Language contact within the speaker: Phonetic variation and crosslinguistic influence

Khia A. Johnson and Molly Babel

Department of Linguistics, University of British Columbia

Abstract

A recent model of sound change posits that the direction of change is determined, at least in part, by the distribution of variation within speech communities (Harrington, Kleber, Reubold, Schiel, & Stevens, 2018; Harrington & Schiel, 2017). We explore this model in the context of bilingual speech, asking whether the less variable language constrains phonetic variation in the more variable language, using a corpus of spontaneous speech from early Cantonese-English bilinguals (Johnson, 2021). As predicted, given the phonetic distributions of stop obstruents in Cantonese compared to English, intervocalic English /b d g/ were produced with less voicing for Cantonese-English bilinguals and word-final English /t k/ were more likely to be unreleased compared to spontaneous speech from two monolingual English control corpora (Pitt, Johnson, Hume, Kiesling, & Raymond, 2005; Swan, 2016). Cantonese phonology is more gradient in terms of voicing initial obstruents (Clumeck, Barton, Macken, & Huntington, 1981; W. Y. P. Wong, 2006) than permitting releases of final obstruents, which is categorically prohibited (Bauer & Benedict, 2011; Khouw & Ciocca, 2006). Neither Cantonese-English bilingual initial voicing nor word-final stop release patterns were significantly impacted by language mode. These results provide evidence that the phonetic variation in crosslinguistically linked categories in bilingual speech is shaped by the distribution of phonetic variation within each language, thus suggesting a mechanistic account for why some segments are more susceptible to cross-language influence than others in studies of mutual influence.

Language contact within the speaker: Phonetic variation and crosslinguistic influence

Introduction

Phonetic variation in bilingual speech stems from multiple sources. Crosslinguistic influence is a major source of variation, as the dynamics of parallel activation and inhibition of competing linguistic units can influence pronunciation patterns. While these pronunciation patterns are typically discussed as *interference* in speech production (Grosjean, 2011; Simonet & Amengual, 2019), the information is beneficial to listeners, as bilinguals use this phonetic variation in pronunciation patterns to efficiently predict and process an upcoming code-switch (Fricke, Kroll, & Dussias, 2016; Shen, Gahl, & Johnson, 2020). Crosslinguistic influence can also be a stable signature of a multilingual or non-native accent, and is termed *transfer* in such cases (Grosjean, 2011). In these cases, the resulting phonetic variation reflects a more codified or systematic linguistic pattern shaped by the bidirectional nature of influence in bilinguals' linguistic systems, as opposed to the arguably more localized and temporary effects of parallel activation. As a pattern becomes more systematic, it can be challenging to distinguish more ephemeral variation from a dialect. These points likely form a continuum, given that new language varieties develop out of language contact (Schneider, 2003).

We are interested in *how* early bilinguals' language systems influence one another—specifically, how crosslinguistic influence at the level of the speech community may lead to the development of novel language varieties. Certainly, on the community level, contact-induced language change is a well-studied entity (see, for example, the seminal book by Thomason & Kaufman, 1988). Moreover, decades of research on second language acquisition has established that bilingual speakers make phonological connections between their languages, presumably providing the cognitive infrastructure that supports mutual influence (e.g., Best & Tyler, 2007; Chang, 2015; Flege, 1995; Flege & Bohn, 2021). In this body of literature, the term “phonological” tends to be used in a psycholinguistic manner, rather than a theoretical phonological one. That is, while some level of abstraction and structural organization is needed to explain phenomena like crosslinguistic influence, it does not necessarily map on to a framework

in phonological theory (see Chang, 2015). This parallels Samuel's (2020) warning that linguistic and psycholinguistic categories should not be conflated with one another in perception, and is further reflected by the revised Speech Learning Model's emphasis on phonetic and surface-level behaviour at the individual level, rather than on a phonological space shared across languages (Flege & Bohn, 2021). The explicit connection of psycholinguistically-informed mutual influence on community patterns and language change has been also made (Tse, 2019; Van Coetsem, 2000; Winford, 2007). Yao and Chang (2016) recently presented evidence that language contact within the bilingual's spheres of knowledge guides sound change.

In this paper—following Yao and Chang's lead—we bring together methods from the study of crosslinguistic influence and concepts from sound change to examine Cantonese influence on English in spontaneous speech from early Cantonese-English bilinguals, using the corpus described in Johnson, Babel, Fong, and Yiu (2020). For the purposes of the current study, the primary delineator between the sound change and crosslinguistic influence literatures is whether language contact occurs between groups of people (e.g., as in the agent-based model by Harrington et al., 2018) or within the linguistic system of a bilingual (for a recent review, see Fricke, Zirnstein, Navarro-Torres, & Kroll, 2019). Given this, we ask, how do the sound systems within an individual accommodate each other in light of this mutual crosslinguistic influence? Our goal is to become more specific about the nature of the influence by adopting frameworks from sound change (Harrington et al., 2018; Harrington & Schiel, 2017).

Bilingualism and mutual influence

A large body of psycholinguistic research consistently finds evidence of mutual influence in bilingual speech production at various levels of linguistic structure (Kroll & Navarro-Torres, 2018). As noted above, mutual influence can take the form of interference, which is the result of both languages being simultaneously activated during speech production. Interference has been found in cognate production (Amengual, 2012), in language-switching paradigms (Goldrick, Runnqvist, & Costa, 2014), by inducing a bilingual mode through the experimental context

(Simonet & Amengual, 2019), and in corpus work examining code-switching behaviour (Fricke et al., 2016). Clear evidence of transfer is harder to come by, in part because there are challenges with ensuring that bilinguals operate in a strictly monolingual mode, but also because the same linguistic items can be simultaneously affected by transfer and interference (Grosjean, 2011).

The focus in this body of work is on the degree (i.e., is it quantifiable and perceptible?; e.g., Sancier & Fowler, 1997) and direction of crosslinguistic influence (i.e., is it the dominant language that influences the non-dominant language?; e.g., R. K.-Y. Tsui, Tong, & Chan, 2019). Varying degrees of phonetic convergence across linguistic systems have been found for vowels (Baker & Trofimovich, 2005; Simonet, 2011; Simonet & Amengual, 2019), /l/ (Barlow, 2014) and stops (Antoniou, Best, Tyler, & Kroos, 2010, 2011; Bullock & Toribio, 2009; Flege, 1987; Fricke et al., 2016; Goldrick et al., 2014; Y. Kang, George, & Soo, 2016; Major, 1992; Olson, 2016; Sancier & Fowler, 1997; Sundara, Polka, & Baum, 2006). This body of literature varies, of course, in terms of the languages studied, the dominance and nature of the bilinguals, and the paradigms exploited. For example, in a single-word language-switch paradigm with Cantonese-English bilinguals, R. K.-Y. Tsui et al. (2019) found that the voice onset time (VOT) patterns of the dominant language carried over to influence the production of stops in the non-dominant language on switch trials for both self-reported Cantonese-dominant and English-dominant bilinguals. Self-reported balanced bilinguals showed no effects of language switching. Chang (2012), however, finds that the learning of a new language (L2 Korean) affects pronunciation patterns of the first—or native—language (L1 English), which was the clear dominant language in the population under study.

While factors like language dominance and mode clearly impact bilingual speech (Fricke et al., 2019; R. K.-Y. Tsui et al., 2019), they do not fully account for why some segments are more susceptible to influence than others (e.g., Bullock & Toribio, 2009; Olson, 2016), or why simultaneous bilinguals sometimes exhibit asymmetrical influence in speech production (e.g., Sundara et al., 2006). The notion of some segments being more susceptible is best illustrated with an example. Some researchers speculate that initial English long-lag stops allow for a wider range

of acceptable VOT values than initial Spanish short-lag stops, and argue that the typical direction of influence from Spanish to English could be accounted for—in part—by this asymmetry (Bullock & Toribio, 2009; Olson, 2016). A similar observation can be made with early French-English bilinguals, who are substantially more likely to produce lead voicing for the English voiced stop series compared to monolinguals (Sundara et al., 2006). Lead voicing for initial “voiced” stops in North American English varieties is often absent, though not always. French, on the other hand, has consistent lead voicing for the equivalent sounds. For bilinguals, the more variable language—English—appears to be influenced by the more regular lead voicing found in French. This influence is asymmetrical, as the same group of early bilinguals continue to produce monolingual-like lead voicing in their French (Sundara et al., 2006). In this paper, we argue that asymmetries like these arise from crosslinguistic differences in category variability. We make this connection by borrowing concepts from the sound change literature.

While the tacit implication of mutual influence literature review here may read as though category convergence or maintenance are the only options, categories can also actively diverge (or dissimilate) from one another. A variety of different studies document divergence in bilingual phonetic systems (e.g., Bullock & Toribio, 2009; Flege, 1987; Flege & Eefting, 1987; Guion, 2003; Polinsky, 2018), though in some cases the distinction between divergence and maintaining contrast may be fuzzy. Early bilinguals show evidence of reorganizing their sound systems to maintain phonological contrasts across their languages. This occasionally results in phonetic distributions that split the difference between monolingual norms in either language, maintain distinctions between close contrasts in their languages, or reflect a phonetic distribution of a phonological contrast shifted somewhat in phonetic space. Any of these processes could be characterized as a kind of divergence. Guion (2003) describes this kind of scenario as the result of balancing inventory economy and phonetic dispersion in the case of the Quichua-Spanish bilingual vowel inventory. Characterizing a contrast as diverging, however, bakes in the assumption that whatever feature they use for the contrast is dynamic (e.g., subject to processes such as interference). While there are many examples of contrasts that early bilinguals have

maintained (e.g., Guion, 2003; K.-H. Kang & Guion, 2006; MacLeod & Stoel-Gammon, 2005, 2009; MacLeod, Stoel-Gammon, & Wassink, 2009; Sundara et al., 2006), there are fewer clear cases of systematic divergence. One such case could be the one described by Bullock and Toribio (2009)—which mirrors the extensive literature on bilingual VOT—where Spanish teachers, who were late Spanish learners, both shorten Spanish short-lag stops and lengthen English long-lag stops when they are in close proximity to one another in elicited lab speech. Bullock and Toribio (2009) argue that this is part of bilinguals’ rich repertoire of forms that can be deployed for any number of sociolinguistic reasons. Essentially, VOT in this case is a case of something that can dynamically diverge or converge at a speaker’s discretion, as a means of showcasing meta-linguistic awareness. In some ways, this underscores the importance of using conversational, spontaneous speech as the way of characterizing an individual’s linguistic behaviour, as it mitigates the role that performative speaking styles play in shaping our understanding of speech processes in general.

Connecting models of sound change and bilingualism

Whether characterizing the consequences of language contact as an indicator of language loss or the development of a new dialect, there has been relatively limited discussion of the underlying cognitive and psycholinguistic mechanisms that goad the linguistic transfer process (Winford, 2007). Extant models like the one described in Van Coetsem (2000) place the motivation for transfer on the dominant language, or are designed around accounting for categorical language behaviour at the morpheme level (Myers-Scotton & Jake, 2017). Prominent second language acquisition models like the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021)—a further development of the seminal Speech Learning Model (SLM; Flege, 1995)—are designed to account for the ways in which bilingual phonetic systems (re)organize over the lifespan, as first and second language input waxes and wanes. The SLM-r focuses on individual differences. While the SLM-r covers a lot of ground, the distributional precision of L1 categories is a core factor in accounting for whether an individual forms composite distributions

of “linked” categories across their languages or keeps distinct categories for each language. Distributional precision is viewed in the SLM-r as an individual-specific factor that is potentially related to auditory acuity. The prediction within the literature is that individuals with less precise categories will be more likely to have composite phonetic categories that exhibit strong(er) mutual influence. This mechanism was inspired by a series of experiments on vowel perception and production by Kartushina and colleagues, who find that individuals with compact L1 vowel categories tend to also have compact L2 categories, even following initial L2 exposure in the lab (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015). Emphasizing the role this plays in the domain of individual differences, the SML-r specifically states that L1 category precision is not shaped by language-specific phonetic factors, but is rather an exclusively endogenous factor.

As the SLM-r is intended to account for second language acquisition, the category precision mechanism for discrete or combined categories does not necessarily fit in with what is known about individuals whose bilingualism arises from infancy and early childhood. Infant bilinguals often have imprecise phonetic categories and high auditory acuity (Werker & Byers-Heinlein, 2008), which does not neatly line up with the presumed relationship between category precision and auditory acuity for adults in SLM-r (Flege & Bohn, 2021). Such early bilinguals are well-equipped to establish similar, yet distinct phonetic categories, that presumably last into adulthood (e.g., Guion, 2003; MacLeod et al., 2009; Sundara et al., 2006). What accounts for mutual influence in this population of bilinguals? The answer to this question would almost certainly shed light on the the nature of linked categories in a multilingual system. And crucially, how might that influence collectively shape a language variety under contact? These questions guide our present inquiry.

