Chapter 4

The structure of voice onset time variation in bilingual long-lag stop categories

4.1 Introduction

A consequence of bilingualism is that individuals must navigate segment inventories that exist in a shared phonetic space, in which sound categories may or may not share aspects of their mental representations (Flege and Bohn, 2021). One of the primary goals in this chapter is to address what languages share, if anything, in the mental representation of speech sound categories. The idea of representation is intended here in the manner typically meant by psycholinguists (e.g., Llompart and Reinisch, 2018) and exemplar theory proponents (e.g., Amengual, 2018). Flege and Bohn (2021) take a similar approach in the revised Speech Learning Model (SLM-r). They describe the units of a multilingual segment inventory as categories comprising input distributions of exemplars: "the sensory stimulation associated with...speech sounds that are heard and seen during production by others...in meaningful conversations" (Flege and Bohn, 2021, p. 32). So, if sound categories from different languages exist in the same phonetic space and are repre-

sented by distributions of exemplars, how, then, can the extent to which languages share representation(s) be assessed?

There are many pieces to this puzzle, and the literature has already addressed some of them. The introduction to this chapter proceeds as follows—section 4.1.1 addresses which sound categories are candidates for shared representation in the first place. Section 4.1.2 reviews the crosslinguistic influence literature, addressing assimilation, dissimilation, and how they reflect the idea of shared representation. Section 4.1.3 identifies the limitations of existing paradigms in crosslinguistic influence and proposes an adaptation of the uniformity framework. Section 4.1.4 introduces the focus of this particular study—long-lag stops in Cantonese and English—and outlines the specific research questions and hypotheses.

4.1.1 Identifying "links" across languages

At first glance, the best candidates for shared representation are sound categories that are linked together. The idea of "links" can be frustratingly vague in the multilingualism literature, but nonetheless represents a crucial concept. In its most basic sense, links exist between sound categories that exert influence on one another under some set of circumstances. The links must exist because crosslinguistic influence can be observed. In a handbook chapter on bilingual phonetics and phonology, Simonet (2016) describes "links or connections of one sort or another between the phonetic categories" (p. 10). Simonet then notes that "these connections...are transiently strengthened in contexts that induce the activation of both languages and inhibited in contexts that favor the use of only one of the languages" (2016, p. 10). Flege and Bohn (2021) offer a bit more detail in the SLM-r, arguing that sound categories will be linked to the closest category in the other language. Determining which categories to pair up remains an empirical challenge from the perspective of speech production, and as a result, Flege and Bohn (2021) rely on perceptual metrics.

The reason for this challenge is because perception and production do not always line up neatly. Flege and Bohn assert that similarity "must be assessed

perceptually rather than acoustically because acoustic measures sometimes diverge from what listeners perceive" (2021, p. 33). This assertion is echoed in an overview chapter on crosslinguistic segment similarity, where Chang (2015) argues that the notion of similarity is best captured abstractly. Chang states that crosslinguistic influence at the segmental level tends to occur between sounds that share "(1) similar positions in the respective phonemic inventories (when considering the contrastive feature oppositions—or, more broadly, the 'relative phonetics'—of the sounds in relation to other sounds in the inventory), and (2) similar distributional facts" (2015, p. 201). This approach to similarity emphasizes a general role for abstraction but does not necessarily invite a formal phonological analysis. Developing such an analysis would likely be a dissertation in itself. Mielke (2012) highlights the challenges of applying phonological features across languages, given the sheer variety of phonetics-phonology mappings in the world's languages.

While Flege and Bohn (2021) and Chang (2015) take different approaches, relative phonetics and perceptual ratings accomplish a similar goal—accounting for abstraction and nonlinearity in listeners' mental representations. In sum, abstract similarity seems to be a prerequisite for the emergence of a link between two sound categories, given how it does a better job of accounting for when and where crosslinguistic influence occurs. The presence of abstract similarity does not address what happens next. It does not entail any particular outcome, and it does not directly address how representation manifests for the sound categories in question.

