Chapter 5

Discussion & Conclusion

What are the consequences of a shared phonetic space in the linguistic systems of bilinguals? What is shared? What is kept separate? And, how can methods couched in the study of crosslinguistic influence provide insight into these areas? These questions sit at the core of this dissertation and are approached from two different angles in Chapters 3 and 4, using the data set described in Chapter 2. While this dissertation focuses on describing and understanding the bilingual speech signal in production, the uniting motivation comes from how the signal is perceived. As such, this chapter proceeds as follows. Section 5.1 recapitulates the main points of the content chapters of this thesis, emphasizing the conclusions that are unique to each chapter. Section 5.2 dives into a more general discussion, highlighting how the studies conspire together to inform a broader understanding of how variation is structured in bilingual speech production. Additionally, implications for perception are considered. Section 5.3 makes note of limitations, and Section 5.4 highlights some of the hypotheses that this dissertation generates but does not answer. Lastly, Section 5.5 concludes by summarizing the key contributions of this dissertation to the fields of phonetics, psycholinguistics, and bilingualism.

5.1 Recap

Chapter 2 introduces a new speech corpus, developed as a part of this dissertation. The SpiCE corpus of Speech in Cantonese and English comprises high-quality recordings and transcripts of sentence reading, storyboard narration, and conversational interviews in each language. All talkers in the corpus were early Cantonese-English bilinguals and members of the heterogeneous bilingual speech community in Vancouver, BC, Canada. Chapter 2 documents the motivation, design, and procedures used in the creation of SpiCE. Additionally, a detailed description of the talkers is provided. SpiCE is an open-access corpus freely available to anyone interested in the data—researchers, developers, hobbyists, and the general public (Johnson, 2021b). On its own, SpiCE represents a major contribution to the study of bilingual speech production.

Chapter 3 describes a study on the structure of acoustic voice variation within and across languages for the talkers in the SpiCE corpus. Using a wide array of source and filter-based acoustic measurements on voiced speech in the conversational interviews, Chapter 3 investigates crosslinguistic similarity in three ways. First, the distributions of each measurement were compared across languages on a by-talker basis using Cohen's *d*. The vast majority of comparisons resulted in trivial differences, indicating that talkers were, for the most part, internally consistent. Where consistent differences emerged, they mostly aligned with prior work—Cantonese tended to have lower fundamental frequency and be associated with breathier (or less creaky) voice quality than English (Ng et al., 2012). Second, a series of principal components analyses (PCAs) were run for each talker and language pair. In broad terms, the PCAs bore remarkable similarities in component structure and variance accounted for, regardless of talker and language, given prior work in this domain (Lee et al., 2019; Lee and Kreiman, 2019, 2020). The PCAs were then subjected to canonical correlation analyses to elucidate how much of the lower dimensional structure in one PCA could be accounted for by the other PCA—that is, how much *redundancy* there is between two PCAs—and vice versa. The result of this analysis clearly demonstrates that talkers bear the most similarity

to themselves across languages, compared to across-talker comparisons within or across languages. While there is some variation in the degree of similarity, the takeaway from this chapter is that voices can largely be thought of as "auditory faces."

Chapter 4 presents a second corpus study, focused on describing and analyzing the structure of phonetic category variation within and across languages for long-lag stops in Cantonese and English. Leveraging the uniformity framework (Chodroff and Wilson, 2017), Chapter 4 demonstrates that there is some structure to the relationship between voice onset time (VOT) patterns, but that the account is far less compelling than prior work on English (Chodroff and Wilson, 2017; Chodroff and Baese-Berk, 2019). Talkers were wildly inconsistent with respect to the expected ordinal relationships between means for the stop categories within languages (Chodroff and Wilson, 2017; Cho and Ladefoged, 1999; Lisker and Abramson, 1964). That is, very few talkers produced /p/ with shorter VOT than /t/ or /k/, and likewise between /t/ and /k/. The second phase of the analysis considered pairwise correlations of category means for VOT within and across languages. Again, the results were far from compelling. While there were consistent moderate correlations for within-language comparisons—especially for English—the across-language correlations were weaker or non-significant compared to prior work (Chodroff and Wilson, 2017; Chodroff and Baese-Berk, 2019). Their presence indicates some degree of structure but does not make for tidy conclusions.

Chapter 4 ends with a Bayesian linear mixed-effects model with two primary goals. First, the model estimates the effect of language while accounting factors known to influence VOT, such as local speaking rate and position. Second, the model allows for evaluating the sources of variability within the model. There was a small but consistent effect of language, such that English stops are produced with longer VOT than their Cantonese counterparts. The model also indicates that differences between talkers and words account for far more variation than differences in language and place of articulation effects. Along with the ordinal relationships and correlation analysis, Chapter 4 depicts both talker and language influences on

the structure of VOT. The results corroborate the ideas in the uniformity framework, albeit in a somewhat weaker manner than anticipated.

5.2 General discussion

Each chapter in this dissertation includes a discussion that deals with topics central—and unique—to the study at hand. This general discussion focuses on the bigger picture and some of the implications that the two studies have for perception, as outlined in Chapter 1. While there is substantial similarity across languages in both voice variability and the structure of long-lag stops, each study leaves room for both language and talker-indexical influences.

5.2.1 Talker-indexical and linguistic influences

In Chapter 3, the dual influences of talker and language show up most clearly in the canonical correlation analysis. While talkers exhibit the greatest degree of redundancy (i.e., similarity metric for comparing two PCAs) when compared to themselves across languages, no talker exhibits perfect redundancy. That there is variability in this metric suggests influence from *non*-talker-indexical sources, which could be linguistic or reflect social factors. Yet, while Chapter 3 does not rule out any particular type of influence, the high decree of within-talker similarity across languages emphasizes a clear role for talker-indexical components. This observation is further reflected in the comparisons of each acoustic measurement via Cohen's *d*.

Just over a third of the talkers in the SpiCE corpus maintain a crosslinguistic difference—in the same direction—for measures associated with pitch and non-modal voice quality. Some prior work has attempted to account for differences in fundamental frequency via the presence or absence of lexical tone. These studies, however, are not consistent with one another—some suggest lexical tone leads to lower F0 (Ng et al., 2012), and others to higher F0 (Lee and Sidtis, 2017; Keating and Kuo, 2012). It may be the case that a particular tone system impacts a

talker's F0 profile, but the evidence is not yet compelling for this argument. If the differences of the present study were due to linguistic reasons, then it would be surprising that only a third of talkers exhibited the difference. A more likely account is one that invokes social factors and how individuals express their identity in each language (Loveday, 1981; Voigt et al., 2016). The lack of a clear role for linguistic influences here is further compounded by the behavior of the acoustic measurements typically associated with linguistic attributes. For example, while the patterning of F2 differences across languages largely reflects expectations around the distributions of back vowels in English and Cantonese, the direction of the effect is not consistent across talkers. While the expected difference is present for some, variability on this front and the uncontrolled nature of spontaneous speech make it challenging to draw a conclusion that reflects the group as a whole.