Contemporary within-language sound change research provides a mechanism—distinct from category precision—to account for why a change goes in a particular direction. Harrington and Schiel (2017) and Harrington et al. (2018) argue that when different versions of a category (e.g., within-category phonetic variation for older and younger speakers) come into contact with each other, sound change follows when one group’s distribution falls in the path of the other. That

is, the skewness and the direction of that skew are paramount, and not simply the variance or compactness of the distribution (c.f., Kartushina & Frauenfelder, 2014; Kartushina, Frauenfelder, & Golestani, 2016; Kartushina et al., 2015; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2016). To be specific, let's say we have two language models, A and B, which realize a category differently. If the direction of the variation is asymmetric such that A's distribution is skewed towards B more than B's towards A, then A will shift in the direction of B. Harrington and Schiel (2017) test this model with /u/-fronting data from older and younger speakers of British English, demonstrating that the change-in-progress dynamics of /u/-fronting can be accounted for within a cognitive model that is sensitive to distributional information without the need for social motivations to advance the change. Harrington and colleagues (Harrington et al., 2018; Harrington & Schiel, 2017) argue that the structure of variability can (at least partially) explain why a sound change occurs, and why it occurs in a particular direction, placing more explanatory power in the actuation problem (Chen & Wang, 1975; Weinreich, Labov, & Herzog, 1968).

We extend this concept to the phonetic variation underlying systems of sound categories across languages as follows. The literature on crosslinguistic phonetic influence has generally focused on the role of language dominance or mode as a means of accounting for which language undergoes more change (i.e., convergence). The malleability of segments can be modelled in a framework like the SLM-r (Flege & Bohn, 2021), but for cross-language sound linkages, the SLM-r does not necessarily account for asymmetries in the direction of influence in early bilinguals. The explanatory mechanism explored here is that—following the age-based variation exploited in Harrington and Schiel (2017)—segment variability in each of the respective languages matters for linked segments across languages within a bilingual, and that some segments are more susceptible to variation than others (as observed in Bullock & Toribio, 2009; Olson, 2016; Sundara et al., 2006). We posit that this parallels the sound change observations, and that the structure of phonetic variation is key for understanding what happens to linked sounds, whether the comparison is made in the distributions between groups or in the distributions between languages within a bilingual's mind.

Harrington and colleagues' framework hinges on perception. The perception-production feedback loop within a bilingual is likely more complicated than that of a monolingual, as bilinguals can make connections at the sub-phonemic (e.g., Barlow, 2014), phonemic (e.g., Sundara et al., 2006), lemma (e.g., Linck, Kroll, & Sunderman, 2009), and social levels (e.g., Szakay, Babel, & King, 2016). For example, Yao and Chang (2016) provide evidence for a reversal of the merger of /e/ and /ɛ/ in Shanghainese, which they argue was necessarily supported by perceptual connections at the lexical-phonetic level, not only at the allophonic level, as the unmerging was not a wholesale shift of the sounds in every environment, but tied to particular lexical sets. The crosslinguistic connections at the word level facilitated and spurred this particular unmerging. In bilinguals, the perceptual links between segments may not actually be at a sensory-based perceptual level—at least not exclusively—but rather comprise part of the cognitive organization of linguistic structure (e.g., see discussion of abstraction and linkage in Chang, 2015). The voiced/unaspirated stop series in English and Cantonese, for example, are linked by their surface-level similarities in onset position, where they can both surface as short-lag stops but participate in different phonological patterns within each language, constrained by language-specific phonotactics (Bauer & Benedict, 2011). In a separate example, /p t k/ can occur in English in initial and final positions, as well as in consonant clusters, whereas the equivalent long-lag Cantonese stop series /p^h t^h k^h/ is licit syllable-initially in Cantonese and does not occur in consonant clusters. Equivalent is intended here in the sense of Chang's (2015) relative phonetics, as the two series occupy the long VOT niche in each language, regardless of how they might be theoretically specified. Additionally, Cantonese stops occur in final position, though the underlying form is ambiguous as all final stops are unreleased at the surface level. What is crucial here, however, is that the locus of these connections is within the bilingual and the consequence of the distributional dynamics à la Harrington and colleagues' model may come to fruition.

Cantonese and Cantonese-English bilingualism

In this study, we consider spontaneous speech from a corpus of early Cantonese-English bilinguals—the corpus is described in more detail below, and in Johnson et al. (2020). Here we provide a brief summary of Cantonese and English in Hong Kong and Canada, as these speech communities are most connected to the bilinguals studied here.

Cantonese is a Yue language with over 80,000,000 speakers; the largest Cantonese-speaking population is centered in the Guangdong province in southern China (Eberhard, Simons, & Fennig, 2020). A snapshot of the obstruent contrasts in word-initial position receiving our focus in this paper is presented in Table 1. The IPA symbols used to denote these obstruents are taken from the literature on English and Cantonese (Bauer & Benedict, 2011; Zee, 1999), though there are relevant deviations from these symbolic choices for Cantonese (Matthews & Yip, 2013). The Cantonese word-initial series denoted here by /p t k/ is described as voiceless unaspirated, and Bauer and Benedict (2011) explicitly connect this series to the phonologically voiced stop series in English: /b d g/. They highlight that while the English voiced stops are variably voiced, the Cantonese equivalent series is never voiced. Matthews and Yip (2013) opt to use the symbols /b d g/ to describe the voiceless unaspirated series in Cantonese. Together, this suggests that the Cantonese voiceless unaspirated word-initial stop series and English's voiced stop series may be linked for bilingual speakers. Such linkages are more often informed by abstract, relative phonetic similarity than raw acoustic-auditory similarity (see discussion in Chang, 2015), though both abstract and physical similarity seem to align for Cantonese and English in this case. The voiceless aspirated series /p^h t^h k^h/ in Cantonese and the voiceless series in English /p t k/ align in a more straightforward manner, given that in both languages these sounds are described as voiceless and aspirated (Matthews & Yip, 2013; Mielke & Nielsen, 2018), and occupy similar relative positions in their respective inventories, regardless of the phonological features ascribed to the series in any given theoretical framework. Cantonese phonology permits stops in syllable-final position and these are described categorically as unreleased in grammars (Bauer & Benedict, 2011; Matthews & Yip, 2013). A caveat for the coronal series is that the Cantonese

coronals are typically described as more dental (Bauer & Benedict, 2011), whereas English is more alveolar. Cantonese also has a labialized velar series which does not have a phonological match in English and is therefore neither discussed nor addressed further in this work.

Table 1

Stop places of articulation in North American English and Cantonese that are relevant for a cross-inventory comparison represented with IPA symbols typical in the literature.

	Labial	Coronal	Dorsal
North American English	b p	d t	g k
Cantonese	p p ^h	t t ^h	k k ^h

Multilingual competence is characteristic of Cantonese speakers within and outside the linguistic homeland of Cantonese. First language speakers of Cantonese also often speak Mandarin (Putonghua) and English, particularly those in Hong Kong, which was colonized and occupied by the British for nearly 100 years. Due in part to its colonial history, English is a lingua franca in Hong Kong. Hong Kong English is typically characterized as an emergent variety (Schneider, 2003; Sung, 2015), and while there is no consensus on whether Hong Kong English can yet be labelled as a codified variety, scholars assert there is an identifiable native Hong Kong English accent (Hung, 2000; Sung, 2015). The literature on Hong Kong English consistently describes a contracted phoneme inventory compared to Received Pronunciation (RP), the British English variety to which it is most often compared (Deterding, Wong, & Kirkpatrick, 2008; Sewell & Chan, 2010; Sung, 2015). While it is frequently noted that English voiced fricatives are often devoiced in Hong Kong English (Bolton & Kwok, 1990; Hung, 2000; Leung & Brice, 2012; Sewell & Chan, 2010; Sung, 2015), there is less discussion of the stop series. Going back several decades, Bolton and Kwok (1990) note that word-final stop consonants are sometimes deleted, but also include an example of post-vocalic final /t/. In the context of final consonant clusters in Hong Kong English, Deterding et al. (2008) report that final /t/ is deleted 47% of the time and /k/ 59% of the time. Comparing Hong Kong English to RP, Hung (2000) uses word list data from a

small sample of university students ($n = 15$) to characterize various aspects of Hong Kong English phonology. He describes word-initial English voiced stops as not being truly voiced in Hong Kong English, noting that the voiced stop series is distinguished from the voiceless stop series in terms of the VOT lag, as is the case in many varieties of English. Hung (2000) also characterizes Hong Kong English voiced and voiceless stops in intervocalic position as being differentiated by VOT lag, as opposed to the voicing contrast typical of this environment across many English varieties. For stops in word-final position, he describes Hong Kong English voiced stops as being unreleased and voiceless stops including an aspirated release in his word list data. This, he describes, is a systematic phonology with influence from Cantonese, but not one that can be attributed to wholesale application of Cantonese phonology into an English system. Setter, Wong, and Chan (2010) describe Hong Kong English as heterogeneous, though it is worth highlighting that all languages exhibit variation (Laver & Trudgill, 1979).

The Cantonese-speaking population in Canada is a diaspora that has extended roots (Yu, 2013), but the evidence as to whether the English of Cantonese-English bilinguals in Canada carries any unique signatures comparable to Hong Kong English is somewhat mixed. Controlled laboratory-elicited read sentences suggest pronunciation differences, if extant, are not perceptible to listeners (Babel & Russell, 2015), but more natural semi-spontaneous speech from early Cantonese-English bilinguals is categorized as associated with Chinese identity at levels well above chance (P. Wong & Babel, 2017). Hoffman and Walker (2010) examined the speech patterns of ethnically Chinese individuals in Toronto who were either first generation Cantonese-dominant speakers ($n = 10$) or second or third generation English-dominant speakers from historically Cantonese-speaking families ($n = 23$). Speech data were collected via interviews conducted by individuals who were “themselves members of the relevant communities” (Hoffman & Walker, 2010, p. 45). T/d deletion was examined in word-final position using a sub-sample of their speakers and those individuals’ speech samples such that they had 50-100 tokens per talker in select word-final consonant clusters. T/d deletion was categorically coded as deleted or present. Both first and second/third generation ethnically

Chinese individuals showed more deletion than the control white monolingual speakers descended from British Isles populations ($n = 20$). Hoffman and Walker (2010) interpret their results as suggesting language transfer in the speech of first-generation speakers, based on the likelihood of patterns in particular phonological environments. For second/third generation ethnically Chinese individuals, the overall rate of t/d deletion is higher than the other demographic groups. They consider this weak evidence for ethnolects, as the patterns do not persist across generations.

Research question

Returning to our focus, we ask: when linked cross-language segment categories differ in the degree of phonetic variation for a given phonetic dimension, does the less variable category constrain or guide production of the equivalent—but more variable—category in a different language? This question hinges on there being an asymmetry in the shape of distributions for linked sounds, and is thus notably distinct from the notion of compactness (as formulated in Flege & Bohn, 2021; Kartushina & Frauenfelder, 2014), for which the compact shape in the L1 is mirrored by a compact shape in the L2. Thus adapting the framework from Harrington and colleagues (2018; 2017), we examine whether the structure of similar cross-language categories affects how crosslinguistic influence is manifested for a bilingual; and, ultimately, how pronunciation variants may be introduced to a community through its bilingual members. To assess whether phonetic variation in stop categories is susceptible to more transient cross-language effects, we consider effects of proximity to a language switch (from Cantonese to English) and language mode.

We consider two variable patterns across the studies reported in this paper. These studies are united in that they focus on phonetic variation in stop categories in English, but have conceptually different sources of Cantonese *invariance*. We acknowledge that it is unfortunate that we can only address variation in English and not in Cantonese, but our abilities to examine variation in Cantonese are stymied by the lack of an appropriate, age- and style-matched Cantonese language corpus—that lacks English influence—to which we can compare early

Cantonese-English bilinguals. We leave such a study to future work.

The first study focuses on sub-phonemic variation in the realization of voiced stops in English. The distribution of phonation during the closure for English /b d g/ is illustrated in Figure 1a, where phonation is visualized here as a unitless less-to-more dimension. The English voiced stop series is highly variable in whether the voiced stop series truly shows phonation during the closure, ranging from voiceless unaspirated to containing phonation during the entire closure (Davidson, 2016; Jacewicz, Fox, & Lyle, 2009). Conversely, the distribution for the equivalent series in Cantonese is more restricted. While the Cantonese stops /p t k/ are consistently described as voiceless unaspirated in single word production with short-lag VOT (Lisker & Abramson, 1964; I. Y. H. Tsui & Ciocca, 2000), and are usually transcribed as /p t k/, the actual realization of these stops is more variable. Cantonese-acquiring children produce highly variable VOT, including many negative values, indicating phonation during the closure (Clumeck et al., 1981). In adults, Cantonese syllable fusion—a gradient prosodically-conditioned reduction process—can lead to variable amounts of phonation in the unaspirated stop series as well (W. Y. P. Wong, 2006). Thus, while the descriptive literature on Cantonese states that these stops are produced in a voiceless unaspirated fashion (Bauer & Benedict, 2011), such descriptions may be more prescriptive than descriptive; hence, we illustrate phonation during these stops in Cantonese as a distribution centered around the no voicing end of the continuum, as opposed to a single discrete peak at no voicing.

