The International
Journal of Speech, Language and the Law

IJSLL (PRINT) ISSN 1748-8885
IJSLL (ONLINE) ISSN 1748-8893

Article

How similar are the formants in the speech of bilingual speakers?

Peggy P. K. Mok, Holly S. H. Fung, Grace W. L. Cao and Chun Wai Leung

Abstract

This study investigated whether a stronger accent in the L2 of Cantonese-English bilingual speakers with high levels of English proficiency correlates with greater convergence of corresponding vowels (four monophthongs (/i a (ɑ) ɔ u/ and five diphthongs /aɪ aʊ eɪ ɔɪ oʊ (əʊ)/) between their two languages in identical phonetic environments. No consistent effect of accent rating was found in either acoustic or perceptual similarity measures. Much individual variation and vowel-specific patterns were observed. The results demonstrate that the formant patterns in one language or one vowel cannot predict those in another language or another vowel, even with highly comparable materials and speakers with a relatively strong accent. Possible reasons and implications for the lack of correlation between accentedness and vowel convergence are discussed.

KEYWORDS: BILINGUAL VOWELS, ACCENT, CONVERGENCE

Affiliations
Chinese University of Hong Kong, Hong Kong
emails: peggymok@cuhk.edu.hk hollyfung@link.cuhk.edu.hk gracecao@cuhk.edu.hk chunwai.leung@cuhk.edu.hk

Received: 25 April 2022 Accepted after revision: 29 May 2023

IJSLL VOL 30.1 2023 52–86 https://doi.org/10.1558/ijsll.22946

Introduction

There have been many studies focusing on the speech patterns of bilingual speakers in recent years for both adults (e.g. Kirkham and Nance 2017; Schertz, Kang and Han 2019) and children (e.g. Mok 2011; Yang and Fox 2017) alike, but very few studies have investigated bilingual speech patterns from a forensic phonetic perspective (Lo 2021a; Mok, Xu and Zuo 2015). Forensically, can patterns in one language predict those in another of the same speaker? A core research question is whether the two languages of the bilingual speakers would converge, or whether language-specific patterns are maintained for similar sounds in the two languages. Both consonants (especially voice onset time, VOT) and vowels have been investigated, but with mixed findings for both convergence and divergence. Previous studies have mainly focused on static features, e.g. VOT duration and midpoint formant measurements of monophthongs, while very few studies have examined dynamic features which can better illustrate both linguistic and individual differences. The current study continues such enquiry by exploring both static (monophthongs) and dynamic (diphthongs) features of corresponding sounds in identical phonetic environments in the two languages of Cantonese-English bilingual speakers with high levels of English proficiency. It examines the similarities between the formant patterns of the two languages in the same bilingual speakers. It is hoped that this study will shed some light on the validity of using cross-linguistic formant data in forensic comparisons.

Bilingual speech patterns

Bilingualism (or even multilingualism) is common in many parts of the world. Alongside this large bilingual population, there has been a wealth of research on bilingual speech patterns, especially over the last three decades. Many studies have shown the phonetic differences between monolingual and bilingual speakers. Bilingual speakers can maintain the contrasts in their two languages in a similar way to monolingual speakers of the respective languages, but their actual realisations can still differ from those of monolingual speakers (Guion 2003; Sundara, Polka and Baum 2006). However, not many studies have investigated bilingual speech patterns from a forensic perspective. How alike are the phonetic realisations of similar sounds in the two languages of bilingual speakers? Can patterns in one language predict those of corresponding sounds in another language? Does someone speaking with a stronger accent produce more similar vowels in both languages than someone speaking with a weaker accent? Such questions have been largely unexplored before.

Previous studies have also clearly demonstrated that the two languages of bilingual speakers can interact in different ways resulting in variegated patterns.

Cross-linguistic influence in both directions, L1-L2 and L2-L1, is found. Unsurprisingly, the effects of L1 on L2 are ubiquitous in studies of both early and (especially) late learners (e.g. Antoniou, Best, Tyler and Kroos 2011; Guion 2005; MacLeod, Stoel-Gammon and Wassink 2009). Reverse influence from L2 onto L1 is not as commonly reported, but it clearly exists (e.g. Chang 2012; Guion 2003; Sundara et al., 2006). The general consensus among the many previous studies is that the two languages of bilingual speakers are separate but non-autonomous systems which can interact in different ways.

How might the two languages interact? Among the prominent speech acquisition models, only the Speech Learning Model (SLM) makes explicit predictions about bilingual production patterns (Flege 1995). SLM focuses on sequential bilinguals who already possess an established L1 phonetic system. It posits that the processes and mechanisms for establishing new categorical representations for speech sounds remain intact and accessible throughout the life span, although the ability to create new categories decreases with age. SLM proposes that the phonetic elements that make up the L1 and L2 phonetic subsystems coexist in a common phonological space, and mutually influence each other. According to SLM, the more different an L2 sound is perceived to be from the nearest L1 sound, the easier it is for the learners to perceive and produce the L2 sound accurately, and the more likely it is that a new category will be formed. When a category is not established for an L2 sound due to its perceptual similarity to an L1 counterpart, the L1 and L2 categories assimilate, leading to a merged L1-L2 category, i.e. equivalence classification, even for experienced learners. In addition, Flege (1995) stated that the L2 phones were perceptually related to the closest positionally defined allophones in the L1.

Many of the above points are retained in the updated SLM-r (Flege and Bohn 2021), although the emphasis on age is replaced by the 'category precision' hypothesis which predicts that individuals who have more precise L1 phonetic categories will be better able to discern the differences between L1 and L2 sounds. Stronger emphasis is also placed on individual differences than on group differences. Given SLM's predictions, similar or corresponding sounds (e.g. vowels with a similar quality, transcribed with the same IPA symbols, occurring in identical phonetic environments) in a bilingual's two languages are very likely to be produced in the same way.

Many factors can modulate the interaction between the two languages. Age of acquisition and language dominance are two natural and important factors, as they affect both the quantity and quality of L2 input. Many studies have investigated the effects of these factors on bilingual speech. As mentioned above, SLM hypothesised that separate phonetic categories are more likely to be established by early than late bilinguals (Flege 1995, 1999). Guion (2003), Kang and Guion

(2006), MacKay, Flege, Thorsten and Schirru (2001) and MacLeod et al. (2009) all demonstrated that early bilinguals had an advantage over late bilinguals in developing monolingual-like categories for both stop consonants and vowels, even for phonemes that are very similar in the acoustic space. Flege, Schirru and MacKay (2003) compared the production of Canadian English /eɪ/ and Italian /e/ by Italian-English bilinguals differing in age of arrival and frequency of continued L1 use (as a measure of language dominance). They found that the early bilinguals with low English usage, i.e. ones that were more dominant in Italian, produced English /eɪ/ with exaggerated movement as dissimilation from Italian /e/, while the late bilinguals produced /eɪ/ with little movement as a merged category with Italian /e/. More recently, Amengual and Chamorro (2015) demonstrated that language dominance was a strong predictor in both the production and perception of Galician mid vowel contrasts (/ɛ/-/e/ and /ɔ/-/o/) in Spanish-Galician bilinguals. Only the Galician-dominant speakers maintained these contrasts in both production and perception.

While the effects of age of acquisition and language dominance on bilingual speech are well demonstrated, another relevant factor highly related to both age and dominance, i.e. accentedness, has received much less research attention. To be sure, foreign accent has been extensively studied for decades (e.g. Derwing and Munro 1997; Flege 1981; Levis, Derwing, and Munro 2020; Moyer 2004), but while many studies have investigated how various factors affect native listeners' perception of accentedness, far fewer have assessed how degree of accentedness correlates with bilinguals' production of similar contrasts in their two languages. In general, the earlier one starts to acquire a second language, the weaker the perceived accent is (Flege, Munro and MacKay 1995; Moyer 2004; Piske, MacKay and Flege 2001). Nevertheless, even very proficient bilinguals often retain a discernible foreign accent, especially late bilinguals or even heritage speakers (Kupisch, Barton, Hailer, Klaschik, Stangen, Lein and van de Weijer 2014). Thus, it seems reasonable to hypothesise that corresponding sounds in the two languages of bilingual speakers will be more similar to each other, or even merged, for speakers with a stronger accent in the L2.

Major (1987) demonstrated that global foreign accent significantly correlated with VOT produced by 53 Brazilian Portuguese learners of English: the weaker the perceived accent, the closer the VOT conformed to the American English norm. However, Lein, Kupisch and Van de Weijer (2016) compared the VOT produced by 14 French-German bilinguals who grew up in France and Germany with respect to the accent ratings for the minority languages of their childhood environment (i.e. German for those growing up in France, and French for those growing up in Germany). They failed to find any systematic relationship between the perceived foreign accent and VOT. There were speakers with a native-like accent and a deviant VOT, and speakers with a non-native accent but distinct VOT categories. Such findings highlight the importance of individual differences in bilingual speech patterns, which have important forensic implications. The bilingual speakers in Lein et al. (2016) were simultaneous bilinguals living in immersion or heritage language environments who may have behaved differently from L2 learners. Therefore, the relationship between accentedness and the similarity of bilingual speech, although reasonably hypothesised, is still unclear.

Forensic investigation of bilingual speech patterns

Compared to the vast literature on phonetic studies of bilingual speech patterns, there are far fewer studies investigating bilingual patterns from different forensic perspectives, e.g. speaker comparison for cases where the suspect is a bilingual speaker with samples in different languages. Forensic investigation of bilingual patterns can be grouped under several themes. Although earlier studies all showed that language familiarity affects speaker identification by monolingual listeners, only two speaker identification studies have involved bilingual speakers and bilingual listeners. Goggin et al. (1991) found that English-Spanish bilingual listeners performed similarly in both languages and did better than monolingual English listeners, but it was unclear if bilingual listeners could identify bilingual speakers in one language based on a memory of their voices in another language, as there was no cross-language condition in their study. Mok et al. (2015) investigated bilingual listeners' identification of voices of bilingual speakers in voice line-ups in two language conditions: same-language and cross-language. They found that Cantonese-English listeners perform significantly better in the same-language than in the cross-language condition, suggesting that some indexical information about speaker identity is language-specific. In addition, studies on automatic speaker recognition involving bilingual or multilingual data have consistently identified the 'cross-language problem', which causes a reduction in performance in many speaker-recognition systems (e.g. Lu, Dong, Zhao, Liu and Wang 2009). Nevertheless, Künzel (2013) demonstrated that, given the right settings and normalisation, cross-language voice comparisons can be equal to or even slightly better than same-language comparisons.