The dual influences of talker and language are perhaps more apparent in Chapter 4. This outcome is not surprising, considering that Chapter 3 examined voice quality, while Chapter 4 homes in on speech sound categories. Voice can be used in both linguistic and non-linguistic manners, while the speech sound categories are a decidedly a linguistic level of structure. For the latter, influence from both talker and language shows up in the juxtaposition between moderate correlations and the difference in VOT across languages. The moderate correlations for homorganic cross-language pairs indicate some level of uniformity in production—talkers with longer VOT in one language tend to also have longer VOT in the other language. Yet, at the same time, the crosslinguistic difference whereby English is characterized by longer VOT reflects a language-based influence.

5.2.2 Shared structure and consequences for perception

Chapter 1 framed this dissertation as being concerned with the consequences of a shared phonetic space, particularly with regard to how the speech signal facilitates processing and identifying multilingual talkers. If talkers are to be consistently identified before and after a language switch, then the speech signals in each language must resemble one another to some extent. Chapters 3 and 4 grapple with

how that resemblance plays out in production. In terms of perception, recall the summary of Orena et al. (2019) in Chapter 1, in which bilingual listeners outperformed monolingual listeners in an identification study featuring bilingual talkers. Orena et al. proposed several accounts for this bilingual advantage, even if performance was above chance across the board. Two of the proposed accounts appeal to listener access to systematicity across languages—Orena et al. (2019) suggest that bilingual listeners are sensitive to systematic *changes* in talker-indexical information and systematic *consistencies* in the linguistic signal.

How, then, do these accounts fare in the context of this dissertation? Put differently, is the assumed (or proposed) systematicity present in the bilingual speech signal? Chapter 3 presents a strong case for the structure of acoustic voice variation and for a variety of specific acoustic measurements. Talkers show a high degree of internal redundancy across languages and overall similarity for the various acoustic measurements. Even in cases where a subset of talkers show a non-trivial difference across languages, few if any differences are large, especially when compared to across-talker differences. In essence, Chapter 3 supports the talker-indexical account, albeit one where there are more likely to be systematic similarities rather than changes. It does not, however, discount the linguistic account.

Chapter 4 offers support for both accounts, though again, it seems to be one of systematic consistencies in talker-indexical aspects and systematic differences in the linguistic system. Note, however, that while the small across-language difference is meaningful regarding what it tells us about the production system, it is not necessarily meaningful to listeners (needs a citation and maybe more elaboration). Such a small difference thus further supports the talker-indexical view of what listeners use to identify talkers across languages.

Returning to the broad question of what language share in phonetic space, this dissertation provides a peek behind the curtain. Languages appear to share a lot in voice quality, despite distributional differences in the segment inventories and different roles for suprasegmental linguistic components (Matthews et al., 2013). While such differences may be more readily apparent in shorter stretches of

speech—as suggested by the passage length analysis in Section 3.2.7—over time, talkers appear to cover their full range. This coverage indicates that different languages make similar use of an individual's full range of acoustic voice variation and exhibit similar patterns of variation in the long run. The task of matching a new utterance up with a familiar talker, then, is one of asking whether or not the new utterance is likely to have arisen from the known range of variation.

Languages also share some aspects of phonetic category structure, albeit to a lesser extent. Chapter 4 demonstrates that talkers with longer VOT in one language tend to have longer VOT in the other language, even though a small distinction between the two languages was maintained. This outcome echoes results for speech rate, in which late bilinguals who are fast in their first language also tend to be fast talkers in their second language (Bradlow et al., 2017). While this relationship between languages is not as simple as saying individuals use the same underlying category in each language, it does demonstrate a certain degree of shared structure; simultaneously, it highlights the complexity of factors conspiring together to produce the acoustic signal.

What is clear from Chapters 3 and 4 is that there is ample shared structure for listeners to use in identifying bilingual talkers. The bilingual advantage in this domain could stem from the variable degree of similarity for talkers across languages and bilinguals' familiarity with how voices might deviate due to social and linguistic reasons. Similarly, it could stem from bilinguals' familiarity with the variety of forms in which a particular category can be produced. While there are clear examples of this kind of sociolinguistically informed variation for initial stops in other language pairs (Bullock and Toribio, 2009), there is also evidence of metalinguistic knowledge for different sound categories in Cantonese-English bilinguals. A recent example juxtaposes the production of word-final stops by the talkers in the SpiCE corpus (Johnson and Babel, 2021a) and a lab-based study of a similar population (Polinsky, 2018). The corpus study demonstrates variability that skews towards Cantonese-like unreleased stops. The lab-based study, conversely, gives evidence for hypercorrection towards longer releases. By adopting the perspective

of Bullock and Toribio (2009), this discrepancy is readily explainable via metalinguistic awareness and how bilinguals use their language in different ways when talking to their peers versus speaking in formal, monolingual, lab-based settings. The point of bringing this example up is to illustrate that bilingual listeners are not only sensitive to the fine-grained acoustics (Ju and Luce, 2004), they are also sensitive to how form varies by communicative context. In sum, there are both systematic similarities and differences available in the signals for listeners—and bilingual listeners in particular—to use in tasks like talker identification.

5.3 Limitations

As with any study, the results presented in this dissertation are necessarily tempered by some limitations and leave a substantial amount of variation unaccounted for. The simplest form of limitation arises from methodological decisions and is touched on in Chapters 3 and 4. Both studies use corpus methods with exclusionary criteria and minimal manual inspection. In Chapter 3 this takes the form of using an automated approach to identify voiced portions of speech and a set of exclusionary criteria to discard likely errors. Chapter 4 relies on forced alignment, refinement via automated methods, and exclusionary criteria. Such approaches allow for the studies to be done at a larger scale but also mean that some degree of error is inevitable. The samples may include items that do not reflect the target. For example, some erroneous VOT measurements may have evaded the exclusionary criteria in Chapter 4—without a rigorous manual check or manual transcription of the SpiCE corpus, the true extent of this problem will remain unknown. While it is outside the scope of this dissertation to perform such a check, prior corpus work with similar exclusionary criteria indicates that the error rate is relatively small (5%: Chodroff and Baese-Berk, 2019).

The population studied here also presents a limitation on the extent to which the results of this dissertation can be generalized to other groups. As summarized in Section 1.1, there is enormous variation between and among bilingual populations—the talkers in the SpiCE corpus are no exception. As described in

detail in Chapter 2, the population studied in this dissertation represents a heterogeneous group of early Cantonese-English bilinguals. While some factors were carefully controlled for in recruitment, others were not. On the one hand, talkers were 18-35 years old at recording, comfortable conversing in Cantonese and English, and began learning both languages before age five. On the other hand, most talkers had some knowledge of at least one additional language (e.g., Mandarin, French, etc.) and varied in their family's current or historical roots in a Cantonese-speaking homeland.

