Uniformity and crosslinguistic influence in Cantonese-English bilingual stops

Khia A. Johnson (khia.johnson@ubc.ca), & Molly Babel (molly.babel@ubc.ca)

Linguistics, University of British Columbia, Vancouver, BC, Canada

Session 12: Poster presentations 3 — 13:00-13:45 CET

Introduction

- Crosslinguistic influence is widely attested for phonologically similar, yet phonetically distinct speech sounds [e.g., 6]
- Observing crosslinguistic influence hinges on presence of variable phonetic difference across languages, being able to *tell things apart*
- Frameworks like the revised Speech Learning Model (SLM-r) posit similar sounds are linked; in composite categories [5]
- Constraints from the perceptual and production systems: don't get too close to each other in perception, and don't get too complicated in production [5, 7]
- Best candidates for shared representation are those with greatest similarity, such as English and Cantonese long-lag stop series [4, 10]
- Articulatory uniformity—systematic implementation of "features", broadly construed—and framework developed by [3, 2] facilitates analyzing *already similar* speech sounds; and here is extended across languages to *tell things together*

Research Question

Question: Do Cantonese-English bilinguals uniformly produce long-lag stops within and across their languages?

Hypothesis: Following the predictions of the SLM-r [5] and evidence of uniformity within L2 English [2], we predicted that long-lag stops would assimilate given proximity in phonetic space and that uniformity would appear comparable within and across languages.

Data

- **SpiCE: Speech in Cantonese and English** is a sizable open-access corpus of conversational bilingual speech [8]
- Heterogeneous group of 34 early Cantonese-English bilinguals in Vancouver, BC (19–34; 17 male, 17 female)
- More information: `https://spice-corpus.rtfd.io`

Segmentation & Measurement

- Prevocalic word-initial /p t k/ from both languages
- MFA force-aligned [11] transcripts refined with AutoVOT [9]
- Exclusionary criteria removes likely errors, extreme outliers, and instances of "to" [other high frequency words retained; total of 30% removed; following 3]
- Stop counts by language used in the analysis:

Language	/p/	/t/	/k/
Cantonese	374	1376	1687
English	1129	1497	3395

- The higher number of English stops is likely due to language-specific lexical distributions

Results

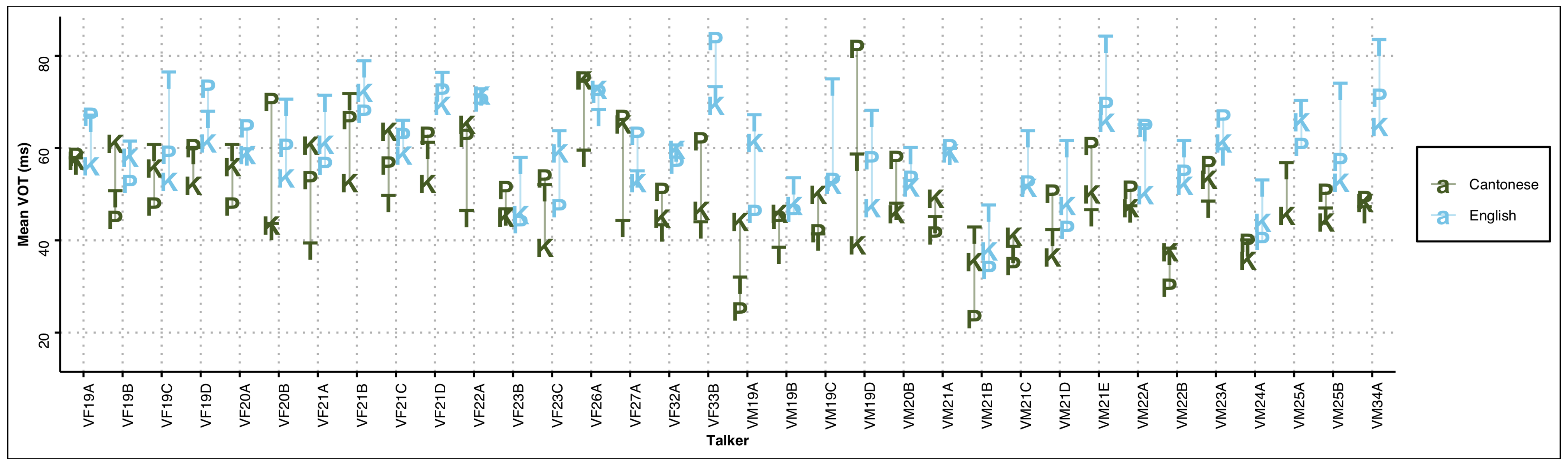

Figure 1:There was very low adherence to the expected ordinal relationship of /p/</t/</k/, and only the relationship for English /p/</t/ reaches anything close to the 80-90% adherence reported in prior work. In this analysis, the following percentages of talkers adhered to the expected pairwise ordinal relationships for their mean VOT: **Cantonese**: 27% /p/</t/, 61% /t/</k/, and 40% /p/</k/; **English**: 74% /p/</t/, 18% /t/</k/, and 41% /p/</k/. This may be due to multilingualism-induced variation, the spontaneous nature of the speech, or something else.

