Original Article

Language *and* Speech

Language *and* Speech
1–25

DOI: 10.1177/0023830919879471
journals.sagepub.com/home/las
SAGE

The Perception of Cantonese Vowel Length Contrast by Mandarin Speakers

Jingxin Luo
Vivian Guo Li
Peggy Pik Ki Mok
The Chinese University of Hong Kong, Hong Kong

Abstract
The study investigates the perception of vowel length contrasts in Cantonese by native Mandarin speakers with varying degrees of experience in Cantonese: naïve listeners (no exposure), inexperienced learners (~1 year), and experienced learners (~5 years). While vowel length contrasts do not exist in Mandarin, they are, to some extent, exploited in English, the second language (L2) of all the participants. Using an AXB discrimination task, we investigate how native and L2 phonological knowledge affects the acquisition of vowel length contrasts in a third language (L3). The results revealed that all participant groups could discriminate three contrastive vowel pairs (/aː/–/ɐ/, /ɛː/–/e/, /ɔː/–/o/), but their performance was influenced by the degree of Cantonese exposure, particularly for learners in the early stage of acquisition. In addition to vowel quality differences, durational differences were proposed to explain the perceptual patterns. Furthermore, L2 English perception of the participants was found to modulate the perception of L3 Cantonese vowel length contrasts. Our findings demonstrate the bi-directional interaction between languages acquired at different stages, and provide concrete data to evaluate some speech acquisition models.

Keywords
Vowel length contrast, speech perception, Cantonese, speech acquisition, cross-linguistic influence

1 Introduction

Determining the roles of the native language (L1) and second language (L2) is an important focus in language acquisition research. The present study addresses how L1 and L2 phonological knowledge affects the perception of vowel length contrasts in a third language (L3), taking L3 Cantonese

Corresponding author:
Peggy Pik Ki Mok, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Leung Kau Kui Building, Shatin, Hong Kong.
Email: peggymok@cuhk.edu.hk

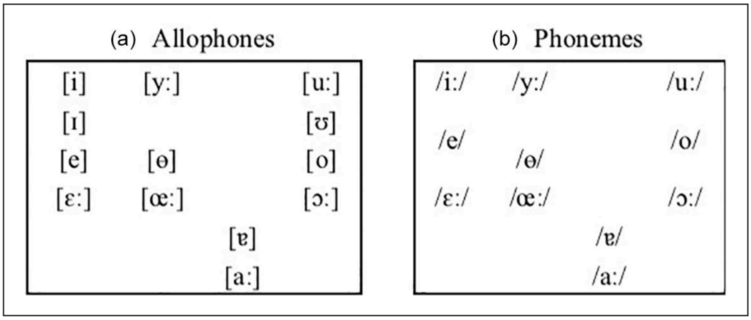

Figure 1. The 11-phoneme system of Cantonese (Bauer & Benedict, 1997).

learned by native Mandarin speakers with L2 English experience as an example. Vowel length contrasts are of interest in this case, as they do not exist in Mandarin (Duanmu, 2007), but are used in Cantonese (Bauer & Benedict, 1997) and partially used in English (Giegerich, 1992). The study examines how native Mandarin speakers with varying amounts of Cantonese experience (naïve, inexperienced, and experienced) perceive Cantonese vowel length contrasts. The influence of L2 English is also discussed.

1.1 Cantonese vowel length contrast

Vowel length contrasts refer to the phonological oppositions between long and short vowels, which can signal different word meanings (Odden, 2011). Languages with vowel length contrasts may utilize vowel quantity (duration) or vowel quality (formant frequencies) to mark the distinctions. Vowels length contrasts in some languages, such as Japanese and Finnish, differ only in quantity (Iivonen & Harnud, 2005; Vance, 2008). In other languages, such as English and Swedish, however, vowel length contrasts involve both vowel quantity and quality (Odden, 2011), although vowel quality functions as the primary cue in English (Giegerich, 1992; House, 1961) and vowel quantity differences are more prominent in Swedish (Behne, Arai, Czigler, & Sullivan, 1999; Hadding-Koch & Abramson, 1964).

Vowel length contrasts in Cantonese are less straightforward. Cantonese has 13 vowel phones, including seven long vowels [iː yː uː ɛː œː ɔː aː] and six short vowels [ɪ ʊ e ɵ o ɐ] (Cheung, 1972; Kao, 1971; Yue-Hashimoto, 1972; Zee, 1999). Of the different analyses that have been proposed for the Cantonese vowel system,[1] the 11-phoneme system recognizes the contrasting pattern between long and short vowel phones, with three long high vowels and four pairs of mid and low vowels contrasting in length (Bauer & Benedict, 1997; Cheung, 1986; Kao, 1971; Lee, 1983, 1985; Li, Huang, Shi, Mai, & Chen, 1995), as shown in Figure 1. In this system, [ɪ] and [e] are allophones of /e/ that contrast with the long vowel /ɛː/; [ʊ] and [o] are allophones of /o/ that contrast with /ɔː/; and the short vowel /ɵ/ contrasts with /œː/, and /ɐ/ with /aː/ (Bauer & Benedict, 1997). These contrastive pairs involve differences in both vowel quantity and quality.

Previous acoustic studies have confirmed systematic durational differences between the contrastive pairs, indicating that vowel quantity is a significant feature of vowel length in Cantonese (Kao, 1971; Lee, 1983, 1985; Shi & Liu, 2005). Lee (1983, 1985) observed and Shi and Liu (2005) further confirmed that, at a similar speech rate, the durational ranges of the long and short vowels of a vowel pair did not overlap, although there was a large degree of spectral overlap (Zee, 2003). The large spectral overlap implied that vowel quality is not the only distinctive feature in signaling the contrast. Thus, they suggested that vowel quantity also functions as a crucial distinctive feature

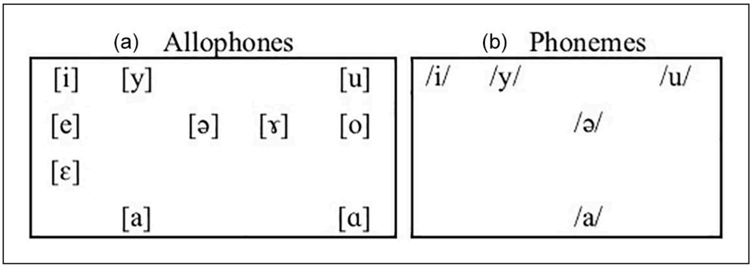

Figure 2. The Mandarin vowel system (Lin, 2007).

(Lee, 1983, 1985; Shi & Liu, 2005). Previous perception studies have concluded that both vowel quality and vowel quantity serve as important cues in the perception of the contrastive pairs (Shi & Liu, 2002; Shi & Mai, 2003). From an acquisition perspective, children process differences in both vowel quality and quantity in order to discriminate between the minimal pairs of vowel length contrasts (Chen, 2011).

This study adopted the 11-phoneme system in which there are three long high vowels and four pairs of mid and low vowels contrasting in length (Bauer & Benedict, 1997; Cheung, 1986; Kao, 1971; Lee, 1983, 1985; Li et al., 1995). However, the mid-vowel pair /œː/–/ɵ/ was not included in the experiment because of its larger quality difference compared to the other three vowel pairs (Zee, 2003) and the limited number of minimal pairs involving these two vowels. Therefore, this study focused on two mid-vowel pairs (/ɛː/–/e/, /ɔː/–/o/) and a low-vowel pair (/aː/–/ɐ/). While most of the previous studies of Cantonese vowel length contrasts have used native speakers with a L1 perspective, very few have examined the L2/L3 acquisition of this feature. The present study aims to fill this gap and to provide some insight into the L2/L3 acquisition of vowel quantity contrasts.

1.2 Mandarin vowel system

Compared to Cantonese, Mandarin has a smaller vowel inventory, as illustrated in Figure 2. There are three contrastive high vowel phonemes, namely /i/, /y/, and /u/, also present in Cantonese. The four mid-vowel phones, on the other hand, occur in different environments and do not distinguish word meanings (Duanmu, 2007; Lee & Zee, 2003; Lin, 2007). Thus, there is only one underspecified mid-vowel phoneme (/ə/) with four mid-vowel allophones appearing in different contexts. With regard to the low vowels, phonologically, Mandarin only has one low-vowel phoneme (/a/), which has different phonetic realizations depending on the context (Duanmu, 2007; Lin, 2007). Unlike Cantonese, Mandarin has no long–short distinctions. Although Mandarin listeners have been found to use duration cues for tonal categorization (Blicher, Diehl, & Cohen, 1990; Liu & Samuel, 2004), this categorization integrates with salient pitch information for a different function.

Since Mandarin lacks vowel length contrasts and has fewer vowels, how do Mandarin learners perceive and produce this feature when learning L3 Cantonese? Currently, most L3 models address issues in morpho-syntax and the lexicon, while research on L3 phonetics and phonology remains limited in number and scope. For example, both the Cumulative Enhancement Model (CEM) (Flynn, Foley, & Vinnitskaya, 2004) and the Typological Primacy Model (Rothman, 2010, 2015) predict positive transfer from both the L1 and L2 to L3, but these models are based on syntactic data. Currently, only the Phonological Permeability Hypothesis (Cabrelli Amaro & Rothman, 2010) addresses issues in L3 phonological acquisition. It proposes that a L2 acquired after the critical period is more vulnerable than the L1 to L3 influence. Since this hypothesis focused more on general issues of directionality of cross-linguistic influence than on making precise predictions on how speech sounds will be perceived and learned, the current study relies on models of L2 speech acquisition to provide concrete predictions.