While variability is an inherent part of the Cantonese-speaking community in Vancouver, BC, Canada—and thus justifiably included in the corpus—it does make comparisons with other bilingual communities somewhat challenging. Such comparisons are also complicated by the rather unique position of the speech community. Cantonese is reported as the "mother tongue" of some eight percent of Metro Vancouver census respondents (Statistics Canada, 2017). Given the overall population, Cantonese is an incredibly visible minority language in the region. As argued in Chan et al. (2020), this population likely has more access to Cantonese than bilingual communities in other English-dominant societies (e.g., Bruggeman and Cutler, 2019). While there is much more to say on this topic, such detail is left for future work. Further, while this corpus could be used to explore sources of variation based on different aspects of the talkers' demographics, a lack of control on many of the potentially relevant parameters renders such approaches speculative.

5.4 Current and future directions

There is another kind of "limitation" baked into the framing of this dissertation. While the focus remains on describing and accounting for variation in speech production, the motivation arises from perception. This disjunction means that questions relating to speech perception are not answered here. Rather, the speech production results presented in this dissertation generate hypotheses for how bilingual talkers are perceived and identified. This section outlines what some of those hypotheses are. Each of the following paragraphs in this section begins with an ital-

icized hypothesis and is followed up by the result that motivates it. Any current research being done on the question will be noted. The first two hypotheses derive from Chapter 3 and the latter three from Chapter 4.

Increased within-talker canonical redundancy across languages will facilitate multilingual talker identification and discrimination. Concurrently, greater redundancy between talkers will lead to a higher chance of false alarms or confusability. This hypothesis emerges directly from the finding that some talkers are more similar to themselves than others in the structure of their acoustic voice variation. Variability in production is thus hypothesized to be mirrored in perception. Prior perceptual work in this area often only considers a handful of talkers and uses a handful of coarse checks on the voices (needs citation). As a result, spurious results may be treated as being unique to the talker. This hypothesis seeks to add a concrete account of why talkers differ in this way. Ongoing research on bilingual talker identification and discrimination in the Speech-in-Context Lab aims to address this hypothesis, along with an additional goal of better understanding the listeners' role (Lloy et al., 2020, 2021).

Global shifts in voice quality dimensions across languages will not disrupt talker identification and discrimination when they mirror consistent (when present) patterns in the speech signal. Conversely, shifts not present in the range of voice variation patterns will be disruptive. This hypothesis arises from the assumption that listeners are better at processing variation when they have experience with it. In the context of bilingualism, this is hinted at in a few ways. First, Orena et al. (2019) found that experience with code-switching led to increased performance at generalizing talker identification across languages. Second, anticipatory interference in the speech signal facilitates the processing of a switch from one language to another by listeners (Fricke et al., 2016b). In both cases, experience with how languages differ and experience hearing the two languages in close proximity led to improved processing of the multilingual speech signal. This hypothesis extends the assumption that experience with variation will lend yet another advantage in processing variation.

While all listeners will benefit from congruent VOT within-talker across languages, bilingual listeners will be more adept at learning systematic differences. This hypothesis emerges from Chapter 4, in which VOT differed slightly across languages but was also highly variable between talkers. While this hypothesis ultimately echoes the proposal suggested in Orena et al. (2019), it calls for more explicit manipulation of VOT (e.g., as in the experimental study in Fricke et al., 2016b).

Uniformity will decrease as speech style becomes progressively more formal. While this hypothesis is supported both by prior work (Chodroff and Wilson, 2017) and its comparison with the results in Chapter 4. However, a direct comparison of spontaneous speech to either of the styles represented in Chodroff and Wilson (2017) remains lacking, as the studies report on rather different populations and include many possible confounds.

Full category assimilation in early bilingual speech is exceedingly rare and possibly even non-existent. This hypothesis arises from the small difference maintained between long-lag stop categories across languages; additionally, it is supported by the arguments against compromise categories for bilinguals in Casillas (2021). This hypothesis merely extends that argument to long-lag stops that exhibit an even greater degree of crosslinguistic similarity from the outset.

5.5 Conclusion

Speech is variable, and learning a new talker can be characterized as learning how that talker varies. This dissertation focuses on comparing systematicity across languages to understand how such structure might facilitate processes like multilingual talker identification. The results presented here demonstrate the presence of systematicity at two levels—acoustic voice variation and how long-lag stop series manifest. This structure shows evidence of both talker-indexical and linguistic influences and generates a multitude of hypotheses for future work.

There is a balance between variation and structure at every level—talker, language, linguistic units, voice quality, and more. Working with spontaneous speech

corpora is one of the better ways to gain an appreciation for this observation. It would not be possible to do this kind of research without data, and one of the largest and lasting contributions of this dissertation is the SpiCE corpus. While this dissertation just scratches the surface of what can be done with SpiCE, making the data available will help push our understanding of bilingual speech production forward.

Bibliography

Afouras, Triantafyllos, Chung, Joon Son, and Zisserman, Andrew. Now you're speaking my language: Visual language identification. In *Proceedings of Interspeech 2020*, pages 2402–2406, 2020. doi:10.21437/Interspeech.2020-2921. → page 72

Alderete, John, Chan, Queenie, and Yeung, H. Henny. Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, 191:103952, 2019. doi:10.1016/j.cognition.2019.04.021. → page 11

Altenberg, Evelyn P. and Ferrand, Carole T. Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice*, 20(1):89–96, 2006. doi:10.1016/j.jvoice.2005.01.005. → pages 43, 44, 46, 54, 73

Amengual, Mark. Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism*, 23(5):954–970, 2017. doi:10.1177/1367006917741364. → pages 3, 13

Amengual, Mark. 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, 2018. doi:10.1016/j.wocn.2018.04.002. → page 77

Antoniou, Mark, Best, Catherine T., Tyler, Michael D., and Kroos, Christian. Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38(4):640–653, 2010. doi:10.1016/j.wocn.2010.09.005. → page 82

Antoniou, Mark, Best, Catherine T., Tyler, Michael D., and Kroos, Christian. Inter-language interference in VOT production by L2-dominant bilinguals:

Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39(4): 558–570, 2011. doi:10.1016/j.wocn.2011.03.001. → page 82

Ardila, Rosana, Branson, Megan, Davis, Kelly, Kohler, Michael, Meyer, Josh, Henretty, Michael, Morais, Reuben, Saunders, Lindsay, Tyers, Francis, and Weber, Gregor. Common voice: A massively-multilingual speech corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4218–4222, Marseille, France, 2020. https://www.aclweb.org/anthology/2020.lrec-1.520. → page 11

Audacity Team. Audacity (R): Free audio editor and recorder, 2018. https://www.audacityteam.org/. → page 19