4.1.2 Crosslinguistic influence and representation

The next step in the puzzle is understanding what happens to linked sound categories. The SLM-r outlines two primary outcomes for sound categories in a shared system—assimilation and dissimilation. The motivation for these outcomes arises from two simple constraints from the productive and perceptual systems. Effectively, don't get too close to each other in perception, and don't get too complicated

in production (Guion, 2003; Lindblom and Maddieson, 1988; Flege and Bohn, 2021). These constraints lead SLM-r to posit that proximity leads to instability, even if what counts as close remains unclear, and that the outcome of instability is assimilation and dissimilation. Considering how bilinguals are fully capable of maintaining subtle distinctions for similar sound categories across languages (e.g., Sundara et al., 2006; Casillas, 2021), this is not a trivial point to make. It may thus be more appropriate to characterize outcomes as assimilation, dissimilation, or contrast maintenance. Each of these outcomes will be discussed in turn.

Assuming that what SLM-r posits is true, then a relatively simple take is that dissimilation for similar sound categories would lead to distinct representations of those categories. A similarly straightforward take is that assimilation leads to a shared representation. The picture is complicated, however, by the idea of imperfect assimilation and what Flege and Bohn term *composite categories*. In the SLM-r, if sounds from two languages are phonetically too close to each other, they will remain linked in a composite category "defined by the statistical regularities present in the combined distributions of the perceptually linked...sounds." (Flege and Bohn, 2021, p. 41). This scenario might be characterized as an imperfectly shared representation, where certain dimensions are kept apart, and others overlap. Alternatively, this lack of clear-cut examples of assimilation in the literature may instead indicate that assimilation and dissimilation might be better cast as ends of a spectrum for a gradient, context-sensitive phenomenon.

There are a few potential reasons for the lack of clear-cut assimilation. First, true assimilation might just be rare in bilingual speech. This reason is supported by a recent meta-analysis of crosslinguistic influence for Spanish and English initial stop consonants (Casillas, 2021). In this environment, English long-lag stops and Spanish short-lag stops are linked to one another. Casillas found that early bilinguals did not produce "compromise" stop categories. That is, early Spanish-English bilinguals did not produce voice onset time that was somehow intermediate to canonical productions by monolinguals of either language. Instead, the production of each category was influenced by task demands and factors such as social

context. This finding echoes arguments made by Bullock and Toribio (2009) on the sophistication and control that bilingual exert over their possible forms. So while there is clear evidence of a link between the two sounds, there is no compromise category. Instead, bilinguals produce a wide range of forms appropriate to and influenced by different contexts. Without considering task and context factors, it is perhaps not surprising that the two sound categories masquerade as a single composite category.

A second reason for the rarity of complete assimilation arises from the experimental and corpus-based approaches typically used to study crosslinguistic influence. In these approaches, the ability to examine crosslinguistic influence for any given pair of sounds hinges on the presence of an observable difference under some set of conditions. I posit that for this reason, most prior work in crosslinguistic influence has focused on sounds that are phonologically similar (i.e., abstract, relative phonetics) yet phonetically distinct. A common example of this arises from languages that differ in their initial stop voicing contrasts. North American English contrasts long- and short-lag stops in initial position; conversely, Spanish contrasts short-lag and prevoiced initial stops. Despite the clear difference in voice onset time, there is strong evidence for a crosslinguistic link between English long-lag and Spanish short-lag stops (Casillas, 2021; Fricke et al., 2016; Goldrick et al., 2014; Bullock and Toribio, 2009; Olson, 2016).