1.3 Models of L2 speech acquisition

Previous studies have revealed that not all non-native contrasts are equally difficult (Best et al., 1988; Polka, 1991; Werker & Tees, 1984). Various models have been proposed to account for the variation in the extent to which individuals learn to accurately perceive and produce phonetic segments in a L2. This study mainly considers three models, namely the Speech Learning Model (SLM) (Flege, 1989, 1995, 1999a; Flege, Bohn, & Jang, 1997), the Perceptual Assimilation Model (PAM) (Best, 1995; Best & Tyler, 2007), and the Second Language Linguistic Perception (L2LP) model (Escudero, 2005, 2009), focusing on the importance of the perceived relationship between native and non-native sounds in the way in which these sounds are discriminated and eventually learned.

The SLM posits that the processes and mechanisms for establishing new categorical representations for speech sounds remain intact and accessible throughout the life span, although phonetic category formation for L2 speech sounds becomes less likely with increasing age[2] (Flege, 1995, 2003). The SLM proposes that the phonetic elements that make up the L1 and L2 phonetic subsystems exist in a "common phonological space" and thus mutually influence each other (Flege, 1995). According to the SLM, the further a L2 sound is from the nearest L1 sound, the more likely a new category will be developed, making it easier for learners to accurately perceive and produce the sound (Flege, 1995). When a L2 sound is perceived as its L1 counterpart, the L1 and L2 categories will assimilate, leading to a "merged" L1–L2 category (Flege, 1995).

A number of studies by Flege examined the English /i/ versus /ɪ/ learned by L2 learners with different L1 backgrounds, for example, Spanish (see reviews by Flege, 1995, 2003). Although this vowel pair differs in both vowel quality and vowel duration, and durational measurements were included in his studies, Flege mainly considered the formant patterns of these two vowels in relation to L1 vowels in his predictions and interpretation of the data. He even suggested that learners could detect the *auditory* durational difference between /i/ and /ɪ/ tokens without *categorizing* these vowels as different—that is, phonetic category formation was blocked by equivalence classification (treating the durationally different tokens as the same, Flege, 1995, p. 248). Thus, although the SLM, as a theory, could possibly incorporate vowel length contrasts involving durational dimensions (i.e., different feature weights), in practice, it is safe to assume that the main concern of the SLM is vowel quality and that it is agnostic about what would happen to long and short vowels in the L2 when there is no L1 length contrast.

The PAM (Best, 1995) and the PAM-L2 (Best & Tyler, 2007) are other relevant models of how listeners perceive L2 contrasts and assimilate them based on L1 categories. Six assimilation types have been proposed based on the articulatory similarities between L1 and L2 sounds, each of which entails expectations about discrimination performance. The PAM was developed under the direct realism approach and assumes articulatory gestures to be the perceptual primitives. As duration is an intrinsic aspect of articulatory gestures (Browman & Goldstein, 1989, 1992), it is reasonable to expect that vowel length contrasts involving both quality and quantity differences could be handled well by this model. Nevertheless, no explicit exposition of durational effects has been given. Most PAM illustrations focus on consonant contrasts. Studies that have tested the prediction of the PAM on vowels have only considered formant patterns even when vowel length is used in the target language, as in Dutch (Escudero & Williams, 2011). So far, no study has specified the gestural constellation for different vowels. Thus, based on the PAM, if the durational feature is not used in the L1, it is unclear how durational differences in L2 vowels can be assimilated into L1 categories.

Neither the SLM nor the PAM has explicitly considered vowel duration in their expositions, but it could be included in their predictions. Our predictions below are based on allophonic qualities (see Figures 1 and 2), following previous studies adopting these models.

Due to the similarity in vowel quality and the environments in which they occur, the SLM predicts that the Cantonese contrastive low vowels /aː/–/ɐ/ might be perceived as similar to the Mandarin low-vowel /a/ (with [a] and [ɑ] allophones). It is not easy for learners to establish a new category for these L2 sounds. Poor discrimination is predicted. Likewise, for the /ɛː/–/e/ pair, the Cantonese /e/ has two allophones ([e] and [ɪ]). They are likely to be perceived as [e] by Mandarin learners. Moreover, [e] and [ɛ] are allophones of the same underspecified mid-vowel /ə/ in Mandarin. It is possible that no new categories would be developed for Mandarin learners, so poor discrimination is also predicted for the /ɛː/–/e/ pair. In the same vein, for the /ɔː/–/o/ pair, the two allophones [o] and [ʊ] of the Cantonese /o/ are likely to be perceived as the Mandarin allophone [o]. However, there is no similar Mandarin vowel to which the Cantonese long vowel /ɔː/ can assimilate. A new category may be developed for the Cantonese /ɔː/ sounds, allowing the learners better separation of the /ɔː/–/o/ pair than the other two vowel pairs.

According to the PAM, two assimilation types are possible for the /aː/–/ɐ/ and /ɛː/–/e/ pairs, as assimilation types can vary among individuals for any given non-native vowel contrast (Tyler, Best, Faber, & Levitt, 2014). Firstly, since both vowels in each Cantonese vowel length pair are likely to assimilate to the same phoneme in Mandarin, Single-Category (SC) Type assimilation may occur with poor discrimination predicted. Alternatively, Category-Goodness (CG) Type assimilation may occur if one vowel in the Cantonese pair is perceived as a better exemplar of the Mandarin phoneme than the other, resulting in intermediate discrimination. As for the /ɔː/–/o/ pair, Uncategorized versus Categorized (UC) Type assimilation is predicted, in which /o/ is categorized and /ɔː/ is uncategorized. Very good discrimination may result.

Both the SLM and the PAM have similar predictions for the three vowel length pairs in Cantonese: better discrimination for the /ɔː/–/o/ pair and poorer discrimination for the /aː/–/ɐ/ and /ɛː/–/e/ pairs.

Unlike the SLM and the PAM, the L2LP model (Escudero, 2005, 2009) does include durational properties in its predictions. It proposes that listeners are optimal perceivers of their L1, and that L2 learners will initially duplicate their L1 perceived categories in their L2 perception. In our case, as Mandarin lacks vowel length contrasts, there is no duration category in their L1 perception. The duplication of the L1 Mandarin categories will result in perceiving the two contrastive Cantonese vowels as a single Mandarin vowel. The Cantonese /aː/–/ɐ/ contrast is predicted to be perceived as the Mandarin low-vowel /a/, and the two Cantonese mid-vowel pairs (/ɛː/–/e/, /ɔː/–/o/) will probably be perceived as the only Mandarin mid-vowel /ə/. Thus, poor discrimination on the part of the naïve listeners is predicted for these three Cantonese vowel pairs.

The L2LP model also propounds that learners will have access to a L1-like learning mechanism when adjusting their initial L2 perception to develop into optimal L2 listeners. As a result of the L1-like development, L2 learners will create new categories along previously unused dimensions to classify sounds in their L1 via auditory-driven learning (Boersma, Escudero, & Hayes, 2003). Thus, the durational distributions of the Cantonese vowel length pairs will lead to a categorization of the vowel duration continuum as two vowel length categories. The L2LP model proposes that duration categories will be developed to address the new durational distribution in the L2. Mandarin learners of Cantonese will create perceptual mappings that link vowel duration values to the newly formed vowel length categories. The new length categories will then become abstract representations via phonological abstraction.

With regard to the end state, the L2LP model hypothesizes that the L1 and L2 are governed by two separate grammars, both of which can be optimal. If Mandarin learners receive sufficient Cantonese input, they are likely to perceive the three vowel pairs accurately in the end state of acquisition.

Predictions based on the SLM and the PAM, which do not explicitly include duration in their expositions, are predicated solely on spectral properties. Since Mandarin learners of Cantonese may not be sensitive to durational differences and may rely more on the spectral cues to discriminate between the contrastive pairs, poor discrimination is predicted for the /aː/–/ɐ/ and /ɛː/–/e/ pairs, and better discrimination for the /ɔː/–/o/ pair. By contrast, the L2LP model predicts that, although Mandarin listeners have no vowel duration category in their L1 perception, learners will gradually develop a new duration category to cope with the new vowel durational distributions in Cantonese. This model predicts that the Mandarin listeners will initially have poor discrimination of the three Cantonese vowel length pairs, but with sufficient input, they will be able to perceive these pairs accurately. Alternatively, the large durational difference between the long and short vowels in Cantonese may be very salient to the learners already, and this alone may be sufficient for good auditory discrimination, regardless of how the vowels are categorized.

1.4 The influence of English as a second language

At present, English is the L2 for most educated Mandarin speakers in Mainland China. Whether English plays a role when Mandarin speakers learn the vowel length contrasts in Cantonese is a question that remains unanswered. As in Cantonese, vowel length is also used to an extent in the English vowel system. Nevertheless, vowel length alone is rarely used contrastively in any variety of English, and for vowel pairs where there is a difference in duration, there is generally also a distinction in vowel quality (Davenport & Hannahs, 2010). Previous studies (Giegerich, 1992; House, 1961, among others) have suggested that vowel quality plays a more important role in perceiving and producing English vowel length distinctions. Specifically, the vowel quality of these contrastive vowel pairs differs in both vowel height and position (front and back) (Jia, Strange, Wu, Collado, & Guan, 2006). Thus, the vowel length contrast in English is usually referred to as the tense-lax difference. The duration of the tense vowels is longer than that of their lax counterparts (Giegerich, 1992). Moreover, there is variation among different varieties of English. For example, the vowels in the pair STRUT/START, which are transcribed by Bauer and Warren (2004) as /ɐ/ and /ɐː/, respectively, appear to be a phonemic quantity pair in New Zealand English, but are not contrasted by quantity alone in many other varieties of English (Warren, 2018). There are some pairs of vowel contrasts differing in both duration and quality in English, most notably /i/–/ɪ/ (e.g., *beat* vs. *bit*) and /u/–/ʊ/ (e.g., *food* vs. *foot*) (Giegerich, 1992).