Balukas, Colleen and Koops, Christian. Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, 19 (4):423–443, 2015. doi:10.1177/1367006913516035. → page 82

Barr, Dale J., Levy, Roger, Scheepers, Christoph, and Tily, Harry J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278, 2013. doi:10.1016/j.jml.2012.11.001. → page 101

Bates, Douglas, Kliegl, Reinhold, Vasishth, Shravan, and Baayen, Harald. Parsimonious mixed models. *ArXiv Preprints*, pages 1–21, May 2018. http://arxiv.org/abs/1506.04967. → page 101

Bauer, Robert S. and Benedict, Paul K. *Modern Cantonese Phonology*. De Gruyter Mouton, Berlin, 1997. doi:10.1515/9783110823707. → page 88

Belin, Pascal, Fecteau, Shirley, and Bédard, Catherine. Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3): 129–135, 2004. doi:10.1016/j.tics.2004.01.008. → page 36

Boersma, Paul and Weenink, David. Praat: Doing phonetics by computer, 2021. http://www.praat.org/. Version 6.1.38. → page 49

Bolton, Kingsley, Bacon-Shone, John, and Lee, Siu-lun. Societal multilingualism in Hong Kong. In *Multilingual Global Cities*, pages 160–184. Routledge, 2020. doi:10.4324/9780429463860-12. → page 15

Bradlow, Ann R, Ackerman, Lauren, Burchfield, L Ann, Hesterberg, Lisa, Luque, Jenna, and Mok, Kelsey. Language- and talker-dependent variation in global features of native and non-native speech. In *Proceedings of the 17th International Congress of Phonetic Sciences*, pages 356–359, Hong Kong, 2011. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Bradlow/Bradlow.pdf. → pages 10, 87

Bradlow, Ann R., Kim, Midam, and Blasingame, Michael. 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, 2017. doi:10.1121/1.4976044. → pages 46, 96, 110, 7

Brehm, Laurel and Alday, Phillip M. A decade of mixed models: It's past time to set your contrasts. *OSF Preprints*, July 2021. https://osf.io/3tgq6/. → page 102

Brown, Esther L. and Amengual, Mark. Fine-grained and probabilistic cross-linguistic influence in the pronunciation of cognates: Evidence from corpus-based spontaneous conversation and experimentally elicited data. *Studies in Hispanic and Lusophone Linguistics*, 8(1):59–83, 2015. doi:10.1515/shll-2015-0003. → page 82

Brown, Esther L. and Harper, David. Phonological evidence of interlingual exemplar connections. *Studies in Hispanic and Lusophone Linguistics*, 2(2): 257–274, 2009. doi:10.1515/shll-2009-1052. → page 3

Bruggeman, Laurence and Cutler, Anne. No L1 privilege in talker adaptation. *Bilingualism: Language and Cognition*, pages 1–13, 2019. doi:10.1017/S1366728919000646. → page 9

Bullock, Barbara E. and Toribio, Almeida Jacqueline. Trying to hit a moving target: On the sociophonetics of code-switching. In Isurin, Ludmila, Winford, Donald, and deBot, Kees, editors, *Studies in Bilingualism*, volume 41, pages 189–206. John Benjamins Publishing Company, Amsterdam, 2009. doi:10.1075/sibil.41.12bul. → pages 4, 42, 47, 81, 82, 83, 7, 8

Burkner, Paul-Christian. brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):1–28, 2017. doi:10.18637/jss.v080.i01. → page 100

Burton, A. Mike, Kramer, Robin S. S., Ritchie, Kay L., and Jenkins, Rob. Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, 40(1):202–223, 2016. doi:10.1111/cogs.12231. → pages 38, 39, 61

Casillas, Joseph V. Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. *Languages*, 6(1):9, 2021. doi:10.3390/languages6010009. → pages 80, 81, 83, 11

Ćavar, Malgorzata, Ćavar, Damir, and Cruz, Hilaria. Endangered language documentation: Bootstrapping a Chatino speech corpus, forced aligner, ASR. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 4004–4011, Portorož, Slovenia, 2016. https://aclanthology.org/L16-1632/. → page 28

Chan, Alice Y. W. and Li, David C. S. English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum*, 13(1):67–85, 2000. doi:10.1080/07908310008666590. → page 88

Chan, Leighanne, Johnson, Khia A., and Babel, Molly. Lexically-guided perceptual learning in early Cantonese-English bilinguals. *The Journal of the Acoustical Society of America*, 147(3):EL277–EL282, 2020. doi:10.1121/10.0000942. → pages 3, 9

Chang, Charles B. Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity. In Raimy, Eric and Cairns, Charles E., editors, *The Segment in Phonetics and Phonology*, pages 199–217. John Wiley & Sons, Inc., Chichester, UK, 1 edition, 2015. doi:10.1002/9781118555491.ch9. → page 79

Cheng, Andrew. Cross-linguistic F0 differences in bilingual speakers of English and Korean. *The Journal of the Acoustical Society of America*, 147(2): EL67–EL73, 2020. doi:10.1121/10.0000498. → pages 45, 46

Cho, Taehong and Ladefoged, Peter. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(2):207–229, 1999. doi:10.1006/jpho.1999.0094. → pages 91, 95, 107, 3

Chodroff, Eleanor. *Structured variation in obstruent production and perception*. PhD dissertation, Johns Hopkins University, Baltimore, MD, 2017. https://jscholarship.library.jhu.edu/handle/1774.2/44696. → page 87

Chodroff, Eleanor and Baese-Berk, Melissa. Constraints on variability in the voice onset time of L2 English stop consonants. In Calhoun, Sasha, Escudero, Paola, Tabain, Marija, and Warren, Paul, editors, *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 661–665, Melbourne, Austrailia, 2019. https://assta.org/proceedings/ICPhS2019/papers/ICPhS_710.pdf. → pages 87, 89, 92, 95, 109, 3, 8

Chodroff, Eleanor and Wilson, Colin. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61:30–47, 2017. doi:10.1016/j.wocn.2017.01.001. → pages 86, 87, 88, 89, 90, 91, 95, 96, 98, 99, 102, 105, 107, 109, 3, 11

Chodroff, Eleanor and Wilson, Colin. Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard*, 4(s2), 2018. doi:10.1515/lingvan_2017_0047. → page 89

Chodroff, Eleanor, Golden, Alessandra, and Wilson, Colin. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America*, 145(1): EL109–EL115, 2019. doi:10.1121/1.5088035. → page 99

Clumeck, Harold, Barton, David, Macken, Marlys A., and Huntington, Dorothy A. The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics*, 9(2):210–225, 1981. https://www.jstor.org/stable/23753507. → pages 88, 109

Deuchar, Margaret, Davies, Peredur, Herring, Jon Russell, Parafita Couto, M. Carmen, and Carter, Diana. Building bilingual corpora. In Thomas, Enlli M. and Mennen, Ineke, editors, *Advances in the Study of Bilingualism*, pages 93–110. Multilingual Matters, 2014. doi:10.21832/9781783091713_008. → pages 9, 35

