratings.

2.3 Procedure

In this discrimination task, the participants were told that they would hear a pair of syllables and have to judge whether the sounds were the same or different. The stimuli presented in each pair were always physically different tokens, even when they were both instances of a single sound (e.g., [mu55] … [mu55]). The purpose of the phonetic variation in stimuli was to encourage participants to abstract away from phonetic detail and to induce "phonological" listening, at least for native and heritage Cantonese speakers.

The participants were each seated at a computer that was connected to a 5-button response box, with up to four participants taking part in the study at a time. The participants listened to the stimuli through headphones and were asked to indicate whether each pair of sounds they heard was the same or different by pressing the 'same' or 'different' button (i.e., the button corresponding to 'same' was labeled as such on the response box, and so was the button for 'different') on the response box. In one condition, button 1 served as the 'same' button while button 5 served as the 'different' button; in the other condition, the assignment of buttons was reversed. Each participant was arbitrarily assigned to one of the two conditions. The testing phase of the experiment contained 360 trials, with 180 'same' trials (6 monosyllables × 6 'same' tone pairs × 5 repetitions [3 AX ordering and 2 XA orderings]) and 180 'different' trials (6 monosyllables × 15 'different' tone pairs × 2 orders of A and X in a single trial). The whole set of trials was presented in a different random order for each participant, using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Participants were given a chance to take a short break after every 120 trials. Within each trial, the two stimuli were separated by 500 ms of speech-shaped noise (i.e., the same one used to mask the stimuli), such as [mu55] <500 ms noise> [mu25]. The syllable segments were the same for the two stimuli within a trial so that the only difference within each trial was the lexical tone. For each trial, participants were given a 2000-ms window after the onset of the second stimulus to register their responses, but they could respond throughout the whole trial (so they could already respond before the end of the second stimulus). Feedback as to the accuracy of their response, their average percent correct overall, and their reaction time (RT) in seconds was subsequently shown on the screen. This feedback was used

to promote both heightened accuracy and shorter RTs. The next trial started immediately after the feedback.

Prior to the testing phase, participants were given a brief task-training session consisting of four practice trials. The task was the same as in the testing phase, but the practice trials were constructed from the syllable [nui], which was not included in the real experiment stimuli. The accuracy and RT for each testing trial were recored and used for the subsequent analysis.

2.4 Quantifying predictors

Two types of potential predictors are used to model the perceptual similarity. Section 2.4.1 describes the method used to quantify the acoustic distance between the stimulus syllables used in the perception experiment. In section 2.4.2, methods and results of quantifying functional load are provided. Individual values for both predictors are given in Appendix A.

2.4.1 Acoustic distance

Acoustic distance aims to capture the degree of difference between each tone pair, such that a pair of tones that are acoustically similar should have a small distance between them. Following Tsui (2012), the current study uses the root-mean-square distance between the F0 trajectories of two tones as the acoustic distance between them. To calculate the acoustic distance between two tones, F0 was estimated using Praat's (Boersma and Weenink 2018) autocorrelation periodicity detection algorithm[2] and manually-checked. Fifteen equidistant points between 15% to 85% of the sonorous portion in the target syllable for each tone were then extracted after pitch smoothing with a bandwidth of 20 Hz. The initial and final 15% of the syllable were excluded because F0 in these two regions is subject to the influence from onset and/or coda consonants and because F0 measurement in these two regions tends to be less accurate. Then the root-mean-square value was computed on the basis of the 15 F0 differences between the corresponding equidistant points for the two tones in question. The resulting acoustic distance values following this procedure are summarized in Figure 2 below. It is important to note that although using root-mean-square as an approximation to acoustic distance does capture overall F0 difference between the two tones in question, it does not take into consideration acoustic properties such as duration, voice quality, direction of contours, and differences between direction of contours. Because this measure uses equidistant points of each tone as the starting points for calculating root-mean-square, durations across all the tones are normalized, and therefore the durational cue for each tone is completely ignored in this measure. This measure also does not take into consideration voice quality like creakiness, which often serves as an important cue for Cantonese Tone4. Further, differences between direction of contours, which could function as a useful cue for identifying tone, are neutralized in this measure in the sense that two pairs of tones can end up having similar acoustic distance measures even if one pair consists of two level tones and the other two rising tones. For instance, the acoustic distance of the tone pair of Tone2 versus Tone5 tends to be numerically similar to the tone pair of Tone1 versus Tone3, even though the two tones in

[2]The setting for each parameter of the algorithm is as follows: time step = 0.001 s, pitch floor = 30.0 Hz for Tone4 tokens and 80 Hz for non-Tone4 tokens, maximum number of candidates = 15, silence threshold = 0.03, voicing threshold = 0.45, octave cost = 0.01, octave-jump cost = 0.35, voiced/unvoiced cost = 0.14, pitch ceiling = 300 Hz. The value for the time step was set to be smaller than the default so that the algorithm returned more pitch values. The values for the pitch floor and pitch ceiling were adjusted based on tonal identity and speaker voice F0. The other values were all algorithm defaults.

the former pair are both rising and the two in the latter level. This root-mean-square-based metric for acoustic distance was still chosen because of its computational simplicity and the advantage of having the same physical unit as the signal itself (i.e., Hertz in this case), which facilitates the interpretation of the measure.

2.4.2 Functional load

As mentioned in section 1, functional load measures phonological contrastiveness in terms of the "work" done by individual pairs of sounds at the lexical level, i.e., how many lexical items a pair of sounds distinguishes. A number of ways to quantify functional load have been proposed in the literature (e.g., Hockett 1955; Surendran and Niyogi 2006; Wedel et al. 2013b). In this study, the measure based on the concept of entropy is used, which estimates the amount of uncertainty associated with predicting the outcome of some variable in a system (Shannon 1948).