More relevant to the current study, there have been some studies comparing vowel production by bilingual speakers. Filled pauses, e.g. *uh* and *um*, are shown to be speaker-specific, and speakers are quite consistent in the vowels of filled pauses. Language-specific patterns are also found in filled pauses (de Leeuw 2007; Hughes, Wood and Foulkes 2016; McDougall and Duckworth 2017). Recently, de Boer and Heeren (2020) and Lo (2020) compared the filled pauses of bilingual speakers, with de Boer and Heeren investigating sequential English-Dutch bilinguals and Lo investigating simultaneous German-French bilinguals. Both studies found that the vowel formant patterns of filled pauses produced by the bilingual speakers were language-specific, and that the L1 or the dominant language influenced the L2 or the weaker language. Moreover, speakers also differed in the extent of cross-language influence, so there is a complex interplay between language- and speaker-specificity. De Boer and Heeren (2020) concluded that filled pauses may only be of limited use as a feature for forensic cross-language speaker comparisons.

On a more global level of vowel production, Lo (2021b) compared the long-term formant distribution (LTFD) of English-French bilinguals, as LTFD has been proposed to be an acoustic-phonetic speaker discriminant (Nolan and Grigoras 2005). Lo found systematic differences in LTFDs between the two languages in LTF2 to LTF4, together with a high degree of within-speaker consistency across languages. His findings show that the discriminatory potential of LTFDs is diminished across languages, similar to the findings for filled pauses above, although Lo contended that LFTDs are still potentially useful for speaker discrimination cross-linguistically, despite the deterioration in performance.

The above studies were not based on specific segmental features like vowel phonemes. These types of features are very common in linguistic phonetic studies of bilingual speakers, e.g. the phonetic realisations of the same or corresponding vowels in the two languages. Although not directly a forensic investigation, Zuo and Mok (2015) investigated speech similarity in the diphthong /ua/ within the two languages of eight pairs of Shanghainese-Mandarin bilingual identical twins. The results showed clear differences in formant dynamics between identical twins. The twin speakers were more similar to their respective twin's productions in their dominant language than in their non-dominant language. As their focus was on whether identical twin speakers could be distinguished by their diphthong production in the two languages, they only compared the diphthongs in the same language between the identical twin speakers, but did not compare the diphthongs between the two languages of the same speaker. Nevertheless, visual inspection of their formant patterns does suggest language-specific differences in formant dynamics of the comparable diphthong /ua/ in the two languages.

The outline of studies above demonstrates that bilingual data pose considerable challenges to forensic investigations in various aspects, although the other factors contributing to such challenges are under-researched. The Code of Practice of the International Association for Forensic Phonetics and Acoustics (IAFPA 2020) also highlights that 'Members should exercise particular caution with cross-language comparisons.' Lo (2020) and Zuo and Mok (2015) show that language dominance is an important factor influencing similarities between the two languages of bilingual speakers. However, it is not clear how the highly related concept of accentedness may affect bilingual forensic investigation. Thus, it is worth examining how accentedness may affect the similarities of bilingual speech patterns in specific vowels across the two languages for a better understanding of the effects of bilingualism in forensic investigation.

The present study

The present study investigated whether degree of accentedness is correlated with the similarity (convergence) of corresponding monophthongs and diphthongs in identical phonetic environments in the two languages produced by Cantonese-English bilingual speakers. We hypothesised higher similarity for speakers with a stronger accent in their L2 English, because a stronger accent means that the L1 has a stronger influence on the L2, thus rendering the L2 more similar to the L1. The relationship between accentedness and vowel realisations is a bit of a chicken and egg situation. Non-native vowel realisation is an important factor contributing to the impression of foreign accent (Chan and Hall 2019; Sidaras, Alexander and Nygaard 2009). Nevertheless, foreign accent is a global impression involving many other speech features, such as consonants and prosody (e.g. Anderson-Hsieh, Johnson and Koehler 1992). Thus, most studies have obtained accent ratings using sentences produced by speakers. It is not always clear which feature is more prominent in such impressions, neither is the relative weighting of different features. Moreover, the influence could be vowel-specific, as shown by Chan and colleagues (Chan and Hall 2019; Chan, Hall and Assgari 2017). The present study explores this relationship, hypothesising a generally higher vowel similarity with a stronger accent, while acknowledging the possibility that the effect may not be uniform.

While language proficiency, age of exposure and dominance are three factors which are closely related to accentedness, in practice they are often highly correlated (Luk and Bialystok 2013). In order to focus on accentedness alone, we controlled for the three factors by using Cantonese-English bilingual university students in Hong Kong who were dominant in Cantonese with varying degree of L1 accent in their English. They had all started learning English in a school setting from an early age and had a high level of proficiency in English (see more in the Method section). By doing so, we hoped that we could isolate accentedness for our study.

Studies have shown that there is an identifiable form of English used in Hong Kong with observable phonetic as well as other linguistic features (McArthur 2002; Setter, Wong and Chan, 2010). Setter et al. (2010) described the typical Hong Kong English (HKE) accent as being spoken by educated individuals who have not spent much time outside of Hong Kong, and it is clearly influenced by features of Cantonese phonology. Bolton and Kwok (1990) provided a brief overview of some phonological features, while Hung (2000) and Deterding, Wong and Kirkpatrick (2008) gave a detailed and thorough account of the segmental phonetics and phonology of HKE, in terms of it being an emergent new variety of English. A detailed overview of HKE phonology and grammatical features can be found in Setter et al. (2010).

Some common features of HKE vowels were identified by the studies above. The HKE vowel inventory is simpler than those of British or American English as there is no clear tense/lax contrast for vowel pairs like /iː ɪ/ and /uː ʊ/. HKE speakers find [æ] difficult to distinguish and often use [ɛ] instead, which is a vowel phoneme in Cantonese. There is only a small difference between [ɒ] and [ɔː] which are likely merged. The [uː] vowel is fronted like in British English. The onset of the GOAT diphthong [oʊ] in HKE is further back than that of British English [əʊ], while the FACE [eɪ] diphthong is similar in quality to the one in British English. The centring diphthongs /ɪə/, /ɛə/ and /ʊə/ in open syllables are often produced as two distinct syllables. Finally, reduced vowels in British or American English are often produced as full vowels in HKE. Readers can refer to the above studies for more details of HKE phonology.

In terms of the development of new varieties of English, Schneider (2007) put HKE at Phase 3 in his Dynamic Model of the Development of Postcolonial Englishes, i.e. at 'Nativisation'. This indicates that HKE is in a state of 'cultural and linguistic transformation' (Schneider 2007: 40) and that there is a move towards independence from the distant country of origin politically, linguistically and culturally. It should be pointed out that, as HKE is an emerging new variety of English, its phonological features are not always stable, and there can be wide variation among its speakers, even among educated (university or postgraduate level) high-proficiency speakers. For example, Sewell and Chan (2010) examined the consonantal features of 25 such HKE speakers, and found individual variation in their consonantal realisations, although systematicity could also be found.

Individual variation can also be found in terms of how strong the accent is. Some speakers have a typical HKE accent, while some speak with an accent more akin to the traditional varieties of English (e.g. British, American or Australian). Despite the long colonial history of Hong Kong before 1997, HKE speakers nowadays exhibit a mix of English accents. The influence of non-British accents is increasing, particularly American English, probably because of media exposure and the Native-speaking English Teacher (NET) Scheme, which since 1998 has brought in many native English teachers from different parts of the world to all government-aided primary and secondary schools in Hong Kong. While Bolton and Kwok (1990) reported that about 10% of their respondents preferred an American accent, Deterding et al. (2008) found that 40% of their speakers exhibited some clear American influences (e.g. using [æ] instead of [ɑ] for the word *last*) in their speech, although the influences are not necessarily consistent (e.g. a varying degree of r-colouring). It is not uncommon to find some features of both British and American accents in the speech of a single individual.

A recent study by Chen, Ng and Li (2012) is highly relevant to the present work. They compared 11 English monophthongs /i, ɪ, e, ɛ, æ, ʌ, u, ʊ, o, ɔ, ɑ/ produced by Cantonese-English bilingual speakers, with college education or above, with those produced by American English monolingual speakers, testing whether familiar or new sounds were produced more accurately by either group. Two native speakers of American English listened to the words produced by the bilingual speakers and rated them on a seven-point scale for foreign accent. The authors used F1 and F2 formant frequencies and the derived Euclidean distance to compare the vowel production of the two groups of speakers, and found that the bilingual speakers, especially female speakers, exhibited more compact vowel spaces than the monolingual English speakers did, and there were significant differences in terms of formant frequencies between the two groups. The authors suggested that the higher accent rating (i.e. more accented) for the female bilingual speakers may account for the vowel production difference compared with native English speakers, although there was no statistical analysis to confirm this observation. They concluded that the bilingual speakers produced familiar vowels more accurately than new vowels. The Chen et al. (2012) study only compared the English vowels produced by Cantonese-English bilingual speakers with those produced by native English speakers, so it is still unclear how similar the vowels are in the two languages of bilingual speakers.

Few previous studies have examined bilingual diphthong realisation. Diphthongs involve formant dynamics which change during the time-course of a vowel. These formant dynamics allow much freedom for speaker-specific behaviour in production, e.g. variability in the amount of spectral change and the rate of change. Diphthongs are useful for forensic comparisons (McDougall 2004, 2006). Chanethom (2011) examined the phonetic interactions in the production of English /aɪ/ and French /aj/ by four French-English bilingual children to see whether they maintained separate categories. The results showed distinct phonetic patterns for each category with individual variation. Children with reduced French input showed more overlapping acoustic properties. As mentioned above, Zuo and Mok (2015) investigated speech similarity in the diphthong /ua/ in the two languages of Shanghainese-Mandarin bilingual identical twins. Their results showed clear differences in formant dynamics between identical twins. These two studies demonstrate that it is valuable to include diphthongs in our investigation of bilingual language convergence, and that language dominance is an important factor that should be controlled.

In summary, the present study examines phonetic convergence in corresponding monophthongs and diphthongs in identical phonetic environments between the two languages of Cantonese-English bilingual speakers who have high levels of English proficiency but are dominant in Cantonese. The Chen et al. (2012) study reviewed above, although highly relevant, differs from our study in several important ways. First and foremost, the aims of our studies are different. Chen et al. (2012) were concerned whether Cantonese-English bilingual speakers could produce familiar or new vowels in English similarly to native English speakers, while our focus is on whether bilingual speakers with a stronger accent have more convergence between the two languages. Thus, they only compared the English vowels produced by bilingual and monolingual speakers without collecting any Cantonese vowels, while we compare the corresponding vowels in the two languages of the Cantonese-English bilingual speakers and evaluate their similarity, i.e. no comparison with native English speakers is made. In the same vein, although their findings give some support to our hypothesis that a stronger accent can affect how the vowels are realised, their main point was that a stronger accent would result in more deviation from the native speakers' targets, while our aim is to evaluate whether a stronger accent is correlated with stronger mutual convergence. Deviation from native targets does not necessarily entail more convergence between the two languages, as deviation can be in different directions. Chan and Hall (2019) also showed that deviation in all directions did not have the same effect on accent perception. Finally, Chen et al. (2012) only included monophthongs, while our study includes both monophthongs and diphthongs for a more comprehensive investigation. It is hoped that our study can better illustrate the interaction between the two languages of bilingual speakers.