While the L1 is the only source of transfer in L2 acquisition, L3 acquisition can have transfer from the L1, the L2, or both (Flynn et al., 2004; Rothman, 2011). In particular, the previous experience of a L2 may also affect the subsequent learning process of an additional language (Tremblay, 2006). Until fairly recently, only a handful of studies had addressed L3 phonetics and phonology. While recent L3 studies have focused mainly on the production of segments (Cabrelli Amaro, 2012; Wrembel, 2012; Wrembel, Gut, & Mehlhorn, 2010), very few studies have examined the perceptual aspects of L3 phonological acquisition (Cabrelli Amaro & Wrembel, 2016). García (2013) conducted an experiment to investigate the production by L3 learners of Portuguese with a L1 English, L2 Spanish background. The results reflected that the L3 learners with a higher proficiency in the L2 performed better than did those with limited proficiency. In a similar vein, Qin and

Jongman (2016) investigated the perception of unfamiliar L3 tones (Cantonese) by 15 English-speaking L2 learners of Mandarin, compared to native speakers of Mandarin and English. They concluded that both the L1 and L2 experience modulated the perception of lexical tones in the L3.

Considering that vowel length contrasts (partially) exist in both English and Cantonese phonological systems but is absent in Mandarin, and that the L2 can be a source of transfer in the acquisition of an additional language, the influence of English as a L2 will be taken into consideration in our study.

2 The present study

The present study investigates how Cantonese vowel length contrasts are perceived by native Mandarin speakers who had learned English as a L2 and had various degrees of L3 Cantonese exposure. As a previously acquired language is also a source of transfer, and very few studies have addressed how a L2 affects perceptual aspects of L3 phonological acquisition, the present study also examines the influence of L2 English on the perceptual learning of Cantonese vowel length contrasts by Mandarin speakers.

Considering the phonological systems of Cantonese, Mandarin, and English, the following research questions were raised: firstly, can Mandarin speakers perceive the distinction between the long and short vowel pairs in Cantonese? If so, how does their experience in Cantonese affect their perceptual discrimination? Secondly, does the perception of L2 English vowel length contrast influence the perceptual learning of L3 Cantonese vowel length contrasts by native Mandarin speakers?

3 Method

3.1 Participants

The participants in this study were native Mandarin speakers born and raised in northern China, thus having little exposure to Cantonese. For all participants, Mandarin was the most commonly used language in their daily lives. The participants had learned English as a L2 and had comparable English proficiency but had no experience of learning any other languages apart from Cantonese. No participant reported any hearing problems.

These native Mandarin speakers were divided into three groups based on their varying amount of exposure to Cantonese. The 27 participants (12 male, 15 female) in the naïve group had no experience with Cantonese. Those in the inexperienced and experienced groups had lived, worked, or studied in a Cantonese-dominant environment (Hong Kong). Specifically, the 20 inexperienced learners (eight male, 12 female) had less than one year of Cantonese exposure, and the 20 experienced learners (eight male, 12 female) had over five years of experience in a Cantonese-dominant environment. A control group consisting of 14 native speakers of Cantonese (four male, 10 female) was also included in the study. These native Cantonese speakers had been born and raised in Hong Kong, and had received education in either English or in Cantonese, but not in Mandarin.

Participants rated their use of Cantonese and English in different circumstances on a language background questionnaire. Compared with the naïve and inexperienced groups, the experienced learners reported that they listened to and spoke Cantonese much more often, particularly in the work place. In terms of English, all the participants had learned English for over 10 years. They reported that they had received their tertiary education either in English or bilingually in English and Mandarin. Most participants reported that they only used English in certain situations, such as when taking classes, delivering presentations, or communicating with English-speaking customers. Their

Table 1. Background information about the participants.

Group	No. of participants	Age in years (SD)	Cantonese experience years (SD)	English experience years (SD)	IELTS overall score (SD)
Naïve	27	18.70(1.41)	0	12.67(1.47)	6.61(.53)
Inexperienced	20	20.85(2.35)	0.79 (0.26)	13.20(1.36)	6.48(.47)
Experienced	20	23.15(3.47)	5.46 (0.78)	14.15(3.42)	6.67(.54)
Native	14	20.86(1.70)	N/A	15.29(3.00)	6.58(.34)

IELTS: International English Language Testing System.

English language proficiency was comparable. Most of the participants had taken the IELTS (International English Language Testing System). The years of English-learning experience and the overall IELTS scores are listed in Table 1. There was no significant difference in the IELTS scores, $F(2, 64) = 2.818, p > 0.05$, or in the years of English exposure, $F(2, 64) = 2.592, p > 0.05$, among the three learner groups.

3.2 Perception experiment

An AXB discrimination task was conducted to assess the participants' perception of Cantonese vowel length contrasts. The three Cantonese vowel pairs were combined with different coda consonants following Cantonese phonotactics. The low-vowel pair /aː/–/ɐ/ can occur in all coda environments. The experimental materials included all the minimal pairs of these coda environments, except /aːk̚/–/ɐk̚/, for which we failed to find appropriate minimally contrastive words. The two mid-vowel pairs /ɛː/–/e/ and /ɔː/–/o/ can only form minimal pairs with the velar codas [ŋ] and [k̚]. For the short vowels /e/ and /o/, only the allophone [ɪ] of /e/ and the allophone [ʊ] of /o/ can occur with velar codas. Therefore, the two mid-vowel pairs could only include the minimal pairs contrasting between [ɛː] and [ɪ], [ɔː], and [ʊ]. Thus, the low-vowel pair /aː/–/ɐ/ included seven coda environments, and the two mid-vowel pairs /ɛː/–/e/ and /ɔː/–/o/ had two coda environments each, resulting in 11 vowel pairs. There were three minimal pairs for each vowel pair, which resulted in a total of 33 Cantonese target word pairs (11 vowel pairs $\times$ 3 minimal pairs). Some examples are given in Table 2.

These 33 word pairs were presented in four possible combinations (AAB, ABB, BAA, BBA), resulting in a total of 132 trials. In addition, 66 filler trials were added, including minimal pairs of the contrasts between the level tones (T1, T3, T6), rising tones (T2, T5), final stops /p̚, t̚, k̚/, and final nasals /m, n, ŋ/ in Cantonese. The 198 trials were presented in six blocks of 33 trials, with all trials presented in a random order. In addition, the same two words in each AXB combination were always two physically different stimulus tokens. The stimuli were produced by a female native speaker of Cantonese and had a similar volume and pitch. The stimulus words were randomly allocated to six lists, which the speaker read three times each. In order to avoid any intonational variations between the same words, the members of a minimal pair were placed in similar positions in different lists, and three extra words were added at the beginning and at the end of each set, respectively, thereby minimizing boundary effects on the stimuli.

With regard to the English materials, two vowel pairs, /i/–/ɪ/ and /u/–/ʊ/, were included as the target contrasts. We included four minimal pairs for each contrast, which were presented with the four possible combinations, resulting in 32 trials. Together with 12 fillers, a total of 44 English trials were presented after the Cantonese session. The English stimuli were produced by a female native speaker of British English with a similar volume and pitch among all the tokens.

Table 2. Examples of Cantonese minimal pairs used in the perception experiment.

Vowel	Long	Short
/aː/–/ɐ/	街 [kaːj55] "street"	雞 [kɐj55] "chicken"
	考 [haːw35] "test"	口 [hɐw35] "mouth"
	三 [saːm55] "three"	心 [sɐm55] "heart"
	班 [paːn55] "class"	賓 [pɐn55] "guest"
	棚 [pʰaːŋ21] "shed"	朋 [pʰɐŋ21] "friend"
	沓 [taːp̚2] "pile"	揸 [tɐp̚2] "beat"
	滑 [waːt̚2] "slip"	核 [wɐt̚2] "fruit stone"
/ɛː/–/e/	贏 [jɛːŋ21] "win"	型 [jɪŋ21] "type or model"
	石 [sɛːk̚2] "rock"	食 [sɪk̚2] "eat"
/ɔː/–/o/	方 [fɔːŋ55] "square"	風 [fʊŋ55] "wind"
	度 [tɔːk̚2] "measure"	讀 [tʊk̚2] "read aloud"

The AXB experiment was conducted using DMDX Display Software (Foster & Foster, 2003). Participants were instructed to make responses on a keyboard using their index fingers to press "Z" if they thought the second item was the same as the first or "M" if the second item sounded like the third. Participants were instructed to respond as quickly and as accurately as possible. The interstimuli interval was 750 ms. Once the participant responded, the next trial was triggered. The time-out time was 4 seconds. Breaks were given between each block. Before the actual test, eight trials were given to familiarize the participants with the task. All the participants performed the experiment individually in a quiet environment. Their accuracy and reaction times (RTs) were collected.

3.3 Acoustic properties of the stimuli

Averaged formant frequencies and duration for the long and short vowels in Cantonese and English stimuli are given in Table 3. As the low-vowel pair /aː/–/ɐ/ can appear as monophthongs and in diphthongs (i.e., with /j/ and /w/ codas), data for this vowel pair are given separately in Table 3. In addition, the duration for this vowel pair appearing in diphthongs includes the duration of the whole diphthong, as it was difficult to separate the vowel and the glide coda. Formant frequencies were averaged over the whole vowel for all monophthongs in the three vowel pairs, while they were taken at the first one-third of the diphthongs for /aː/–/ɐ/ to minimize the influence of the glide codas.