Following Hockett (1955), the functional load of a binary opposition can be expressed as the normalized change in entropy after the complete neutralization of the opposition. Surendran and Niyogi (2006) show this definition can be generalized to any phonological contrast and computed mathematically at various levels, using data from a corpus. The general formula for functional load of T unit in language L as estimated from corpus S is given in (1). The function f represents the loss of the contrast we are interested in finding the functional load of, the non-negative integer k is the order of Markov process[3], and $f(L)$ and $f(S)$ stand for the language L and corpus S after the loss of the contrast. The term $H_{TkS}(L)$ in (1) is the approximated entropy of language L and can be calculated by using (2), where $p(t)$ is the probability of unit t of type T in some inventory (e.g., every word in a language's lexicon or every syllable in a language's syllable inventory). Similarly, the term $H_{Ukf(S)}(f(L))$ is the entropy of the new language $f(L)$ where the pair of contrastive sounds in question is neutralized.

$$FL_{TkS}(f) = \frac{H_{TkS}(L) - H_{Ukf(S)}(f(L))}{H_{TkS}(L)} \tag{1}$$

$$H_{TkS}(L) = \frac{1}{k+1}\left(-\sum_{t \in T} p(t) \log_2 p(t)\right) \tag{2}$$

Functional load based on a change in entropy was calculated from the Hong Kong Cantonese Corpus (Luke and Wong 2015), a collection of about 230,000 Cantonese words from spontaneous speech and radio programs, which should represent colloquial Cantonese better than text-based corpora. The phonological contrast in question is lexical tone (i.e., T = lexical tone in the formula above), the calculation is based on the word level over type frequency and over token frequency, and the order of Markov process is zero (i.e., $k = 0$). The decision to calculate functional load over type frequency, in additional to just over token frequency as in Surendran and Niyogi (2006), was motivated by the finding from Hall and Hume (2015) that the type-based measures are more predictive of perceived similarity for French vowel than the token-based measures. The order k was set to zero (i.e., not using any context at all) because the occurrence of a lexical tone was assumed not to depend on the previous lexical tone. Appendix A tabulates all pairs of lexical tones and their functional loads.

[3] A k-order Markov process means that the probability distribution on any phoneme depends on the k phonemes occurring before it.

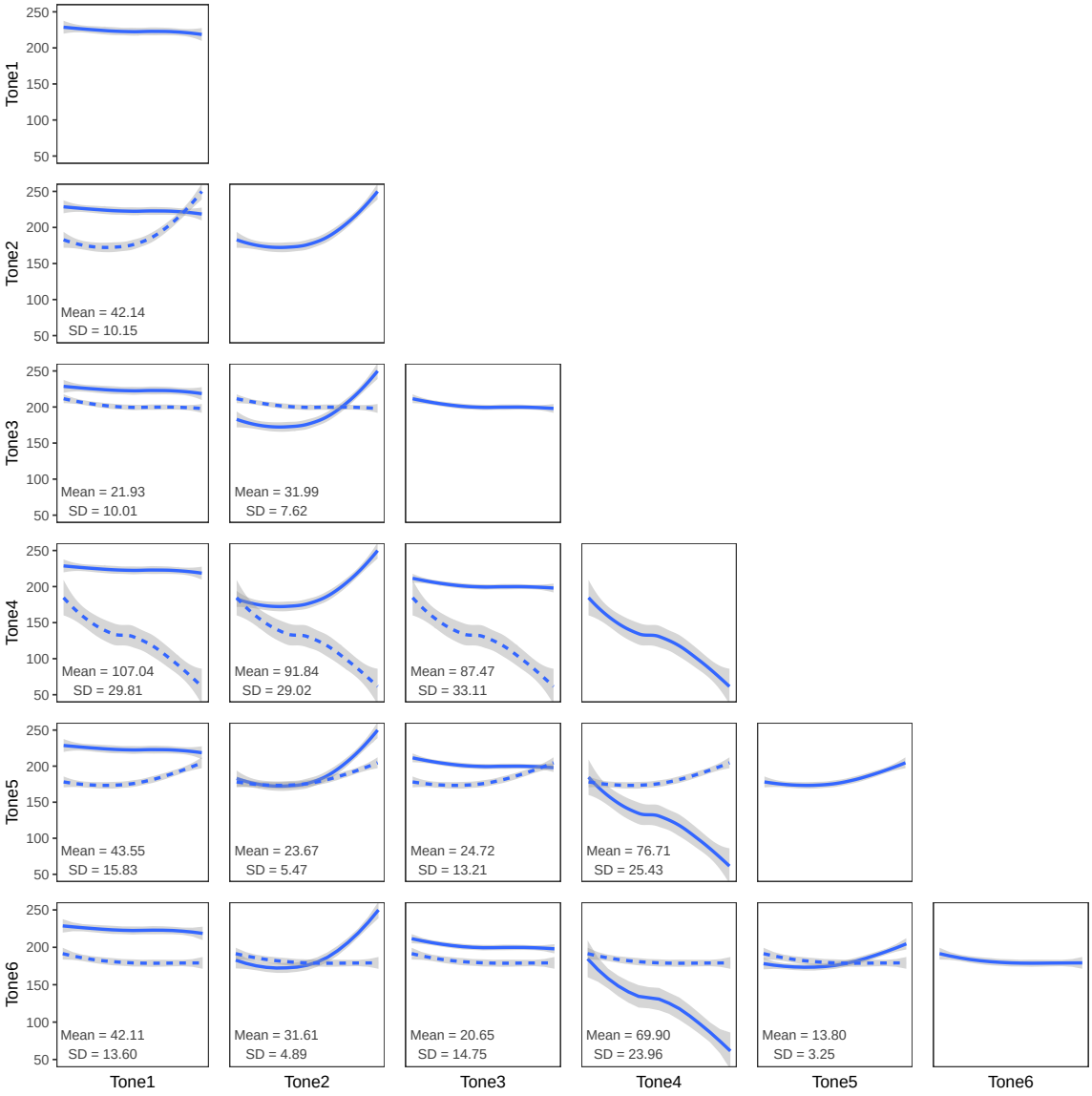

Figure 2: The means and standard deviations of the acoustic distance for each tone pair, whose tonal contours are also depicted. Both estimates were calculated over the six acoustic distance values within each pair. For instance, the mean (42.14 Hz) and standard deviation (10.15 Hz) for the Tone1-Tone2 pair were computed based on the acoustic distance values for *chem1* / *chem2* (23.296 Hz), *fau1* / *fau2* (52.135 Hz), *keng1* / *keng2* (41.139 Hz), *mu1* / *mu2* (41.388 Hz), *te1* / *te2* (46.431 Hz), and *yon1* / *yon2* (48.479 Hz), as documented in Appendix A.