In our study, emphasis will be placed on individual variation in the realisations of both types of corresponding vowels. Many previous studies on bilingual vowel production have not controlled for the phonetic environments in the two languages in which the target vowels/diphthongs occur, probably because of phonotactic and/or other practical constraints. In order to ensure high comparability of the materials across the two languages, we use only real words with identical phonetic environments in terms of segmental context and syllable structure, e.g. Cantonese 靠 'depend' [kʰaʊ] and English *cow* (see more in the Method section). Thus, the materials are as similar as they could be in the two languages. In addition, both acoustic and perceptual similarities of the corresponding vowels are assessed. It is of interest to see if formant similarity is higher across the two languages for the bilingual speakers with a stronger L1 accent than for those with a weaker L1 accent.

Method

Production

Speakers

Ten Cantonese-English bilingual speakers (five males; aged 18–25), who were university students at Chinese University of Hong Kong, were recorded. All of them were native Cantonese speakers who grew up in Hong Kong. They had started learning English in early childhood in local school settings, and were considered to have high proficiency in English based on their IELTS (or IELTS-equivalent)[1] scores (range: 7–8; mean: 7.45). They were all Cantonese-dominant. They were chosen because they exhibited varying degrees of Cantonese accent in their English as judged auditorily by the authors. More information about their accent ratings is given below.

Materials

The target items consisted of real monosyllabic words in Cantonese and English. These words contained either a monophthong (/i a (ɑ) ɔ u/) or a diphthong (/aɪ aʊ eɪ ɔɪ oʊ (əʊ)/) in open syllables (see the Appendix for the full list). The target vowels were chosen based on several criteria. They are transcribed with the same (or closely related) IPA symbols in the two languages; they are phonemes; they can appear in identical phonetic environments (all open syllables to minimise formant transitions) in the two languages; and they are in different positions of the vowel quadrilateral. For easy reference, they are called the 'corresponding vowels' in this study. Altogether 19 monophthongal words in Cantonese, 21 monophthongal words in English and 31 diphthongal words in each language were included. As mentioned above, there is a mixture of British and American influences in HKE speakers. Therefore, two more English monophthongal words for /a (ɑ)/ with no 'r' in the spelling were included because of the uncertainty of whether or not the bilingual speakers had a rhotic accent. All the words were carefully chosen for identical phonetic environments in the two languages as far as possible. Phonotactic constraints can explain the uneven number of words selected for each monophthong and diphthong.

Given that Cantonese is a tone language, tone was also taken into consideration while selecting the Cantonese words: only words with level-tones were included. Tone 1 (the high-level tone ˥) was chosen whenever possible. When a real word was not available in Tone 1, another real word in either Tone 3 (the mid-level tone ˧) or Tone 6 (the low-level tone ˨) was chosen.

The target words were embedded in short carrier phrases: ŋɔ tʊk ___ tsi (我讀 ___ 字 'I read ___ word') in Cantonese; 'I read ___ to you now' in English.

The Rainbow Passage (Fairbanks 1960: 127) was also included for evaluating the degree of Cantonese accent in the speakers' English.

Procedure

All recording sessions were conducted individually in a sound-proofed recording booth at the Chinese University of Hong Kong. The materials were recorded using a portable solid-state recorder with a built-in microphone placed approximately 30 cm away from the speaker's mouth with a sampling frequency of 44,100 Hz. Breaks were given whenever needed.

The Cantonese and English sessions were conducted consecutively using a similar procedure. Instructions were given by the experimenter in Cantonese for the Cantonese session. The speaker was given a list of the Cantonese items arranged in a randomised order divided into four blocks. They were asked to familiarise themselves with the items before the recording. Three repetitions of the whole list were recorded. Any problematic tokens were recorded again at the end of the session.

A break was scheduled between the Cantonese and English sessions. The experimenter switched to using English to communicate with the speakers for the English session. They were first given the Rainbow Passage printed on a piece of paper and were allowed some time to read the passage. They were then recorded reading the Rainbow Passage naturally once. Next, the speakers were given the list of English items arranged in a randomised order, also divided into four blocks. They were also recorded reading the list of English items three times. Any problematic tokens were recorded again at the end of the session. After finishing the English list, the speakers were recorded reading the Rainbow Passage again.

Data analysis

For each speaker, there were 306 tokens altogether ((19 Cantonese + 21 English monophthongal words + 31 Cantonese diphthongal words + 31 English diphthongal words) × 3 repetitions). All tokens were manually segmented in Praat. For monophthongal words, the F1 and F2 frequencies at the midpoint of each vowel token were measured using a Praat script and checked manually. The measured F1 and F2 frequencies were averaged separately across the tokens of the same monophthong irrespective of the onset consonants in each of the two languages. These values were then used for constructing the speakers' vowel spaces in the two languages and for further statistical analyses. Euclidean distance was calculated between the mean formant frequency in Bark of a Cantonese vowel and its paired English vowel for each corresponding monophthongal pair.

For diphthongal words, the F1 and F2 frequencies were measured at ten equidistant points within the interval of each vowel token using FormantPro (Xu and Gao 2018), excluding the start and end points. All automatically generated formant values were checked manually for anomalies. The measured F1 and F2 frequencies (in Bark) of all tokens of the same diphthong in each language, irrespective of onset consonants, were analysed using smoothing spline analysis of variance (SS ANOVA) (Gu 2014) and generalised additive mixed models (GAMMs).

Accent rating

Participants

To investigate the potential relationship between vowel patterns and the degree of accentedness, 46 native Cantonese-speaking students (14 males; aged 17–27) at the Chinese University of Hong Kong rated the ten bilingual speakers for the level of Cantonese accent in their L2 English in an accent rating task. The participants were also bilingual in Cantonese and English and had grown up in Hong Kong, but their English proficiency varied more than that of the ten speakers.

Materials and procedure

As the study's aim was to evaluate how strong the Cantonese accent was in the speakers' English, the participants were asked to rate how likely the speakers were to be from Hong Kong based on two utterances (each 8 to 14 words long) from their Rainbow Passage recordings, using a nine-point scale (a higher rating means more likely). It would not have been appropriate to ask native English speakers to do the accent rating because our aim was not to evaluate how similar the bilingual speakers were to native English speakers, but how strong their Cantonese accent was. The selected utterances produced by the ten speakers were all different. To widen the spectrum of accents, the voices of two native Canadian English speakers and two native Hong Kong Cantonese secondary school students with a strong Hong Kong accent (reading the North Wind and the Sun passage) were also included. There were altogether 28 test utterances (2 utterances × (10 + 4 speakers)).

The participants did the accent-rating task individually in a language laboratory. Each participant was seated in front of a computer screen on which an online response form was presented. To familiarise the participants with the task, a practice trial was administered prior to the actual trials. In the practice trial, utterances produced by one Canadian speaker and one Hong Kong secondary school student were used as the stimuli to provide the participants with the entire range of accentedness they may expect. Each speaker's two utterances were played back-to-back via headphones, once only for each utterance. The participants were asked to write down what they had heard and rate their accent.

By-speaker means of the accent ratings were calculated. The means were taken as indices of how local the non-native English accent of each speaker was to native Cantonese listeners. A higher rating indicates a more prominent accent typical of a native Cantonese speaker in Hong Kong.

Similarity judgement

Participants

To investigate the perceptual similarity between the Cantonese and English vowels produced by the ten bilingual speakers, 35 native Cantonese listeners (12 males, aged 17–24; 34 of whom also did the accent-rating task) participated in a similarity judgement task. They listened to pairs of corresponding Cantonese and English words (e.g. 佳 [kaɪ˥] 'good' vs. guy) produced by the ten speakers via headphones in a language laboratory and gave a similarity rating for each pair on a five-point scale. A higher number means the words were considered to be more similar.

Materials

Given practical constraints, a number of compromises were made. First, only the first repetitions of the Cantonese and English words by each of the ten speakers were used in this judgement task. Secondly, there were two versions of this experiment, each containing around half of the items produced by all the ten speakers; there were 280 Cantonese-English word pairs in each version (including 60 word pairs appearing in both versions to check for consistency). This resulted in 560 word pairs altogether ((17 diphthongal pairs + 11 monophthongal pairs) × 10 speakers × 2 versions). Seventeen participants (five males) did the first version and 18 participants (seven males) the second version. Third, each word pair was rated once only, i.e. there was no repetition.

Experimental design

The judgement task was run in Praat's ExperimentMFC environment. For each judgement, the participants heard a pair of corresponding words in Cantonese and English produced by one of the ten speakers. They were reminded to focus on the vowel quality and ignore any pitch differences. A pair of sounds with an ISI of 500 ms was played in a randomised order via headphones. After the playing of both words, a five-point scale appeared on the screen (1 = extremely dissimilar; 5 = extremely similar), and the participants were required to indicate how similar/dissimilar the vowels of the two sounds were by using the number pad on the keyboard or clicking on the scale with the mouse. The participants could replay the pair of sounds as many times as they wished before making a judgement. Twenty practice trials were given prior to the actual experiment. Breaks were allowed after every 40 trials. Similarity ratings were averaged across word pairs for the same vowel produced by the same speaker.