Ethnologue. Chinese, Yue. In Eberhard, David Mm, Simons, Gary F., and Fennig, Charles D., editors, *Ethnologue: Languages of the World*. SIL

International, Dallas, TX, 24 edition, 2021. http://www.ethnologue.com. Online version. → page 11

Faytak, Matthew Donald. *Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese*. Doctoral dissertation, University of California, Berkeley, 2018. https://escholarship.org/uc/item/0jr0010h. → pages 86, 87, 88

Flege, James Emil and Bohn, Ocke-Schwen. The revised speech learning model (SLM-r). In Wayland, Ratree, editor, *Second Language Speech Learning: Theoretical and Empirical Progress*, pages 3–83. Cambridge University Press, 2021. doi:10.1017/9781108886901.002. → pages 1, 3, 77, 78, 79, 80, 81, 85, 88

Fricke, Melinda, Baese-Berk, Melissa M., and Goldrick, Matthew. Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience*, 31 (5):639–645, 2016a. doi:10.1080/23273798.2015.1130234. → page 3

Fricke, Melinda, Kroll, Judith F., and Dussias, Paola E. Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language*, 89:110–137, 2016b. doi:10.1016/j.jml.2015.10.001. → pages 48, 81, 82, 83, 10, 11

Fricke, Melinda, Zirnstein, Megan, Navarro-Torres, Christian, and Kroll, Judith F. Bilingualism reveals fundamental variation in language processing. *Bilingualism: Language and Cognition*, 22(1):200–207, 2019. doi:10.1017/S1366728918000482. → pages 3, 86

Gahl, Susanne, Yao, Yao, and Johnson, Keith. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789–806, 2012. doi:10.1016/j.jml.2011.11.006. → pages 8, 109

Garellek, Marc. The phonetics of voice. In Katz, William F. and Assmann, Peter F., editors, *The Routledge Handbook of Phonetics*. Routledge, 2019. doi:10.4324/9780429056253_5. → page 50

Gelman, Andrew, Simpson, Daniel, and Betancourt, Michael. The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, 2017. doi:10.3390/e19100555. → pages 100, 103

Gertken, Libby M., Amengual, Mark, and Birdsong, David. Assessing language dominance with the Bilingual Language Profile. In Leclercq, Pascale, Edmonds, Amanda, and Hilton, Heather, editors, *Measuring L2 proficiency: Perspectives from SLA*, pages 208–225. Multilingual Matters, Bristol, UK, 2014. doi:10.21832/9781783092291-014. → page 2

Godfrey, J.J., Holliman, E.C., and McDaniel, J. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992. doi:10.1109/icassp.1992.225858. → page 11

Goldrick, Matthew, Runnqvist, Elin, and Costa, Albert. Language switching makes pronunciation less nativelike. *Psychological Science*, 25(4):1031–1036, 2014. doi:10.1177/0956797613520014. → pages 81, 82, 83

Google. Cloud speech-to-text, 2019. https://cloud.google.com/speech-to-text/. v1. → pages 11, 25

Grieve, Jack. Observation, experimentation, and replication in linguistics. *Linguistics*, 0, 2021. doi:10.1515/ling-2021-0094. → page 10

Grosjean, François. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1):3–15, 1989. doi:10.1016/0093_934X(89)90048_5. → pages 1, 2, 3

Grosjean, François. An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16(1):11–21, 2011. doi:10.1177/1367006911403210. → pages 82, 84, 85

Guion, Susan G. The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, 60(2):98–128, 2003. doi:10.1159/000071449. → page 80

Haines, Nathaniel, Kvam, Peter D., Irving, Louis H., Smith, Colin, Beauchaine, Theodore P., Pitt, Mark A., Ahn, Woo-Young, and Turner, Brandon M. Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox. *PsyArXiv Preprints*, August 2020. doi:10.31234/osf.io/xr7y3. → page 99

Hillenbrand, J, Cleveland, R A, and Erickson, R L. Acoustic correlates of breathy vocal quality. *Journal of speech and hearing research*, 37(4):769–778, 1994. doi:10.1044/jshr.3704.769. → page 51

IEEE. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3):225–246, 1969. doi:10.1109/TAU.1969.1162058. → page 21

Ingvalson, Erin M., Ettlinger, Marc, and Wong, Patrick C. M. Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, 18(1):35–47, 2014. doi:10.1177/1367006912456586. → page 4

Iseli, Markus, Shue, Yen-Liang, and Alwan, Abeer. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, 121(4):2283–2295, 2007. doi:10.1121/1.2697522. → page 50

Jadoul, Yannick, Thompson, Bill, and de Boer, Bart. Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71:1–15, 2018. doi:10.1016/j.wocn.2018.07.001. → page 49

Järvinen, Kati, Laukkanen, Anne-Maria, and Aaltonen, Olli. Speaking a foreign language and its effect on F0. *Logopedics Phoniatrics Vocology*, 38(2):47–51, 2013. ISSN 1401-5439. doi:10.3109/14015439.2012.687764. https://doi.org/10.3109/14015439.2012.687764. → pages 44, 46, 48, 58

Johnson, Keith. Massive reduction in conversational American English. In Yoneyama, K. and Maekawa, K., editors, *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, pages 29–54, Toyko, Japan, 2004. The National International Institute for Japanese Language. https://linguistics.berkeley.edu/~kjohnson/papers/Massive.pdf. → page 5

Johnson, Khia A. Probabilistic reduction in Spanish-English bilingual speech. In Calhoun, Sasha, Escudero, Paola, Tabain, Marija, and Warren, Paul, editors, *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 1263–1267, Melbourne, Australia, 2019. → page 109

Johnson, Khia A. Leveraging the uniformity framework to examine crosslinguistic similarity for long-lag stops in spontaneous Cantonese-English bilingual speech. In *Proceedings of Interspeech 2021*, pages 2671–2675, June 2021a. doi:10.21437/Interspeech.2021-1780. → page vi

Johnson, Khia A. SpiCE: Speech in Cantonese and English, 2021b. https://doi.org/10.5683/SP2/MJOXP3. V1. → pages 6, 8, 2

Johnson, Khia A. and Babel, Molly. Language contact within the speaker: Phonetic variation and crosslinguistic influence. *OSF Preprints*, 2021a. doi:10.31219/osf.io/jhsfc. https://osf.io/jhsfc/. → pages 83, 7

Johnson, Khia A. and Babel, Molly. Language contact within the speaker: Phonetic variation and crosslinguistic influence. Technical report, OSF Preprints, 2021b. https://osf.io/jhsfc/. → page 8

Johnson, Khia A., Babel, Molly, Fong, Ivan, and Yiu, Nancy. 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, 2020a. https://www.aclweb.org/anthology/2020.lrec-1.503. → page vi