2.5 Results

The summary statistics for accuracy and reaction time, as measured from the offset of the second stimulus, are provided in Table 3. A set of mixed effects models was fit in order to determine which of the independent variables had a significant effect on predicting the Cantonese lexical tone perceptual similarity data, measured in terms of accuracy and RT. The full statistical models of the analyses can be found in Appendix B. The rationale behind using accuracy and RT as an estimation for perceived similarity is that speakers should be able to distinguish two sounds more accurately and faster, if the two sounds are perceived as more distinct than if they are perceived as more similar. These two measures capture these two characteristics respectively and therefore should also reflect the perceived similarity of the two sounds in question.

In the analyses, all the models were fit either with the functional load over type frequency or the functional load over token frequency. The models with both types of measures give similar results because the two measures are highly correlated. In this section, only the results from the models with type-based frequency are reported; the interested reader is referred to Appendix B for the detail of the models fit with token-based frequency.

This section is organized as follows: section 2.5.1 presents the results of the model fit with accuracy of the "different" pairs as the dependent variable, and section 2.5.2 summarizes the statistics from the model using RT for the correctly-identified "different" pairs as the dependent variable.

Table 3: Summary statistics for accuracy and RT across the three linguistic groups.

Group	**Accuracy**		**Reaction time (ms)**		
	Mean	SD	Mean	Median	SD
Cantonese	0.81	0.39	356.32	309.51	279.50
Heritage	0.79	0.41	287.86	239.46	263.24
English	0.73	0.45	315.07	273.76	277.25

2.5.1 Accuracy

The mean results and 95% confidence intervals for the "different" pairs are depicted in Figure 3. Accuracy for these "different" pairs is taken to be a measure of perceived similarity among lexical tones, where lower accuracy indicates increased perceived similarity and higher accuracy decreased perceived similarity. The accuracy data from trials without any response or with responses before the onset of the second stimulus were removed, but otherwise all the data, including outliers, were included in the following analysis. Altogether the results from 118 trials were removed, constituting about 1% of the original data.

In the logistic mixed effects model, the fixed effects included Linguistic Group (categorical), Acoustic Distance (continuous numerical value, centered and scaled), and Functional Load (continuous numerical value, centered and scaled), as well as their two-way and three-way interactions. The Linguistic Group effect has three levels (i.e., Cantonese, Heritage, and English) and was coded using treatment (dummy) coding, with the reference level being Cantonese. The random effects structure was as maximally specified as possible, with Participant and Syllable as random effects. The structure also includes a by-Participant random slope for Acoustic Distance, Functional Load, their inter-

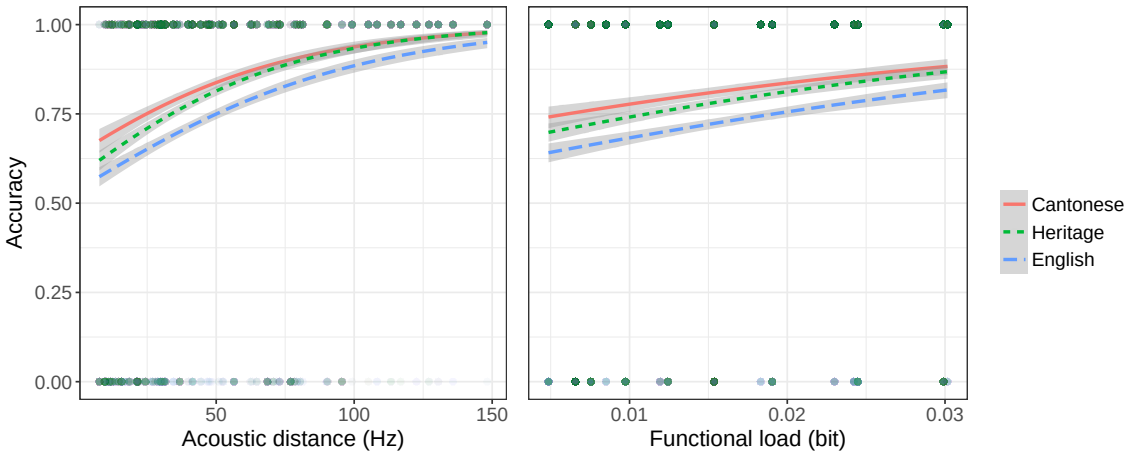

Figure 3: Accuracy of responses as a function of acoustic distance and functional load across the three linguistic groups.

action, and a by-Syllable random slope for Linguistic Group, Acoustic Distance, Functional Load, and their interactions. A collinearity test confirmed that there was no collinearity among the three fixed effects (i.e., variance inflation factor (VIF) are all smaller than 2), so they are all kept in the model.

There was a significant effect for the intercept ($\beta = 2.19$, $SE = 0.44$, $p < 0.01$), indicating that native Cantonese speakers could correctly discriminate the two stimuli in a trial above the chance level. A significant fixed effect was found for Linguistic Group of English versus Cantonese ($\beta = -0.92$, $SE = 0.32$, $p < 0.01$), as native English speakers were less accurate in discriminating Cantonese lexical tones. The fixed effect for Linguistic Group of Heritage versus Cantonese was not significant ($\beta = -0.33$, $SE = 0.22$, $p = 0.13$), indicating that heritage speakers' ability to discriminate different lexical tones was comparable to that of native speakers. As expected, there was a significant main effect of Acoustic Distance ($\beta = 1.34$, $SE = 0.44$, $p < 0.01$), such that tone pairs with relatively larger acoustic distance were perceived as less similar and therefore discriminated more accurately than pairs with smaller acoustic distance. Crucially, Functional Load emerged as a significant fixed effect ($\beta = 0.84$, $SE = 0.22$, $p < 0.01$), so native Cantonese speakers perceived tone pairs with high functional load as more distinct and thus could discriminate them more accurately than tone pairs with low functional load. There was also a marginal interaction between Functional Load and Linguistic Group of English versus Cantonese ($\beta = -0.45$, $SE = 0.26$, $p = 0.08$), suggesting that functional load increased accuracy for native Cantonese speakers but only played a smaller role for native English speakers. In addition, there were a significant two-way interaction Functional Load $\times$ Acoustic Distance ($\beta = 0.81$, $SE = 0.33$, $p < 0.05$) and a three-way interaction Functional Load $\times$ Acoustic Distance $\times$ Linguistic Group of English versus Cantonese ($\beta = -0.71$, $SE = 0.35$, $p < 0.05$), implying that the influence of functional load might be conditioned on acoustic salience, and that the dependence of the two factors was not uniform across groups of different linguistic backgrounds. The positive two-way interaction indicates that the effect of functional load becomes larger when the acoustic distance between two tones increases. The three-way interaction is illustrated in Figure 4, where the acoustic distance measures are shown in High and Low categories using median splitting. As can be seen in Figure 4, this three-way inter-

action appears to arise from functional load having a stronger effect (i.e., the slope is steeper) in the low acoustic distance condition for native English speakers, but not for native or heritage speakers of Cantonese.