Results

Formant frequency patterns

Figure 1 shows the Cantonese and English vowel spaces of the monophthongs /i a (ɑ) ɔ u/ for each speaker plotted based on the mean midpoint F1 and F2 frequencies in Bark.[2] The first letter in the speaker ID represents the sex of the speaker, where F stands for female and M for male. These plots show that the vowel /i/ was the most similar between the two languages across speakers out of the four corresponding monophthong pairs. As noted in previous HKE literature, the /u/ in HKE is generally fronted, but speakers differed in how

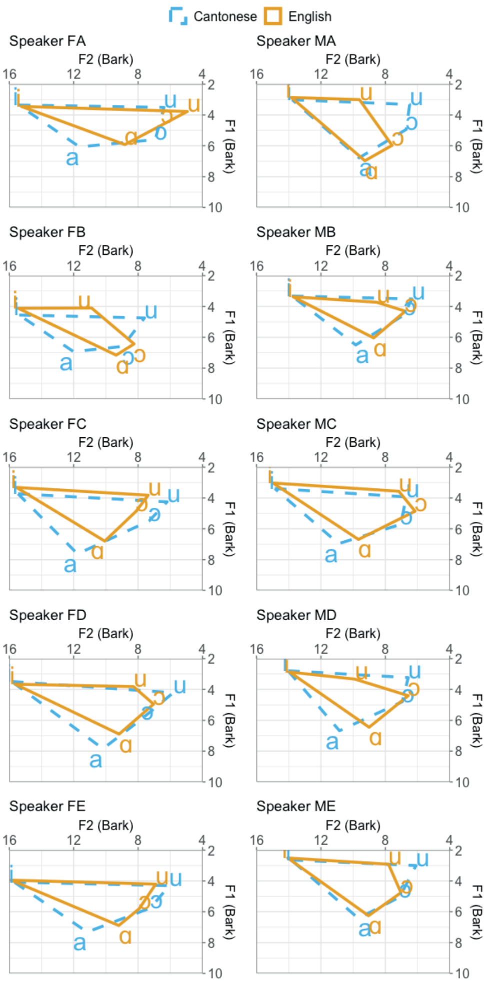

Figure 1: Mean F1 and F2 formant frequencies (in Bark) in Cantonese and English for each bilingual speaker.

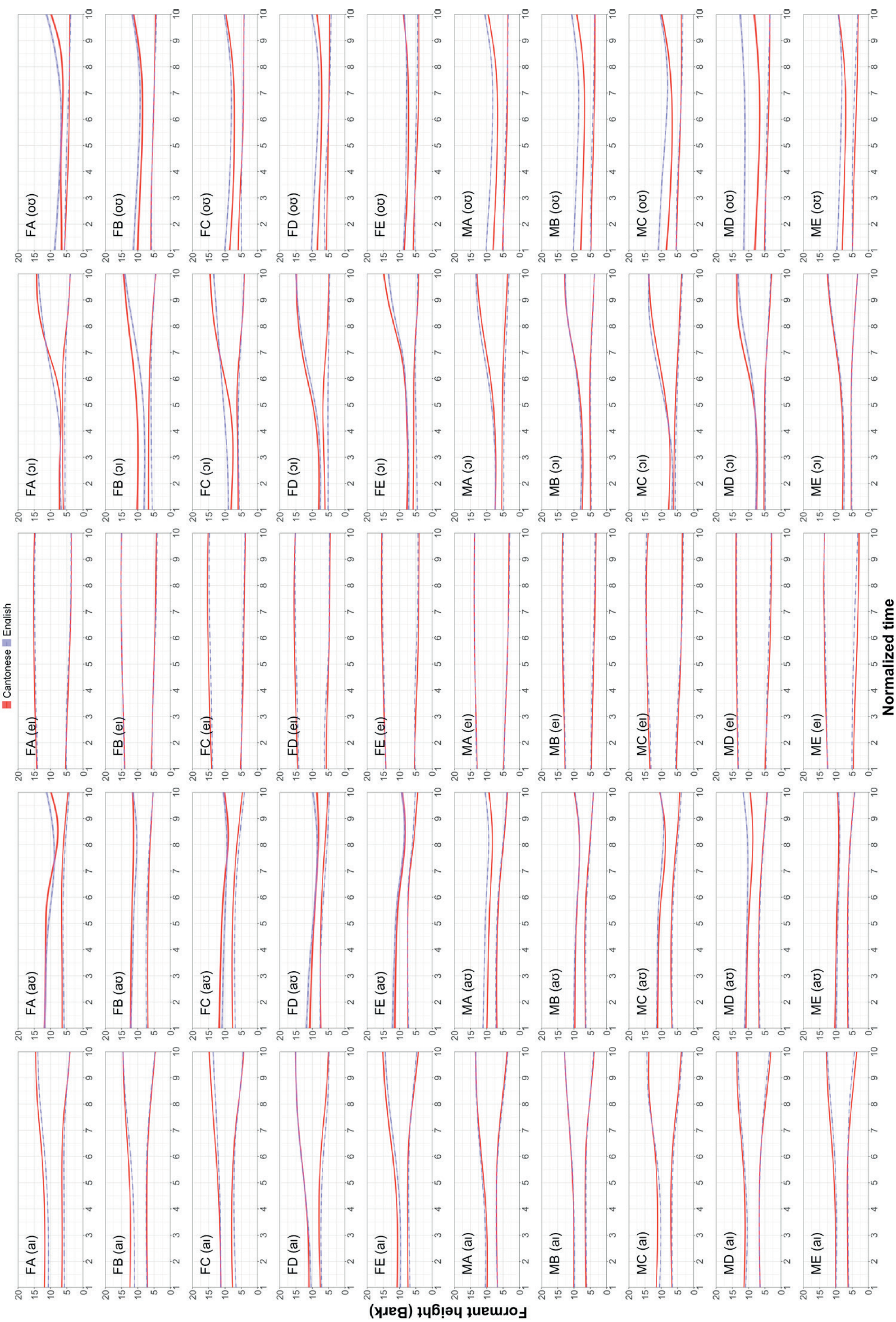

Figure 2: SS ANOVA results of the F1 and F2 formant frequencies of the five diphthongs in Cantonese and English produced by each speaker.

fronted the English /u/ was in relation to their Cantonese /u/, with Speaker FA having the opposite pattern. The Cantonese /a/ was fronter and lower than the English /ɑ/ for most speakers, while the two vowels were much more similar for Speakers MA and ME. Likewise, speakers differed in the frontness and height of these two vowels. The data show that there were some common patterns among the corresponding monophthongs in the two languages across speakers, but individual differences were also noticeable. Overall, none of the speakers exhibited very similar Cantonese and English vowel spaces. Convergence/divergence in one monophthong cannot predict that of another monophthong. The patterns are vowel-specific.

Each column in Figure 2 shows the F1 and F2 formant trajectories in Bark generated by SS ANOVA for each of the five diphthongs (/aɪ aʊ eɪ ɔɪ oʊ (əʊ)/) produced by the ten bilingual speakers (each row shows data from the same speaker). The output of the SS ANOVA was individual graphs for each diphthong produced by each speaker, showing an estimation of the diphthongs' trajectories based on ten value points. Significant difference is indicated by the absence of overlap between two curves. Although both F1 and F2 frequencies of each diphthong in both languages produced by each speaker are given in the same graph for ease of visual presentation, it should be noted that two separate analyses were performed for each vowel, one for F1 and one for F2. F1 and F2 contours should not be compared with each other.

The diphthong /eɪ/ exhibited the highest similarity between the two languages, while the diphthong /oʊ (əʊ)/ exhibits noticeable differences, especially in F2, for most speakers, as noted by previous studies on HKE discussed above. This could be explained by the different vowel qualities of the corresponding diphthongs in the two languages, as they were also transcribed with different IPA symbols. The speaker MB had very similar formant realisations in the two languages for /aɪ aʊ eɪ ɔɪ/ but a markedly lower F2 for Cantonese /oʊ/ than for its English counterpart. Some speakers like MA and ME had quite similar formant dynamics in the two languages (the two curves are nearly parallel) although the formant frequencies were still significantly different (with the two curves not overlapping), while other speakers like FA and FC had different formant dynamics (i.e., the shape of the two curves were different) (see Figure 2).

Accent rating and perceptual similarity judgement

Recall that the ten bilingual speakers had been rated by 46 Cantonese listeners for accentedness in their English, i.e. how likely it was that the speaker was from Hong Kong, on a scale of 1 (extremely unlikely) to 9 (extremely likely). By-speaker means and standard deviations of the accent ratings are given in Table 1. Figure 3 is a violin plot of the accent ratings for easy reference. The corresponding Cantonese and English words have also been rated auditorily for their perceptual

Table 1: The mean accent rating and perceptual similarity judgement of the nine vowels of the ten bilingual speakers (with standard deviations)

Speaker	Accent rating (1–9)	Monophthong (1–5)				Diphthong (1–5)				
		a (ɑ)	i	ɔ	u	aɪ	aʊ	eɪ	ɔɪ	oʊ (əʊ)
FA	5.39 (2.09)	2.9 (1.42)	3.85 (1.24)	3.48 (1.3)	3.21 (1.4)	3.36 (1.35)	3.75 (1.31)	4.05 (1.16)	3.6 (1.31)	3.8 (1.16)
FB	6.54 (1.85)	2.63 (1.34)	3.34 (1.4)	3.62 (1.4)	3.07 (1.5)	3.43 (1.34)	3.56 (1.49)	4.22 (1.06)	3.11 (1.41)	3.97 (1.18)
FC	5.67 (2.2)	2.1 (1.28)	3.47 (1.39)	2.5 (1.3)	2.83 (1.42)	2.85 (1.47)	2.64 (1.44)	3.63 (1.29)	3.16 (1.29)	2.98 (1.36)
FD	3.57 (2.3)	2.82 (1.38)	4.17 (1.06)	2.94 (1.34)	2.17 (1.43)	2.69 (1.36)	3.49 (1.46)	4.25 (1.07)	2.5 (1.39)	3.65 (1.29)
FE	3.7 (1.99)	3.13 (1.35)	3.47 (1.37)	3.79 (1.24)	3.7 (1.3)	3.41 (1.41)	3.64 (1.35)	3.75 (1.28)	2.63 (1.44)	3.34 (1.41)
MA	3.57 (2.12)	3.59 (1.29)	3.54 (1.39)	3.24 (1.33)	2.46 (1.3)	3.11 (1.34)	3.75 (1.31)	4.14 (1.13)	3.17 (1.25)	3.3 (1.23)
MB	5.3 (2.21)	3.97 (1.22)	4.11 (1.16)	4.14 (1.1)	3.2 (1.34)	4.16 (1.1)	4.38 (0.94)	4.11 (1.13)	4.2 (1.16)	3.52 (1.24)
MC	5.76 (2.29)	3.46 (1.43)	3.64 (1.32)	2.82 (1.47)	2.9 (1.35)	3.62 (1.34)	3.88 (1.21)	4.09 (1.13)	3.71 (1.19)	2.98 (1.37)
MD	3.46 (1.96)	2.99 (1.33)	3.61 (1.36)	2.49 (1.52)	1.87 (1.3)	3.49 (1.32)	3.21 (1.36)	3.79 (1.26)	3.8 (1.31)	2.31 (1.24)
ME	5.98 (2.29)	3.62 (1.37)	3.98 (1.26)	4.2 (1.09)	2.11 (1.19)	3.74 (1.27)	3.93 (1.17)	4.23 (1.06)	3.69 (1.32)	3.21 (1.35)

similarity by 35 Cantonese listeners on a scale of 1 (extremely dissimilar) to 5 (extremely similar). The by-speaker means and standard deviations of the perceptual similarity judgements are also given in Table 1.