Table 3 shows that there was no overlap in duration between the long and short vowels for the three Cantonese vowel pairs (except for /aː/–/ɐ/ appearing in diphthongs, which includes the duration of the glide codas). The F1 and F2 formant frequencies are quite similar for each vowel pair as well, indicating a large degree of spectral overlap. In contrast, the English vowel pairs differ more in formant frequencies than in duration. These observations concur with patterns described in the literature for the two languages.

4 Results

4.1 Perception of Cantonese vowel length contrast

A mixed-effects regression approach was performed to analyze response accuracy (logistic regression) and RT (linear regression).

Table 3. Formant frequencies (Hz) and duration (ms) of the vowel stimuli in the perception experiment.

Vowel	Cantonese								English			
	/aː/ (diphthong)	/ɐ/ (diphthong)	/aː/	/ɐ/	/ɛː/	/e/	/ɔː/	/o/	/iː/	/ɪ/	/uː/	/ʊ/
F1 mean (SD)	817.64 (43.77)	780.88 (19.75)	829.16 (53.23)	778.67 (27.32)	726.21 (30.48)	641.74 (54.58)	728.43 (25.72)	704.26 (34.95)	483.33 (32.23)	554.47 (16.19)	447.93 (24.58)	581.19 (79.7)
F2 mean (SD)	1517.8 (134.79)	1541.99 (171.65)	1554.88 (104.26)	1489.17 (102.04)	1958.21 (156.03)	1986.39 (167.22)	1182.62 (85.39)	1171.26 (197)	2662.59 (158.82)	2049.34 (167.11)	1636.03 (335.66)	1455.91 (180.94)
Duration mean (SD)	525.07 (88.83)	486.46 (63.39)	293.1 (75.94)	130.41 (18.41)	276.52 (80.7)	141.26 (26.74)	271.26 (96.46)	155.47 (35.28)	238.65 (92.23)	147.96 (39.88)	231.84 (103.71)	146.84 (30.31)

Table 4. Mean accuracy (%) for the three Cantonese contrasts.

Vowel pairs	*Naïve* (n = 27) % correct (SD)	*Inexperienced* (n = 20) % correct (SD)	*Experienced* (n = 20) % correct (SD)	*Native* (n = 14) % correct (SD)
/aː/–/ɐ/	88.01 (6.19)	95.06 (2.19)	96.55 (2.35)	99.40 (0.77)
/ɛː/–/e/	87.35 (11.29)	91.88 (5.32)	96.25 (3.28)	98.48 (2.11)
/ɔː/–/o/	86.14 (10.91)	92.73 (5.28)	94.73 (3.36)	98.46 (2.15)
Overall	87.16 (7.07)	93.22 (2.94)	95.84 (2.02)	98.78 (1.16)

4.1.1 Perception accuracy. The accuracy results were compared across the four participant groups, namely naïve, inexperienced, experienced, and native. All four groups performed well above chance level for all the pairs. The means and standard deviations are shown in Table 4. The naïve group had the lowest accuracy for all three vowel pairs compared with the other three groups, but their mean accuracy for the three pairs was still higher than 85%. Similarly, the experienced group performed better than the inexperienced group for all the three vowel pairs /aː/–/ɐ/, /ɛː/–/e/, and /ɔː/–/o/, with 96.55%, 96.25%, and 94.73% accuracy, respectively. The native group had the highest score for all the three pairs among the participant groups, with over 98% accuracy for all contrasts. The overall pattern showed that the perception accuracy of the three Cantonese vowel length pairs increased in conjunction with increased experience in Cantonese.

To further examine the differential performance across the four groups, a mixed-effects logistic regression was performed using the glmer function in the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2018). Responses with a 4 second or longer RT were discarded. The initial model included Group (naïve, inexperienced, experienced, native) and Vowel, aa (/aː/–/ɐ/), ee (/ɛː/–/e/), oo (/ɔː/–/o/), as fixed factors, along with by-participant and by-word (the item used as stimulus X in the AXB sequence) as random intercepts.

Comparisons with models that included English exposure (measured in years, scaled and centered), coda (/j, w, p, t, k, m, n, ŋ/), by-speaker mean accuracy for English contrasts (scaled and centered), or the two-way interactions between Group and other factors, or models that included by-participant and by-word random slopes, showed that Coda ($\chi^2 = 26.84$, $p < .001$), accuracy for English contrasts ($\chi^2 = 10.089$, $p = .001$), as well as the interaction between Group and accuracy for English contrasts, significantly improved the model ($\chi^2 = 11.839$, $p = .007$). Models were compared via the anova() function in lme4, which was based on likelihood and deviance. Other factors, such as English exposure, its interaction with Group, and random slopes, did not improve the model.

The final model (Table 5) therefore included Group, Vowel, Coda, accuracy for English contrasts, two-way interaction between Group, and accuracy for English contrasts, with by-participant and by-word intercepts as predictors. Contrasts were treatment-coded, with the native group and the aa (/aː/–/ɐ/) pair with the /j/ coda as reference levels.

Table 5 suggests that for the reference condition (the /aː/–/ɐ/ contrast with the /j/ coda), accuracy of non-native groups differed significantly from that of the native group. For the native group, tokens with the /w/ coda were significantly lower in accuracy than tokens with the /j/ coda; different vowel pairs did not differ significantly in accuracy. We explored the more general patterns related to group difference, vowel contrasts, and coda conditions (e.g., patterns of group differences averaged over all vowel contrasts and coda conditions) by performing post-hoc pairwise comparisons. Terms related to accuracy for English contrasts suggested that, for the native group, an increase in accuracy for English contrasts is associated with a slight and insignificant decrease in accuracy for Cantonese contrasts, but for all non-native groups, increase in accuracy for English

Table 5. Fixed effects in the final model for perception accuracy.

	Estimate	Std. error	z-value	Pr(>|z|)
(Intercept)	5.99	0.57	10.54	<0.0001***
Group: experienced	−1.98	0.52	−3.79	0.0001***
Group: inexperienced	−2.50	0.52	−4.81	<0.0001***
Group: naïve	−2.98	0.51	−5.79	<0.0001***
Coda:k	−0.45	0.43	−1.05	0.29
Coda:m	0.01	0.37	0.03	0.98
Coda:n	−0.41	0.35	−1.17	0.24
Coda:ng	−0.22	0.36	−0.61	0.54
Coda:p	−0.30	0.36	−0.84	0.40
Coda:t	−0.11	0.36	−0.32	0.75
Coda:w	−1.52	0.33	−4.54	<0.0001***
English accuracy	−0.94	0.61	−1.55	0.12
Vowel: ee	−0.41	0.32	−1.29	0.20
Vowel: oo	−0.50	0.32	−1.56	0.12
Group: experienced × English accuracy	1.05	0.61	1.71	0.09
Group: inexperienced × English accuracy	1.13	0.62	1.83	0.07
Group: naïve × English accuracy	1.31	0.61	2.13	0.03*

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1.
glmer(response~Group+Coda+EngAccuracy+VowelContrast+Group:EngAccuracy+(1|ID)+(1|stimulusX), data=data, family=binomial, control = glmerControl(optimizer = "bobyqa")).

contrasts is positively associated with their accuracy for Cantonese contrasts. For the naïve group (but not the experienced or inexperienced groups), the increase in accuracy for Cantonese contrasts that comes with a unit increase in accuracy for English contrasts reached statistical significance in comparison with the native group. Figure 3, generated with the jtools package (Long, 2018), illustrates the Group × English accuracy interaction. Further analysis regarding Cantonese and English accuracy is provided below.

Post-hoc pairwise comparisons (see Tables 1–3 in the Online Supplemental Material) using lsmeans from the emmeans package (Lenth, 2018) with Tukey correction indicated that accuracy of the native group was higher than that of the experienced group, which was in turn higher than that of the inexperienced group; the naïve group had the lowest accuracy; all pairwise comparisons among the groups were statistically significant ($p < 0.05$). Accuracy for the /aː/–/ɐ/ pair was higher than that for other vowel pairs, but the difference was not statistically significant. Accuracy for tokens with the /w/ coda were significantly lower than that for tokens with all other codas except /k/ (i.e., tokens with /j, p, t, m, n, ŋ/ as coda) ($p < 0.05$).

4.1.2 Reaction time. The mean RT for the three vowel pairs across the four groups is shown in Figure 4. The data illustrate that RTs decreased in conjunction with the degree of experience in Cantonese. Specifically, the naïve group demonstrated the longest RTs for all three pairs when compared with the two learner groups and the native group.

A mixed-effects linear regression was performed using the lmer function in the lme4 package (Bates et al., 2015) in R (R Core Team, 2018). Responses of 4 seconds or longer were discarded. The initial model included Group (naïve, inexperienced, experienced, native) and Vowel, aa (/aː/–/ɐ/), ee (/ɛː/–/e/), oo (/ɔː/–/o/), as fixed factors, with by-participant and by-word (the item used as stimulus X in the AXB sequence) as random intercepts.

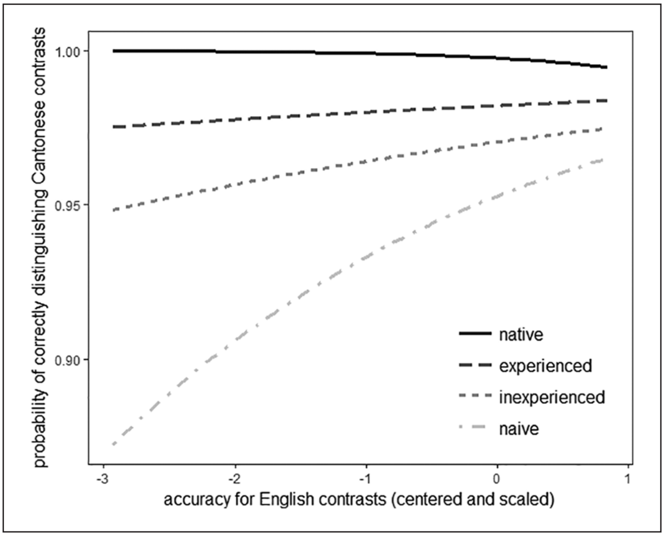

Figure 3. Group × Accuracy for English contrasts interaction.