These studies demonstrate phonetic convergence—or variable assimilation—in two ways. First, VOT is shorter for English initial stops produced by bilinguals when compared to monolingual control groups. This result is attributed to the influence on English long-lag stops from the short-lag category in the other language (Olson, 2016). Second, bilinguals appear more likely to produce lead voicing in initial English voiced stops compared to English monolinguals (Sundara et al., 2006). Evidence of crosslinguistic influence arises from comparing bilinguals to monolinguals comparing bilinguals to themselves across different circumstances. For example, Fricke et al. (2016) use a spontaneous speech corpus to demonstrate that Spanish-English bilinguals produce shorter, more Spanish-like VOT in the

lead up to an English-to-Spanish code switch (Fricke et al., 2016). While this body of work makes the presence of a link clear, it also highlights that there are distinct aspects of how these sound categories are represented in the bilingual mind (Casillas, 2021). In the SLM-r, these examples might be considered composite categories. Alternatively, they might be examples of contrasts being maintained in the face of proximity.

In any case, this focus presents a conundrum. By using methods where observing similarity hinges on the ability to detect a difference, researchers preemptively exclude the best candidates for shared representation—those that share both abstract and acoustic similarity. One example comparing highly similar sound categories in the early bilingualism literature comes from a lab-based study of Mandarin-English bilingual children (Yang, 2019). The authors found that highly proficient bilingual 5 to 6-year-olds produced equivalent VOT for Mandarin and English long-lag stops, even though the monolingual comparison groups were consistently different. Yang's result suggests that the difference is either too small to maintain or that 5 to 6-year-old children have not yet mastered it. These claims should be tempered, however, as Yang (2019) did not control for language mode, and adult bilingual behavior was not considered.

Despite some inroads, there is nonetheless a distinct paucity of work examining highly phonetically similar speech sounds across languages, even when such a connection would make sense. A recent study of crosslinguistic influence in Cantonese-English bilinguals compares English long-lag and Cantonese short-lag stops in the context of a language switching paradigm (Tsui et al., 2019). While this comparison reflects the need for stimuli to be acoustically distinct beforehand, it glosses over the fact that both languages contrast short-lag and long-lag VOT in initial position. The best candidates for links—and accompanying crosslinguistic influence—should be the long-lag stops in each language. The null result with balanced bilinguals is thus unsurprising. I do not mean to suggest that the (Tsui et al., 2019) would have gotten more insightful results by comparing Cantonese long-lag stops to English long-lag stops. Instead, my aim is to highlight the design con-

straints of the paradigms used in crosslinguistic influence research. Such methods are better suited for detecting dissimilation or contrasts that have been maintained. Conversely, methods for assessing assimilation must not rely on the presence of detectable acoustic differences.

To summarize, most work in crosslinguistic influence has focused on phonologically-similar yet phonetically-distinct pairs of segments, which are not strong candidates for shared mental representation (as defined at the beginning of this chapter). This focus of this work on telling sounds apart is likely a result of commonly-used paradigms requiring differences to detect influence, and the possibility that assimilation may in fact be rare in early bilingual speech. Additionally, comparisons of categories that already show strong evidence of both abstract and phonetic similarity may be taken for granted and not considered an interesting problem to focus on, despite the nature of mental representation of sound categories being a key focus of psycholinguistics in general (Samuel, 2020). In the interest of understanding mental representation, the best candidates would be the hardest to distinguish using surface forms in the first place.

4.1.3 Adapting the uniformity framework

The study described in this chapter focuses on how to assess whether a pair of sounds is indeed similar, and thus cannot rely on detecting differences in the way that prior studies do. In this way, rather than tell sound categories apart, this chapter aims to tell sound *together*. To this end, I extend the articulatory uniformity framework to the study of multilingual segment inventories. Articulatory uniformity is conceptualized as a constraint on within-talker phonetic variation, in which phonological primitives (e.g., features) are implemented systematically in speech production (Chodroff and Wilson, 2017; Faytak, 2018; Ménard et al., 2008). Put differently, if a set of segments share a phonological feature, that feature should be implemented with the same phonetic target or articulatory gesture (which may or may not have an acoustic consequence). This systematicity has been observed for in vowel height (Ménard et al., 2008), tongue shape (Faytak, 2018), fricative peak

frequency, and stop consonant VOT (Chodroff and Wilson, 2017). In the case of VOT, the relationship between laryngeal gesture and acoustic consequence is clear. While there are straightforward ties to theoretical phonology from articulatory uniformity, the selection of a particular framework is not a straightforward task in a bilingual context. English and Cantonese stops are typically analyzed with different distinctive features—[voice] and [spread glottis], respectively—despite surfacing with long-lag VOT in initial position and occupying the same relative position. The study reported here focuses only on the relative phonetics and sidesteps theoretical phonology for the time being. This is consistent with the argument that theoretical linguistic descriptions do not always neatly map onto psycholinguistic phenomena (Samuel, 2020).