Table 1 and Figure 3 show that the speakers FB and ME had the strongest accent, and the speakers FD and MD had the weakest accents for the two groups of speakers. If we compare their patterns in Figures 1 and 2, it is not clear that the strength of their accents had any consistent effect on their vowel spaces or formant trajectories. For the speaker MB, who had similar formant trajectories between the two languages for four diphthongs, his accent rating (5.30) was not too high; while the speaker FA with a very similar accent rating (5.39) exhibited quite different formant dynamics in the two languages.

In addition to the above descriptive reports, further analyses were conducted to assess the relationship between accent rating, perceptual similarity and formant patterns. For *monophthongs,* the Euclidean distance between a Cantonese vowel and its corresponding English vowel was calculated for each participant. The mean F1 and F2 values (in Bark) of Cantonese vowels were calculated for each participant. The reason for averaging the formant values of Cantonese vowels

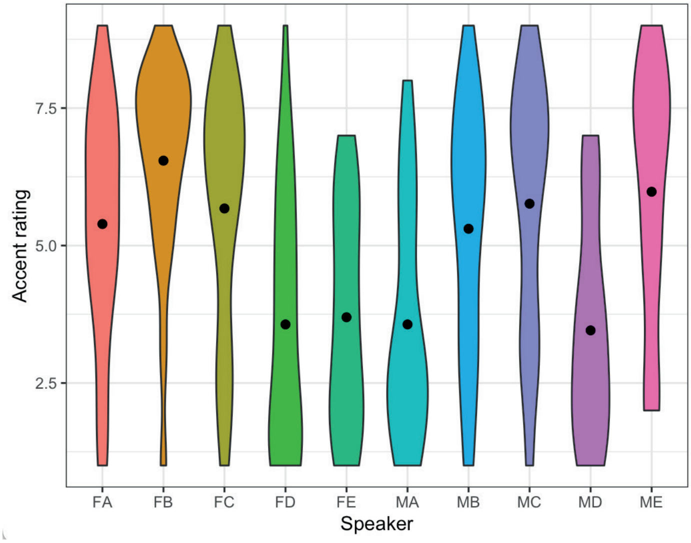

Figure 3: Violin plot for accent ratings. The black circle represents the mean. A higher rating means higher likelihood to be from Hong Kong.

instead of the English ones is that Cantonese was the L1/dominant language of the participants, so less variance was expected in Cantonese compared to English. Figure 4 shows the boxplot of the Euclidean distance by vowels.

The lmerTest package (Kuznetsova, Brockhoff and Christensen 2017) was used to run the linear mixed-effect models in R. Using Euclidean distance as the dependent variable, the model contained accent rating, perceptual similarity and

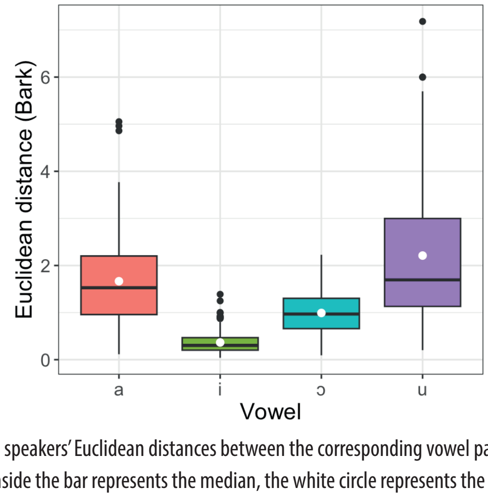

Figure 4: Boxplots of the speakers' Euclidean distances between the corresponding vowel pairs in Cantonese and English. The black line inside the bar represents the median, the white circle represents the mean.

Table 2: Summary of the mixed effects model for monophthongs

	Estimate	SE	Df	*p*
(Intercept)	3.533	0.44	19.12	<.001*
Accent rating	0.043	0.07	9.93	.56
Perceptual similarity	−0.659	0.08	−8.06	<.001*
Vowel (i)	−0.93	0.12	−7.61	<.001*
Vowel (ɔ)	−0.56	0.11	−4.99	<.001*
Vowel (u)	0.26	0.14	1.87	.37

The estimates represent comparisons against a reference value (vowel [a] for vowel). '*' indicates a significant effect at .05 level.

vowel as the fixed effects, and random intercepts by speaker and by word. Results are shown in Table 2, which suggested that accent rating was not a significant predictor for the model.

As perceptual similarity was found to be significant in the model, four Pearson correlation tests were run to further explore this variable. The dataset was divided into four subsets by vowels. For each subset, the Pearson correlation between perceptual similarity (x-axis) and Euclidean distance (y-axis) was calculated and plotted in Figure 5. A significant correlation was found for the vowels

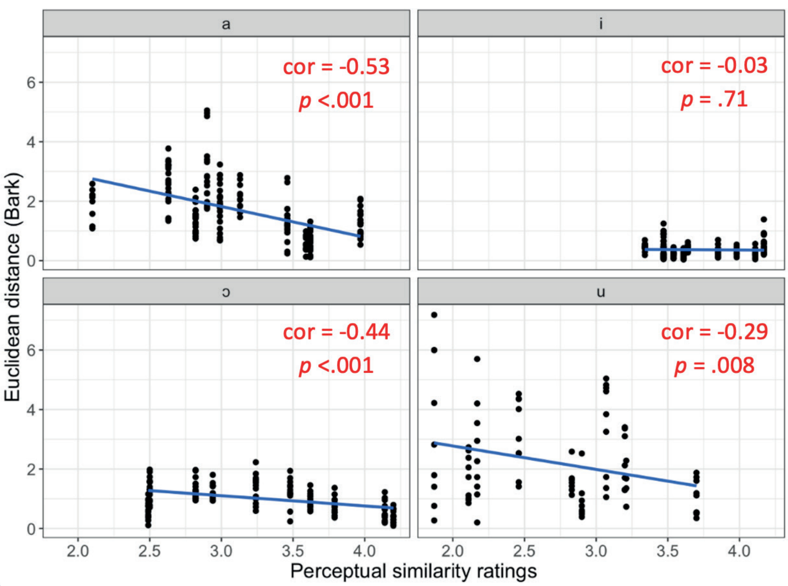

Figure 5: Correlation between perceptual similarity and Euclidean distance for each of the vowels. A higher similarity rating means that the words were considered to be more similar.

[a] (r (189) = –0.53, p <.001), [ɔ] (r (147) = –0.44, p <.001) and [u] (r (80) = –0.29, p = .008), but not for [i] (r (143) = –0.03, p = .71). The results indicate that the similarity judgements by the listeners were supported by acoustic differences.

As for **diphthongs**, the dataset was divided into ten subsets by diphthongs and formants (5 diphthongs × 2 formants = 10 subsets). For each subset, a Generalised Additive Mixed Model (GAMM) was run using the bam function in the mgcv package in R (Wood 2011). The diphthongs were time-normalised. The dependent variables were F1 (or F2) values in Bark. For each GAMM, the parametric factor was language (Cantonese vs English), non-linear smooths (s) included measurement points (10 points for each diphthong), the interaction between point and language, the tensor interaction (ti) between point, accent rating and language. Following Wieling (2018) and Sóskuthy (2021), the random effects structure included a random slope over word by speaker, a random slope over repetition by word, and a random smooth over point with a grouping variable of the combination of speaker and language. The variable language was transferred to an ordered factor for comparison. For each model,[3] an autocorrelation parameter was included, and the rho value was calculated separately using an initial simpler model.

A summary of the parametric effects is shown in Table 3. A significant effect was found for the F2 of [oʊ], indicating that the F2 contour of Cantonese [oʊ] was significantly higher than the contour of English [oʊ] (əʊ). As the present study has a greater focus on the shape of the contours, Table 4 summarises the results of the non-linear smooths.

As shown in the upper panel of Table 4, the shape of the Cantonese contour was significantly different from the English contour (i.e. s (point) by language) for F1 of [aɪ] and [ɔɪ], F1 and F2 of [eɪ] and [oʊ]. Figure 6 shows the visualisation of the F1 and F2 contours for Cantonese and English in each diphthong across speakers. These results confirm the individual SS ANOVA observations above, i.e. that the corresponding diphthongs were realised differently in the two languages.

Table 3: Parametric coefficients of the GAMMs for each diphthong

	F1				**F2**			
Language (English vs Cantonese)	**Est**	**SE**	**t**	***p***	**Est**	**SE**	**t**	***p***
[aɪ]	–0.171	0.207	–0.82	.41	–0.253	0.293	–0.87	.39
[aʊ]	–0.159	0.194	–0.82	.41	0.316	0.322	0.98	.33
[eɪ]	0.203	0.217	0.94	.35	–0.133	0.319	–0.42	.68
[ɔɪ]	–0.354	0.205	–1.73	.08	–0.306	0.263	–1.16	.25
[oʊ]	0.141	0.188	0.75	.45	1.223	0.382	3.20	**.001***

Table 4: Smooth functions and tensor interaction for the non-linear smooths of the GAMMs. "*" indicates the significance at 0.05 level or better

	F1			F2		
***s* (point): language English vs Cantonese**	**edf**	**F**	***p***	**edf**	**F**	***p***
[aɪ]	5.834	2.316	**.023***	1.01	1.108	.29
[aʊ]	1	0.228	.63	1.5	0.562	.64
[eɪ]	5.884	8.040	**<.001***	4.745	8.269	**<.001***
[ɔɪ]	2.456	4.173	**.01***	1.432	0.187	.803
[oʊ]	5.77	13.474	**<.001***	1.00	7.208	**.007***
		F1			**F2**	
***ti* (point, accent rating): language English vs Cantonese**	**edf**	**F**	***p***	**edf**	**F**	***p***
[aɪ]	3.157	2.159	.057	3.42	4.679	**.01***
[aʊ]	3.514	3.257	**.008***	62.189	5.675	.437
[eɪ]	2.027	0.933	.41	2.337	1.283	.278
[ɔɪ]	2.474	2.427	.064	2.499	0.309	0.79
[oʊ]	3.583	1.054	.49	6.454	1.518	0.14

When assessing the effect of accent rating on the contour difference between Cantonese and English (i.e. *ti* (point, accent rating) by language, lower panel of Table 4), a significant effect was found for F2 of [aɪ] and F1 of [aʊ]. Although the accent rating was significant for these two diphthongs, it is difficult to interpret the results because the significance was only observed in either F1 or F2 of two different diphthongs with no common pattern. It is unlikely that listeners would correlate a general perceptual judgement (accent rating) with difference of a specific formant (e.g. F2 of [aɪ]). Moreover, since the same set of accent ratings were used repeatedly in the ten GAMMs, if we adopt a more conservative approach by adjusting the alpha for multiple comparisons (0.05/10 = 0.005), the two results would not be significant.