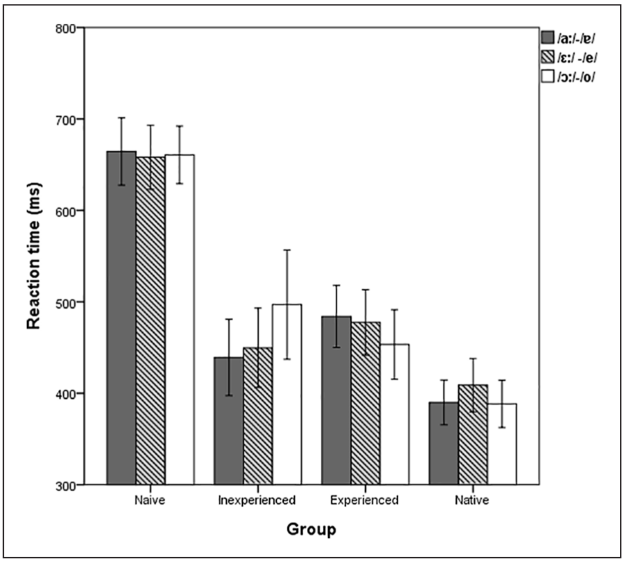

Figure 4. Reaction time on three Cantonese vowel contrasts. Error bars represent ±1 standard error.

Comparisons with models that included English exposure (measured in years, scaled and centered), Coda (/j, w, p, t, k, m, n, ŋ (ng)/), by-speaker mean RT for English contrasts, or the two-way interactions between Group and other factors, or models that included by-participant and by-word random slopes, showed that Coda ($\chi^2 = 31.494$, $p < .001$) and mean RT for English contrasts ($\chi^2 = 61.647$, $p < .001$), as well as the interaction between Group and Coda ($\chi^2 = 48.826$, $p = .001$), significantly improved the model. Models were compared via the anova() function in lme4, which was based on likelihood and deviance. Other factors, such as English exposure, Group × Vowel interaction, and random slopes, did not improve the model.

Table 6. Fixed effects in the final model for reaction time.

Predictor	Coef	SE	*t*-value	*p*
(Intercept)	60.71	57.37	1.06	0.29
Group: experienced	168.23	61.38	2.74	0.01*
Group: inexperienced	99.68	61.34	1.63	0.10
Group: naïve	69.82	62.16	1.12	0.26
Coda:k	142.58	55.61	2.56	0.01*
Coda:m	63.42	56.76	1.12	0.26
Coda:n	13.73	56.84	0.24	0.81
Coda:ng	45.68	49.53	0.92	0.36
Coda:p	123.77	56.69	2.18	0.03*
Coda:t	101.39	56.61	1.79	0.07
Coda:w	56.74	56.61	1.00	0.32
English RT	0.74	0.08	9.33	<0.0001***
Vowel: ee	−24.96	28.18	−0.89	0.38
Vowel: oo	−31.45	28.43	−1.11	0.27
Group: experienced × Coda:k	−158.60	59.62	−2.66	0.01*
Group: inexperienced × Coda:k	1.43	59.65	0.02	0.98
Group: naïve × Coda:k	68.35	56.55	1.21	0.23
Group: experienced × Coda:m	−171.60	68.94	−2.49	0.01*
Group: inexperienced × Coda:m	−82.02	68.94	−1.19	0.23
Group: naïve × Coda:m	30.39	65.49	0.46	0.64
Group: experienced × Coda:n	−25.47	69.09	−0.37	0.71
Group: inexperienced × Coda:n	−10.97	69.00	−0.16	0.87
Group: naïve × Coda:n	41.71	65.59	0.64	0.52
Group: experienced × Coda:ng	−79.73	56.44	−1.41	0.16
Group: inexperienced × Coda:ng	29.67	56.44	0.53	0.60
Group: naïve × Coda:ng	6.70	53.55	0.13	0.90
Group: experienced × Coda:p	−140.16	68.91	−2.03	0.04*
Group: inexperienced × Coda:p	−89.31	68.88	−1.30	0.19
Group: naïve × Coda:p	−1.11	65.39	−0.02	0.99
Group: experienced × Coda:t	−134.57	68.91	−1.95	0.05
Group: inexperienced × Coda:t	−17.41	68.82	−0.25	0.80
Group: naïve × Coda:t	49.08	65.25	0.75	0.45
Group: experienced × Coda:w	−41.77	68.85	−0.61	0.54
Group: inexperienced × Coda:w	4.95	68.85	0.07	0.94
Group: naïve × Coda:w	129.89	65.27	1.99	0.05*

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1.
lmer(reactionTime~Group+coda+EngRT+VowelContrast+Group:Coda+(1|ID)+(1|stimulusX), data=data).
p-values were based on Wald *z*-tests.
RT: reaction time.

The final model therefore included Group, Vowel, Coda, RT for English contrasts, two-way interaction between Group and Coda, and by-participant, by-word intercepts as predictors. Contrasts were treatment-coded, with the native group and the aa (/aː/–/ɐ/) pair with the /j/ coda set as reference levels.

The results (Table 6) suggest that for the reference condition, the RTs of the experienced group were significantly longer than those of the native group. For the native group, tokens with /p/ and

/k/ codas had significantly longer RTs than tokens with the /j/ coda; the vowel pairs did not differ significantly. An increase in mean RT for English contrasts by 1 ms predicted a 0.74 ms increase in RT for Cantonese contrasts. There were also a number of significant interaction terms with regard to Group × Coda conditions, indicating complex patterns in the differences among groups under varied coda conditions. Post-hoc pairwise comparisons were performed to explore the more general patterns related to group difference, vowel contrasts, and coda (e.g., patterns of group differences averaged over all vowel contrasts and coda conditions).

Post-hoc pairwise comparisons using lsmeans from the emmeans package (Lenth, 2018) with Tukey correction indicated that when averaged over different codas and Vowel (see Table 5 in the Online Supplemental Material), although RTs increased from native to experienced, to inexperienced, then to naïve, the differences were not statistically significant. The difference in RT between the native group and the naïve group approached significance ($p = 0.06$). Pairwise comparisons between groups, split by Coda conditions and averaged over Vowel conditions (see Table 6 in the Online Supplemental Material), showed that in almost all cases (except with the /m/ coda), the native group was the fastest to respond. Patterns for non-native groups varied depending on the coda condition. However, the differences in RT were significant only in a few cases: with the /j/ coda, the experienced group was significantly slower than the native group ($p = 0.03$). When the coda was /k/, the naïve group took a significantly longer time to respond than the experienced ($p = 0.03$) and the native ($p = 0.04$) groups. When the coda was /ŋ/, the inexperienced group was significantly slower than the native group ($p = 0.03$). When the coda was /w/, the naïve group was significantly slower than the native group ($p = 0.01$). The remaining comparisons were not statistically significant.

The lack of a significant interaction between Group and Vowel in the mixed-effects model suggests that all groups responded to the vowel contrasts in a similar fashion in terms of RT. Pairwise comparisons (see Table 6 in the Online Supplemental Material) revealed that RTs for the three vowel contrasts were not significantly different from each other.

4.2 Association between the perception of Cantonese and English

As discussed previously, English was the L2 of all the participants in this study, and L2 experience has been shown to influence the acquisition of an additional language (Tremblay, 2006). Moreover, the above linear mixed-effect analyses illustrate that the accuracy for English contrasts had significant effects on both accuracy and RT. Thus, further exploration of the association between the perception of Cantonese and English contrasts is warranted.

The mean accuracy of Cantonese and English vowel pairs across the four groups is shown in Figure 5. A Pearson product-moment correlation was run to determine the relationship between the perception accuracy of Cantonese and English vowel length pairs. A significant positive correlation was found between the accuracy of Cantonese and English for the non-native listeners ($r = .454$, $n = 67$, $p < 0.001$), while the correlation was not significant for native Cantonese speakers ($p > 0.05$), probably due to reaching a ceiling accuracy for the Cantonese materials. Specifically, the naïve group demonstrated a stronger positive correlation between the performance of Cantonese and English ($r = .566$, $n = 27$, $p < 0.005$), while the inexperienced group had a weaker correlation ($r = .480$, $n = 20$, $p < 0.05$). There was no significant correlation for the experienced group ($r = .169$, $n = 20$, $p > 0.05$). In other words, the association between English and Cantonese accuracy decreased in conjunction with Cantonese experience.

Figure 6 illustrates the mean RTs across the four groups in both Cantonese and English. There was a significant positive correlation between the RT of Cantonese and English for all non-native participants ($r = .473$, $n = 67$, $p < 0.001$), as well as for the native participants ($r = .694$, $n = 14$, $p < 0.001$). Focusing on the results for individual groups, the experienced group demonstrated the

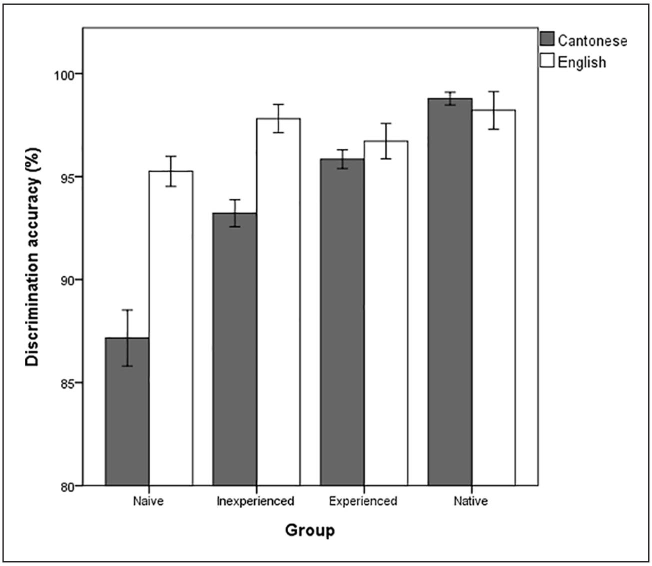

Figure 5. Mean accuracy of Cantonese and English vowel pairs. Error bars represent ±1 standard error.