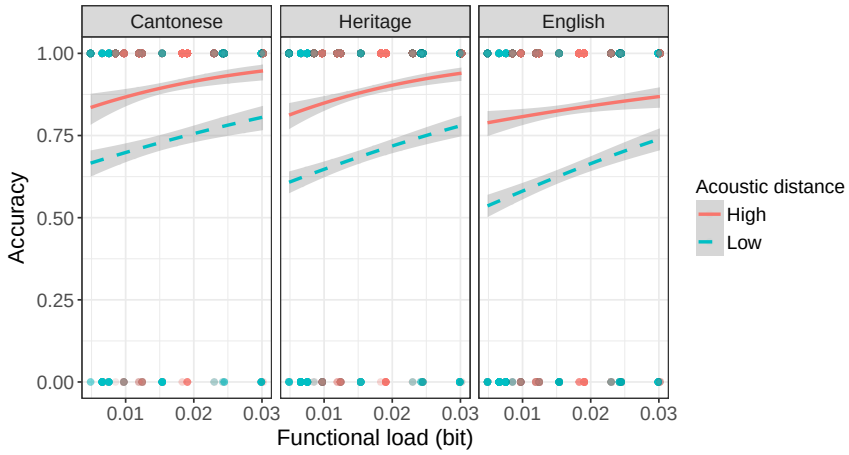

Figure 4: Three-way interaction between Linguistic Group, Functional Load, and Acoustic Distance. The acoustic distance measures were partitioned into High and Low categories using median splitting in order to better visualize the interaction. As can be seen, the effect of functional load within High and Low categories is different across the three linguistic groups.

2.5.2 Reaction time

The results for RT are shown in Figure 5, with the means plotted along with their 95% confidence intervals. Here perceived similarity among lexical tones is measured as a function of RT accumulated from the offset of the second stimulus from only the correctly-identified 'different' trials, with longer RTs indicating increased perceived similarity and shorter RTs decreased perceived similarity. Similar to the accuracy data, RT data from the trials without responses or with responses registered before the start of the second stimulus were excluded from the analysis. This procedure removed 27 trials or 0.32% of the data.

As with the model for accuracy data, the fixed effects in the mixed effects model were Linguistic Group (categorical: Cantonese, Heritage, English), Acoustic Distance (continuous numerical value, centered and scaled), and Functional Load (continuous numerical value, centered and scaled), as well as their interaction terms. The random effects structure had Participant and Syllable as random intercepts as before, a by-Participant random slope for a linear combination of Acoustic Distance and Functional Load, and a by-Syllable random slope for a linear combination of Linguistic Group, Acoustic Distance, and Functional Load. The interaction terms were excluded in the random slopes because including them led to a lack of model convergence. There was also no collinearity among the three fixed effects.

There was a significant effect for the intercept ($\beta = 343.29$, $SE = 32.75$, $p < 0.01$), indicating that hypothetically it would take 343.29 ms for native Cantonese speakers to respond when the effects of acoustic distance and functional load were left out. The only significant main effect revealed by the model is Functional Load ($\beta = -17.63$, $SE = 7.08$, $p < 0.05$), and the effect of Acoustic Distance is marginally significant ($\beta = -26.55$, $SE = 12.24$, $p = 0.07$). As expected, both main

effects indicate that RT decreases as functional load and acoustic distance increase.

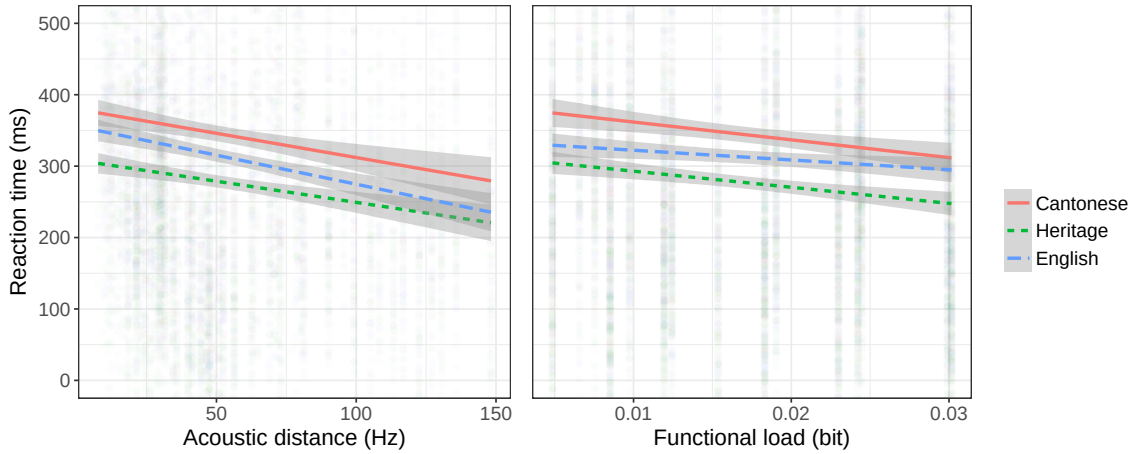

Figure 5: Reaction time of correct responses as a function of acoustic distance and functional load across the three linguistic groups.