The second study examines the release behaviour of voiceless final stops in English. Cantonese is described as categorically not releasing voiceless stops in syllable-final position, but oral occlusion is assumed (Bauer & Benedict, 2011; Matthews & Yip, 2013), while English release rates and degree of constriction are highly variable (Byrd, 1993; Davidson, 2011; Y. Kang, 2003). We expand on the discussion of final stop variability in the introduction for Study 2. This is modeled in Figure 1b, where the horizontal axis represents a selection of release behaviours—including deletion, unreleased, and released outcomes. The vertical dimension reflects categorical outcomes for the portion of the stop where closure occurs. Neither of these

dimensions are to be interpreted as continuous, rather we intend to show categories of patterns. The offset between Cantonese and English squares is to aid in visualization of the two patterns and is not intended to suggest shifts in the patterns. We model Cantonese /p t k/ as falling more heavily into the non-released outcome bins—though, given the nature of spontaneous speech, quasi-releases likely occur at least some portion of the time. This schematic glosses over the outcome of the presumed patterns in syllable fusion which includes deletion and other forms of lenition. English, along both of these dimensions, is highly variable. Thus, our examination of this pattern probes how a more categorical phonological pattern from one language (Cantonese) may constrain a more variable pattern from another language (English). Further discussion of these phenomena is provided in the introductions to Study 1 and Study 2, respectively.

While the schematic distributions for phonation and word-final release rates for Cantonese differ in terms of their described categoricity or discreteness within Cantonese, they are both substantially less variable than English. Following the sound change models of Harrington and Schiel (2017) and Harrington et al. (2018), we predicted that the English of Cantonese-English bilinguals would be shifted towards the Cantonese distribution, as the more variable distribution encroaches onto the less variable distribution. That is, we predicted that English /b d g/ as produced by Cantonese-English bilinguals would have less voicing, and English /p t k/ in word-final position would have lower release rates as a result of mutual influence within the bilingual mind.

Materials

The studies described in the paper make use of three spontaneous speech corpora. The first comprises Cantonese-English bilingual speech, and the latter two are monolingual control corpora.

Bilingual SpiCE corpus

The SpiCE corpus of bilingual speech in Cantonese and English includes high-quality lab recordings of conversational interviews with 34 early bilinguals in both languages (Johnson,

Figure 1. Schematizations of the distributions for (a) stop voicing during closure, and (b) voiceless stop closure and release rates in word-final position. In panel (a), the English distribution covers a much wider range of values compared to the Cantonese series. In panel (b) English is characterized by a wide range of variability for degree of both closure and release with boxes of the same size in each category, while Cantonese is assumed to primarily exhibit lenition-related variability and categorically lack a release. Given the highly variable nature of spontaneous speech, there would likely be occasional releases—as such, the distribution across discrete categories for Cantonese is depicted as heavily skewed. The squares in (b) are offset to ease visualization and do not suggest shifted patterns.

2021), and is described in greater detail in Johnson et al. (2020).¹ The Cantonese and English interviews were conducted by a pair of bilingual undergraduate research assistants, and lasted approximately 30 minutes each. The order of interviews was counterbalanced across participants within a two-interview session. SpiCE recordings are accompanied by orthographic, force-aligned transcripts in English and Cantonese.²

¹ Documentation for the corpus can be found at <https://spice-corpus.readthedocs.io/>.

² Cantonese transcription was primarily done in Cantonese characters, with Jyutping romanization restricted to cases for which there is no character.

Half of the 34 speakers in the corpus reported their gender as female, and the other half as male. Speakers were generally university-aged ($Mdn = 21$, $M = 22.4$, $SD = 4$), and learned both Cantonese ($Mdn = 0$, $M = 0.03$, $SD = 0.17$) and English ($Mdn = 0$, $M = 1.32$, $SD = 1.77$) from a young age. Twenty-one individuals report being simultaneous bilinguals, while the remaining 13 report sequential bilingualism. The sequential bilinguals all reported Cantonese as their first language, with English acquisition beginning between the ages of two and five. Speakers reported high speaking proficiency in both English (Good/Excellent, $n = 33$; Fair, $n = 1$) and Cantonese (Good/Excellent, $n = 29$; Fair, $n = 5$). Self-ratings of Cantonese and English speaking proficiency were equivalent for 16 speakers. Six speakers provided higher speaking proficiency ratings for Cantonese, and 12 gave a higher rating for their English. The linguistic profiles of the individuals in the SpiCE corpus are thus variable, but all are high proficiency bilinguals.

Most speakers reported knowledge of Mandarin Chinese, and a variety of other languages—proficiency and age of acquisition for these other languages varied widely. For all speakers in the SpiCE corpus, English and Cantonese were among the earliest learned and highest proficiency languages. . Most of the SpiCE bilinguals categorize their English variety as North American ($n = 16$),³ Hong Kong English ($n = 4$), or having influence from both varieties ($n = 9$). The rest report other varieties and combinations (e.g., British or Singaporean; $n = 5$). The varieties of English are also reflected in where the speakers grew up. Most report being raised in Canada ($n = 11$), Hong Kong ($n = 9$), or a combination of the two ($n = 11$). The remaining speakers grew up in Singapore ($n = 1$), Canada and Singapore ($n = 1$), or Hong Kong and Mainland China ($n = 1$). The sample of early bilinguals in SpiCE is therefore heterogeneous, but is nevertheless representative of the Cantonese-speaking speech community in the region.

³ North American here includes those who responded with “American” or “Canadian”—a more fine-grained self-reporting of English variety was not included.

Monolingual Comparison Corpora

We compare SpiCE with two monolingual comparison corpora: a collection of sociolinguistic interviews with Vancouver, BC English speakers (Swan, 2016), and the Buckeye Corpus of Conversational Speech (Pitt et al., 2005). Both collections comprise high-quality recordings of spontaneous monolingual English speech. These two comparison corpora were selected to give a comparison to a widely-studied variety of English (Pitt et al., 2005), as well as the regional variety of English that many of the SpiCE bilinguals speak and all are exposed to (Swan, 2016). In both cases, it is not particularly important that the individuals are monolingual, *per se*, but rather that they (non-exhaustively) represent local and non-local English-speaking communities that do not have early multilingual experience with languages that would confound our line of inquiry.

The local spontaneous speech corpus collected by Swan (2016) includes sociolinguistic interviews with 40 speakers ranging in age from 18 to 36, and was recorded in 2014. We refer to this corpus as the Corpus of Transnational West Coast English (CTWCE). While this corpus also includes the speech of speakers from Seattle, WA, we only examined the speech of monolingual speakers from Vancouver ($n = 15$). Of this group, a majority were female ($n = 9$), and the ages of the speakers were broadly similar to those in the SpiCE corpus (19–36 years of age). This corpus has been orthographically transcribed and force-aligned with the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017).

The Buckeye corpus is a collection of conversational interviews with 40 Caucasian monolingual English speakers from central Ohio, and data collection was completed in 2000 (Pitt et al., 2005). The interviews are transcribed and also include hand-corrected phonetic annotations. For this paper, we examine only the speech of the younger half of the Buckeye corpus (under age 30; $n = 20$). Half of the speakers were female, and half were male.

The inclusion of both these corpora allows us to be clear about what is a local English pattern that SpiCE and CTWCE share, versus what is a bilingual pattern that is not shared between SpiCE and either of the monolingual corpora. We note, however, that while these

corpora all comprise spontaneous speech, they nonetheless have some key differences. Buckeye participants were interviewed by postdoctoral researchers and not told in advance that the purpose of the recording was for linguistic analysis. CTWCE, on the other hand, features sociolinguistic interviews recorded by Swan (2016), who was a graduate student at the time. Additionally, these interviews included meta-linguistic discussion about dialect and language ideology. SpiCE is perhaps the most peer-to-peer of the three corpora, and while some aspects of the design were modeled after the Buckeye corpus, participants knew the purpose for recording (i.e., creating a bilingual corpus). In any case, all three corpora comprise conversational interview speech produced by younger adults, which renders them comparable.

General Method

We consider two cases studies where some degree of Cantonese influence is predicted to occur—intervocalic /b d g/ voicing (Study 1) and word-final /p t k/ release rates (Study 2). In both cases, we restrict the analysis to the set of environments where we are *least* likely to see the predicted result in spontaneous speech, but where a difference between monolinguals and bilinguals is likely to be compelling, if present.

Using the transcripts from each corpus, we identified target instances of /b d g/ and /p t k/ using the PolyglotDB Python package (McAuliffe, Stengel-Eskin, Socolof, & Sonderegger, 2017), and compared the SpiCE bilinguals' productions to each of the monolingual comparison corpora.⁴ The analyses described in the following studies report the results of mixed effects models for each study—cross-corpus comparisons, and a within-SpiCE analysis examining factors that influence language mode.

⁴ SpiCE and the sociolinguistic interviews from Swan (2016) were force-aligned using the Montreal Forced Aligner with the built-in English acoustic model and pronunciation dictionary (McAuliffe, Socolof, et al., 2017)

Study 1: Intervocalic /b d g/ voicing

The production of voicing for /b d g/ is highly variable within and across varieties of North American English, and across phonological environments within a variety or dialect (Davidson, 2016; Jacewicz et al., 2009). While these segments typically lack voicing in phrase-initial position (Keating, 1984), they are likely to exhibit full or partial voicing in other environments (Davidson, 2016), with the proportion of voicing above 90% in intervocalic position for some dialects (Jacewicz et al., 2009). Cantonese, on the other hand, lacks any phonemically voiced obstruents (Matthews & Yip, 2013). Bauer and Benedict (2011, p. 17) describe the initial stops in Cantonese as categorically voiceless and unaspirated, and suggest that American English speakers learning Cantonese “keep in mind the stop consonants of American English [b], [d], and [g] when they occur in word-initial position and refrain from voicing them.”

In this sense, English exhibits a wide range of acceptable variability, including production of /b d g/ without voicing, which parallels the voiceless unaspirated stop series in Cantonese. Given this disparity in variability, we predicted that bilinguals would produce less voicing during closure compared to monolinguals in English word-medial intervocalic productions of /b d g/—one of the environments where voicing is most likely (Davidson, 2016). This follows from the parallels drawn with Harrington and colleagues (2018; 2017) in the introduction.

Methods

We identified all instances of phrase-medial intervocalic /b d g/, and used manual inspection to identify gross errors and instances where the target segment was deleted.⁵ This is an important step, considering that we used force-aligned transcripts for two of the three corpora. Additionally, by using the hand-corrected phonetic transcripts of the Buckeye corpus, instances of deletion in that corpus were not included in the initial sample. To this end, a single research assistant coded

⁵ The Buckeye’s more narrow surface transcription was used, compared to the force-aligned transcripts for SpiCE and CTWCE. This means that what would likely be called a vowel-nasal sequence in SpiCE and CTWCE is a nasalized vowel in Buckeye and thus included in the analysis.

all target items as: (1) good, (2) deletion, or (3) transcript error. Additionally, the research assistant indicated their confidence, identifying if an item needed checking. Items that needed checking were coded by the first author (574 of 17,093). Prior to analysis, items were discarded if the original coder and author disagreed (1.4%), and if the coder labelled it confidently as a deletion (2.6%) or transcript error (1.9%). A total of 16,621 items were coded as “good” and used in the final analysis. Table 2 summarizes the counts used in the final set.

Table 2

Counts of /b d g/ across the three corpora included in the final analysis.

	/b/	/d/	/g/
SpiCE	1736	2958	1344
CTWCE	549	757	270
Buckeye	2656	4451	1900

With this data set, we measured the proportion of voiced frames between the onset and offset, using Praat’s *Voice Report* (Boersma & Weenink, 2020) implemented with the *Parselmouth* Python package (Jadoul, Thompson, & de Boer, 2018). For both male and female talkers, the pitch floor was set to 50 Hz. The ceiling was set to 400 and 500 Hz, respectively.

Results

We analyzed the data using a generalized linear mixed effects model with a beta response distribution and logit link function. The model was implemented with the *glmmTMB* R package (Brooks et al., 2017). Beta regression is one of the recommended choices for modeling a proportional response variable that derives from a bounded continuous process (see discussion in Smithson & Verkuilen, 2006). Additionally, the selection of the logit link function facilitates a log-odds interpretation analogous to logistic regression. The model formula used was: *Proportion Voiced* ~ *Corpus* × *Place* × *Duration* + *Position* + *Stress Precedes* + *Stress Follows* + *Word*

Frequency + (1 | Word) + (1 + Place + Duration | Talker).⁶ The variables in the model were calculated and coded as indicated below. Many of the variables were implemented with weighted effect coding, which facilitates the interpretation of effects and interactions when the levels are not balanced, as is often the case with corpus data (Nieuwenhuis, te Grotenhuis, & Pelzer, 2017). In weighted effect coding, levels are compared against a weighted mean for that variable.