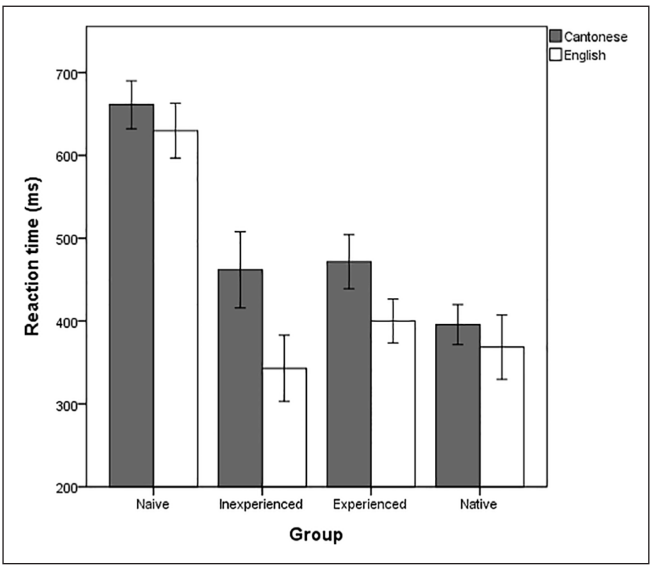

Figure 6. Averaged reaction time of Cantonese and English vowel pairs. Error bars represent ±1 standard error.

strongest correlation ($r = .836$, $n = 20$, $p < 0.001$), while the inexperienced and naïve groups had weaker correlations ($r = .793$, $n = 20$, $p < 0.001$; $r = .451$, $n = 27$, $p < 0.05$). For the non-native listeners, the correlation of the RT between English and Cantonese seemed to increase in conjunction with Cantonese experience.

In short, there were positive correlations between the perception of Cantonese and English vowel length contrasts in terms of both accuracy and RT for non-native listeners. The results indicated that English as a L2 was probably instrumental in the perception of Cantonese vowel length for the non-native speakers. Moreover, the correlation varied according to different stages of acquisition, which suggests that the degree of L2 influence may differ in the perceptual learning process of an additional language.

5 Discussion

This study investigated the perception of three Cantonese contrastive vowel length pairs, /ɛː/–/e/, /ɔː/–/o/, /aː/–/ɐ/, by Mandarin native speakers with varying degrees of exposure to Cantonese: naïve listeners with no Cantonese experience, inexperienced learners with less than one year of experience, and experienced learners with over five years of experience in a Cantonese-dominant environment. The differences and similarities of vowel length contrasts in Mandarin, English, and Cantonese enabled the investigation of the influence of the L1 and L2 on L3 phonological acquisition.

5.1 The role of Cantonese experience

The results revealed different perceptual performances among the four groups. The naïve participants had significantly lower accuracy and reacted significantly more slowly to the three vowel pairs than the other three participant groups. In terms of accuracy, the inexperienced group achieved significantly higher accuracy than the naïve group, but was significantly less accurate than the experienced and native groups. The experienced group also had significantly lower accuracy than the native group. In other words, the experimental results revealed that discrimination accuracy increased in conjunction with exposure to Cantonese.

Cantonese exposure may play different roles in the perception of non-native contrasts at various stages of the acquisition process. In particular, the significant difference between the naïve listeners and the learners with Cantonese experience clearly showed that language experience did indeed facilitate the discrimination of the non-native contrasts. Moreover, the difference in accuracy between the naïve and inexperienced groups was greater than the difference between inexperienced and experienced learners, which indicated that Cantonese exposure might play a more important role at the initial stage. Increased language experience seemed to be less important for the discrimination of non-native contrasts at a later stage of acquisition.

These findings suggest that the learners had probably developed the ability to discriminate the non-native contrasts rapidly at the initial stage of acquisition. Our results support the idea that the majority of perceptual learning may develop fairly early in the L2 acquisition process. Learners with as little as six to 12 months of L2 exposure have been found to perform differently from learners with zero to six months of experience (Best & Tyler, 2007; Flege, 1988; Riney & Flege, 1998). Some studies have also revealed that additional experience after the initial period does little to benefit L2 perception for most learners (Flege & Liu, 2001; Jia et al., 2006). Learners may be able to categorize non-native sounds with a limited amount of L2 experience. Once categories have been established, they are unlikely to change significantly even with more language experience, which may explain why the experienced participants did not perform significantly better than the inexperienced learners, despite their higher accuracy scores.

Although the learners achieved very high accuracy, they did not reach the same high rates as the native speakers. These results demonstrate the difficulty for non-native learners to achieve native-like perception, even with a relatively long period of Cantonese exposure (over five years). While it is possible that native-like perception could be developed with an even longer period of Cantonese exposure, it is also possible that L2 learners may never attain native-like perception (Schmid, Gilbers, & Nota, 2014). Further studies are needed to assess these possibilities.

Based on the overall high accuracy, it seems the discrimination of the Cantonese vowel length contrasts was not particularly difficult for native Mandarin speakers. One possible reason for the relatively high accuracy could be due to the experimental design. As the durations of the long and short vowels do not overlap (Lee, 1983, 1985; Shi & Liu, 2005), the sizable auditory durational differences may have helped the listeners to distinguish the vowel contrasts. However, the significant group differences suggest that this cannot be the only reason. Similar results across groups would be expected if the listeners were just performing a non-linguistic auditory task.

5.2 Perceptual learning of the three Cantonese vowel pairs

With regard to the effect of the L1 phonological system on the perception of non-native contrasts, the speech acquisition models discussed in Section 1 provided different predictions for the perception of Cantonese vowel length contrasts by native Mandarin speakers. Based on the L2LP model, the three pairs of contrasting Cantonese vowels were each predicted to be perceived as one single vowel in the initial stage, which would result in poor discrimination. The learners would gradually develop duration categories to cope with the new vowel durational distributions in Cantonese. Moreover, the learners would integrate several perceptual cues (such as F1, F2, and duration) to adjust the category boundaries. With sufficient Cantonese input, contrasts could be perceived in a more native-like manner. According to the SLM and the PAM, the /aː/–/ɐ/ and /ɛː/–/e/ pairs would be discriminated more poorly than the /ɔː/–/o/ pair.

However, the results did not conform to any of these predictions: all the participants, including naïve listeners, had relatively high levels of accuracy for the discrimination of the three vowel pairs. Some factors might have influenced the perceptual learning of the three contrastive vowel length pairs. Firstly, as mentioned above, the durational differences between the contrastive vowels would probably make the contrast quite salient. Neither the SLM nor the PAM addressed how duration could be incorporated into assimilation patterns; thus, their predictions were made based solely on vowel quality, ignoring durational differences. However, the large durational differences probably served as an important cue in non-native perception. As the vowel length pairs overlapped considerably in vowel quality, the non-native listeners probably relied more on the differences in vowel quantity than on vowel quality.

McAllister, Flege, and Piske (2002) proposed that a L2 contrastive category would be difficult for learners if it were based on a feature that was not used in their L1. They investigated the perception and production of Swedish vowel quantities by native speakers of American English, Latin American Spanish, and Estonian. Their results indicated that the native Estonians, whose L1 has quantity distinctions based on duration, were the most successful in learning the Swedish quantity contrast. The English and Spanish participants, who had no comparable quantity distinctions in their L1s, did not perform as well. Interestingly, however, the English participants did perform slightly better than their Spanish counterparts, despite the absence of pure duration-based quantity contrasts in English. McAllister et al. attributed this to the partial use of the temporal dimension in English, a feature not utilized in Spanish. Their results suggested that the duration feature can be difficult for L2 learners whose L1 does not use this feature.

Nevertheless, in our study, the Mandarin speakers did quite well despite lacking the duration feature in their L1, which seems to contradict McAllister et al.'s claim. One possibility could be that, although vowel length is not contrastive in Mandarin, durational difference is a secondary cue to Mandarin tones (Blicher et al., 1990; Liu & Samuel, 2004). Mandarin participants may not be totally unfamiliar with durational difference phonetically, even though the durational difference is not a phonological feature and is used as a secondary cue for a totally separate category (tone). Bohn's (1995) finding of naïve Mandarin listeners relying primarily on durational cues to differentiate /i/–/ɪ/ and /ɛ/–/æ/ in English corroborates our findings. He proposed that whenever spectral differences are insufficient to differentiate vowel contrasts because of L1 background (i.e., when listeners are desensitized), a general speech perception strategy would take over and listeners would use duration differences to differentiate these vowel contrasts.

The better performance of the English-speaking participants in McAllister et al.'s (2002) study does, however, echo our finding of the correlation between the English and Cantonese vowel length contrasts. Nevertheless, it is unclear whether the correlation is due to the experience in L2 English in particular, or if experience in any language would have a similar facilitation. Ideally, a control group with a L2 other than English would be needed, but such participants are very difficult to find. Further studies could investigate the particular roles that vowel quality and quantity play in the perception of Cantonese vowel length contrasts by learners with different L2 backgrounds.