Within-language uniformity has been observed for initial stops in non-native English, such that the relationship between stops for an individual is clear even if between-talker variability is larger than for native speakers (Chodroff and Baese-Berk, 2019). However, the uniformity framework has not yet been extended to early bilingual speech, in particular as a mechanism for comparing how bilinguals produce phonetically similar sounds in each of their languages. Extending the framework in this way, however, follows the conceptualization of uniformity arising from articulatory reuse (Faytak, 2018).

ADD MORE HERE

4.1.4 Long-lag stops in Cantonese and English

English and Cantonese initial long-lag stops are strong candidates for shared underlying representation, because they exhibit both acoustic and relative phonetic *similarity* akin to the difference for Mandarin and English in (Yang, 2019). Consider the initial stop [k^h] with a mean VOT of 80 ms in American English (Lisker and Abramson, 1964) and 91 ms in Hong Kong Cantonese (Clumeck et al., 1981). While these values are objectively different—though based on small sample sizes—it seems that using the same laryngeal timing gesture in this case would be advantageous given the small difference across monolingual populations, that may or

may not be perceptible. While this remains an empirical question, it follows the finding that bilingual Mandarin-English children did not distinguish between languages in VOT (Yang, 2019). Following the predictions of the SLM-r (Flege and Bohn, 2021), this work suggests that long-lag items of minimally distinct VOT would assimilate or dissimilate, but not be stable in such close proximity.

Thus, the present study asks: Do Cantonese-English bilinguals uniformly produce long-lag stops within and across each of their languages? Leveraging the methodology from Chodroff and colleagues (Chodroff and Wilson, 2017, 2018; Chodroff and Baese-Berk, 2019) allows for a new perspective on the structure of variation and nature of representation in bilinguals, and facilitates the study of phonetically similar speech sounds, in ways that other paradigms do not. As may be clear from the framing of the introduction, the hypothesis was that bilinguals would indeed exhibit crosslinguistic uniformity.

4.2 Methods

4.2.1 Corpus

This study uses conversational interview recordings from the SpiCE corpus of speech in Cantonese and English (Johnson et al., 2020). The corpus includes recordings of 34 early Cantonese-English bilinguals (half female, half male) in both languages, with the order of languages counterbalanced. SpiCE also includes hand-corrected orthographic and force-aligned phone level transcripts. The design of the SpiCE corpus is well-suited to the present study, as it includes comparable samples of spontaneous speech from the same set of individuals in two languages, though it differs from prior studies that use larger read speech corpora (Chodroff and Wilson, 2017; Chodroff and Baese-Berk, 2019).

4.2.2 Segmentation & measurement

All instances of prevocalic word-initial /p t k/ were identified from the SpiCE corpus' force-aligned TextGrid transcripts ($n = 13,488$). VOT estimates were refined using AutoVOT (Keshet et al., 2014), with the minimum allowed VOT value set to 15 ms. AutoVOT identifies the onset and offset of positive VOT within a specified window (here, force-aligned boundaries +/- 31 ms). If stops were too close for a 31 ms buffer, the onset of the second stop's window was set as the offset of the preceding window, as TextGrids do not permit overlapping intervals. After running AutoVOT, instances of /p t k/ were subjected to exclusionary criteria to catch errors. Items were excluded if there was substantial enough misalignment that the AutoVOT offset did not fall within the original force-aligned boundaries of the word ($n = 600$), if the previous word was unknown (i.e., unintelligible or in a different language; $n = 268$), if VOT was equal to the minimum value of 15 ms ($n = 618$), or if items had a VOT more than 2.5 s.d. above the grand mean (> 127.8 ms; $n = 249$). Lastly, following (Chodroff and Wilson, 2017), instances of the English word "to" were excluded from the analysis given its propensity for reduction and extremely high frequency ($n = 2295$).