Therefore, given that only two sporadic differences were found in diphthongs and no difference was found for monophthongs, accent ratings do not seem to be a reliable predictor for cross-language difference.

Discussion

Our study investigated the relationship between accentedness and the realisation of corresponding vowels in the two languages of Cantonese-English bilingual speakers. Specifically, we examined whether a stronger accent would be correlated with higher similarity between the vowels in the two languages in both acoustic

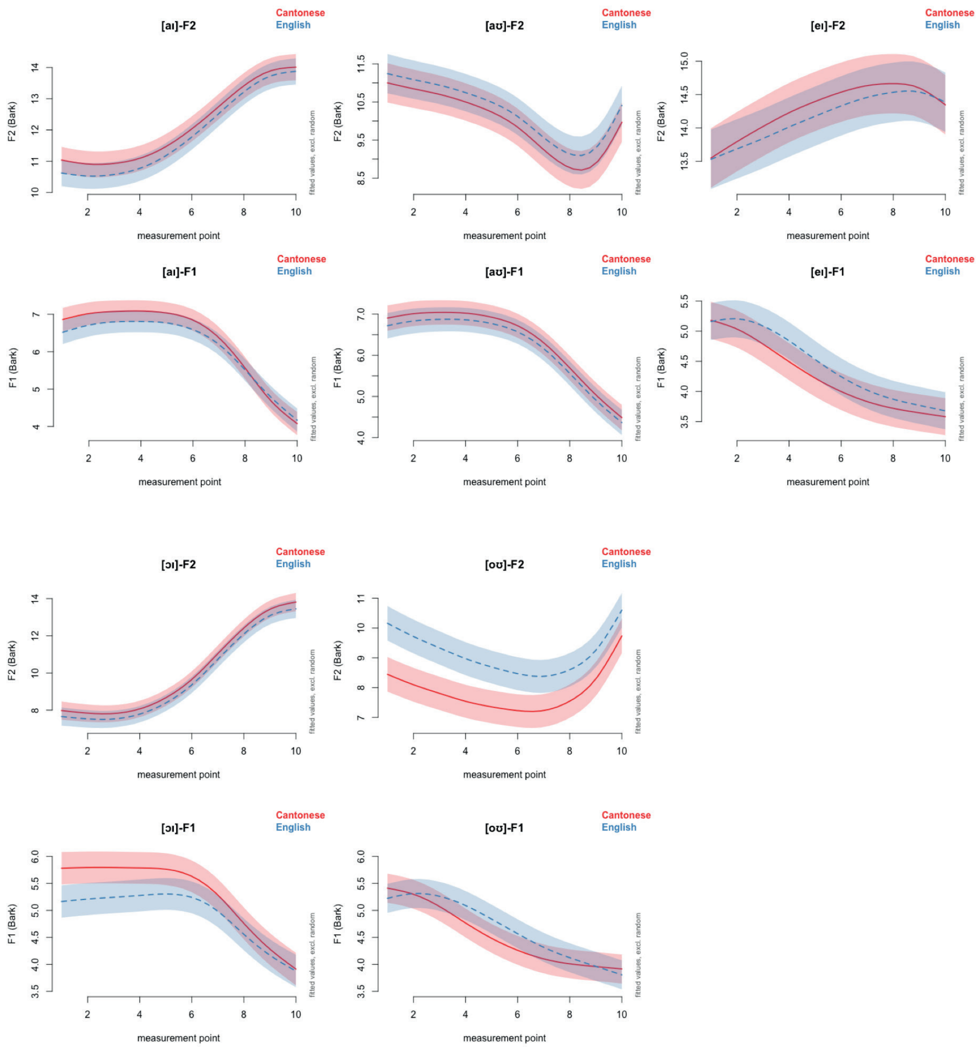

Figure 6: F1 and F2 contours of Cantonese (solid line) and English (dashed line) by diphthongs.

and perceptual dimensions. However, our results showed no correlation between them. Individual variations were observed in monophthong and especially in diphthong realisations. Furthermore, the convergence or divergence in one corresponding vowel could not predict that of another vowel. The monophthong /i/ was the most similar between the two languages while the other three monophthongs exhibited more differences. Formant trajectories in the two languages were quite similar for /eɪ/ (although still with significant differences in more sensitive and powerful GAMMs) while they differed quite substantially for /oʊ (əʊ)/ (especially in F2) across speakers regardless of how strong the accent was. Vowel-specific patterns in convergence/divergence were observed in our data.

Why did the degree of accentedness not have an effect as hypothesised? One likely reason is that, as mentioned in the Introduction, the impression of global foreign accent incorporates many speech features, e.g. segmental (consonants and vowels) and suprasegmental (speech rate, stress, rhythm, intonation, etc.) (Anderson-Hsieh et al. 1992; Sidaras et al. 2009). The relative weighting of the contribution of each feature to foreign accent is unclear, although recently Sereno, Lammers and Jongman (2016) suggested that deviations in segments contribute to the perceived accentedness, comprehensibility and intelligibility of Korean-accented English substantially more than intonation. However, other studies like de Mareuil and Vieru-Dimulescu (2006) clearly demonstrated the important role of prosodic factors in accent perception and intelligibility. A very recent study (van Maastricht, Zee, Krahmer and Swerts, 2021) investigated the relative contribution of intonation, rhythm and speech rate to foreign accent in L2 Dutch speech produced by Spanish L1 learners and showed that intonation and speech rate contributed more than speech rhythm to foreign accent. Thus, prosody does have a strong effect on accent perception. It is possible to have a dissociation between overall accentedness (clearly affected by prosodic patterns) and vowel convergence, especially given that our data demonstrated vowel-specific patterns of convergence. Of course, logically, if the accent rating was based on individual words or vowels rather than on sentences produced by the speakers (e.g. Chan and Hall 2019; Franklin, Oksanen and Gilfert 2016; Munro, Flege and MacKay 1996), a closer mapping between accentedness and vowel convergence may emerge. Nevertheless, most previous studies of the effects of foreign accent (e.g. Flege et al. 1995) collected accent ratings using sentences with good reason, as accent is a global feature that is perceptually salient but difficult to pin down. It would also be much easier for listeners (especially naïve listeners) to give accent ratings for longer stretches of speech than for single words/vowels.

Nevertheless, the dissociation between accentedness and vowel convergence was not simply caused by a mismatch between what speech features listeners paid attention to for the purposes of accent rating (such as prosody) and what we focused on (vowels). Even if vowel realisations were the only important factor for accentedness, vowel-specific patterns would also contribute to the dissociation. Chan and colleagues (Chan and Hall 2019; Chan et al. 2017) demonstrated that listeners responded differently to the vowels /ɔ ʌ e ɑ æ/ with the same type of spectral manipulation. They concluded that neither the direction nor the magnitude of spectral deviations from native norms could account for their findings.

Accent rating scores increased only when the deviations led to more overlap and confusability with neighbouring vowel categories. They argued that deviations away from the crowded regions of the vowel space may instead reduce confusability, and thus reduce perceived accentedness and comprehensibility. Moreover, our results also illustrate that convergence/divergence of one corresponding vowel cannot predict that of another, especially for diphthongs. Such vowel-specific variation is a likely reason for the dissociation. The relationship between accentedness and vowel realisation is not uniform.

Vowel-specific convergence patterns were also reported in Mayr, Morris, Mennen and Wiliams (2017). They compared the monophthongs in the two languages of Welsh-English bilinguals. Given the long-term language contact between the two languages, they found large-scale convergence in most of the monophthongs they investigated. Interestingly, language-specific vowel divergence patterns were also reported: English NURSE (*herd*), FOOT (*hood*) and SQUARE (*hared*) were distinct from all Welsh categories (words in bracket show the Welsh targets), and Welsh *hwd* and *hêd* were distinct from all English categories. They suggested that the lack of corresponding vowels for assimilation may account for some of the findings (e.g. English NURSE (*herd*)), but not all. English FOOT (*hood*) was much more fronted than its Welsh counterpart *hwd*. The lack of corresponding vowels for assimilation should not be a reason for our findings as the materials were as similar as possible in the two languages in terms of IPA transcription, vowel distribution, auditory vowel quality and phonetic environment. English /u/ was also fronter than Cantonese /u/ for most speakers (see Figure 1). Thus, both vocalic convergence and divergence have been found in a single language contact situation (Mayr et al. 2017) and individual speakers with a similar language background (our study).

Another possible reason for the lack of correlation between accentedness and vowel convergence could be related to the small number of speakers (10) in our study, which resulted in low statistical power. While this remains a possibility, the findings in Lein et al. (2016), together with our findings, reveal a more fundamental issue. Lein et al. (2016) also had a similar number of German and French bilingual participants (14), who demonstrated much individual variation in terms of accentedness and VOT realisation. There were speakers with native-like accents and a deviant VOT, and speakers with a strong accent but clear VOT categories. They also found no systematic relationship between accentedness and segmental realisation. Their findings on a consonantal feature corroborate our findings on vowels very well. The speakers with the strongest accent did not exhibit the most acoustic difference among the corresponding vowels in our study. Thus, besides the small sample size, individual variation is likely a more fundamental reason for the lack of correlation. Further study with more speakers is warranted to evaluate the correlation between accentedness and vowel convergence more comprehensively.

The findings in our studies echo the emphasis on individual differences over group differences in the recent SLM-r (Flege and Bohn 2021). Flege and Bohn measured the mean VOT values produced by native Spanish speakers who arrived in the US at or after 16 years of age (see their figure 1.3). Wide variations in VOT values were reported for the late learners, some with Spanish-like VOT, some with English-like VOT, while others had values in between. They argued that just giving labels to arbitrarily selected subgroups of these late learners cannot explain the inter-speaker variability. Such variability highlights the need to understand why individual L2 learners differ so much from one another. They said that it is important to know if the participants have formed new phonetic categories for the target L2 sounds, and whether the presence or absence of category formation is influenced by individual differences in language processing abilities.