3 Discussion

In summary, the results point to the effects of acoustic distance and functional load on the perceived similarity of Cantonese lexical tones, measured in terms of accuracy and RT. Not surprisingly, acoustic distance emerges as an important predictor for all three linguistic groups, with a similar effect size for each group. Crucially and rather unexpectedly, functional load also plays a significant role for *all* three groups of speakers. The effect size of functional load is comparable for native and heritage Cantonese speakers ($\beta = 0.85$, $SE = 0.21$, $p = 0.42$), but is marginally smaller for native English speakers ($\beta = 0.39$, $SE = 0.26$, $p = 0.08$).

While functional load of Cantonese lexical tone pairs indeed speaks to its role in the perceived similarity in native and heritage Cantonese speakers, one question that arises from the results is how functional load should also appear as a significant predictor for native English speakers' perceived similarity. After all, the implementation of functional load in this project is directly calculated over the Cantonese lexicon, which native English speakers lack. I argue that the causes for this apparent effect of functional load for English speakers might be attributed to the nature of the acoustic distance measure used in the current project. Specifically, it is possible that the measure, defined as the root-mean-square of 15 equidistant F0 values over the central 70% portion of two tones, does not reflect the true psychoacoustic distance as perceived by the human auditory system. Therefore, part of perceived similarity that should have been accounted for by a proper measure of acoustic distance might still be present in the dependent measure, and this was in turn (erroneously) captured by functional load. It is reasonable to assume that, even if functional load does affect perceived similarity, it should not have any effect for native English speakers who do not have any knowledge about Cantonese. The fact that functional load emerges as a significant predictor even for native English speakers suggests that the functional load measure might capture some acoustic characteristics.[4] In other words, the result that functional load is a significant predictor for native English speakers

[4]One possible objection to the statement that functional load measure might also capture certain acoustic as-

might be due to some acoustic aspects inherent in functional load. The results may become more clear with a better acoustic distance measure. The ways to improve the acoustic distance measure for subsequent studies include calculating the acoustic distance based on a different psychoacoustic scale, such as Bark, mel, semitone, or equivalent rectangular bandwidth (ERB), which should approximate perception of F0 better than the Hertz scale. Another way to fine-tune the measure is to include the initial and final 15% of pitch tracking into the calculation. The reason why the initial and final 15% of the syllable was left out is to bypass the inaccuracies in F0 measurement in these two regions. However, there is no reason to believe that speakers would not use phonetic information in these two regions to distinguish between different tones; excluding F0 information in these two regions might therefore leave important phonetic cues out of the acoustic characterization of Cantonese lexical tones. One may legitimately argue that, after the measure of acoustic distance is refined, even the influence of functional load seen in native and heritage Cantonese speakers might turn out to be insignificant. While this is certainly possible, I argue that the stronger effect of functional load observed for native and heritage Cantonese speakers on the one hand, and the reduced effect for native English speakers on the other hand, is suggestive of the active role functional load plays in shaping perceived similarity across different linguistic groups.

One factor that can potentially complicate the results is the asymmetric effect of noise-masking for Cantonese versus for non-Cantonese speakers. Given that the non-words used in the experiment resemble real Cantonese words and that the participants were not explicitly told that the stimuli were non-words, native and heritage speakers of Cantonese might attempt to make sense of the stimuli during the experiment. That is, adding noise to non-words could have created different tasks for the language groups because of a listener's natural tendency to look for signal in noise. This asymmetry might in turn result in difference in accuracy or RT that was not accounted for in the statistical models. It is impossible to evaluate the impact of this factor on the experiment results with the current experimental design. Therefore, I did not address this issue further but left it as a word of caution for future research using noise masking.

Using accuracy to characterize perceived similarity might also not be ideal for the purpose of the current study. While Hall and Hume (2015) and Martin and Peperkamp (2017) also made use of accuracy, the accuracy in Hall and Hume (2015) was from an identification task and was used to derive a vowel confusion matrix, and that in Martin and Peperkamp (2017) was from an ABX discrimination paradigm. Utilizing raw accuracy in the context of an AX discrimination paradigm is therefore less motivated on view of previous research. Given that raw accuracy might not be the most sensitive measure for perceived similarity, future research can benefit from experimenting with different measures for perceived similarity.

Another potential caveat is that the functional load estimated using the Hong Kong Cantonese Corpus (Luke and Wong 2015) might not necessarily reflect the phonological knowledge of heritage speakers of Cantonese tested in the experiment. This is because the lexicon of heritage speakers could be quite different from the one represented by the corpus due to the difference in the kind of linguistics input received and domains of use between heritage and native speakers. So far there are still no publicly-available corpora of Cantonese as a heritage language. This barrier is therefore hard

---

pects can be made based on the statistical test result that shows the non-collinearity between functional load and acoustic distance in the current study. However, it is possible that functional load and acoustic distance are correlated, but the collinearity is measure-dependent: it just so happens that the acoustic distance measure using root-mean-square does not correlated with functional load. The collinearity may show up with the "right" acoustic distance measure.

to overcome at the moment until a dedicated corpus comes into being.

The current finding that links functional load to perceived similarity is consistent with one approach that views phonological generalizations as emerging indirectly via abstractions over word-forms, instead of directly from the phonetics (Pierrehumbert 2003). Word-forms, which are abstractions over sequences of phonetic material, are central in this approach, as Pierrehumbert (2003) observes that phonological generalizations are often those that emerge from a dataset that has the size and quality of the lexicon, as opposed to word-specific generalizations (i.e., words having specific phonetic detail, which does not necessarily spread to other words), which tend to be those that can be made directly from running speech. Given the way it is generally defined and computed in the current study, functional load provides a measure of a contrast's service to distinguishing words in a language and is therefore word-based and over the whole lexicon. As such, functional load can be conceptualized as a way to quantify some phonological aspect of a language that is determined by taking the whole lexicon into consideration. The finding that functional load is predictive of perceived similarity is indirect evidence that it is the word-level that is influencing perceptual representations.