Of the initial sample, 29.9% was excluded, resulting in 9,458 long-lag stops, with Cantonese /p/: $n = 374$, /t/: $n = 1376$, and /k/ $n = 1687$; and English /p/ $n = 1129$, /t/ $n = 1497$, and /k/ $n = 3395$. Talkers had a median of 97 Cantonese stops (range: 59-194) and 166 English stops (range: 69-574). The higher number of English stops is likely due to lexical distributional reasons. The SpiCE corpus has a similar amount of recorded speech in each language, and while Cantonese stops were culled at a slightly higher rate in the exclusions specified above, they made up a smaller proportion to begin with (33% of initial sample vs. 29% of sample before excluding "to"). English also seems to have more highly frequent /k/-initial word types. Conversely, Cantonese /p/ occurs in fewer, less frequent word types in the final sample ($n = 60$, max frequency of 97) than English ($n = 185$, max frequency of 214).

4.3 Analysis & Results

The articulatory uniformity framework offers strong theoretical grounds for interpreting the structure of VOT variation within and across talkers. The analysis qualifies and quantifies that structure from a few different perspectives. In all cases, the pattern of results is depicted by Figure XX, which plots individuals' mean and standard errors for each of the three stops by language—showcasing both variability and commonalities.

4.3.1 Ordinal relationships

Prior work with lab and read speech strongly suggests an expected ordinal relationship for VOT across places of articulation: /p/ $<$ /t/ $<$ /k/. One of the major contributions of (Chodroff and Wilson, 2017) is that these relationships are tighter that would be expected from a purely ordinal perspective. While ordinal relationships are a starting place, they represent just one piece of the puzzle.

The results for the SpiCE corpus suggest that *puzzle* is an appropriate characterization, as talkers largely did not adhere the expected order. Table 4.1 reports the proportion of talkers whose mean VOT values followed the expected /p/ $<$ /t/ $<$ /k/ relationships. Prior work on connected speech reports rates of adherence in the 80-90% range (Chodroff and Baese-Berk, 2019), with the exception of English /t/ $<$ /k/ being drastically lower for native English speakers. While the /t/ $<$ /k/ comparison is also low here (18%), only the English /p/ $<$ /t/ proportion (0.74) is at all close to previous work. This lack of adherence is apparent in... many cases the standard errors overlap, suggesting that a strict ordering by means may not be appropriate. Additionally, many talkers are not internally consistent across languages... this is depicted by... LOTS TO ADD HERE

4.3.2 Pairwise correlations

To examine the relationship between stops within and across languages, 15 pairwise Pearson's *r* correlations were calculated across talker means and are reported

Table 4.1: Proportion of talker means that adhered to expected ordinal relationship for VOT: /p/ < /t/ </ k/ VOT durations. Note that talker VM25A has no instances of Cantonese /p/ in the sample.

Language	**p<t**	**t<k**	**p<k**	n
Cantonese	0.27	0.61	0.40	33
English	0.74	0.18	0.41	34

along with Holm-adjusted p-values where significant. In each case, means were calculated over *residual* VOT values from a simple linear regression in which VOT was predicted by average phone duration within the word—a proxy for speech rate calculated as the the difference between the AutoVOT-estimated onset and the force-aligned word offset, divided by the number of segments in the canonical form of the word. Using residual VOT means mitigates the impact of talker- and language-specific speech rate for these comparisons. This is important, as speech rate is known to influence VOT (Chodroff and Wilson, 2017), and because prior work demonstrate talker and language effects on speech rate (Bradlow et al., 2017).