Secondly, with regard to the learners with Cantonese exposure, the communicatively relevant pressure to detect the difference between minimally contrasting non-native words is a factor that might affect the perceptual learning of a new phonological category (Best & Tyler, 2007). In the case of the /aː/–/ɐ/ contrast, the two vowels contain a large number of minimal pairs, as they can occur in all coda environments. Many of the minimal pairs are often encountered in daily life, as in /kaːj/ "street" versus /kɐj/ "chicken." Thus, this contrast occurs frequently in the input data that the learners received. These factors may jointly increase the communicatively relevant pressure to learn to perceive /aː/–/ɐ/ contrast (Best & Tyler, 2007). It is possible that the better perceived length contrast between /aː/–/ɐ/ could in turn facilitate the perception of the other two less frequently occurring contrasts with an established length feature.

5.3 The relationship between English and Cantonese

Given that vowel length contrasts exist in Cantonese and (partially) exist in English, and that all the participants had over 10 years of English-learning experience, it is essential to consider whether L2 English plays a role in the perceptual learning of L3 Cantonese vowel length contrasts. The L2LP model makes predictions based on the presupposition that the learner has a single L1 and learns a single target L2. However, the Mandarin learners in this study had already learned English as a L2. It is possible that, when they duplicated their L1 categories for their Cantonese perception at the initial state, they also copied their L2 English categories to perceive Cantonese. Some recent L3 acquisition models, such as the CEM (Flynn et al., 2004) and the Typological Primacy Model (Rothman, 2010, 2015), also assume transfer from all previously learned languages to a L3. Given that vowel length contrasts exist (partially) in English, duration categories may have been developed when learning English. The existing duration categories are likely to facilitate the perception of Cantonese vowel length, even for the naïve group.

While there was no significant difference in the IELTS scores and the years of exposure to English among the three learner groups (Table 2), a significant positive correlation between Cantonese and English was found for all the non-native participants as a group in terms of both accuracy and RT. The statistical results suggested that the learners who performed better in English also performed better in Cantonese. The good perception of L2 English vowel length contrasts seems to have facilitated the perception of Cantonese vowel length contrasts. Alternatively, it could be interpreted that the listeners who performed better in Cantonese also performed better in English (see more discussion on this alternative interpretation below).

Different degrees of correlation between Cantonese and English were found in the three non-native groups, showing the varied effects of English perception on the perception of Cantonese vowel length contrasts at different stages. In terms of accuracy, the correlation between English and Cantonese decreased with an increase in Cantonese experience. The accuracy correlation was relatively strong for the learners with no Cantonese experience ($r = .566$, $n = 27$, $p < 0.005$), while there was no significant correlation for the learners with over five years of Cantonese exposure ($r = .169$, $n = 20$, $p > 0.05$). This result suggests that the L2 English perception may play a stronger role and become a source of transfer at the initial stage, which may generate more perceptual benefits for naïve listeners, which may have declined gradually as Cantonese experience increased. This result agrees with the findings of some morpho-syntactical studies that the transfer in the L3 acquisition stems from L2 properties when the L2 and the L3 are structurally similar (Leung, 2007; Montrul, Dias, & Santos, 2011). This result also supports the tenets of the CEM, which proposes that language acquisition proceeds in a gradual fashion, and that transfer can originate from any prior language (Flynn et al., 2004). Nevertheless, after being exposed to Cantonese for some time, the facilitative transfer from English seems to diminish over time, and learners may instead rely more on the Cantonese input data itself.

A similar trend was also observed in the RT results. The correlation between English and Cantonese increased with Cantonese experience. The experienced group had the strongest positive correlation ($r = .836$, $n = 20$, $p < 0.001$), while the inexperienced and naïve groups had weaker correlations ($r = .793$, $n = 20$, $p < 0.001$; $r = .451$, $n = 27$, $p < 0.05$). The results suggest that the learners who responded faster in Cantonese also responded faster in English. Furthermore, the data showed a significant group effect on the RT for the English vowel pairs. The naïve group responded significantly more slowly than the two learner groups, even though all the participants had relatively similar general English proficiency. Based on the results, we speculate that Cantonese experience may have also modulated the perception of English. Although very few phonetic-phonological studies have found that the L3 has facilitative influence on the L2, some morpho-syntactic evidence has indicated that L3 acquisition has a trickle-down effect that results in L2 acquisition (Cabrelli Amaro, 2017b; Hui, 2010). In addition to progressive transfer (L1→L3, L2→L3), it is also possible for the L3 to have a facilitative regressive transfer (L3→L1, L3→L2). That is, the L3 Cantonese experience or proficiency may increase the perceptual sensitivity to L2 English, particularly at a later stage of L3 acquisition. Furthermore, this result supports the predictions yielded by the Phonological Permeability Hypothesis (Cabrelli Amaro, 2013, 2017a), which proposes that a L2 acquired later is more vulnerable to L3 influence. Some previous studies have found that the degree of transfer could have a strong correlation with language use and relative proficiency (Cabrelli Amaro, 2017a; Llama & Lopez-Morelos, 2016). The Cantonese learners in our study were immersed in a Cantonese-dominated environment. They listened and spoke Cantonese on a daily basis, especially the experienced learners. In contrast, none of the participants ever had extensive exposure to English in a naturalistic setting. Although English is widely used in Hong Kong, the participants in this study reported that they only used English in some particular circumstances. This resulted in their L2 English having a relatively more vulnerable status compared to their L3 Cantonese. As Cantonese experience and proficiency increase, the degree of regressive transfer (L3→L2) may become stronger.

5.4 Limitations and further directions

This study addressed the perception of Cantonese vowel length contrasts by Mandarin speakers with a L2 English background. Some limitations in the present study provide directions for further investigations. Firstly, the particular roles of quantity and quality in the non-native perception and production of Cantonese vowel length contrasts remain unclear. Further studies could continue to examine the relative importance of quantity and quality in their perception and production. Secondly, this study only addressed L1 Mandarin speakers. It remains unclear whether the results reported in this study could apply more generally to speakers of other languages without vowel length contrasts. Future studies could include other L1 language groups, such as Spanish.

Alternative interpretation of the L2–L3 interaction found in this study is possible. The correlation between the L2 and L3 could also be a reflection of the individual differences in language learning abilities, in addition to inter-language interaction per se. Also, the correlation in accuracy is found more in the inexperienced group. This might be due to the fact that the experienced group has already reached the ceiling. Furthermore, the correlation between Cantonese and English perception abilities could be simply due to the fact that the learners have gained more non-L1 learning experiences, instead of the fact that English partially uses duration for vowel distinction. In order to tease apart these two possibilities, further study could test L1 English speakers with various exposures of L2s when acquiring a L3 with length contrast. The results could potentially distinguish the roles of more general language learning experience versus durational distinction in English vowels. It will also be useful to test a control group with a L1 other than Mandarin (e.g., L1 Spanish–L2 English–L3 Cantonese) to see whether Cantonese experience can indeed modulate the perception of English. Thus, more study is needed to corroborate our findings of L2–L3 interaction.

McAllister et al. (2002) proposed a feature hypothesis suggesting that L2 features not used to signal phonological contrast in L1 will be difficult for L2 learners to perceive, although Bohn (1995) demonstrated otherwise. McAllister et al.'s (2002) idea could be interpreted from a different perspective: would it be easier for learners who have the same features in their L1 to perceive the same L2 contrastive features? For example, Lee and Mok (2018) investigated the acquisition of Japanese vowel and consonant quantity contrasts by Cantonese learners. Vowel quantity contrasts are used in both Cantonese and Japanese, while consonant quantity contrasts are only used in Japanese. The results indicated that both beginning and advanced learner groups were able to distinguish short versus long vowels and consonants, but only the native groups enhanced the quantity contrasts in their own production in slower speech. In addition, the learner groups were able to lengthen the vowel before a geminate (i.e., a long consonant) in some cases. To test this possibility, further studies could include participants with different L1 backgrounds, for example, speakers of Spanish (without contrastive vowel length), Japanese (with contrastive vowel length purely differing in quantity), and English (with contrastive vowel length differing in both quality and quantity). A more thorough understanding of the transfer of vowel length contrasts could thus be obtained.

Acknowledgements

The authors would like to thank all the participants in this study.

Funding

This work was supported by the Department of Linguistics and Modern Languages of The Chinese University of Hong Kong.

ORCID iD

Peggy Pik Ki Mok https://orcid.org/0000-0002-9284-6083

Supplemental material

Supplemental material for this article is available online.

Notes

1. Some linguists have proposed an eight-phoneme system (/i y u ɛ œ ɔ ɐ a/), in which vowel quality plays an essential role in terms of Cantonese vowel contrasts, and duration differences are not recognized as a distinctive feature (Cheung, 1972; Liu, 1987, 2003; Wang, 1999; Wong, 1941; Yuan, 1960; Yue-Hashimoto, 1972).
2. Please note that all learner participants in the current study were adults.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bauer, L., & Warren, P. (2004). New Zealand English: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A handbook of varieties of English, 1: Phonology* (pp. 580–602). Berlin, Germany: Mouton de Gruyter.

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

Behne, D., Arai, T., Czigler, P., & Sullivan, K. (1999). Vowel duration and spectra as perceptual cues to vowel quantity: A comparison of Japanese and Swedish. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th international congress of phonetic sciences* (pp. 857–860).

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange, (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 171–203). Timonium, MD: York.

Best, C. T., McRoberts, G. W., & 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*, 345–360.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the mandarin tone 2/tone3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, *18*, 37–49.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1013–1016). Barcelona, Spain: Causal Productions.

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 279–304). Timonium, MD: York Press.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*, 201–251.

Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 55–80.