While it is impossible for us to know if our speakers differed in language processing abilities because we did not collect such data, it is reasonable to ask if they had formed new phonetic categories. According to SLM-r (Flege and Bohn 2021: 40), new L2 categories may shift away from neighbouring L1 categories to maintain phonetic contrast (i.e. divergence). If no new category is formed for an L2 sound, SLM-r predicts a merger of the phonetic properties of an L1 and L2 sound (i.e. convergence). Our data, particularly the diphthong data, showed significant differences in formant frequencies and formant dynamics for the corresponding vowels in the two languages. The existence of these language-specific differences can be taken as evidence for separate phonological systems within bilinguals. While MacLeod et al. (2009) showed that early bilingual speakers maintained separate categories for similar high vowels in Canadian English (/i ɪ u ʊ/), and Canadian French (/i u/ [ɪ ʊ]), our results demonstrated that L2 learners could also establish separate categories for similar corresponding vowels, both monophthongs and diphthongs. Thus, we can conclude that new L2 phonetic categories were formed for our bilingual speakers, even for those with a stronger accent.

The success in establishing new L2 phonetic categories does not mean that the bilingual speakers were native-like in their production, however. Rather, as maintained by Flege and Bohn (2021) for SLM-r, it is virtually impossible for L2 learners to produce and perceive an L2 sound exactly like mature monolingual native speakers of the target L2. They argued that it is no longer of theoretical interest to compare whether L2 leaners perform the same as native speakers. This concurs well with the bi/multilingual turn (Ortega 2013) which emphasises understanding of the cognitive, linguistic and psycholinguistic mechanisms and consequences of becoming bi/multilingual later in life over the question of why bilinguals are not like monolingual speakers. Our study contributes to the bi/multilingual turn by showing that degree of overall accentedness in the L2 cannot predict vowel convergence between the two languages in a straight forward manner. Both common cross-linguistic patterns and individual variations are present among bilingual speakers with a similar language background. Further study is needed to investigate what other factors can modulate the interaction between the two languages.

Finally, our findings have some implications for forensic speaker comparison. As reviewed in the Introduction, there have been very few forensically oriented studies involving bilingual materials, and those few studies generally show that having two languages reduces the reliability of various forensic comparisons by both human and machine alike. An important forensic question asked in the present study is whether speech patterns of highly similar materials in one language can predict those in another language spoken by the same bilingual speaker. Our results (see Figures 1 and 2) show that for English and Cantonese vowels the answer is no. Some vowels (e.g. /i eɪ/) were more similar across the two languages than others (e.g. /u ɔɪ/), although still with significant differences in formant frequencies, and individual speakers varied in similarity. Moreover, no one single bilingual speaker exhibited the same type of variation across monophthongs and diphthongs. The cross-linguistic difference is not simply about a shift in formant frequency ranges: the formant dynamics can also be different in the two languages. Thus, our findings clearly demonstrate that the formant patterns in one language cannot predict those in another language even with highly comparable materials and/or speakers with a relatively strong accent. Our findings add to the small but crucial literature demonstrating the challenges and unreliability of bilingual forensic comparisons between one language and another. Our data and those in the previous bilingual literature have not included bilingual speakers who sound very different in their two languages, but anecdotal experience does suggest that it is not uncommon to encounter such bilingual speakers. They add extra complexity to bilingual forensic investigation. Indeed, much caution needs to be exercised when dealing with forensic materials in bilingual settings, echoing the concern raised by the Code of Practice of the IAFPA (2020) that 'Members should exercise particular caution with cross-language comparisons'. Nonetheless, it is still important to do further forensic investigation of bilingual speech patterns because knowing what should not be done and what could possibly be useful can guide proper comparison. Further research should explore whether there are any features which may be stable across the two languages of bilingual speakers who sound different in their two languages, particularly those that are less linguistically dependent, and also whether speech styles (e.g. more spontaneous speech vs. reading sentences) can influence the convergence between corresponding vowels. Much awaits further investigation.

In conclusion, our study investigated whether there was any correlation between degree of accentedness and convergence between corresponding vowels in the two languages of Cantonese-English bilingual speakers with high levels of English proficiency. Both monophthongs and diphthongs were included. The results showed no such correlation. Both vowel-specific and individual variation was found with no uniform effect of accentedness. Further study with more speakers and different language pairs is needed to corroborate these findings.

Acknowledgements

The authors would like to thank the speakers for participating in the experiment. They would also like to thank the anonymous reviewers and editors for their comments and suggestions which helped improve the manuscript in multiple ways.

About the authors

Peggy Mok is a Professor at the Department of Linguistics and Modern Languages, Chinese University of Hong Kong. She is interested in both speech production and perception, particularly with cross-linguistic and psycholinguistic perspectives. She investigates both segmental and prosodic aspects of speech. Speech acquisition in different contexts is an important theme in her research. Additionally, she is interested in forensic phonetics and the bilingual mental lexicon.

Holly Fung received her B.A. and M.Phil. degrees in Linguistics from the Chinese University of Hong Kong. Her main research interest is in speech prosody and its relationship with co-speech gestures. She co-authored a paper on the temporal coordination between focus prosody and pointing gesture in Cantonese.

Grace Wenling Cao is a postdoctoral fellow working on forensic phonetics with a trilingual perspective under the RGC Postdocotral Fellowship Scheme at the Chinese University of Hong Kong. She received her PhD degree in sociophonetics from University of York. Her recent publications include 'L2 phonetic convergence of Hong Kong English' in *Language and Speech*. She is interested in L2 sociophonetics, forensic phonetics of trilingual speech and language attitudes.

Chun Wai Leung is an Assistant Lecturer at the English Language Teaching Unit at The Chinese University of Hong Kong, teaching and coordinating various English for Academic Purposes and English for Professional Purposes courses. He obtained his BA in Linguistics from University College London and MPhil in Theoretical and Applied Linguistics from the University of Cambridge. His academic interests include Cantonese linguistics, English language pedagogy, and the interaction between Cantonese and English in bilinguals.

Notes

1. https://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/
2. All the patterns and results of the statistical analyses are very similar in Hz. Therefore, we only report the results in Bark.
3. An example of the full model for the F1 of [aɪ] is shown below:
   F1 ~language.ord + s (point, bs='cr') + s (point, by=language.ord, bs='cr') +
   ti (point, accent rating, by=language.ord) +
   s (speaker, word, bs='re') + s(word, repetition, bs='re') +
   s (point, SpeakerLanguage.ord, bs= 'fs', m=1, k=9),
   data=ai, method="fREML", discrete = T, AR.start = ai$AR.start, rho=0.64)

References

Amengual, M. and Chamorro, P. (2015) The effects of language dominance in the perception and production of the Galician mid vowel contrasts. *Phonetica* 72: 207–236. https://doi.org/10.1159/000439406

Anderson-Hsieh, J., Johnson, R. and Koehler, K. (1992) The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42: 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Antoniou, M., Best, C. T., Tyler, M. D. and Kroos, C. (2011) Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics* 39: 558–570. https://doi.org/10.1016/j.wocn.2011.03.001

Bolton, K. and Kwok, H. (1990) The dynamics of the Hong Kong accent: social identity and sociolinguistic description. *Journal of Asian Pacific Communication* 1: 147–172.

Chan, K. Y. and Hall, M. D. (2019) The importance of vowel formant frequencies and proximity in vowel space to the perception of foreign accent. *Journal of Phonetics* 77: 100919. https://doi.org/10.1016/j.wocn.2019.100919

Chan, K. Y., Hall, M. D. and Assgari, A. A. (2017) The role of vowel formant frequencies and duration in the perception of foreign accent. *Journal of Cognitive Psychology* 29: 23–34. https://doi.org/10.1080/20445911.2016.1170746

Chanethom, V. (2011) Dynamic differences in child bilinguals' production of diphthongs. In the *4th Tutorial and Research Workshop on Experimental Linguistics* (ExLing 2011) 55–58. Paris.

Chang, C. B. (2012) Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–268. https://doi.org/10.1016/j.wocn.2011.10.007

Chen, Y., Ng, M. L. and Li, T. S. (2012) English vowels produced by Cantonese – English bilingual speakers. *International Journal of Speech-Language Pathology* 14: 557–568. https://doi.org/10.3109/17549507.2012.718360

de Boer, M. M. and Heeren, W. F. L. (2020) Cross-linguistic filled pause realization: the acoustics of uh and um in native Dutch and non-native English. *Journal of the Acoustical Society of America* 148: 3612–3622. https://doi.org/10.1121/10.0002871 de Leeuw, E. (2007) Hesitation markers in English, German and Dutch. *Journal of Germanic Linguistics* 19: 85–114. https://doi.org/10.1017/s1470542707000049

de Mareuil, P. B. and Vieru-Dimulescu, B. (2006) The contribution of prosody to the perception of foreign accent. *Phonetica* 63: 247–267. https://doi.org/10.1159/000097308

Derwing, T. M. and Munro, M. J. (1997) Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16. https://doi.org/10.1017/s0272263197001010

Deterding, D., Wong, J. and Kirkpatrick, A. (2008) The pronunciation of Hong Kong English. *English World-Wide* 29: 148–175. https://doi.org/10.1075/eww.29.2.03det

Fairbanks, G. (1960) *Voice and Articulation Drillbook* (2nd edition). New York: Harper & Row.

Flege, J. E. (1981) The phonological basis of foreign accent. *TESOL Quarterly* 15: 443–455. https://doi.org/10.2307/3586485

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 Research* 229–273. Timonium, MD: York Press.

Flege, J. E. (1999) Age of learning and second language speech. In D. Birdsong (ed.) *Second Language Acquisition and the Critical Period Hypothesis* 101–131. Mahwah, NJ: Lawrence Erlbaum Associates.

Flege, J. E. and Bohn, O. S. (2021) The Revised Speech Learning Model (SLM-r). In R. Wayland (ed.) *Second Language Speech Learning: Theoretical and Empirical Progress* 3–83. Cambridge: Cambridge University Press.