Finally, the lack of statistical difference for the results from native and heritage Cantonese speakers suggests that heritage Cantonese speakers pattern more closely with native Cantonese speakers than with native English speakers. This finding provides support for the claim that previous experience with a heritage language confers an advantage in the perception of that language, such that their perception performance is comparable to that of native speakers (e.g., Oh et al. 2003). However, it is worth pointing out that it is an area where there are huge individual differences, and that whether heritage language speakers show an advantage over non-native speakers just in perception or in both perception and production of the heritage language seems to depend on the nature of their heritage language experience. For instance, Oh et al. (2003) found that while heritage Korean speakers who spoke, as opposed to just hearing, Korean regularly during childhood are measurably more native-like than L2-learners in both perception and production of Korean, those who only had childhood experience with overhearing Korean but not with speaking or being spoken to were more native-like than L2-learners only in perception. Given that the experiment in the current project is a perceptual one, we should expect the results from our heritage Cantonese speakers to be more native-like, regardless of whether they are childhood hearers or childhood speakers, and this prediction is indeed borne out.

4 Conclusion

This work investigates the potential influence of acoustic distance and functional load on the perceived similarity of lexical tones in Cantonese by three groups of participants of different linguistic backgrounds. As expected, tone pairs with greater acoustic distance are perceived as being more distinct than pairs with the opposite characteristic for all three linguistic groups. Functional load also emerges as a significant predictor for tonal perceived similarity such that increased functional load is correlated with decreased perceived similarity. Rather strangely, however, the effect of functional load also holds for native English speakers with no knowledge about Cantonese. I argue that this "phantom" effect of functional load in native English speakers might be due to the fact that the measure for acoustic distance here does not thoroughly capture the actual psychoacoustic distance associated with tone pairs. The leaked acoustic distance effect is in turn erroneously accounted for in the model by the measure for functional load, causing a misleading significant effect.

The results are also consistent with a model of language where phonological generalizations are abstracted over word-level units, rather than over segmental units. In addition, the results lend support to the claim that early exposure to language grants an advantage in perception of a currently non-dominant language.

References

Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3-4):129–181.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3):345–360.

Boersma, P. and Weenink, D. (2018). Praat: Doing phonetics by computer. [Computer program].

Boomershine, A., Hall, K. C., Hume, E., and Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In Avery, P., Dresher, B. E., and Rice, K., editors, *Contrast in phonology: Theory, perception, acquisition*, pages 145–171. Mouton de Gruyter, Berlin.

Bouchard-Côté, A., Hall, D., Griffiths, T. L., and Klein, D. (2013). Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences of the United States of America*, 110(11):4224–4229.

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

Hall, K. C. and Hume, E. (2015). Modeling perceived similarity: The influence of phonetics, phonology and frequency on the perception of French vowels. submitted.

Hockett, C. F. (1955). *A manual of phonology*. Waverly Press, Inc., Baltimore, MD.

Huang, T. and Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, 67:243–267.

Johnson, K. and Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, 38(1):127–136.

Kazanina, N., Phillips, C., and Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103(30):11381–11386.

Kirby, J. R. and Yu, A. C. L. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. In Trouvain, J. and Barry, W. J., editors, *Proceedings of The 16th International Congress of Phonetic Sciences*, pages 1389–1392, Saarbrücken.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606–608.

Levine, G. S. (2015). *Incomplete L1 acquisition in the immigrant situation: Yiddish in the United States*. Max Niemeyer Verlag, Berlin.

Luke, K. K. and Wong, M. L. (2015). The Hong Kong Cantonese corpus: Design and uses. *Journal of Chinese Linguistics*, 25:312–333.

Martin, A. and Peperkamp, S. (2017). Assessing the distinctiveness of phonological features in word recognition: Prelexical and lexical influences. *Journal of Phonetics*, 62:1–11.

Oh, J. S., Au, T. K.-F., and Jun, S.-A. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language*, 37(5):1123–1132.

Oh, J. S., Jun, S.-A., Knightly, L. M., and Au, T. K.-F. (2003). Holding on to childhood language memory. *Cognition*, 86:B53–B64.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2–3):115–154.

Polka, L. and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2):421–435.

Saadah, E. (2011). *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Labs Technical Journal*, 27(3):379–423.

Stevenson, S. (2015). *The strength of segmental contrasts: A study on Laurentian French*. PhD thesis, University of Ottawa, Ottawa, ON.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In Bohn, O.-S. and Munro, M. J., editors, *Language experience in second language speech learning: In honor of James Emil Flege*, volume 17 of *Language learning & language teaching*, chapter 3, pages 35–55. John Benjamins Publishing, Amsterdam.

Sun, K.-C. and Huang, T. (2012). A cross-linguistic study of Taiwanese tone perception by Taiwanese and English listeners. *Journal of East Asian Linguistics*, 21:305–327.

Surendran, D. and Niyogi, P. (2003). Measuring the usefulness (functional load) of phonological contrasts. Technical report, Department of Computer Science, University of Chicago.

Surendran, D. and Niyogi, P. (2006). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In Thomsen, O. N., editor, *Competing models of linguistic change: Evolution and beyond*, pages 43–58. John Benjamins Publishing Company, Amsterdam.

Tsui, T.-H. (2012). Tonal variation in Hong Kong Cantonese: Acoustic distance & functional load. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 48, pages 579–588.

Wedel, A., Jackson, S., and Kaplan, A. (2013a). 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.

Wedel, A., Kaplan, A., and Jackson, S. (2013b). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186.

Appendix A: Acoustic distance and functional load for each tone pair

In the following chart, each tone pair is listed along with its acoustic distance and functional load values. Acoustic distance is the root-mean-square in Hertz of F0 differences between 15 equidistant points of the central 70% of the two syllables. Functional load is measured as the change in entropy of the corpus (in bits) upon the merger of the two tones, calculated over word type and token. Both acoustic distance and functional load measures are symmetrical, in that the order of the tones in the pair is irrelevant. Note also that the functional load values for each tone pair across the six syllables are the same as they are based on the merging of the corresponding tones in real words in the corpus, whereas the acoustic distance values are calculated based on the actual stimuli.