Cabrelli Amaro, J. (2012). L3 Phonology: An understudied domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 33–60). Amsterdam, The Netherlands: John Benjamins.

Cabrelli Amaro, J. (2013). The Phonological Permeability Hypothesis: Measuring regressive L3 influence to test L2 phonological representations. Unpublished doctoral dissertation, University of Florida.

Cabrelli Amaro, J. (2017a). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualis*, *21*, 698–717.

Cabrelli Amaro, J. (2017b). Regressive transfer in L3 syntactic development. In A. Hahn & T. Angelovska (Eds.), *L3 syntactic transfer: Models, new developments and implications*. Amsterdam, The Netherlands: John Benjamins.

Cabrelli Amaro, J., & Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of non-native phonological systems. *International Review of Applied Linguistics in Language Teaching*, *48*, 275–296.

Cabrelli Amaro, J., & Wrembel, M. (2016). Investigating the acquisition of phonology in a third language – a state of the science and an outlook for the future. *International Journal of Multilingualism (Special Issue on Advances in the Investigation of L3 Phonological Acquisition)*, *7*, 395–409.

Chen, H. (2011). *The acquisition of vowel length in Cantonese* (MPhil Thesis). The Chinese University of Hong Kong, Hong Kong.

Cheung, H. N. S. (1972). *Xianggang Yueyu yufa de yanjiu (Cantonese as Spoken in Hong Kong)*. Hong Kong: The Chinese University Press.

Cheung, K. H. (1986). *The phonology of present-day Cantonese* (PhD dissertation). University College London, UK.
Davenport, M., & Hannahs, S. J. (2010). *Introducing phonetics and phonology*. London, UK: Hodder Education.
Duanmu, S. (2007). *The phonology of standard Chinese* (2nd edition). Oxford, UK: Oxford University Press.
Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht, The Netherlands: LOT.
Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology and phonetics: Phonology in perception* (pp. 151–190). Berlin, Germany: Mouton De Gruyter.
Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *Journal of the Acoustical Society of America*, *129*, EL1–EL7.
Flege, J. E. (1988). Factors affecting degree of perceived FA in English sentences. *Journal of the Acoustical Society of America*, *84*, 70–79.
Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/ - /d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, *86*, 1684–1697.
Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 233–277). Timonium, MD: York Press.
Flege, J. E. (1999a). Age of learning and second-language speech. In D.P. Birdsong (Ed.), *New perspectives on the critical period hypothesis for second language acquisition* (pp. 101–132). Hillsdale, NJ: Erlbaum.
Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–355). Berlin, Germany: Mouton de Gruyter.
Flege, J. E., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.
Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527–552.
Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative enhancement model for language acquisition: Comparing adults and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, *1*, 3–16.
Foster, K. I., & Foster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124.
García, A. (2013). The effects of L2 proficiency on L3 phonological acquisition: A preliminary test of the L2 proficiency hypothesis. In E. Voss, S. J. D. Tai, & Z. Li (Eds.), *Selected proceedings of the 2011 second language research forum* (pp. 173–186). Somerville, MA Cascadilla.
Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge, UK: Cambridge University Press.
Hadding-Koch, K., & Abramson, A. S. (1964). Duration versus spectrum in Swedish vowels: Some perceptual experiments. *Studia Linguistica*, *18*, 94–107.
House, A. S. (1961). On vowel duration in English. *Journal of Acoustical Society of America*, *33*, 1174–1182.
Hui, B. (2010). Backward transfer from L3 French to L2 English production of relative clauses by L1 Cantonese speakers in Hong Kong. *Hong Kong Journal of Applied Linguistics*, *12*, 45–60.
Iivonen, A., & Harnud, H. (2005). Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *Journal of the International Phonetic Association*, *35*, 59–71.
Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, *119*, 1118–1130.
Kao, D. (1971). *The structure of the syllable in Cantonese*. The Hague, The Netherlands: Mouton.
Lee, A., & Mok, P. (2018). Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research*, *34*, 419–448.
Lee, T. H.-t. (1983). The vowel system in two varieties of Cantonese. *UCLA Working Papers in Phonetics*, *57*, 97–114.

Lee, T. H.-t. (1985). Guangzhou hua de yinzhi ji chang duan duili (Opposition of vowel quality and length in Cantonese). *Fangyan (Dialect)*, *1*, 28–38.

Lee, W., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association: Illustrations of the IPA*, *33*, 109–112.

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means. R package version 1.2.*3*. Retrieved from https://CRAN.R-project.org/package=emmeans

Leung, Y.-K. I. (2007). Third language acquisition: Why it is interesting to generative linguists. *Second Language Research*, *23*, 97–114.

Li, X., Huang, J., Shi, Q., Mai, Y., & Chen, D. (1995). *Guangzhou fangyan yanjiu (Studies on Guangzhou Dialects)*. Guangzhou, China: Guangdong People's Publishing House.

Lin, Y. H. (2007). *The sounds of Chinese*. Cambridge, UK; New York, NY: Cambridge University Press.

Liu, S. (1987). Guangzhou hua de chang duan yuanyin wenti (Problems on vowel length in Guangzhou Cantonese). In Editorial board of the Chinese Language Department of Nankai University (Ed.), *Yuyan yanjiu luncong* (*Essays on Language Research*) (p. 3). Tianjin, China: Tianjin People Press.

Liu, S. (2003). Guangzhou hua yuanyin yinwei de liangge wenti (Two problems on vowel phonemes in Guangzhou Cantonese). In Zhang Bohui (Ed.), *Di ba jie guoji yue fangyan taolun hui lunwen ji (Proceedings of the 8th International Conference on Yue Dialects)*. Beijing, China: Zhongguo Sheke Chubanshe (Chinese Social Science Publishers).

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when f0 information is neutralized. *Language and Speech*, *47*, 109–138.

Llama, R., & López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, *13*, 444–458.

Long, J. A. (2018). jtools: Analysis and presentation of social scientific data. R package version 1.1.*0*. Retrieved from https://cran.r-project.org/package=jtools

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*, 229–258.

Montrul, S., Dias, R., & Santos, H. (2011). Clitics and object expression in the L3 acquisition of Brazilian Portuguese: Structural similarity matters for transfer. *Second Language Research*, *27*, 21–58.

Odden, D. (2011). The representation of vowel length. In M. van Oostendorp, C. J. Ewen, E. H. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 465–490). Oxford, UK: Wiley-Blackwell.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *Journal of the Acoustical Society of America*, *89*, 2961–2977.

Qin, Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech*, *59*, 318–338.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*, *20*, 213–244.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *International Review of Applied Linguistics in Teaching*, *48*, 245–273.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*, 107–27.

Rothman, J. (2015). Linguistic and cognitive motivations for the typological primacy model of third language (L3) transfer: Considering the role of timing of acquisition and proficiency in the previous languages. *Bilingualism: Language and Cognition*, *18*, 179–190.

Schmid, M. S., Gilbers, S., & Nota, A. (2014). Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch-English bilingualism. *Second Language Research*, *30*, 129–157.

Shi, F., & Liu, Y. (2002). Xianggang yueyu chang duan yuanyin de tingbian shiyan (A perceptual experiment on vowel length in Hong Kong Cantonese). In P. Wuyun (Ed.), *Dongfang yuyan yu wenhua* (*Oriental Languages and Cultures*) (pp. 98–109). Shanghai, China: Orient Press.

Shi, F., & Liu, Y. (2005). Guangzhou hua yuanyin de zai fenxi (Revisit vowels in Guangzhou Cantonese). *Fangyan* (Dialects), *1*, 1–8.

Shi, F., & Mai, Y. (2003). Guangzhou hua chang [a] he duan [ɐ] yuanyin de tingbian shiyan (A perceptual experiment on the long vowel [a] and short vowel [ɐ] in Guangzhou Cantonese). *Zhongguo Yunwen Yanjiu* (Chinese Language Studies), *2*, 50–60.

Tremblay, M. C. (2006). Cross-Linguistic influence in third language acquisition: The role of L2 proficiency and L2 exposure. *Ottawa Papers in Linguistics*, *34*, 109–119.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*, 4–21.

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge, UK: Cambridge University Press.

Wang, Z. (1999). Guangzhou hua de yuanyin xitong (The vowel system in Guangzhou Cantonese). In Xu Liejiong (Eds.), *Gongxing yu gexing: Hanyu yuyanxue zhong de zhengyi* (Universals and Particulars - Controversies in Chinese Linguistics), 255–287. Beijing: Beijing Language and Culture University Press.

Warren, P. (2018). Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association*, *48*, 305–330.

Werker, J., & Tees, R. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*, 1866–1878.

Wong, S. L. (1941). *Yue Yin Jun Hui* (*A Chinese syllabary pronounced according to the dialect of Canton*). Shanghai, China: Zhonghua Book Company.

Wrembel, M. (2012). Foreign accentedness in third language acquisition: The case of L3 English. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 281–309). Amsterdam, The Netherlands: John Benjamins.

Wrembel, M., Gut, U., & Mehlhorn, G. (2010). Phonetics/phonology in third language acquisition—Introduction. *International Journal of Multilingualism (Special Issue on L3 phonology)*, *7*, 1–3.

Yuan, J. (1960). *Hanyu Fangyan Gaiyao* (*A synopsis of Chinese dialects*). Beijing, China: The Writing Reform Publishing House.

Yue-Hashimoto, A. O. (1972). *Cantonese (the Yue dialects I)*. Cambridge, UK: Cambridge University Press.

Zee, E. (1999). Chinese (Hong Kong Cantonese). In F. Nolan & J. Esling (Eds.), *Handbook of the International Phonetic Association* (pp. 58–60). London, UK: Cambridge University Press.

Zee, E. (2003). Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1117–1120).