Table 4.2 summarizes the output of the significant correlations. While there is some evidence for both within- and across-language structured variation, the correlations reported here are considerably lower compared to prior work on English connected speech, where similar within-language comparisons had $r > 0.7$ (Chodroff and Wilson, 2017; Chodroff and Baese-Berk, 2019). With the exception of the English /p/ $\sim$ /k/ ($r = 0.75$, $p < 0.001$), all of the correlations were either moderate ($0.5 < r < 0.7$; $p < 0.01$) or not significant. Within-language correlations more consistently occurred (5 of 6 significant), compared to the across-language comparisons (3 of 9). Notably, most of the comparisons involving /t/ in either language, were not significant. While these relationships seem to indicate some degree of articulatory reuse, the overall picture is not particularly compelling.

Table 4.2: Correlations based on mean residual VOT by talker and language. Each row indicates the comparison, Pearson's *r*, and Holm-adjusted p-value.

Comparison	***r***	***p***
Cantonese /p/ ∼ Cantonese /t/	0.59	0.004
Cantonese /p/ ∼ Cantonese /k/	0.54	0.009
Cantonese /t/ ∼ Cantonese /k/	0.33	0.28
English /p/ ∼ English /t/	0.58	0.004
English /p/ ∼ English /k/	0.75	<0.001
English /t/ ∼ English /k/	0.57	0.005
Cantonese /p/ ∼ English /p/	0.57	0.006
Cantonese /t/ ∼ English /t/	0.31	0.29
Cantonese /k/ ∼ English /k/	0.55	0.006
Cantonese /p/ ∼ English /t/	0.23	0.33
Cantonese /p/ ∼ English /k/	0.35	0.29
Cantonese /t/ ∼ English /p/	0.43	0.08
Cantonese /t/ ∼ English /k/	0.31	0.29
Cantonese /k/ ∼ English /p/	0.56	0.006
Cantonese /k/ ∼ English /t/	0.24	0.33

4.3.3 Linear mixed effect model

In an effort to better account for variation due to known factors such as speech rate and the presence of a preceding pause, a linear mixed effect model was fit with the *lme4* R package (Bates et al., 2015). The aims of the model were two-fold: estimating the effect of language by segment, and elucidating the sources of variation in the random effect structure. The dependent variable, VOT (centered) was predicted by Average Phone Duration (standardized), Preceding Pause (False$= -0.32$, True$= 1$), Language (Cantonese$= -1.75$, English$= 1$), Place of Articulation (Place T: /p/$= -1.91$, /t/$= 1$, /k/$= 0$; Place K: /p/$= -3.38$, /t/$= 10$, /k/$= 1$), and the Language $\times$ Place interaction. As likely apparent from the parenthetical values, all categorical fixed effects were weighted effect coded (following Chodroff and Wilson, 2017). Random intercepts for Talker and Word were in-

cluded, as were by-Talker slopes for Language, Place, and their interaction.[1]

The model returned a significant intercept ($\beta = 3.62$, $SE = 1.22$, $p = 0.004$), significant main effects for Average Phone Duration ($\beta = 7.75$, $SE = 0.23$, $p < 0.001$) and Preceding Pause (True; $\beta = 2.96$, $SE = 0.38$, $p < 0.001$) as well as significant simple effect for Language (English; $\beta = 2.81$, $SE = 0.59$, $p < 0.001$), indicating that VOT was longer at slower speech rates, as well as after pauses and in English, compared to the weighted mean. Neither Place nor its interaction with Language was significant. As one of the mixed effect model analysis goals was to assess the effect of Language across places of articulation, pairwise post-hoc comparisons were computed for Language by Place of Articulation using emmeans, with a confidence level of 0.95, and the Kenward-Roger degrees-of-freedom method. The contrast between languages was significant for /t/ ($\beta = -7.96$, $SE = 2.25$, $p < 0.001$) and /k/ ($\beta = -9.66$, $SE = 2.43$, $p < 0.001$), but not for /p/ ($\beta = -0.81$, $SE = 2.28$, $p = 0.78$). This suggests that VOT is consistently longer in English for /t/ and /k/.