Proportion voiced is a continuous value ranging between zero and one, and represents the proportion of voiced frames in the target (i.e., voicing during closure). While for any given instance the proportion is based on a count of total frames, the underlying process is continuous. In order to facilitate beta regression, values were “squeezed” in the analysis, such that $y'' = \frac{y'(N-1)+0.5}{N}$, ensuring there were no ones and zeroes in the analysis. This is the dependent variable.

Corpus is a categorical fixed effect variable identifying the corpus an instance came from.

Corpus was simple coded with three levels—SpiCE, CTWCE, and Buckeye—with SpiCE as the reference level. This ensures that the comparisons of interest were made (SpiCE vs. CTWCE and SpiCE vs. Buckeye), and that the contrasts sum to zero, which facilitates the interpretation of variables involved in interactions. Corpus is the primary parameter of interest in this study.

Place is a categorical fixed effect variable identifying the place of articulation of the stop that was measured with levels B, D, and G. Place was weighted effect coded, with G as the reference level [*Place B*: B = 1, D = 0, G = −1.41; *Place D*: B = 0, D = 1, G = −2.32]. Place (G) was selected as the reference level because we anticipated that the dorsal stop would behave most similarly across the corpora due to aerodynamic constraints on voicing at that place of articulation.

Duration of the target segment in seconds was centered and standardized—it is a continuous fixed effect variable. Duration is included as a control fixed effect, as the proportion of

⁶ The model failed to converge with a Place × Duration interaction in the random effects structure.

voicing is expected to decrease as duration increases, due to the aerodynamic constraints of maintaining voicing during obstruents.

Position is a binary fixed effect variable indicating whether the target occurred word-initially or not. Position was weighted effect coded [Initial = 1, Not Initial = -1.50]. Position was included as a control fixed effect, as initial stops were expected to surface with less voicing (Jacewicz et al., 2009).

Preceding stress is a binary fixed effect variable indicating whether there was stress on the preceding vowel (primary or secondary stress were both included as stressed). Preceding stress was weighted effect coded [True = 1, False = -1.94].

Following Stress is a binary fixed effect variable indicating whether there was stress on following vowel (again, primary or secondary stress were both included as stressed). Following stress was weighted effect coded [True = 1, False = -2.88].

Word Frequency was tabulated, log-transformed, centered and standardized from the corpora (as in Tanner, Sonderegger, & Wagner, 2017). This was done on a within-speech community basis, such that SpiCE and CTWCE transcripts were considered together (both are local speech), and transcripts from both young and old Buckeye corpus talkers were used (non-local). This ensures that regional vocabulary is not under- or misrepresented. Word frequency is a continuous fixed effect variable. It was included as a control fixed effect, as frequency has been shown to interact with various lenition processes (e.g., Tanner et al., 2017).

Word indicates the word that a target occurred in. Note that full word forms are used. The model includes random intercepts for Word.

Talker indicates the individual who produced the target. The model includes random intercepts for Talker, with by-Talker random slopes for Place and Duration.

While it may be clear from the above list that the analysis will center on the role of Corpus—accounting for Place—we include a variety of other parameters to control for prosodic and probabilistic aspects that are known to impact a wide variety of processes in spontaneous speech. This decision follows from prior corpus-based work, such as work by Tanner and colleagues on the topics of coronal deletion (Tanner et al., 2017) and stop voicing during closure (Tanner, Sonderegger, & Stuart-Smith, 2020). We discuss the expected outcomes of these variables in the context of reporting our results and the interim discussion.

The beta regression model returned with a significant intercept [$\beta = 1.85$], indicating that items in the SpiCE corpus are likely to have a higher proportion of voicing when all other variables are at their zero point—the weighted mean. Recall that the model uses a logit link function, and that results are interpreted on a log-odds scale. All of the model’s fixed effects are reported in Table 3, and we describe significant results in text (at a significance level of $p < 0.05$), reporting parameter estimates. Refer to the table for *SE*, *t*, and *p* values.

There were significant differences between the SpiCE corpus and both monolingual corpora, though the difference emerged in different ways. The estimate for the comparison between SpiCE and Buckeye [$\beta = 0.25$] indicates that Buckeye items were more likely to have a higher proportion of voicing. This was particularly true for Place (B), both across the board [$\beta = 0.16$], and specifically in the Place (B) $\times$ Corpus (Buckeye) interaction [$\beta = 0.10$]. This pattern is apparent in Figure 2, which shows clear differences in proportion voiced at all places of articulation for the SpiCE-Buckeye comparison.

While the effect of Duration indicates that longer stops likely have a lower proportion of voicing [$\beta = -0.56$], this effect was attenuated in the case of the Buckeye corpus [$\beta = 0.18$], and for items with Place (B) more generally [$\beta = 0.13$]. The effects of duration are depicted in Figure 3, with shallower group mean linear smooths for Buckeye stops compared to SpiCE across the board. Together, these results suggest that the Buckeye and labial stops exhibit higher proportions of voicing more generally—a conclusion that is corroborated by Figures 2 and 3.

In the case of the CTWCE corpus, the differences with SpiCE surface only in the

Table 3

Generalized linear mixed effects model output for the /b d g/ comparison across corpora in Study 1. Estimates are to be interpreted as log-odds, given the beta distributed dependent variable and logit link function.

Parameter	Estimate	SE	<i>t</i>	<i>p</i>
Intercept	1.85	0.04	41.17	<0.001
Corpus (CTWCE)	0.10	0.10	0.92	0.36
Corpus (Buckeye)	0.25	0.09	2.74	0.006
Place (B)	0.16	0.03	5.92	<0.001
Place (D)	-0.10	0.02	-5.73	<0.001
Duration	-0.56	0.03	-21.72	<0.001
Position (Initial)	-0.10	0.01	-8.61	<0.001
Stress Precedes (True)	0.03	0.01	3.73	<0.001
Stress Follows (True)	-0.04	0.01	-5.49	<0.001
Word Frequency	0.04	0.01	3.10	0.002
Corpus (CTWCE) × Place (B)	0.13	0.06	2.09	0.04
Corpus (Buckeye) × Place (B)	0.10	0.05	2.22	0.03
Corpus (CTWCE) × Place (D)	-0.11	0.04	-2.77	0.006
Corpus (Buckeye) × Place (D)	-0.05	0.03	-1.66	0.10
Corpus (CTWCE) × Duration	0.09	0.06	1.32	0.19
Corpus (Buckeye) × Duration	0.18	0.05	3.49	<0.001
Place (B) × Duration	0.13	0.02	6.28	<0.001
Place (D) × Duration	0.01	0.01	1.00	0.32
Corpus (CTWCE) × Place (B) × Duration	-0.01	0.06	-0.22	0.82
Corpus (Buckeye) × Place (B) × Duration	0.003	0.03	0.08	0.94
Corpus (CTWCE) × Place (D) × Duration	0.004	0.04	0.11	0.91
Corpus (Buckeye) × Place (D) × Duration	-0.02	0.02	-1.00	0.32

interactions, such that Place (B) in the CTWCE corpus is more likely to have a higher proportion of voicing compared to SpiCE [$\beta = 0.13$]. Place (D) is overall likely to have a lower proportion of voicing compared to the weighted mean [$\beta = -0.10$]. Additionally, the Place (D) $\times$ Corpus (CTWCE) interaction [$\beta = -0.11$] and the simple effect for Place (D) indicate that—as suggested by the empirical results in Figure 2—SpiCE and CTWCE do not differ substantially in the proportion of /d/ voicing. Figure 2 facilitates the interpretation of the Place (D) interaction, and shows a general pattern of CTWCE falling in between SpiCE and Buckeye.

As all of the categorical fixed effect variable contrasts sum to zero, it is possible to get reasonable main effect estimates from the model (Nieuwenhuis et al., 2017). In light of this, more frequent words were slightly more likely to be voiced [$\beta = 0.04$]. The proportion of voicing was also affected by prosodic environment, as anticipated. Stops in initial position were less likely to be voiced [$\beta = -0.10$], and /b d g/ items which preceded a stressed vowel were less likely to be voiced [$\beta = -0.04$], while post-tonic /b d g/ are more likely to be voiced [$\beta = 0.03$].

Figure 2. Proportion voiced for /b d g/ for the three corpora.

SpiCE-only /b d g/ model

A second generalized linear mixed effects model focuses on just the target items occurring in the SpiCE corpus. As before, beta regression with a logit link function was used to model the

Figure 3. Proportion voiced by Duration for /b d g/ for the three corpora. Lines are linear smooths. Each thin line is for a single participant, and the thicker lines represent group means.

proportion of voicing during closure. The goal of this model was to understand how language mode impacts the proportion of voicing for /b d g/. Language mode is captured here in the variables described below—Code-switch Distance and Interview Order. The *glmmTMB* formula was: *Proportion Voiced* $\sim$ *Place* $\times$ *Duration* $\times$ *Code-switch Distance* + *Interview Order* + *Position* + *Stress Precedes* + *Stress Follows* + *Word Frequency* + (*1* | *Word*) + (*1* + *Place* + *Duration* | *Talker*).⁷ Variables were defined in the same way as the 3-corpus model, with the exception of two new fixed effect variables:

Code-switch Distance is a continuous variable that captures proximity to a preceding code-switch into Cantonese. It was measured in seconds, and subsequently standardized. Prior work with spontaneous speech finds that crosslinguistic influence in code-switching is anticipatory, rather than perseverative (Fricke et al., 2016), though both directions of influence have been found in experimental settings (Bullock & Toribio, 2009). For instances of /b d g/ that did not have a preceding switch, we used the onset of the English recording session, which then includes the elicitation of English sentences and the narration of an English story. We acknowledge that the onset of the English recording session is not a

⁷ The model failed to converge with a *Place* $\times$ *Duration* interaction in the random effects structure.

perfect proxy for the nearest code switch, as we have no knowledge of when the speaker would have last spoken Cantonese either as a code-switch or not.

Interview Order is a binary variable indicating whether or not the Cantonese interview preceded the English interview in the SpiCE recording session.⁸ Interview Order was weighted effect coded [Cantonese First = 1, English First = -1.03].

Additionally, weights were re-computed for all other categorical variables with weighted effect coding—Place [*Place B*: B = 1, D = 0, G = -1.29; *Place D*: B = 0, D = 1, G = -2.20], Position [Initial = 1, Not Initial = -1.53], Following Stress [True = 1, False = -2.81], and Preceding Stress [True = 1, False = -2.06].

The model intercept was significant [$\beta = 1.74$], indicating that stops in the SpiCE corpus were likely to have a high proportion of voicing. For the parameters that this model shares with the 3-corpus model, the output is largely the same, and detailed model output is reported in Table 4. This includes significant effects for Place (B) [$\beta = 0.09$], Place (D) [$\beta = -0.05$], Duration [$\beta = -0.62$], and Place (B) $\times$ Duration [$\beta = 0.12$]. There were also significant effects of Position (Initial) [$\beta = -0.14$], Word Frequency [$\beta = 0.08$], and Stress [Precedes: $\beta = 0.03$, Follows: $\beta = -0.06$]. These effects can be interpreted in the same fashion as for the 3-corpus model reported in the previous section.

There was not a significant effect of Interview Order in the model, despite the consistent direction depicted in Figure 4. It is worth noting that this fixed effect was close to the threshold for significance [$\beta = -0.09$, $p = 0.06$], but also that there was considerable between-subject variability. Additionally, there was not a significant effect of Code-switch Distance in the SpiCE corpus, an outcome that is not unexpected given the variation depicted in Figure 5. This is likely due to the relatively long distance from any given token to the nearest code-switch [$n = 372$,

⁸ Note that while the order of interviews was counterbalanced across participants, counterbalancing did not extend to other demographic factors. For example, 10 of the 17 participants who completed the Cantonese interview first were female.

Table 4

Generalized linear mixed effects model output for the [b d g] comparison within the SpiCE corpus in Study 1.

Parameter	Estimate	SE	<i>t</i>	<i>p</i>
Intercept	1.74	0.05	32.25	<0.001
Code-switch Distance	-0.02	0.02	-1.17	0.24
Place (B)	0.09	0.04	2.14	0.03
Place (D)	-0.05	0.03	-1.83	0.07
Duration	-0.62	0.03	-17.88	<0.001
Interview Order (Cantonese First)	-0.09	0.05	-1.88	0.06
Position (Initial)	-0.14	0.02	-7.19	<0.001
Stress Precedes (True)	0.03	0.01	2.43	0.02
Stress Follows (True)	-0.06	0.01	-5.08	<0.001
Word Frequency	0.08	0.03	3.08	0.002
Code-switch Distance × Place (B)	-0.02	0.03	-0.93	0.35
Code-switch Distance × Place (D)	0.03	0.02	1.74	0.08
Code-switch Distance × Duration	0.005	0.02	0.25	0.80
Place (B) × Duration	0.12	0.03	4.49	<0.001
Place (D) × Duration	0.03	0.02	1.59	0.11
Code-switch Distance × Place (B) × Duration	0.02	0.02	0.62	0.53
Code-switch Distance × Place (D) × Duration	-0.01	0.02	-0.73	0.46

Figure 4. Proportion voicing during closure for /b d g/ within the SpiCE corpus, by Interview Order.