Flege, J. E., Munro, M. J. and MacKay, I. R. A. (1995) Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–3134. https://doi.org/10.1121/1.413041

Flege, J. E., Schirru, C. and MacKay, I. R. A. (2003) Interaction between the native and second language phonetic subsystems. *Speech Communication* 40: 467–491. https://doi.org/10.1016/s0167-6393(02)00128-0

Franklin, A. D., Oksanen, K. A. and Gilfert, K. E. (2016) Goodness and accentedness ratings of /hVt/ tokens by aware and naïve listeners. *American Journal of Speech-Language Pathology* 25: 620–633. https://doi.org/10.1044/2016_ajslp-15-0106

Goggin, J. P., Thompson, C. P., Strube, G. and Simental, L. R. (1991) The role of language familiarity in voice identification. *Memory and Cognition* 19: 448–458. https://doi.org/10.3758/bf03199567

Gu, C. (2014) Smoothing spline ANOVA models: R Package gss. *Journal of Statistical Software*, 58: 1–25. https://doi.org/10.18637/jss.v058.i05

Guion, S. G. (2003) The vowel systems of Quichua-Spanish bilinguals: age of acquisition effects on the mutual influence of the first and second languages. *Phonetica* 60: 98–128. https://doi.org/10.1159/000071449

Guion, S. G. (2005) Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition* 27: 503–533. https://doi.org/10.1017/s0272263105050230

Hughes, V., Wood, S. and Foulkes, P. (2016) Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law* 23: 99–132. https://doi.org/10.1558/ijsll.v23i1.29874

Hung, T. (2000) Towards a phonology of Hong Kong English. *World Englishes* 19: 337–356. https://doi.org/10.1111/1467-971x.00183

IAFPA. (2020) Code of practice. Available at https://www.iafpa.net/about/code-of-practice/

Kang, K. H. and Guion, S. G. (2006) Phonological systems in bilinguals: age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America* 119: 1672–1683. https://doi.org/10.1121/1.2166607

Kirkham, S. and Nance, C. (2017) An acoustic-articulatory study of bilingual vowel production: advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English. *Journal of Phonetics* 62: 65–81. https://doi.org/10.1016/j.wocn.2017.03.004

Künzel, H. J. (2013) Automatic speaker recognition with crosslanguage speech material. I*nternational Journal of Speech, Language and the Law* 20: 21–44. https://doi.org/10.1558/ijsll.v20i1.21

Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., Lein, T. and van de Weijer, J. (2014) Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal* 11: 123–150. https://doi.org/10.46538/hlj.11.2.2

Kuznetsova, A., Brockhoff, P. and Christensen, R. (2017) lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82: 1–26. https://doi.org/10.18637/jss.v082.i13

Lein, T., Kupisch, T. and Van de Weijer, J. (2016) Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20: 732–749. https://doi.org/10.1177/1367006915589424

Levis, J. M., Derwing, T. M. and Munro, M. J. (2020) 25 years of intelligibility, comprehensibility and accentedness. Special issue of the *Journal of Second Language Pronunciation* 6: 3. https://doi.org/10.1075/jslp.6.3

Lo, J. J. H. (2020) Between Äh(m) and Euh(m): The distribution and realization of filled pauses in the speech of German-French simultaneous bilinguals. L*anguage and Speech* 63: 746–768. https://doi.org/10.1177/0023830919890068

Lo, J. J. H. (2021a). Cross-linguistic speaker individuality of long-term formant distributions: phonetic and forensic perspectives. In *Interspeech 2021* 416–420. Brno, Czechia.

Lo, J. J. H. (2021b). Seeing the trees in the forest: diagnosing individual performance with acoustic data in likelihood ratio based forensic voice comparison. *Studi AISV* 8 (Speaker individuality in phonetics and speech sciences): 77–96.

Lu, L., Dong, Y., Zhao, X., Liu, J. and Wang, H. (2009) The effect of language factors for robust speaker recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing* 4217–4220. Taipei: IEEE.

Luk, G. and Bialystok, E. (2013) Bilingualism is not a categorical variable: interaction between language proficiency and usage. *Journal of Cognitive Psychology* 25: 605–621. https://doi.org/10.1080/20445911.2013.795574

MacKay, I. R. A., Flege, J. E., Thorsten, P. and Schirru, C. (2001) Category restructuring during second-language (L2) speech acquisition. *Journal of the Acoustical Society of America* 110: 516–528. https://doi.org/10.1121/1.1377287

MacLeod, A. A. N., Stoel-Gammon, C. and Wassink, A. B. (2009) Production of high vowels in Canadian English and Canadian French: a comparison of early bilingual and monolingual speakers. *Journal of Phonetics* 37: 374–387. https://doi.org/10.1016/j.wocn.2009.07.001

Major, R. C. (1987) English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics* 15: 197–202. https://doi.org/10.1016/s0095-4470(19)30560-1

Mayr, R., Morris, J., Mennen, I. and Wiliams, D. (2017) Disentangling the effects of long-term language contact and individual bilingualism: the case of monophthongs in Welsh and English. *International Journal of Bilingualism* 21: 245–267. https://doi.org/10.1177/1367006915614921

McArthur, T. (2002) *The Oxford Guide to World English*. Oxford: Oxford University Press.

McDougall, K. (2004) Speaker-specific formant dynamics: an experiment on Australian English /aɪ/. *International Journal of Speech, Language and the Law* 11, 103–130. https://doi.org/10.1558/sll.2004.11.1.103

McDougall, K. (2006) Dynamic features of speech and the characterisation of speakers: towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law* 13: 89–126. https://doi.org/10.1558/sll.2006.13.1.89

McDougall, K. and Duckworth, M. (2017) Profiling fluency: an analysis of individual variation in disfluencies in adult males. *Speech Communication* 95: 16–27. https://doi.org/10.1016/j.specom.2017.10.001

Mok, P. (2011) The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition* 14: 458–472. https://doi.org/10.1017/s1366728910000453

Mok, P., Xu, B. and Zuo, D. (2015) Bilingual speaker identification: Chinese and English. *International Journal of Speech, Language and the Law* 22: 57–77. https://doi.org/10.1558/ijsll.v22i1.18636

Moyer, A. (2004) *Age, Accent and Experience in Second Language Acquisition*. Clevedon: Multilingual Matters.

Munro, M. J., Flege, J. E. and MacKay, I. R. A. (1996) The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics* 17: 313–334. https://doi.org/10.1017/s0142716400007967

Nolan, F. and Grigoras, C. (2005) A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law* 12: 143–173. https://doi.org/10.1558/sll.2005.12.2.143

Ortega, L. (2013) Ways forward for a bi/multilingual turn in SLA. In S. May (ed.) *The Multilingual Turn: Implications for SLA, TESOL, and Bilingual Education* 32–52. New York: Routledge.

Piske, T., MacKay, I. R. A. and Flege, J. E. (2001) Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29: 191–215. https://doi.org/10.1006/jpho.2001.0134

Schertz, J., Kang, Y. and Han, S. (2019) Cross-language correspondences in the face of change: phonetic independence versus convergence in two Korean-Mandarin bilingual communities. *International Journal of Bilingualism* 23: 157–199. https://doi.org/10.1177/1367006917728389

Schneider, E. W. (2007) *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.

Sereno, J., Lammers, L. and Jongman, A. (2016) The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics* 37: 303–322. https://doi.org/10.1017/s0142716414000575

Setter, J., Wong, C. S. P. and Chan, B. H. S. (2010) *Hong Kong English*. Edinburgh: Edinburgh University Press.

Sewell, A. and Chan, J. (2010) Patterns of variation in the consonantal phonology of Hong Kong English. *English World-Wide* 31: 138–161. https://doi.org/10.1075/eww.31.2.02sew

Sidaras, S. K., Alexander, J. E. D. and Nygaard, L. C. (2009) Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America* 125: 3306–3316. https://doi.org/10.1121/1.3101452

Sóskuthy, M. (2021) Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84: 1–22. https://doi.org/10.1016/j.wocn.2020.101017

Sundara, M., Polka, L. and Baum, S. (2006) Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition* 9: 97–114. https://doi.org/10.1017/s1366728905002403

van Maastricht, L., Zee, T., Krahmer, E. and Swerts, M. (2021) The interplay of prosodic cues in the L2: how intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Communication* 133: 81–90. https://doi.org/10.1016/j.specom.2020.04.003

Wieling, M. (2018) Analyzing dynamic phonetic data using generalised additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70: 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalised linear models. *Journal of the Royal Statistical Society* 73: 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Xu, Y., and Gao, H. (2018) FormantPro as a tool for speech analysis and segmentation. *Revista de Estudos da Linguagem* 26: 1435–1454. https://doi.org/10.17851/2237-2083.26.4.1435-1454

Yang, J. and Fox, R. A. (2017) L1–L2 interactions of vowel systems in young bilingual Mandarin-English children. *Journal of Phonetics* 65, 60–76. https://doi.org/10.1016/j.wocn.2017.06.002

Zuo, D., and Mok, P. (2015) Formant dynamics of bilingual identical twins. *Journal of Phonetics* 52: 1–12. https://doi.org/10.1016/j.wocn.2015.03.003

Appendix: Real words in Cantonese and English used in the production experiment

	/a (ɑ)/				/i/			/u/			/ɔ/		
Consonants	Can		Eng		Can		Eng	Can		Eng	Can		Eng
p	pʰa˥	趴	pa	par	ti˥	啲	D				pʰɔ˧	破	paw
b					pi˥	B*	Bee						
t	tʰa˥	他		tar									
k	kʰa˥	卡		car				kʰu˥	箍	coo			
g								ku˥	菇	goo			
f	fa˥	花	fah	far									
s					si˥	詩	sea				sɔ˥	梳	saw
h	ha˥	哈	ha		hi˥	嘻	he				hɔ˥	呵	haw
m	ma˥	媽	ma								mɔ˥	麼	maw
l					li˥	呢	lee				lɔ˥	囉	law
w								wu˥	污	woo			
Total words	6		8		5		5	3		3	5		5

*as in nicknames or the localised form of 'baby'

	/aɪ/			/aʊ/			/eɪ/			/ɔɪ/			/ou (əʊ)/		
Consonants	**Can**		**Eng**	**Can**		**Eng**	**Can**		**Eng**	**Can**		**Eng**	**Can**		**Eng**
p	pʰaɪ˧	派	pie	pʰaʊ˥	抛	poww	pʰeɪ˥	披	pay						
b				paʊ˥	包	bow	peɪ˥	悲	bay						
t	tʰaɪ˥	呔	tie							tʰɔɪ˥	胎	toy	tʰoʊ˥	滔	toe
d							teɪ˧	地	day				toʊ˥	都	doe
k				kʰaʊ˧	靠	cow	kʰeɪ˥	畸	K	kʰɔɪ˧	鈣	coy			
g	kaɪ˥	佳	guy				keɪ˥	基	gay				koʊ˥	高	go
s	saɪ˥	曬	sigh				seɪ˧	四	say	sɔɪ˥	鰓	soy	soʊ˥	蘇	so
h	haɪ˥	揩	hi	haʊ˥	敲	how	heɪ˥	希	hay	hɔɪ˥	開	hoy	hoʊ˧	耗	hoe
m	maɪ˧	賣	my				meɪ˥	瞇	may						
n				naʊ˧	鬧	now									
l	laɪ˥	拉	lie				leɪ˧	利	lay				loʊ˥	撈	low
Total words	7		7	5		5	9		9	4		4	6		6