Johnson, Khia A., Babel, Molly, and Fuhrman, Robert A. Bilingual Acoustic Voice Variation is Similarly Structured Across Languages. In *Proceedings of Interspeech 2020*, pages 2387–2391, 2020b. doi:10.21437/Interspeech.2020-3095. http://dx.doi.org/10.21437/Interspeech.2020-3095. → page vi

Jolliffe, I. T. *Principal Component Analysis*. Springer-Verlag, New York, 2 edition, 2002. doi:10.1007/b98835. → pages 61, 67

Ju, Min and Luce, Paul A. Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15(5):314–318, 2004. doi:10.1111/j.0956-7976.2004.00675.x. → pages 4, 8

Kawahara, Hideki, Agiomyrgiannakis, Yannis, and Zen, Heiga. Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis. In *Proceedings of the 9th ISCA Speech Synthesis Workshop*, pages 221–228, 2016. doi:10.21437/SSW.2016-36. → page 49

Keating, Patricia and Kuo, Grace. Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2):1050–1060, 2012. doi:10.1121/1.4730893. → pages 41, 43, 44, 4

Keating, Patricia, Kreiman, Jody, and Alwan, Abeer. A new speech database for within- and between-speaker variability. In *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 736–739, Melbourne,

Australia, 2019. https://www.assta.org/proceedings/ICPhS2019/papers/ICPhS_785.pdf. → pages 38, 75

Keshet, J., Sonderegger, M., and Knowles, T. AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction, 2014. https://github.com/mlml/autovot/. Version 0.94. → page 90

Kleinschmidt, Dave F., Weatherholtz, Kodi, and Jaeger, T. Florian. Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science*, 10(4):818–834, 2018. doi:https://doi.org/10.1111/tops.12331. → page 5

Kreiman, Jody, Gerratt, Bruce R., Garellek, Marc, Samlan, Robin, and Zhang, Zhaoyan. Toward a unified theory of voice production and perception. *Loquens*, 1(1):e009, 2014. doi:10.3989/loquens.2014.009. → pages 37, 38, 47, 49, 50, 61

Kreiman, Jody, Lee, Yoonjeong, Garellek, Marc, Samlan, Robin, and Gerratt, Bruce R. Validating a psychoacoustic model of voice quality. *The Journal of the Acoustical Society of America*, 149(1):457–465, 2021. doi:10.1121/10.0003331. → pages 37, 61

Kruschke, John K. Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, 6(3):299–312, 2011. doi:10.1177/1745691611406925. → pages 100, 101

Latinus, Marianne and Belin, Pascal. Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2:175, 2011. doi:10.3389/fpsyg.2011.00175. → pages 39, 73

Lavner, Yizhar, Rosenhouse, Judith, and Gath, Isak. The prototype model in speaker identification by human listeners. *International Journal of Speech Technology*, 4(1):63–74, 2001. doi:10.1023/A:1009656816383. → pages 39, 73

Lee, Binna and Sidtis, Diana Van Lancker. The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing*, 20(3):174–185, 2017. doi:10.1080/2050571x.2016.1273572. → pages 44, 45, 46, 48, 75, 4

Lee, Jackson L. PyCantonese, 2018. https://pycantonese.org/. Version 2.2.0. → pages 11, 28

Lee, Yoonjeong and Kreiman, Jody. Within- and between-speaker acoustic variability: Spontaneous versus read speech. In *The 178th Meeting of the Acoustical Society of America*, San Diego, CA, 2019. doi:10.1121/1.5137431. Poster. → pages 38, 39, 66, 2

Lee, Yoonjeong and Kreiman, Jody. Language effects on acoustic voice variation within and between talkers. In *The 179th Meeting of the Acoustical Society of America*, Acoustics Virtually Everywhere, 2020. doi:10.1121/1.5146847. Poster. → pages 38, 39, 42, 61, 66, 2

Lee, Yoonjeong, Keating, Patricia, and Kreiman, Jody. Acoustic voice variation within and between speakers. *The Journal of the Acoustical Society of America*, 146(3):1568–1579, 2019. doi:10.1121/1.5125134. → pages 36, 38, 40, 41, 47, 48, 49, 52, 54, 61, 66, 70, 72, 73, 75, 2

Lein, Tatjana, Kupisch, Tanja, and van de Weijer, Joost. Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism*, 20(6):732–749, 2016. doi:10.1177/1367006915589424. → page 80

Leung, Man-Tak and Law, Sam-Po. HKCAC: The Hong Kong Cantonese adult language corpus. *International Journal of Corpus Linguistics*, 6(2):305–325, 2001. doi:10.1075/ijcl.6.2.06leu. → page 11

Levi, Susannah V. Methodological considerations for interpreting the language familiarity effect in talker processing. *WIREs Cognitive Science*, 10(2):e1483, 2019. doi:10.1002/wcs.1483. → page 40

Liang, Sihua. *Language Attitudes and Identities in Multilingual China: A Linguistic Ethnography*. Springer International Publishing, 2015. doi:10.1007/978-3-319-12619_7. → page 48

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 74(6):431–461, 1967. doi:10.1037/h0020279. → pages 5, 37

Liberman, Mark Y. Corpus phonetics. *Annual Review of Linguistics*, 5(1): 91–107, 2019. doi:10.1146/annurev-linguistics-011516-033830. → page 10

Lieberman, Philip and Blumstein, Sheila E. *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge Studies in Speech Science and Communication. Cambridge University Press, Cambridge, 1988. doi:10.1017/CBO9781139165952. → page 90

Lindblom, Björn and Maddieson, Ian. Phonetic universals in consonant systems. In Hyman, Larry M. and Li, Charles N., editors, *Language, Speech, and Mind: Studies in Honour of Victoria A. Fromkin*, pages 62–78. Routledge, London, 1988. → page 80

Lisker, Leigh and Abramson, Arthur S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422, 1964. doi:10.1080/00437956.1964.11659830. → pages 88, 107, 109, 3

Lisker, Leigh and Abramson, Arthur S. Some effects of context on voice onset time in english stops. *Language and Speech*, 10(1):1–28, 1967. doi:10.1177/002383096701000101. → page 89

Littell, Patrick. Thank-you notes [Version 1.0: Agent focus], 2010. http://totemfieldstoryboards.org/stories/thank_you_notes/. → page 22

Llompart, Miquel and Reinisch, Eva. Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels. *Journal of Phonetics*, 67, 2018. doi:10.1016/j.wocn.2017.12.001. → page 77

Lloy, Angelina, Johnson, Khia A., and Babel, Molly. Bilingual talker identification with spontaneous speech in Cantonese and English: The role of language-specific knowledge. In *The 179th Meeting of the Acoustical Society of America*, Virtual, 2020. doi:10.1121/1.5147685. Poster. → page 10