Figure 5. Proportion voicing during closure by Code-switch Distance for /b d g/ in the SpiCE corpus. Lines are linear smooths, where thin lines represent participants, and thicker lines represent group means by Place.

$Mdn = 345.7s$]. Prior literature documenting proximity effects typically considers a window well under 10 seconds (e.g., Fricke et al., 2016).

Interim discussion: /b d g/

These results largely follow predictions that individuals in the SpiCE corpus voice phonologically voiced stops less than individuals in the monolingual English corpora, which we interpret as evidence for Cantonese influence on the English spoken by these early Cantonese-English bilinguals. Recall that the selected phonological environment is where we expect the *most* voicing—a voiced stop immediately preceded and followed by vowels in peer-to-peer conversational speech. While the effect reported on here is relatively small, the environment disfavours our result, and thus serves as a proof of concept. Differences in stop voicing may be much greater in magnitude in phonological environments which disfavour voicing and communicative contexts that impact factors like the propensity for hyper- and hypo-articulation.

While we observe what we might call a global effect of Cantonese, we did not find a significant effect of language mode, as Interview Order narrowly missed the significance threshold. Given this proximity, and that the data trended in a direction consistent with our hypothesis, it may be possible that this particular variable was not sensitive to language mode, even if the bilingual individuals had entered into a particular language mode during the interview. Given the high degree of variability across individuals, we posit that the effects of Interview Order may warrant individual-specific exploration at a future point. While the effect of Interview Order might be considered marginal, there was a robust lack of an effect of Code-switch Distance, visualized in Figure 5. Target /b d g/ items that are closer to a Cantonese switch were not less likely to be voiceless, as shown by the flat line, summarizing the group behaviour. The temporal proximity of target /b d g/ and switches was rather long, as switches were infrequent for most of the SpiCE talkers. Thus, this lack of effect is not surprising. Figure 5 also shows individuals' mean lines, which are highly variable, suggesting an investigation of individual differences may

prove fruitful.

The finding that stops in initial position were less likely to be voiced replicates the finding based on read sentences from Jacewicz et al. (2009) that voiced stops in prominent positions or subject to emphasis are less likely to be voiced with spontaneous speech. Like Davidson (2016), stops that followed a stressed vowel were more likely to be voiced than those that followed an unstressed vowel. Stops that preceded an unstressed vowel were more likely to be voiced than stops that preceded a stressed vowel.

Study 2: Word-final /p t k/ release rates

Final consonant deletion and lenition—particularly for coronal stops—has been widely studied within the sociolinguistics literature. Within English, coronal stop deletion and release rates are affected by style and dialect differences, in addition to phonological and morphological conditioning factors (e.g., Guy, 1980; Labov, 1989; Tagliamonte & Temple, 2005; Tanner et al., 2017). Word-final release rates for /p t k/ vary drastically in English across varieties and positions, with reported release rates ranging from 20% to 75% in pre-pausal, pre-nasal, and pre-obstruent positions (Davidson, 2011). Prevocalic positions are cited as a context that encourages stops to release and discourages deletion (see Davidson, 2011). As a result, most researchers examine stop behaviour in consonant environments. Studies of the highly heterogeneous TIMIT corpus (Garofolo et al., 1993) confirm that release rates in American English are quite varied with estimates, which differ based on the corpus sample and exclusionary criteria, ranging from 52% unreleased for sentence-final postvocalic stops (Y. Kang, 2003) to 60% released overall for sentence-final stops (Byrd, 1993). In Byrd's analysis, stop release rates vary considerably by place of articulation: bilabial stops were released 50% of the time, alveolar stops 57% of the time, and velar stops 83% of the time (Byrd, 1993). Together, this suggests that final-stop behaviour is simply highly variable in North American English.

Cantonese has voiceless /p t k/ in syllable-final position. These stops are described as categorically unreleased in coda position (Bauer & Benedict, 2011; Matthews & Yip, 2013),

which is often considered to be a neutralization of the long-lag (aspirated) and short-lag (unaspirated) stops found in syllable-initial position (see Table 1). Because of this, Cantonese-speaking listeners rely heavily on information in the adjacent vowel to identify the place of articulation of final stop consonants (Khouw & Ciocca, 2006).

In this study, we examine prevocalic and pre-approximant word-final stops, as these environments appear to encourage stop releases. The goal is to test whether early Cantonese-English bilinguals show a lower incidence of stop release for word-final /p t k/ in English, compared to monolingual English speakers of two different varieties (local and non-local). Here, English has a wide range of acceptable variability, while Cantonese exhibits relative invariance for a similar set of consonants. In this case, crosslinguistic influence comes from a categorical phonological process in Cantonese. This source of influence differs markedly from the more phonetic and gradient influence found with /b d g/ voicing, as Cantonese stops are categorically unreleased in codas. While the underlying VOT for final stops is ambiguous, Matthews and Yip (2013) describe the process as a neutralization of short- and long-lag stop VOT in final position. Given this description, we predicted that bilinguals would be less likely to release English word-final stops compared to monolinguals.

It is important to highlight that this prediction is situated within the context of spontaneous speech between peers from the same speech community, and that we are predicting a difference in the extent to which different outcomes occur, not the absence of a pattern. Polinsky (2018) reports on an unpublished 2007 study involving a small sample of Cantonese-English heritage bilinguals, which found that heritage speakers were more likely to release final stops in English than a monolingual age-matched group. While this finding runs counter to our predictions, there are some crucial differences in task and communicative context. Polinsky indicates that releasing stops is used as a clear speech strategy, and may have been “amplified by the nature of the task as the subjects [were] asked to read the English material in a formal test setting” (2018: 144). In contrast, SpiCE participants were in conversation with a bilingual peer—a communicative context where clear speech is less often necessary. The hyperarticulation for Cantonese-English heritage

speakers reported in Polinsky (2018) may be a clear speech strategy that aligns with the characterization of bilinguals as sociolinguistically sophisticated in their capacity to use a wide variety of forms (Bullock & Toribio, 2009). Indeed, given Chang (2016)’s finding that early experience with Korean—another language with unreleased final stops—facilitates the perception of unreleased stops, Cantonese-English bilinguals may simply be particularly adept at deploying variation in a way that suites the stylistic and social demands of the context.

Methods

All instances of word-final /p t k/ that preceded a vowel or approximant were identified. The segment preceding the target instance was not restricted, though vowels were by far the most frequent. Two research assistants coded all items by inspecting a waveform and spectrogram (dynamic range set to 65 dB; intensity contour superimposed), and by listening. Items were coded as one of six categories (adapted from Davidson, 2011). These categories were as follows:

Release indicates that stop closure was followed by an abrupt, transitory aperiodic event, regardless of whether or not there was aspiration.⁹

Unreleased indicates that there was a clear stop closure, but no release.

Lenition indicates there was no stop closure, but rather frication or formant structure in place of the stop. For example, this includes instances like “Eat an apple”, where /t/ flaps.

Glottalization is characterized by irregular glottal striations in the preceding segment (if it was a vowel), and throughout the stop closure. While glottalization may accompany other categories, glottalization was only selected if there was evidence of glottal tension throughout.

Deletion was used for cases with no stop closure or evidence of lenition. Research assistants were instructed to default to lenition in cases where they were unsure.

⁹ This differs from Davidson (2011) in that we lump all cases of quasi-release in with release.

Transcript error indicates that the transcript was wrong in some way. This could be due to substantial misalignment, or selection of the wrong pronunciation variant (e.g., “w” transcribed as “double-u” rather than as a phonological fragment [w]). In most cases, this resulted from forced-alignment errors.

Inter-rater reliability (IRR) between the two coders was assessed using Fleiss’ Kappa. While there was substantial agreement between the two raters for the six categories ($\kappa = 0.61$, $p < 0.01$), instances where there was disagreement were coded by a third research assistant ($n = 3,938$). As the primary point of interest in this study is whether or not a word-final stop releases, we collapsed lenition and glottalization into the unreleased category for the purposes of discarding observations. In practical terms, if rater A selected glottalization and rater B selected unreleased for the same item, this was treated as agreement.

From the initial total of 13,630, observations were then discarded if the three raters disagreed (1.9%), or if the consensus was transcript error (2.2%) or deletion (15.9%). Coders identified 163 deletions for Buckeye, 340 for CTWCE, and 1668 for SpiCE. The comparatively smaller number of deletions for Buckeye is expected, as we used the surface transcriptions, which means that many instances of deletion were not included in the initial sample. While stop deletion is an interesting phenomenon, it is not the focus of this study, and we do not analyze it further. Table 5 summarizes the counts of each stop by corpus used in the analysis, from a total of 10,906 items. Table 5 also provides counts for the analysis described in the following section, which focuses specifically on the unreleased category, and uses a slightly smaller subset. While counts for all segments are provided in Table 5, we opted to exclude all /p/ items from the analysis, given the small and potentially non-representative sample—particularly in the case of CTWCE. After removing /p/ items, there was a total of 10,215 items.

Results

We analyzed the data with a logistic mixed effects model using the *lme4* R package (Bates, Mächler, Bolker, & Walker, 2015). The *glmer* formula used was: $Released_{True} \sim Corpus \times$

Table 5

Counts of /p t k/ across the three corpora used in the analysis. The value in parentheses is the subset used for a more fine-grained unreleased analysis described in the following section. Note that while counts for /p/ are noted here, they were excluded from the analyses.

	/p/	/t/	/k/
SpiCE	259 (251)	3341 (3101)	2494 (2397)
CTWCE	53 (49)	859 (789)	444 (409)
Buckeye	379 (374)	1114 (1079)	1963 (1898)

$Place \times Rate + Vowel\ Precedes + Stress\ Precedes + Stress\ Follows + Word\ Frequency + (1 | Word) + (1 + Place \times Rate | Talker)$. Unless a difference is reported below, these variables were computed and coded as in Study 1. The new variables were:

Rate is the average segment duration in the word. It was calculated as the word duration divided by the number of segments in the canonical representation of the word, and is intended as a proxy for speech rate. We opted to use a within-word measure, as some of the words in question are phrase-initial, and preceding context could not be used. This continuous fixed effect variable was centered and standardized.

Place is a binary fixed effect variable that accounts for place of articulation. Place was weighted effect coded [T = 1, K = -1.08].

Vowel Precedes is a binary fixed effect variable that indicates whether the preceding segment was a vowel or not. While the coronal stop deletion literature focusing on varieties of North American English often includes a more fine-grained categorization schema (e.g., Hazen, 2011), we used a simplified comparison that better matches Cantonese phonology (i.e., given that Cantonese does not have consonant clusters, the distinction is better assessed as vowels, which are licit in Cantonese, or not vowels, which are illicit in Cantonese). Vowel Precedes was weighted effect coded [True = 1, False = -3.48].

As in Study 1, Corpus was simple coded with SpiCE as the reference level, and weighted effect coding was used for Preceding Stress [True = 1; False = -10.65] and Following Stress [True = 1; False = -4.03].

The logistic model returned a significant intercept [$\beta = -1.47$], indicating that released stops were the less likely outcome across the board. All model output is reported in Table 6. Parameter estimates are reported in the text when significant (at a level of $p < 0.05$). There was an effect of the Buckeye Corpus [$\beta = 1.33$], but not CTWCE. Both SpiCE and CTWCE had /t k/ release rates of 23% and 24% respectively, whereas Buckeye had a 55% release rate. Figure 6 depicts the proportion of Released (True) for Place by Corpus.

There was an effect of Place (T) [$\beta = -1.29$], indicating that /t/ was less likely to be released, but Place did not interact with Corpus. There was an effect of Rate [$\beta = 0.70$ —slower speech rates were accompanied by more released stops. Additionally, Rate interacted with both Corpus (CTWCE) [$\beta = -0.42$] and Corpus (Buckeye) [$\beta = -0.35$], which indicates that the differences in release rates compared compared to SpiCE are diminished as average phone duration within the word increases. Rate interacted with Place (T) [$\beta = -0.11$], which indicates

Figure 6. The proportion of Released /t k/ items by corpus. Refer to Table 5 for the counts by corpus. Other includes lenition, glottalization, and unreleased outcomes.

Table 6

Logistic mixed effects model comparing release rates for /t k/ across the three corpora in Study 2.