The second goal of the mixed effects analysis was to gain insight into the sources of variation through the random effects structure. Of the random effects, the intercepts for Word ($SD = 11.45$) and Talker ($SD = 6.11$) accounted for the most variation, followed by the by-Talker slope standard deviations for Language ($SD = 1.76$), Place T ($SD = 2.76$), Place T $\times$ Language ($SD = 1.53$), Place K ($SD = 1.80$) and Place K $\times$ Language ($SD = 1.03$). This indicates that talkers and words differ substantially in mean VOT, and that the slopes for Place and Language effects are more consistent across talkers.

4.4 Discussion

This paper reports a study of long-lag stops in Cantonese-English bilingual speech from the SpiCE corpus (Johnson et al., 2020), and uses the uniformity framework to assess VOT similarity within and across languages. In broad strokes, the ev-

[1]Formula: VOT $\sim$ 1 + Place $\times$ Language + Average Phone Duration + Preceding Pause + (Place $\times$ Language | Talker) + (1 | Word).

idence for uniformity both within and across languages was limited. A correlation analysis provides evidence for within-language uniformity and some across-language structure. The magnitudes were mostly moderate, and most did not involve coronal stops. These results are corroborated by the random effects structure of the linear mixed effects model, as more of the variation is attributable to talker intercepts than to the Language and Place slope effects. In this sense, while there is some degree of structure in VOT variation, it seems to be weaker than the evidence in prior work, where strong within-language patterns were observed (Chodroff and Wilson, 2017; Chodroff and Baese-Berk, 2019).

The far more interesting outcomes relate to unexpected results. The ordinal relationships should be interpreted with a grain of salt, as there are a number of potential explanations not immediately relevant to the research question. For example, means were based off of fewer tokens than in prior work (especially for /p/), which may render those proportions less reliable; and, the speech in SpiCE differs in style (conversational vs. read). Lastly, the error often overlaps, potentially making the ordinal relationships unreliable or less meaningful. Another unexpected outcome is that English VOT seems to be consistently longer than in Cantonese—the opposite of what prior work suggested (Clumeck et al., 1981; Lisker and Abramson, 1964). No explanation is offered here other than to reiterate the casual speech style under examination, and that lab and corpus results often differ (Gahl et al., 2012), as do corpus studies of monolingual and bilingual speech (Johnson, 2019).

While the results here do not necessarily provide evidence for a crosslinguistic uniformity constraint, they offer insight into what makes bilingual speech unique, as well as empirical descriptions of bilingual long-lag stop. In terms of describing the relationship between the long-lag stops in each language, talkers seem to maintain a crosslinguistic contrast despite the close proximity of the stops—for many talkers—in the long-lag space. This makes a composite category in SLM-r terms seem plausible (Flege and Bohn, 2021), and merits further investigation.

A lack of strong cross-language uniformity has implications for speech perception, in which tracking a uniformity-like constraint has been proposed as mecha-

nism for rapidly adapting to speech across languages (Reinisch et al., 2013), and in multilingual talker identification (Orena et al., 2019). If the results of this study persist, then such a constraint may have limited use in real communicative contexts, whether or not listeners use it in a lab setting. On the whole, this study highlights the need to study spontaneous speech, and offers a first pass at leveraging the methods of the uniformity framework to better understand crosslinguistic phonetic similarity.