Lloy, Angelina, Johnson, Khia, and Babel, Molly. Examining the roles of language familiarity and bilingualism in talker recognition. In *The 13th International Symposium on Bilingualism*, Virtual, 2021. Poster. → page 10

Loveday, Leo. Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, 24(1):71–89, 1981. doi:10.1177/002383098102400105. → pages 45, 5

Lüdecke, Daniel, Ben-Shachar, Mattan S., Patil, Indrajeet, and Makowski, Dominique. Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, 5(53):2445, 2020. doi:10.21105/joss.02445. → page 61

Luke, Kang Kwong and Wong, May L.Y. The Hong Kong Cantonese corpus: Design and uses. *Journal of Chinese Linguistics Monograph Series*, 25: 312–333, 2015. https://www.jstor.org/stable/26455290. → page 11

Matthews, Stephen, Yip, Virginia, and Yip, Virginia. *Cantonese: A Comprehensive Grammar*. Routledge, 2013. doi:10.4324/9780203835012. → pages 20, 72, 88, 6

McAuliffe, Michael, Socolof, Michaela, Stengel-Eskin, Elias, Mihuc, Sarah, Wagner, Michael, and Sonderegger, Morgan. Montreal Forced Aligner, 2017. https://montrealcorpustools.github.io/Montreal-Forced-Aligner/. Version 1.0.1. → page 28

McElreath, Richard. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, Boca Raton, 2 edition, 2020. doi:10.1201/9780429029608. → pages 101, 102, 103

McMurray, Bob, Tanenhaus, Michael K., and Aslin, Richard N. Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2): B33–B42, 2002. doi:10.1016/S0010-0277(02)00157-9. → page 110

Ménard, Lucie, Schwartz, Jean-Luc, and Aubin, Jérôme. Invariance and variability in the production of the height feature in French vowels. *Speech Communication*, 50(1):14–28, 2008. doi:10.1016/j.specom.2007.06.004. → pages 86, 87

Mielke, Jeff. A phonetically based metric of sound similarity. *Lingua*, 122(2): 145–163, 2012. doi:10.1016/j.lingua.2011.04.006. → page 79

Mielke, Jeff and Nielsen, Kuniko. 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, 2018. doi:10.1121/1.5059493. → page 88

Munson, Benjamin and Babel, Molly. The phonetics of sex and gender. In Katz, William F. and Assmann, Peter F., editors, *The Routledge Handbook of Phonetics*. Routledge, 2019. doi:10.4324/9780429056253_19. → page 50

Munson, Benjamin, Edwards, Jan, Schellinger, Sarah K, Beckman, Mary E, and Meyer, Marie K. Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of vox humana. *Clinical linguistics & phonetics*, 24(4-5):245–260, 2010. → page 42

Myers-Scotton, Carol. The matrix language frame model: Developments and responses. In *Codeswitching Worldwide*, volume 126 of *Trends in Linguistics. Studies and Monographs*. De Gruyter Mouton, 2011. doi:10.1515/9783110808742.23. → page 49

Nagy, Naomi. A multilingual corpus to explore variation in language contact situations. *Rassegna Italiana di Linguistica Applicata*, 43(1-2):65–84, 2011. http://digital.casalini.it/10.1400/190440. → pages 20, 23, 25

Navarro, Danielle. *Learning statistics with R: A tutorial for psychology students and other beginners*. University of Adelaide, Adelaide, Australia, 2015. http://ua.edu.au/ccs/teaching/lsr. Version 0.5. → page 54

Ng, Manwa L., Hsueh, Gigi, and Sam Leung, Cheung-Shing. Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children. *International Journal of Speech-Language Pathology*, 12 (3):230–236, 2010. doi:10.3109/17549501003721080. → pages 43, 54, 72

Ng, Manwa L, Chen, Yang, and Chan, Ellen YK. Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers—a long-term average spectral analysis. *Journal of Voice*, 26(4):e171–e176, 2012. doi:10.1016/j.jvoice.2011.07.013. → pages 43, 44, 45, 47, 54, 57, 72, 2, 4

Ng, Raymond W. M., Kwan, Alvin C.M., Lee, Tan, and Hain, Thomas. ShefCE: A Cantonese-English bilingual speech corpus for pronunciation assessment. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5825–5829, 2017. doi:10.1109/ICASSP.2017.7953273. → page 10

Nieuwenhuis, Rense, Manfred, te Grotenhuis, and Pelzer, Ben. Weighted effect coding for observational data with wec. *The R Journal*, 9(1):477, 2017. doi:10.32614/rj-2017-017. → page 102

Nygaard, Lynne C. and Pisoni, David B. Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3):355–376, 1998. doi:10.3758/bf03206860. → page 40

Olson, Daniel J. The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*, 46(3): 263–285, 2016. doi:10.1017/S0025100315000468. → pages 81, 82, 83

Ordin, Mikhail and Mennen, Ineke. Cross-linguistic differences in bilinguals' fundamental frequency ranges. *Journal of Speech, Language, and Hearing Research*, 60(6):1493–1506, 2017. doi:10.1044/2016_JSLHR-S-16-0315. → page 46

Orena, Adriel John, Polka, Linda, and Theodore, Rachel M. Identifying bilingual talkers after a language switch: Language experience matters. *The Journal of the Acoustical Society of America*, 145(4):EL303–EL309, 2019. doi:10.1121/1.5097735. → pages 4, 5, 40, 41, 111, 6, 10, 11

Panayotov, Vassil, Chen, Guoguo, Povey, Daniel, and Khudanpur, Sanjeev. Librispeech: An ASR corpus based on public domain audio books. In *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015. doi:10.1109/icassp.2015.7178964. → page 11

Perrachione, Tyler K. Recognizing speakers across languages. In Frühholz, Sascha and Belin, Pascal, editors, *The Oxford Handbook of Voice Perception*, pages 514–538. Oxford University Press, 2018. doi:10.1093/oxfordhb/9780198743187.013.23. → page 40

Perrachione, Tyler K., Furbeck, Kristina T., and Thurston, Emily J. Acoustic and linguistic factors affecting perceptual dissimilarity judgments of voices. *The Journal of the Acoustical Society of America*, 146(5):3384–3399, 2019. doi:10.1121/1.5126697. → pages 40, 41, 47, 66

Pitt, Mark A., Johnson, Keith, Hume, Elizabeth, Kiesling, Scott, and Raymond, William. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, 45(1):89–95, 2005. doi:10.1016/j.specom.2004.09.001. → pages 9, 10, 23

Polinsky, Maria. *Heritage Languages and their Speakers*. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge, 2018. doi:10.1017/9781107252349. → page 7

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. http://www.R-project.org/. → pages 54, 61, 95, 100

Reinisch, Eva, Weber, Andrea, and Mitterer, Holger. Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1):75–86, 2013. doi:10.1037/a0027979. → page 111