Tone pair	Acoustic Distance	Functional Load (type)	Functional Load (token)
chem1 / *chem2*	23.296	0.030	0.050
fau1 / *fau2*	52.135	0.030	0.050
keng1 / *keng2*	41.139	0.030	0.050
mu1 / *mu2*	41.388	0.030	0.050
te1 / *te2*	46.431	0.030	0.050
yon1 / *yon2*	48.479	0.030	0.050
chem1 / *chem3*	21.252	0.030	0.089
fau1 / *fau3*	9.762	0.030	0.089
keng1 / *keng3*	18.508	0.030	0.089
mu1 / *mu3*	31.593	0.030	0.089
te1 / *te3*	14.604	0.030	0.089
yon1 / *yon3*	35.836	0.030	0.089
chem1 / *chem4*	90.011	0.018	0.024
fau1 / *fau4*	122.559	0.018	0.024
keng1 / *keng4*	148.206	0.018	0.024
mu1 / *mu4*	127.069	0.018	0.024
te1 / *te4*	81.328	0.018	0.024
yon1 / *yon4*	73.064	0.018	0.024
chem1 / *chem5*	30.145	0.009	0.004
fau1 / *fau5*	56.360	0.009	0.004
keng1 / *keng5*	39.937	0.009	0.004
mu1 / *mu5*	62.433	0.009	0.004
te1 / *te5*	21.512	0.009	0.004
yon1 / *yon5*	50.935	0.009	0.004

Tone pair	Acoustic Distance	Functional Load (type)	Functional Load (token)
chem1 / *chem6*	30.004	0.023	0.043
fau1 / *fau6*	56.378	0.023	0.043
keng1 / *keng6*	44.570	0.023	0.043
mu1 / *mu6*	52.558	0.023	0.043
te1 / *te6*	21.428	0.023	0.043
yon1 / *yon6*	47.722	0.023	0.043
chem2 / *chem3*	27.032	0.024	0.072
fau2 / *fau3*	44.251	0.024	0.072
keng2 / *keng3*	34.498	0.024	0.072
mu2 / *mu3*	22.124	0.024	0.072
te2 / *te3*	34.426	0.024	0.072
yon2 / *yon3*	29.635	0.024	0.072
chem2 / *chem4*	95.559	0.019	0.012
fau2 / *fau4*	90.296	0.019	0.012
keng2 / *keng4*	135.799	0.019	0.012
mu2 / *mu4*	108.290	0.019	0.012
te2 / *te4*	64.671	0.019	0.012
yon2 / *yon4*	56.448	0.019	0.012
chem2 / *chem5*	21.593	0.007	0.008
fau2 / *fau5*	15.658	0.007	0.008
keng2 / *keng5*	24.336	0.007	0.008
mu2 / *mu5*	31.179	0.007	0.008
te2 / *te5*	27.948	0.007	0.008
yon2 / *yon5*	21.283	0.007	0.008
chem2 / *chem6*	29.964	0.024	0.051
fau2 / *fau6*	30.546	0.024	0.051
keng2 / *keng6*	41.495	0.024	0.051
mu2 / *mu6*	29.848	0.024	0.051
te2 / *te6*	28.957	0.024	0.051
yon2 / *yon6*	28.848	0.024	0.051
chem3 / *chem4*	72.910	0.012	0.031
fau3 / *fau4*	113.334	0.012	0.031
keng3 / *keng4*	130.490	0.012	0.031
mu3 / *mu4*	99.226	0.012	0.031
te3 / *te4*	69.056	0.012	0.031
yon3 / *yon4*	39.793	0.012	0.031
chem3 / *chem5*	15.684	0.005	0.008
fau3 / *fau5*	47.435	0.005	0.008
keng3 / *keng5*	24.589	0.005	0.008
mu3 / *mu5*	31.415	0.005	0.008
te3 / *te5*	11.104	0.005	0.008
yon3 / *yon5*	18.076	0.005	0.008

Tone pair	Acoustic Distance	Functional Load (type)	Functional Load (token)
chem3 / *chem6*	9.526	0.015	0.044
fau3 / *fau6*	46.980	0.015	0.044
keng3 / *keng6*	26.272	0.015	0.044
mu3 / *mu6*	21.216	0.015	0.044
te3 / *te6*	7.674	0.015	0.044
yon3 / *yon6*	12.251	0.015	0.044
chem4 / *chem5*	76.976	0.010	0.014
fau4 / *fau5*	78.382	0.010	0.014
keng4 / *keng5*	116.836	0.010	0.014
mu4 / *mu5*	79.235	0.010	0.014
te4 / *te5*	72.026	0.010	0.014
yon4 / *yon5*	36.793	0.010	0.014
chem4 / *chem6*	68.543	0.012	0.029
fau4 / *fau6*	70.663	0.012	0.029
keng4 / *keng6*	105.144	0.012	0.029
mu4 / *mu6*	80.448	0.012	0.029
te4 / *te6*	63.002	0.012	0.029
yon4 / *yon6*	31.606	0.012	0.029
chem5 / *chem6*	11.553	0.008	0.029
fau5 / *fau6*	16.602	0.008	0.029
keng5 / *keng6*	18.759	0.008	0.029
mu5 / *mu6*	13.355	0.008	0.029
te5 / *te6*	12.349	0.008	0.029
yon5 / *yon6*	10.173	0.008	0.029

Appendix B: Detailed statistical modeling results of perceived similarity

Table B1: Results of the logistic mixed effects model for accuracy on Linguistic Group, Acoustic Distance, and Functional Load based on type frequency.

Main model	β	SE	z	$p(>\|z\|)$
Fixed effects				
(Intercept)	2.19	0.44	4.94	< 0.01
English	-0.92	0.32	-2.86	< 0.01
Heritage	-0.33	0.22	-1.52	0.13
Acoustic Distance	1.34	0.44	3.03	< 0.01
Functional Load (type)	0.84	0.22	3.81	< 0.01
English × Acoustic Distance	-0.55	0.36	-1.53	0.13
Heritage × Acoustic Distance	-0.13	0.20	-0.63	0.53
English × Functional Load (type)	-0.45	0.26	-1.76	0.08
Heritage × Functional Load (type)	-0.14	0.20	-0.71	0.48
Acoustic Distance × Functional Load (type)	0.81	0.33	2.42	< 0.05
English × Acoustic Distance × Functional Load (type)	-0.71	0.35	-2.01	< 0.05
Heritage × Acoustic Distance × Functional Load (type)	-0.22	0.26	-0.86	0.39
			Variance	SD
Random effects				
Participant (Intercept)			0.23	0.48
Participant: Acoustic Distance			0.04	0.19
Participant: Functional Load (type)			0.03	0.17
Participant: Acoustic Distance × Functional Load (type)			0.02	0.14
Syllable (Intercept)			1.02	1.01
Syllable: Acoustic Distance			1.08	1.04
Syllable: English			0.40	0.63
Syllable: Heritage			0.04	0.20
Syllable: Functional Load (type)			0.18	0.43
Syllable: Acoustic Distance × English			0.65	0.80
Syllable: Acoustic Distance × Heritage			0.09	0.30
Syllable: Acoustic Distance × Functional Load (type)			0.51	0.71
Syllable: English × Functional Load (type)			0.25	0.50
Syllable: Heritage× Functional Load (type)			0.07	0.27
Syllable: Acoustic Distance × English × Functional Load (type)			0.54	0.74
Syllable: Acoustic Distance × Heritage × Functional Load (type)			0.16	0.40