Bibliography

Amengual, M. (2018). Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics*, 69:12–28. → page 1

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48. → page 13

Bradlow, A. R., Kim, M., and Blasingame, M. (2017). Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts l2 speaking rate. *The Journal of the Acoustical Society of America*, 141(2):886–899. → page 12

Bullock, B. E. and Toribio, A. J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In Isurin, L., Winford, D., and deBot, K., editors, *Studies in Bilingualism*, volume 41, pages 189–206. John Benjamins Publishing Company, Amsterdam. → page 5

Casillas, J. V. (2021). Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. *Languages*, 6(1):9. → pages 4, 5, 6

Chang, C. B. (2015). Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In Raimy, E. and Cairns, C. E., editors, *The Segment in Phonetics and Phonology*, pages 199–217. John Wiley & Sons, Inc., Chichester, UK, 1 edition. → page 3

Chodroff, E. and Baese-Berk, M. (2019). Constraints on variability in the voice onset time of L2 English stop consonants. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors, *Proceedings of the 19th International Congress of*

Phonetic Sciences, pages 661–665, Melbourne, Austrailia. Australasian Speech Science and Technology Association Inc. → pages 8, 9, 11, 12, 15

Chodroff, E. and Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61:30–47. → pages 7, 8, 9, 10, 11, 12, 13, 15

Chodroff, E. and Wilson, C. (2018). Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard*, 4(s2). → page 9

Clumeck, H., Barton, D., Macken, M. A., and Huntington, D. A. (1981). The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics*, 9(2):210–225. → pages 8, 15

Faytak, M. D. (2018). *Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese*. Doctoral dissertation, University of California, Berkeley. → pages 7, 8

Flege, J. E. and Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In Wayland, R., editor, *Second Language Speech Learning: Theoretical and Empirical Progress*, pages 3–83. Cambridge University Press. → pages 1, 2, 3, 4, 9, 15

Fricke, M., Kroll, J. F., and 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. → pages 5, 6

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789–806. → page 15

Goldrick, M., Runnqvist, E., and Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, 25(4):1031–1036. → page 5

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, 60(2):98–128. → page 4

Johnson, K. A. (2019). Probabilistic reduction in Spanish-English bilingual speech. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors,

Proceedings of the 19th International Congress of Phonetic Sciences, pages 1263–1267, Melbourne, Australia. Australasian Speech Science and Technology Association Inc. → page 15

Johnson, K. A., Babel, M., Fong, I., and Yiu, N. (2020). SpiCE: A new open-access corpus of conversational bilingual speech in Cantonese and English. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4089–4095, Marseille, France. European Language Resources Association. → pages 9, 14

Keshet, J., Sonderegger, M., and Knowles, T. (2014). AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction (0.91) [Computer Software]. → page 10

Lindblom, B. and Maddieson, I. (1988). Phonetic universals in consonant systems. In Hyman, L. M. and Li, C. N., editors, *Language, speech, and mind: studies in honour of Victoria A. Fromkin*, pages 62–78. Routledge, London. → page 4

Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422. → pages 8, 15

Llompart, M. and Reinisch, E. (2018). Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels. *Journal of Phonetics*, 67. → page 1

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua*, 122(2):145–163. → page 3

Ménard, L., Schwartz, J.-L., and Aubin, J. (2008). Invariance and variability in the production of the height feature in French vowels. *Speech Communication*, 50(1):14–28. → page 7

Olson, D. J. (2016). The role of code-switching and language context in bilingual phonetic transfer. *International Phonetic Association. Journal of the International Phonetic Association; Cambridge*, 46(3):263–285. → page 5

Orena, A. J., Polka, L., and Theodore, R. M. (2019). Identifying bilingual talkers after a language switch: Language experience matters. *The Journal of the Acoustical Society of America*, 145(4):EL303–EL309. → page 16

Reinisch, E., Weber, A., and Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):75–86. → page 16

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, 111:104070. → pages 7, 8

Simonet, M. (2016). The phonetics and phonology of bilingualism. In *Oxford Handbooks Online*. Oxford University Press. → page 2

Sundara, M., Polka, L., and Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, 9(1):97–114. → pages 4, 5

Tsui, R. K.-Y., Tong, X., and Chan, C. S. K. (2019). Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching. *Applied Psycholinguistics*, 40(1):29–58. → page 6

Yang, J. (2019). Comparison of VOTs in Mandarin–English bilingual children and corresponding monolingual children and adults. *Second Language Research*, page 0267658319851820. → pages 6, 8, 9