Revelle, William. psych: Procedures for psychological, psychometric, and personality research, 2021. https://CRAN.R-project.org/package=psych. Version 2.1.3. → page 95

Ryabov, Rashel, Malakh, Marcella, Trachtenberg, Malka, Wohl, Sherrie, and Oliveira, Gisele. Self-perceived and acoustic voice characteristics of Russian-English bilinguals. *Journal of Voice*, 30(6):772.e1–772.e8, 2016. doi:10.1016/j.jvoice.2015.11.009. → page 45

Samuel, Arthur G. Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, 111:104070, 2020. doi:10.1016/j.jml.2019.104070. → page 86

Sancier, Michele L. and Fowler, Carol A. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4):421–436, 1997. doi:10.1006/jpho.1997.0051. → pages 3, 82

Shue, Yen-Liang, Keating, Patricia, Vicenik, Chad, and Yu, Kristine. VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences*, volume 3, pages 1846–1849, Hong Kong, 2011. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011. → page 49

Simonet, Miguel. The phonetics and phonology of bilingualism. In *Oxford Handbooks Online*. Oxford University Press, 2016. doi:10.1093/oxfordhb/9780199935345.013.72. → page 78

Simonet, Miquel and Amengual, Mark. Increased language co-activation leads to enhanced cross-linguistic phonetic convergence. *International Journal of Bilingualism*, 24(2):208–221, 2019. doi:10.1177/1367006919826388. → pages 3, 23, 82, 83, 85

Sjölander, Kåre. The Snack Sound Toolkit, 2004. https://www.speech.kth.se/snack/. → page 50

Sloetjes, Han and Wittenburg, Peter. Annotation by category: ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, 2008. http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf. → page 25

Soto-Faraco, Salvador, Navarra, Jordi, Weikum, Whitney M., Vouloumanos, Athena, Sebastián-Gallés, Núria, and Werker, Janet F. Discriminating languages by speech-reading. *Perception & Psychophysics*, 69(2):218–231, 2007. doi:10.3758/BF03193744. → page 72

Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*, 2021. https://mc-stan.org. → page 100

Statistics Canada. Proportion of mother tongue responses for various regions in Canada, 2016 Census, 2017. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dv-vd/lang/index-eng.cfm. → pages 13, 9

Stewart, Douglas and Love, William. A general canonical correlation index. *Psychological Bulletin*, 70(3, pt.1):160–163, 1968. doi:10.1037/h0026143. → page 67

Stuart-Smith, Jane, Sonderegger, Morgan, Rathcke, Tamara, and Macdonald, Rachel. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology*, 6(3-4):505–549, 2015. doi:10.1515/lp-2015-0015. → pages 88, 96, 109

Sun, Junyi. jieba, 2020. https://github.com/fxsjy/jieba. Version 0.42.1. → page 28

Sun, Xuejing. Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. In *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*,

volume 1, pages I–333–I–336, 2002. doi:10.1109/ICASSP.2002.5743722. → page 51

Sundara, Megha, Polka, Linda, and Baum, Shari. Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, 9(1): 97–114, 2006. doi:10.1017/S1366728905002403. → pages 80, 81, 82, 83

Tabachnick, Barbara G. and Fidell, Linda S. *Using Multivariate Statistics*. Pearson Education, Inc., 6 edition, 2013. → page 61

Tanner, James, Sonderegger, Morgan, Stuart-Smith, Jane, and Fruehwald, Josef. Toward "English" phonetics: Variability in the pre-consonantal voicing effect across English dialects and speakers. *Frontiers in Artificial Intelligence*, 3, 2020. doi:10.3389/frai.2020.00038. → page 6

Tse, Holman. *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, Pittsburgh, PA, 2019. http://d-scholarship.pitt.edu/35721/. → pages 20, 28

Tsui, Rachel Ka-Ying, Tong, Xiuli, and Chan, Chuck Siu Ki. Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching. *Applied Psycholinguistics*, 40(1):29–58, 2019. doi:10.1017/S0142716418000449. → page 84

Turk, M.A. and Pentland, A.P. Face recognition using eigenfaces. In *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991. doi:10.1109/cvpr.1991.139758. → page 61

Vasishth, Shravan, Nicenboim, Bruno, Beckman, Mary E., Li, Fangfang, and Kong, Eun Jong. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71:147–161, 2018. doi:10.1016/j.wocn.2018.07.008. → pages 100, 103

Voigt, Rob, Jurafsky, Dan, and Sumner, Meghan. Between- and within-speaker effects of bilingualism on f0 variation. In *Proceedings of Interspeech 2016*, pages 1122–1126, San Francisco, CA, 2016. doi:10.21437/Interspeech.2016-1506. → pages 45, 5

Wei, Li. Translanguaging as a practical theory of language. *Applied Linguistics*, 39(1):9–30, 2018. doi:10.1093/applin/amx039. → page 42

Winterstein, Grégoire, Tang, Carmen, and Lai, Regine. CantoMap: A Hong Kong Cantonese MapTask corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2906–2913, Marseille, France, 2020. https://aclanthology.org/2020.lrec-1.355. → page 11

Wong, Wai Yi Peggy. *Syllable fusion in Hong Kong Cantonese connected speech*. Doctoral dissertation, The Ohio State University, Columbus, OH, 2006. http://rave.ohiolink.edu/etdc/view?acc_num=osu1143227948. → pages 26, 28

Xue, Steve An, Hagstrom, Fran, and Hao, JianPing. Speaking fundamental frequency characteristics of young and elderly bilingual Chinese-English speakers: A functional system approach. *Asia Pacific Journal of Speech, Language and Hearing*, 7(1):55–62, 2002. doi:10.1179/136132802805576544. → page 44

Yang, Jing. Comparison of VOTs in Mandarin–English bilingual children and corresponding monolingual children and adults. *Second Language Research*, page 0267658319851820, 2019. doi:10.1177/0267658319851820. → pages 84, 88

Yang, Yike, Chen, Si, and Chen, Xi. F0 patterns in Mandarin statements of Mandarin and Cantonese speakers. In *Proceedings of Interspeech 2020*, pages 4163–4167, 2020. doi:10.21437/Interspeech.2020-2549. → page 45

Yau, Macro. PyJyutping, 2019. https://github.com/MacroYau/PyJyutping. → page 11

Yovel, Galit and Belin, Pascal. A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, 17(6):263–271, 2013. doi:10.1016/j.tics.2013.04.004. → page 39

Yu, Henry. Mountains of gold: Canada, North America, and the Cantonese Pacific. In *Routledge Handbook of the Chinese Diaspora*, pages 108–121. Routledge, 2013. doi:10.4324/9780203100387.ch7. → page 13

Yuan, Jiahong, Ryant, Neville, and Liberman, Mark. Automatic phonetic segmentation in Mandarin Chinese: Boundary models, glottal features and tone. In *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2539–2543, 2014. doi:10.1109/ICASSP.2014.6854058. → page 28