Parameter	Estimate	SE	z	p
Intercept	-1.47	0.20	-7.49	<0.001
Corpus (CTWCE)	0.31	0.22	1.44	0.15
Corpus (Buckeye)	1.33	0.18	7.23	<0.001
Place (T)	-1.29	0.12	-10.70	<0.001
Rate	0.70	0.06	11.78	<0.001
Vowel Precedes (True)	-0.47	0.04	-11.23	<0.001
Stress Precedes (True)	0.003	0.01	0.19	0.85
Stress Follows (True)	-0.11	0.01	-7.87	<0.001
Word Frequency	-0.43	0.10	-4.45	<0.001
Corpus (CTWCE) × Place (T)	0.07	0.18	0.40	0.69
Corpus (Buckeye) × Place (T)	0.24	0.15	1.63	0.10
Corpus (CTWCE) × Rate	-0.42	0.14	-3.02	0.003
Corpus (Buckeye) × Rate	-0.35	0.12	-2.99	0.003
Place (T) × Rate	-0.11	0.06	-2.00	0.045
Corpus (CTWCE) × Place (T) × Rate	0.16	0.14	1.15	0.25
Corpus (Buckeye) × Place (T) × Rate	0.17	0.11	1.45	0.15

Figure 7. Probability of Released (True) by Place and Rate. On the y-axis, one corresponds to released, and zero to not released (which includes the various forms of non-release). Individual items are plotted with translucent circles. The lines are generalized linear smooths (logit), where thick lines correspond to group means and thin lines to individuals.

that increased chance of release at slower speech rates is attenuated for coronals. These results are depicted in Figure 7.

More frequent words were less likely to be released [$\beta = -0.43$]. When /t k/ followed a vowel, the stop was less likely to be released [$\beta = -0.47$] than if that stop was preceded by a consonant. If the word-final /t k/ was followed by a vowel that is stressed, that stop is less likely to be released [$\beta = -0.11$].

Unreleased /t k/ model

The above analysis establishes that SpiCE bilinguals and CTWCE monolinguals release stops at similar levels. While the analysis mirrors the structure of prior work on final stops, it glosses over the specific behaviour that Cantonese-English bilinguals were anticipated to engage in—producing *unreleased* final stops. In light of this, we present an additional more exploratory analysis. We flip the question and model the incidence of unreleased final stops

specifically—clear stop closure, and *no* burst—as opposed to all other phonetic variants. While it may seem counterintuitive to lump glottalization and lention in with released final stops, this formulation addresses the question of how likely stops are to be produced as they are in Cantonese—as *unreleased* stops.

The model was nearly identical to the model of the previous section, with the exception that the dependent variable here is Unreleased, a binary variable with values True and False.

Additionally, as this model is based on the more fine-grained coding scheme, we examine the subset of the data set in which two of the three raters agreed on the unreleased category—totaling 9,673 of the initial 10,215 instances of /t k/. Counts by place of articulation and corpus are reported in parentheses in Table 5. Additionally, weighted effect coding was updated for Place [T = 1, K = −1.06], Vowel Precedes [True = 1, False = −3.52], Preceding Stress [True = 1, False = −11.06], and Following Stress [True = 1, False = −4.03]. The model formula was:

$$\text{Unreleased}_{True} \sim \text{Corpus} \times \text{Place} \times \text{Rate} + \text{Vowel Precedes} + \text{Stress Precedes} + \text{Stress Follows} + \text{Word Frequency} + (1 \mid \text{Word}) + (1 + \text{Place} \times \text{Rate} \mid \text{Talker}).$$

Model results are provided in Table 7. The intercept was significant [$\beta = -1.05$], indicating that unreleased stops were less likely than the alternative surface forms in aggregate. SpiCE items were more likely to be unreleased, compared to both monolingual corpora [Buckeye: $\beta = -0.62$, CTWCE: $\beta = -0.51$]—a pattern that is clearly depicted in 8. There was also a significant Corpus (Buckeye) $\times$ Place (T) interaction [$\beta = 0.33$], suggesting that the comparison between SpiCE and Buckeye differs across the two places of articulation. Rate [$\beta = 0.25$] interacted with Place (T) [$\beta = 0.50$], which suggests that at slower speech rates, /t/ is more likely to be unreleased. This is shown in Figure 9, and we offer a hypothesis for this pattern below.

For the control fixed effects, the presence of a preceding vowel [$\beta = 0.20$] led to an increased probability of being unreleased. The probability of being unreleased also was affected by Stress [Precedes: $\beta = -0.04$, Follows: $\beta = 0.37$], and Word Frequency [$\beta = 0.31$]. That is, a stop was less likely to be unreleased when it occurred before a stressed syllable, and more likely to be unreleased when it followed a stressed syllable. More frequent words were more likely to be

unreleased.

SpiCE-only / t k/

For the same reasons as in Study 1, we examined the role of language mode on / t k/ production in a separate analysis focusing on just the SpiCE corpus. As differences between SpiCE and the local monolingual comparison group (CTWCE) were evident in the unreleased model, we adapted that model to probe effects of language mode. As such, the dependent variable reflects whether or not a particular item was coded as unreleased in the fine-grained coding scheme. The additional variables in the SpiCE-only model were coded in the same fashion as Study 1, with the weighted effect coding updated for Place [$T = 1$, $K = -1.29$], Interview Order [Cantonese First = 1, English First = -1.06], Vowel Precedes [True = 1, False = -5.24], Stress Precedes [True = 1, False = -9.82], and Stress Follows [True = 1, False = -4.37]. The *glmer* formula was: $Unreleased_{True} \sim Place \times Rate \times Code-switch Distance + Interview Order + Vowel Precedes + Stress Precedes + Stress Follows + Word Frequency + (1 | Word) + (1 + Place \times Rate | Talker)$.

Model output is summarized in Table 8. The model returned a significant intercept [$\beta = -0.87$], indicating that unreleased was a less likely outcome than released, lenition, and glottalization considered together.

The primary variables of interest in this model—Code-switch Distance and Interview Order—exerted no significant effect on the response variable and did not participate in any interactions. If Code-switch Distance were to be considered given its marginal status [$\beta = 0.07$, $p = 0.07$], it would run counter to the hypothesis, as a positive effect suggests that stops are more likely to be unreleased as distance from a code-switch increases. Given the extremely flat smooths for group means and high degree of individual variability in Figure 10, it seems more plausible that this is a spurious marginal result. As with the /b d g/ SpiCE-only model, instances of / t k/ were generally quite far from a code-switch ($Mdn = 335.7s$), and the non-significant result is thus not particularly surprising.

Figure 8. The proportion of tokens that are Unreleased by Place and Corpus. In this plot, Other includes released, glottalization, and lenition.

Figure 9. Probability of being Unreleased by Rate. One corresponds to unreleased, and zero to *not* unreleased (which includes released, lenition, and glottalization). Individual items are plotted with translucent circles. The lines are generalized linear smooths (logit), where thick lines capture group means, and thin lines correspond to individuals.

Table 7

Logistic mixed effects model output for the / t k/ comparison for the more fine-grained coding of unreleased.

Parameter	Estimate	SE	z	p
Intercept	-1.05	0.20	-5.38	<0.001
Corpus (CTWCE)	-0.62	0.19	-3.30	<0.001
Corpus (Buckeye)	-0.51	0.16	-3.15	0.002
Place (T)	0.25	0.11	2.20	0.03
Rate	0.25	0.06	4.41	<0.001
Vowel Precedes (True)	0.20	0.04	4.93	<0.001
Stress Precedes (True)	-0.04	0.01	-3.61	<0.001
Stress Follows (True)	0.07	0.01	5.36	<0.001
Word Frequency	0.31	0.10	3.03	0.002
Corpus (CTWCE) × Place (T)	0.15	0.14	1.01	0.31
Corpus (Buckeye) × Place (T)	0.33	0.12	2.72	0.006
Corpus (CTWCE) × Rate	0.20	0.13	1.48	0.14
Corpus (Buckeye) × Rate	-0.02	0.11	-0.18	0.85
Place (T) × Rate	0.50	0.05	10.00	<0.001
Corpus (CTWCE) × Place (T) × Rate	0.001	0.12	0.01	0.99
Corpus (Buckeye) × Place (T) × Rate	-0.15	0.10	-1.50	0.13

The control variables in the model largely mirrored the 3-corpus model in effect size and significance, and can be interpreted in the same fashion. There were significant effects for Rate [$\beta = 0.22$] and a Rate $\times$ Place (T) interaction [$\beta = 0.48$], though the simple effect of Place (T) missed the significance threshold [$\beta = -0.19$, $p = 0.09$]. There were also significant effects for Stress [Precedes: $\beta = -0.04$, Follows: $\beta = 0.07$]. The effect of preceding vowel missed the cutoff for significance [$\beta = 0.05$, $p = 0.11$], but trends in the same direction as the 3-corpus model. Unlike the 3-corpus model, there was no effect of Word Frequency on the probability of an item being unreleased. This could be due to any number of reasons, all of which would highly speculative in nature.

Figure 10. The probability of an item being unreleased is plotted against Code-switch Distance. Translucent circles represent individual tokens. The thin lines represent individuals, and thick lines represent group means. Despite the flat appearance of the group means, the lines are all generalized linear smooths (logit).

Interim discussion: / t k/

The results for / t k/ ultimately follow the prediction that English release patterns would be influenced by Cantonese. This does not surface for release rates, but rather for *unreleased* rates. There are two primary factors contributing to this outcome. First, release rates were surprisingly

low in the local monolingual speech from CTWCE (24%) and comparable to the release rates for SpiCE (23%), compared to the more widely studied non-local variety in the Buckeye corpus 55%. This highlights that our expectations about what the local variety would do were not entirely correct. While prior work on British Columbia (BC) English has not examined stop release rates, there are a variety of documented features unique to the area (e.g., Mellesmoen, 2018), and the fact that CTWCE and Buckeye do not show the same pattern further contributes to this understanding of BC English. Second, talkers in the SpiCE corpus are more likely to produce unreleased / t k/ compared to both groups of monolinguals—local and non-local. Following from descriptions of Cantonese final stops, this outcome is indicative of crosslinguistic influence, as the bilinguals are producing word-final English stops with Cantonese-like settings the most frequently of the three groups.

While these models provide evidence for crosslinguistic influence, it is ultimately a coarse picture, as the nature of the releases and non-releases is under-described in the literature. For example, while we find that SpiCE talkers are more likely to have unreleased / t k/, this makes no comment as to the nature of their released tokens, which include both full releases and quasi-releases. Additionally, there is also the question of what happens in Cantonese, and whether quasi-releases can occur in spontaneous speech. These categorical results build on prior work (e.g., Davidson, 2011), and invite future research in this area.

There was no effect of Interview Order or Code-switch Distance While language mode is widely studied, it is possible that the nature of final releases is more codified as a feature of the bilinguals' English and thus less subject to these more ephemeral forms of interference. It is also possible that our decision to categorize items as released and unreleased obscures the effect of online crosslinguistic influence. This would be the case if degree of release were to pattern gradiently, like proportion voicing for /b d g/. Again, these results raise questions that invite future research, which would additionally involve describing whether and how quasi-releases surface in Cantonese final stops.

In addition to the main findings for / t k/, we also report on the unique behaviour of /k/

across all three corpora. At slower rates /k/ was the most likely to release, which might be due to unique behaviour of frequent words within this class like *like* (Drager, 2011). Indeed, “like” accounts for 73% of /k/ items in the SpiCE corpus, 56% in CTWCE, and 50% in the Buckeye corpus. The unique behaviour of /k/ could also hinge on standard phonetic mechanisms, as the increase in inter-oral pressure behind a velar constriction may essentially require a release more than stops at more anterior places of articulation (Byrd, 1993; Cho & Ladefoged, 1999). This is also evident in how Place patterns in the models—/t/ is less likely to be released. Releases pattern in the same way as /k/ with respect to speech rate. In the case of the unreleased model, /t/ is more likely to be unreleased, even as speech rate slows down (unlike /k/). The increased unreleased rates in this environment likely reflects an increased prevalence of lenition at faster speech rates.

There were a number of significant results for phonological environment and stress. The presence of a preceding vowel corresponds to less releases, and an increased probability of specifically being unreleased, which may be variation afforded by the informative formant transitions anticipating the stop closure. Stress on the preceding syllable reduces the probability of unreleased but had no effect on release rates. Following stress decreases probability of release and increases probability of unreleased. These results all point to prosodic effects structuring phonetic variation and warrant their own empirical investigations.

General discussion

We examined pronunciation differences in English /b d g/ and /t k/ as spoken by early Cantonese-English bilinguals (SpiCE corpus: Johnson, 2021) compared to a local English monolingual control group (CTWCE: Swan, 2016) and a well-studied English monolingual control group from a different North American English speech community (Buckeye corpus: Pitt et al., 2005). The SpiCE corpus consists of spontaneous speech from 34 early Cantonese-English bilinguals, who have high speaking proficiency in both languages. Cantonese speakers world-wide are a multilingual group, making this particular sample representative of the larger Cantonese speech community.