Table B2: Results of the logistic mixed effects model for accuracy on Linguistic Group, Acoustic Distance, and Functional Load based on token frequency.

Main model	β	SE	z	$p(>\|z\|)$
Fixed effects				
(Intercept)	2.52	0.52	4.80	< 0.01
English	-1.08	0.41	-2.67	< 0.01
Heritage	-0.18	0.26	-0.71	0.48
Acoustic Distance	1.70	0.48	3.58	< 0.01
Functional Load (token)	1.12	0.38	2.92	< 0.01
English $\times$ Acoustic Distance	-0.60	0.40	-1.49	0.14
Heritage $\times$ Acoustic Distance	0.08	0.24	0.35	0.72
English $\times$ Functional Load (token)	-0.52	0.36	-1.43	0.15
Heritage $\times$ Functional Load (token)	0.14	0.23	0.59	0.55
Acoustic Distance $\times$ Functional Load (token)	1.19	0.51	2.32	< 0.05
English $\times$ Acoustic Distance $\times$ Functional Load (token)	-0.67	0.44	-1.55	0.12
Heritage $\times$ Acoustic Distance $\times$ Functional Load (token)	0.11	0.26	0.42	0.67
			Variance	SD
Random effects				
Participant (Intercept)			0.27	0.52
Participant: Acoustic Distance			0.07	0.26
Participant: Functional Load (token)			0.03	0.18
Participant: Acoustic Distance $\times$ Functional Load (token)			0.01	0.10
Syllable (Intercept)			1.44	1.20
Syllable: Acoustic Distance			1.21	1.10
Syllable: English			0.69	0.83
Syllable: Heritage			0.06	0.25
Syllable: Functional Load (token)			0.73	0.86
Syllable: Acoustic Distance $\times$ English			0.77	0.88
Syllable: Acoustic Distance $\times$ Heritage			0.09	0.30
Syllable: Acoustic Distance $\times$ Functional Load (token)			1.37	1.17
Syllable: English $\times$ Functional Load (token)			0.59	0.77
Syllable: Heritage$\times$ Functional Load (token)			0.08	0.28
Syllable: Acoustic Distance $\times$ English $\times$ Functional Load (token)			0.89	0.94
Syllable: Acoustic Distance $\times$ Heritage $\times$ Functional Load (token)			0.10	0.31

Table B3: Results of the mixed effects model for reaction time on Linguistic Group, Acoustic Distance, and Functional Load based on type frequency.

Main model	β	SE	t	$p(>\|t\|)$
Fixed effects				
(Intercept)	343.29	32.75	10.48	< 0.01
English	−38.07	38.91	−0.98	0.33
Heritage	−65.02	38.58	−1.69	0.10
Acoustic Distance	−26.55	12.24	−2.17	0.07
Functional Load (type)	−17.63	7.08	−2.49	< 0.05
English × Acoustic Distance	−7.62	7.00	−1.09	0.28
Heritage × Acoustic Distance	3.60	6.84	0.53	0.60
English × Functional Load (type)	−2.06	8.14	−0.25	0.80
Heritage × Functional Load (type)	2.52	8.02	0.32	0.75
Acoustic Distance × Functional Load (type)	5.88	8.10	0.73	0.47
English × Acoustic Distance × Functional Load (type)	−16.43	10.57	−1.55	0.12
Heritage × Acoustic Distance × Functional Load (type)	0.55	10.45	0.05	0.96
			Variance	SD
Random effects				
Participant (Intercept)			12736.05	112.85
Participant: Acoustic Distance			22.41	4.73
Participant: Functional Load (type)			8.44	2.91
Syllable (Intercept)			1193.27	34.54
Syllable: Acoustic Distance			708.15	26.61
Syllable: English			415.82	20.39
Syllable: Heritage			274.58	16.57
Syllable: Functional Load (type)			67.19	8.20

Table B4: Results of the mixed effects model for reaction time on Linguistic Group, Acoustic Distance, and Functional Load based on token frequency.

Main model	β	SE	t	$p(>\|t\|)$
Fixed effects				
(Intercept)	321.33	32.50	9.89	< 0.01
English	-30.87	39.32	-0.79	0.44
Heritage	-61.84	38.83	-1.59	0.12
Acoustic Distance	-61.95	14.97	-4.14	< 0.01
Functional Load (token)	-53.63	8.72	-6.15	< 0.01
English × Acoustic Distance	4.22	9.24	0.46	0.65
Heritage × Acoustic Distance	7.25	9.09	0.80	0.43
English × Functional Load (token)	11.93	9.93	1.20	0.23
Heritage × Functional Load (token)	5.76	9.79	0.59	0.56
Acoustic Distance × Functional Load (token)	-61.896	9.88	-6.26	< 0.01
English × Acoustic Distance × Functional Load (token)	19.84	12.76	1.55	0.12
Heritage × Acoustic Distance × Functional Load (token)	8.74	12.74	0.69	0.49
			Variance	SD
Random effects				
Participant (Intercept)			12716.12	112.77
Participant: Acoustic Distance			42.10	6.49
Participant: Functional Load (token)			18.65	4.32
Syllable (Intercept)			1043.47	32.30
Syllable: Acoustic Distance			1025.21	32.02
Syllable: English			522.67	22.86
Syllable: Heritage			312.20	17.67
Syllable: Functional Load (token)			108.86	10.43