A goal of this work was to bring together the specific predictions from recent theories on sound change (Harrington et al., 2018; Harrington & Schiel, 2017) with the well-established empirical findings that a bilingual's languages exhibit mutual influence (Best & Tyler, 2007; Flege & Bohn, 2021; Kroll & Navarro-Torres, 2018; Simonet, 2016), which credit the existence of mutual influence to parallel activation and connections between linguistic units, but do not offer, to our knowledge, a concrete mechanistic account of asymmetry in crosslinguistic influence for similar, linked segments in early bilingual speech production. There are, of course, a variety of mechanisms that have been proposed to account for what leads segments to interact with one another in the first place in L2 speech perception and production (e.g., Best & Tyler, 2007; Flege & Bohn, 2021). The mechanism proposed here deepens our understanding of what happens within what Flege and Bohn's SLM-r (2021) would term a composite category. SLM-r can account for why a pair of sounds may be linked through notions of similarity, and our approach helps account for what happens within such a linked composite category. Note that we are not implying that a composite category is a unimodal merging of two cross-linguistic categories, though in late L2 learners, that may certainly be the case.

In describing what makes this mechanism unique, we touch on some points made in the introduction. While considerations of category compactness may provide explanatory force for the learning outcomes of late second language learners (Kartushina & Frauenfelder, 2014; Kartushina, Frauenfelder, & Golestani, 2016; Kartushina et al., 2015; Kartushina, Hervais-Adelman, et al., 2016), they do not align with the developmental trajectory of category formation in early bilinguals (Werker & Byers-Heinlein, 2008). Additionally, Olson (2016) and Bullock and Toribio (2009) observe that some sound categories may have more "room for influence" or phonetic space, inviting variability and providing opportunity for bilinguals to exhibit sociolinguistic sophistication of the stylistic deployment of particular linguistic features. This concept of room for influence combined with the work by Harrington and colleagues invites us to ask whether the distribution of phonetic variation in linked sound categories across a bilingual's languages goads pronunciation variation in the direction of the less variable language.

In the two studies we explored here, Cantonese was the less variable language. Study 1 examined the proportion of voicing during the stop closure for English /b d g/. As presented in Table 1, the phonologically linked sounds in Cantonese are the word-initial voiceless unaspirated stops /p t k/, which while described in Cantonese grammars as voiceless unaspirated, are also described as being similar to the English voiced stop series (Bauer & Benedict, 2011; Matthews & Yip, 2013), in addition to showing phonation in acquisition (Clumeck et al., 1981) and in morpho-phonological processes (W. Y. P. Wong, 2006). Thus, while there is some indication that phonation for Cantonese /p t k/ is gradient, it is much less variable than what is permitted in English for /b d g/. We predicted, therefore, that Cantonese-English bilinguals would show effects of the less variable language (Cantonese) pulling on the distribution of the more variable language (English). Indeed, we found that Cantonese-English bilinguals were less likely to produce phonation during /b d g/ than either of the monolingual English control groups. We interpret this as an indication that the English in early Cantonese-English bilinguals shows effects of influence from Cantonese. In /b d g/ voicing patterns we also looked at the role of language mode. While SpiCE bilinguals who completed a Cantonese interview before the English interview showed a pattern of producing less voicing compared to those who completed the English interview first, the effect was marginal. Thus, while voicing may be susceptible to language mode—either operationalized in a more sensitive way or as part of a different task—we do not establish evidence for it here.

The phonological pattern of /p t k/ in word-final position in Cantonese is different. Obstruent codas in Cantonese are categorically described as unreleased (Bauer & Benedict, 2011; Khouw & Ciocca, 2006; Matthews & Yip, 2013), though susceptible to lenition processes in the context of syllable fusion (W. Y. P. Wong, 2006). Cantonese word-final /p t k/ is much less variable than Cantonese word-initial /p t k/ realizations or word-final stop behaviour in English. Thus, while we therefore predicted that, again, the less variable pattern (Cantonese) will pull the pattern of the more variable language (English) towards it, the more categorical nature of the absence of final-stop releases in Cantonese makes this pattern conceptually quite different.

We report initial counts for each of the three segments, but opted to only analyze /t/ and /k/, as there were relatively few instances of /p/ across the three corpora, which raised concerns about the representativeness of the sample. The initial analysis modelling the likelihood of word-final English / t k/ being released established a difference between SpiCE and the non-local monolingual variety (Buckeye), but not between SpiCE and the local CTWCE corpus. Rather, we uncovered a rather surprising regional difference, as both Cantonese-English bilingual and monolingual Vancouver speech patterns had very low rates of stop releases in word-final position. Our more exploratory analysis pivoted the research question to be about the probability of stops having a closure but being unreleased (and thus grouping other realizations together: lenition, glottalization, and full closure with release). In this analysis, we found that SpiCE talkers were more likely to produce unreleased stops than either CTWCE or Buckeye talkers. The unreleased rates of word-final English / t k/ showed no effects of language mode or (interpretable) Code-switch Distance, suggesting that the pattern uncovered in the unreleased model is a codified feature of Cantonese-English bilingual speech and not subject to more transient effects—at least, as coarsely implemented here. Alternatively, if release and non-release patterns are actually gradient, then it is possible the analysis here is not sensitive enough to detect such differences in language mode.

A finding separate from the core questions about bilingual speech that merits attention is the unique differences between / t k/ in word-final unreleased rates. At slower speech rates, the probability of /t/ for all three corpora being unreleased increased. The pattern for /k/, however, is robust: for all three corpora, the likelihood of /k/ being unreleased increases as speech rate decreases. The pattern with /k/ appears robust in Figure 9, but should also be accompanied with the caveat that a very high number of word-final /k/ are due to the word “like”. Whether the unique pattern for /k/ is due to peculiarities with this particularly multifaceted lexical item or due to post-constriction pressure increases that require a release (Cho & Ladefoged, 1999) is an issue we leave to future research.

In considering Cantonese, English, and their mutual influence, one must remember that

languages in contact can take many tracks as a result of the social dynamics between speech communities and the linguistic structure of the languages (Thomason & Kaufman, 1988). A common language contact situation in contemporary Western societies is that of a diaspora speech community, where there is a shift to the societally dominant language, such that younger speakers might be characterized as heritage speakers of their ancestral language. While Polinsky and Scontras (2020) suggest there is little evidence for structural transfer as a result of contact-induced change in heritage languages, Muysken (2020) criticizes this jump to conclusions and cautions that such a characterization is likely due to a focus in the literature on bilingual immigrant communities that undergo rapid language shift in the highly monolingual and English-dominant context of the United States. Bilingual communities with longer histories, Muysken argues, are those more likely to show evidence of structural contact-induced change. The bilingual individuals whose speech is under study here are early bilinguals, who may or may not be more or less dominant in one language or another. Some—but not all—of the speakers in the SpiCE corpus may fall under the designation of heritage speakers in the sense that they were raised in an English-dominant society and primarily use Cantonese with family. As a whole, the Cantonese-English speech community in the Vancouver, BC area is well established, having existed since the latter half of the 19th century (Yu, 2013). Therefore, this speech community is a clear candidate for contact-induced change (Muysken, 2020).

In addition to considering the impacts of social dynamics at the level of the speech community (Thomason & Kaufman, 1988), it is also important to consider social and communicative factors within the speech sample being studied. At first glance, our hypotheses and results might appear incompatible with prior work demonstrating that bilinguals show patterns of divergence in the types of environments where we predict a distributional shift more similar to convergences. For example, Polinsky (2018) finds that Korean-English and Cantonese-English heritage bilinguals are more likely to produce word-final stop releases in English compared to monolinguals in an experimental setting. Following from Bullock and Toribio's (2009) argument that bilinguals have a rich repertoire of forms at their disposal, it is

perhaps unsurprising that bilinguals hyperarticulate in formal, experimental settings. Polinsky (2018, p. 144) speculates, however, “that heritage speakers’ tendency to release final stops...may be a side effect of a more general tendency to enunciate...[as] they are used to communicating in English with nonnative speakers who benefit from clear word boundaries.” Our results suggest a different explanation. If bilinguals are highly likely to produce released stops in a formal experimental setting, but are more likely to produce unreleased stops when conversing with a peer in their speech community, then an account of this behaviour must take communicative context and interlocutors into account. As bilinguals are better at perceiving place of articulation in unreleased stops (Chang, 2016), it follows that a conversation between Cantonese-English bilinguals would be less impacted by the prevalence of unreleased stops from an intelligibility perspective. While this explanation remains somewhat speculative, our results offer yet another cautionary tale for the (implicit) assumption that lab-based findings will replicate with spontaneous speech (e.g., see Gahl, Yao, & Johnson, 2012). This is particularly relevant in the case of the phenomena we study, as our focus is on how broad patterns shape the speech of communities. We identify this as a promising area for future work.

While we consider these results as evidence that the nature of within-category phonetic distributions can drive mutual influence across languages, we must, of course, acknowledge we cannot make this claim strongly. The claim is tempered, as we only investigate the effect of Cantonese on English and not English on Cantonese, as we have no evidence of *mutual* influence. The multilingualism of Cantonese speakers, however, poses a challenge for fully addressing the research question with this particular set of languages, as we do not have access to a comparable spontaneous speech sample from Cantonese speakers whose linguistic knowledge is devoid of potential influence from another language that might confound the results—and in many cases, where the other language is specifically English. Our approach, however, brings new predictive rigor to the study of crosslinguistic influence and can be further tested, refuted, or refined in investigations of additional languages.

Acknowledgments

Thank you to Ivan Fong, Nancy Yiu, Ariana Zattera, Christina Sen and other research assistants in the Speech in Context Lab for assistance in the creation of the SpiCE corpus. Thank you to: Angelina Lloy, Christina Sen, and Stephanie Chung for assistance with data coding; Amanda Cardoso, Gloria Mellesmoen, and the Speech in Context Lab for comments on this work at various stages. This work has been substantially improved by comments from the editors and reviewers.

This work has been supported in part by funding from Canada's Social Sciences and Humanities Research Council (MB), Canada's Natural Sciences and Engineering Research Council (MB), and a UBC Public Scholars Initiative Award (KJ).

Table 8

Logistic mixed effects output modeling the probability of /p t k/ items being unreleased within the SpiCE corpus in Study 2.

Parameter	Estimate	SE	z	p
Intercept	-0.87	0.22	-3.94	<0.001
Place (T)	0.19	0.11	1.72	0.09
Rate	0.22	0.06	3.67	<0.001
Code-switch Distance	0.07	0.04	1.79	0.07
Interview Order (Cantonese First)	0.07	0.09	0.76	0.45
Vowel Precedes (True)	0.05	0.03	1.61	0.11
Stress Precedes (True)	-0.05	0.02	-3.02	0.003
Stress Follows (True)	0.07	0.02	4.57	<0.001
Word Frequency	0.10	0.11	0.93	0.35
Place (T) × Rate	0.48	0.05	8.77	<0.001
Place (T) × Code-switch Distance	-0.05	0.03	-1.47	0.14
Rate × Code-switch Distance	-0.04	0.04	-1.15	0.25
Place (T) × Rate × Code-switch Distance	0.04	0.03	1.13	0.26

References

- Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15(3), 517–530. doi: 10.1017/s1366728911000460
- Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38(4), 640–653. doi: 10.1016/j.wocn.2010.09.005
- Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39(4), 558–570. doi: 10.1016/j.wocn.2011.03.001
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, 137(5), 2823–2833. doi: 10.1121/1.4919317
- Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, 48(1), 1–27. doi: 10.1177/00238309050480010101
- Barlow, J. A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.00288
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). doi: 10.18637/jss.v067.i01
- Bauer, R. S., & Benedict, P. K. (2011). *Modern Cantonese phonology*. De Gruyter Mouton. doi: 10.1515/9783110823707
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (Vol. 17, p. 13-34). John Benjamins Publishing.
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer*. Retrieved 2020-01-20, from <http://www.praat.org/> (Version 6.1.09)

- Bolton, K., & Kwok, H. (1990). The dynamics of the Hong Kong accent: Social identity and sociolinguistic description. *Journal of Asian Pacific Communication*, 1(1), 147–172.
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378–400. doi: 10.32614/RJ-2017-066
- Bullock, B. E., & Toribio, A. J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In L. Isurin, D. Winford, & K. deBot (Eds.), *Studies in Bilingualism* (Vol. 41, pp. 189–206). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/sibil.41.12bul
- Byrd, D. (1993). 54,000 american stops. *UCLA working Papers in Phonetics*, 83, 97–116. Retrieved from <https://escholarship.org/uc/item/4j43d6qj>
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40(2), 249–268. doi: 10.1016/j.wocn.2011.10.007
- Chang, C. B. (2015). Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In E. Raimy & C. E. Cairns (Eds.), *The Segment in Phonetics and Phonology* (1st ed., pp. 199–217). John Wiley & Sons, Inc. doi: 10.1002/9781118555491.ch9
- Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, 19(4), 791–809. doi: 10.1017/S1366728914000261
- Chen, M. Y., & Wang, W. S.-Y. (1975). Sound change: Actuation and implementation. *Language*, 51(2), 255–281. doi: 10.2307/412854
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–229. doi: 10.1006/jpho.1999.0094
- Clumeck, H., Barton, D., Macken, M. A., & Huntington, D. A. (1981). The aspiration contrast in

- Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics*, 210–225. Retrieved from www.jstor.org/stable/23753507
- Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, 53(8), 1042–1058. doi: 10.1121/1.3508799
- Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54, 35–50. doi: 10.1016/j.wocn.2015.09.003
- Deterding, D., Wong, J., & Kirkpatrick, A. (2008). The pronunciation of Hong Kong English. *English World-Wide*, 29(2), 148–175. doi: 10.1075/eww.29.2.03det
- Drager, K. K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics*, 39(4), 694–707. doi: 10.1016/j.wocn.2011.08.005
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2020). *Ethnologue: Languages of the world* (23rd ed.). SIL International. Retrieved from <https://www.ethnologue.com>
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65. doi: 10.1016/s0095-4470(19)30537-6
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research* (Vol. 92, pp. 233–277). York Press.
- Flege, J. E., & Bohn, O.-S. (2021, January). *The revised Speech Learning Model*. (ResearchGate Preprint) doi: 10.13140/RG.2.2.27529.06249
- Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, 6(3), 185–202. doi: 10.1016/0167-6393(87)90025-2
- Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language*, 89, 110–137. doi: 10.1016/j.jml.2015.10.001
- Fricke, M., Zirnstein, M., Navarro-Torres, C., & Kroll, J. F. (2019). Bilingualism reveals

- fundamental variation in language processing. *Bilingualism: Language and Cognition*, 22(1), 200–207. doi: 10.1017/S1366728918000482
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language*, 66(4), 789–806.
- Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993). *TIMIT acoustic-phonetic continuous speech corpus*. Philadelphia: Linguistic Data Consortium. (LDC93S1, Web Download) doi: 10.35111/17gk-bn40
- Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, 25(4), 1031–1036. doi: 10.1177/0956797613520014
- Grosjean, F. (2011). An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16(1), 11–21. doi: 10.1177/1367006911403210
- Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, 60(2), 98–128. doi: 10.1159/000071449
- Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). Academic Press.
- Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science*, 10(4), 707–728. doi: 10.1111/tops.12329
- Harrington, J., & Schiel, F. (2017). /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language*, 93(2), 414–445. doi: 10.1353/lan.2017.0019
- Hazen, K. (2011). Flying high above the social radar: Coronal stop deletion in modern Appalachia. *Language Variation and Change*, 23(1), 105–137. doi: 10.1017/s0954394510000220
- Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, 22(1), 37–67. doi:

10.1017/s0954394509990238

- Hung, T. T. (2000). Towards a phonology of Hong Kong English. *World Englishes*, 19(3), 337–356. doi: 10.1111/1467-971X.00183
- Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of american english. *Journal of the International Phonetic Association*, 39(3), 313–334. doi: 10.1017/s0025100309990156
- Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1–15. doi: 10.1016/j.wocn.2018.07.001
- Johnson, K. A. (2021). *SpiCE: Speech in Cantonese and English*. Scholars Portal Dataverse. (Version 1) doi: 10.5683/SP2/MJOXP3
- Johnson, K. A., Babel, M., Fong, I., & Yiu, N. (2020). SpiCE: A new open-access corpus of conversational bilingual speech in Cantonese and English. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)* (pp. 4089–4095). Retrieved from <https://www.aclweb.org/anthology/2020.lrec-1.503.pdf>
- Kang, K.-H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of korean-english bilinguals. *The Journal of the Acoustical Society of America*, 119(3), 1672–1683.
- Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in korean. *Phonology*, 219–273.
- Kang, Y., George, S., & Soo, R. (2016). Cross-language influence in the stop voicing contrast in heritage Tagalog. *Heritage Language Journal*, 13(2), 184–218. Retrieved from <https://www.heritagelanguages.org/Journal.aspx>
- Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.01246
- Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language*

- Learning*, 66(S2), 155–186. doi: 10.1111/lang.12187
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832. doi: 10.1121/1.4926561
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, 57, 21–39. doi: 10.1016/j.wocn.2016.05.001
- Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60(2), 286–319. doi: 10.2307/413642
- Khouw, E., & Ciocca, V. (2006). An acoustic and perceptual study of final stops produced by profoundly hearing impaired adolescents. *Journal of Speech, Language, and Hearing Research*, 49(1), 172–185. doi: 10.1044/1092-4388(2006/014)
- Kroll, J. F., & Navarro-Torres, C. A. (2018). Bilingualism. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–29). American Cancer Society. doi: 10.1002/9781119170174.epcn307
- Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1(1), 85–97. doi: 10.1017/s0954394500000120
- Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In *Social markers in speech* (pp. 1–30). Cambridge University Press.
- Leung, C. B., & Brice, A. E. (2012). An analysis of phonological processes involved in spoken English of Hong Kong primary pre-service teachers. *Language Testing in Asia*, 2(2), 42–60. doi: 10.1186/2229-0443-2-2-42
- Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, 20(12), 1507–1515. doi:

10.1111/j.1467-9280.2009.02480.x

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops:

Acoustical measurements. *WORD*, 20(3), 384–422. doi:

10.1080/00437956.1964.11659830

MacLeod, A. A. N., & Stoel-Gammon, C. (2005). Are bilinguals different? what VOT tells us

about simultaneous bilinguals. *Journal of Multilingual Communication Disorders*, 3(2),

118–127. doi: 10.1080/14769670500066313

MacLeod, A. A. N., & Stoel-Gammon, C. (2009). The use of voice onset time by early bilinguals

to distinguish homorganic stops in Canadian English and Canadian French. *Applied*

Psycholinguistics, 30(1), 53–77. doi: 10.1017/s0142716408090036

MacLeod, A. A. N., Stoel-Gammon, C., & 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(4), 374–387. doi: 10.1016/j.wocn.2009.07.001

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, 76(2),

190–208. doi: 10.1111/j.1540-4781.1992.tb01100.x

Matthews, S., & Yip, V. (2013). *Cantonese: A comprehensive grammar*. Routledge. doi:

10.4324/9780203835012

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced

aligner: Trainable text-speech alignment using kald. In *Interspeech 2017*. ISCA. doi:

10.21437/interspeech.2017-1386

McAuliffe, M., Stengel-Eskin, E., Socolof, M., & Sonderegger, M. (2017). Polyglot and speech

corpus tools: A system for representing, integrating, and querying speech corpora. In

Interspeech 2017. ISCA. doi: 10.21437/interspeech.2017-1390

Mellesmoen, G. (2018). A remedial path to merger: Merger by phonological transfer in British

Columbia English. *Toronto Working Papers in Linguistics*, 40(1). Retrieved from

<https://twpl.library.utoronto.ca/index.php/twpl/article/view/29162>

Mielke, J., & Nielsen, K. (2018). Voice onset time in English voiceless stops is affected by

- following postvocalic liquids and voiceless onsets. *The Journal of the Acoustical Society of America*, 144(4), 2166–2177. doi: 10.1121/1.5059493
- Muysken, P. (2020). The case for contact induced-change in heritage languages. *Bilingualism: Language and Cognition*, 23(1), 37–38. doi: 10.1017/s1366728919000373
- Myers-Scotton, C. M., & Jake, J. L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, 21(3), 340–366. doi: 10.1177/1367006915626588
- Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2017). Weighted effect coding for observational data with wec. *The R Journal*, 9(1), 477. doi: 10.32614/rj-2017-017
- Olson, D. J. (2016). The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*, 46(3), 263–285. doi: 10.1017/s0025100315000468
- Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, 45(1), 89–95. doi: 10.1016/j.specom.2004.09.001
- Polinsky, M. (2018). *Heritage Languages and their Speakers*. Cambridge: Cambridge University Press. doi: 10.1017/9781107252349
- Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, 23(1), 4–20. doi: 10.1017/s1366728919000245
- Samuel, A. G. (2020, apr). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, 111, 104070. Retrieved from <https://doi.org/10.1016/j.jml.2019.104070> doi: 10.1016/j.jml.2019.104070
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436. doi: 10.1006/jpho.1997.0051
- Schneider, E. W. (2003). The dynamics of New Englishes: From identity construction to dialect birth. *Language*, 79(2), 233–281. doi: 10.1353/lan.2003.0136

- Setter, J., Wong, C., & Chan, B. (2010). *Hong Kong English*. Edinburgh University Press.
- Sewell, A., & Chan, J. (2010). Patterns of variation in the consonantal phonology of Hong Kong English. *English World-Wide*, 31(2), 138–161. doi: 10.1075/eww.31.2.02sew
- Shen, A., Gahl, S., & Johnson, K. (2020). Didn't hear that coming: Effects of withholding phonetic cues to code-switching. *Bilingualism: Language and Cognition*, 1–12. doi: 10.1017/s1366728919000877
- Simonet, M. (2011). Production of a Catalan-specific vowel contrast by early Spanish-Catalan bilinguals. *Phonetica*, 68(1-2), 88–110. doi: 10.1159/000328847
- Simonet, M. (2016). The phonetics and phonology of bilingualism. *Oxford Handbooks Online*. doi: 10.1093/oxfordhb/9780199935345.013.72
- Simonet, M., & Amengual, M. (2019). Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. *International Journal of Bilingualism*, 24(2), 208–221. doi: 10.1177/1367006919826388
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71. doi: 10.1037/1082-989X.11.1.54
- Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, 9(1), 97–114. doi: 10.1017/S1366728905002403
- Sung, C. C. M. (2015). Hong Kong English: Linguistic and sociolinguistic perspectives. *Language and Linguistics Compass*, 9(6), 256–270. doi: 10.1111/lnc3.12142
- Swan, J. T. (2016). *Language ideologies, border effects, and dialectal variation: Evidence from /ae/, /au/, and /a/ in Seattle, WA and Vancouver, BC* (Doctoral dissertation). The University of Chicago, ProQuest Dissertations Publishing. (10158013)
- Szakay, A., Babel, M., & King, J. (2016). Social categories are shared across bilinguals lexicons. *Journal of Phonetics*, 59, 92–109. doi: 10.1016/j.wocn.2016.09.005
- Tagliamonte, S., & Temple, R. (2005). New perspectives on an ol' variable: (t, d) in British

- English. *Language Variation and Change*, 17(3), 281–302. doi: 10.1017/s0954394505050118
- Tanner, J., Sonderegger, M., & Stuart-Smith, J. (2020). Structured speaker variability in Japanese stops: Relationships within versus across cues to stop voicing. *The Journal of the Acoustical Society of America*, 148(2), 793–804. (Publisher: Acoustical Society of America) doi: 10.1121/10.0001734
- Tanner, J., Sonderegger, M., & Wagner, M. (2017). Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1). doi: 10.5334/labphon.96
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.
- Tse, H. (2019). *Beyond the monolingual core and out into the wild: A variationist study of early bilingualism and sound change in Toronto heritage Cantonese* (Doctoral dissertation, University of Pittsburgh). Retrieved from <https://d-scholarship.pitt.edu/35721/>
- Tsui, I. Y. H., & Ciocca, V. (2000). Perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners. *International Journal of Language & Communication Disorders*, 35(4), 507–525. doi: 10.1080/136828200750001269
- Tsui, R. K.-Y., Tong, X., & Chan, C. S. K. (2019). Impact of language dominance on phonetic transfer in Cantonese-English bilingual language switching. *Applied Psycholinguistics*, 40(1), 29–58. doi: 10.1017/s0142716418000449
- Van Coetsem, F. (2000). *A general and unified theory of the transmission process in language contact* (Vol. 19). Winter.
- Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 97–195). University of Texas Press.
- Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and

- comprehension. *Trends in cognitive sciences*, 12(4), 144–151.
- Winford, D. (2007). Some issues in the study of language contact. *Journal of Language Contact*, 1(1), 22–40. doi: 10.1163/000000007792548288
- Wong, P., & Babel, M. (2017). Perceptual identification of talker ethnicity in Vancouver English. *Journal of Sociolinguistics*, 21(5), 603–628. doi: 10.1111/josl.12264
- Wong, W. Y. P. (2006). *Syllable fusion in Hong Kong Cantonese connected speech* (Doctoral dissertation, The Ohio State University). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu1143227948
- Yao, Y., & Chang, C. B. (2016). On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language*, 92(2), 433–467. doi: 10.1353/lan.2016.0031
- Yu, H. (2013). Mountains of gold: Canada, North America, and the Cantonese Pacific. In *Routledge handbook of the chinese diaspora* (pp. 124–137). Routledge. doi: 10.4324/9780203100387.ch7
- Zee, E. (1999). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics*, 120–167. Retrieved from www.jstor.org/stable/23756746