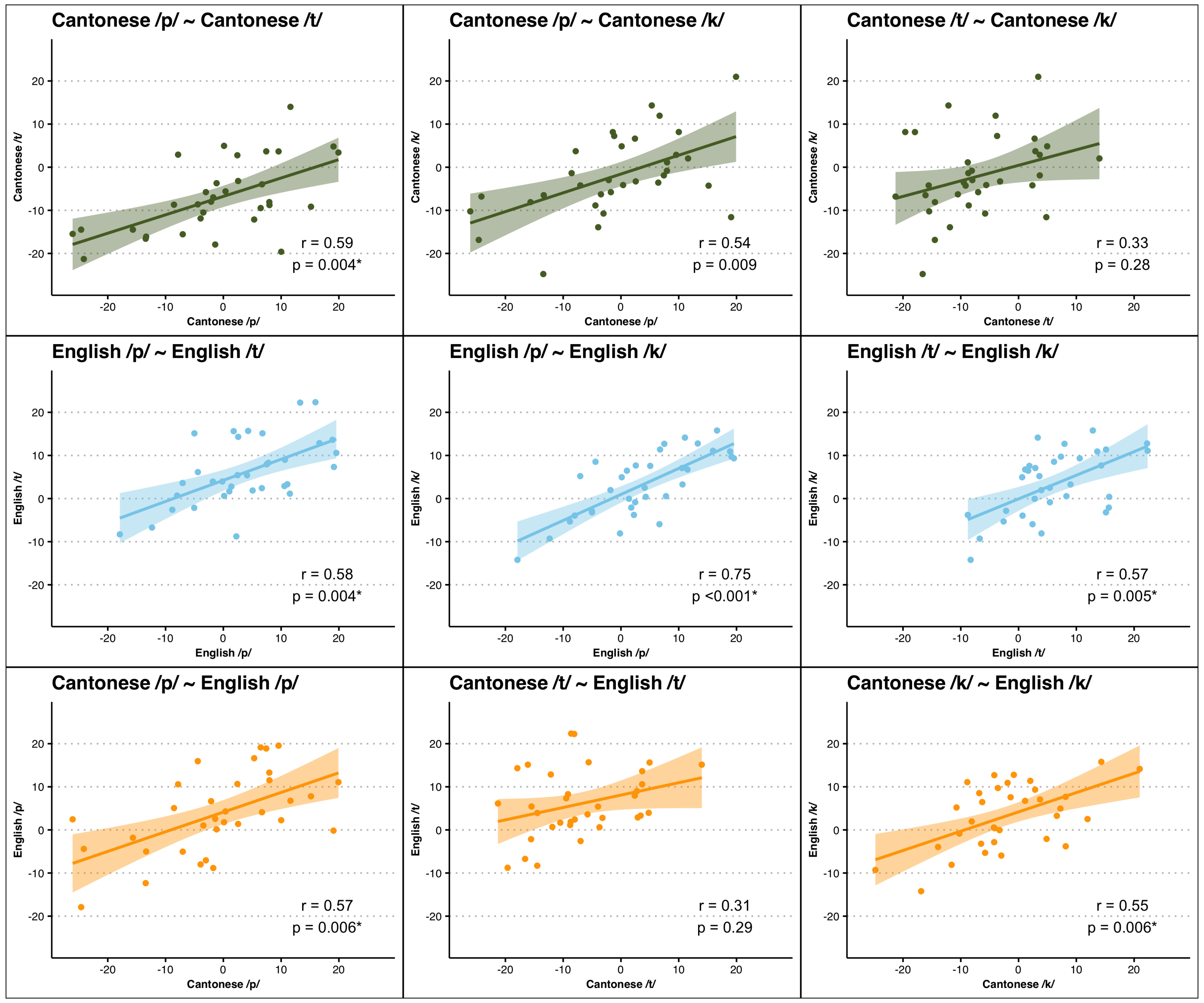

Figure 2:Pairwise correlations of mean residual VOT. Residual VOT was calculated via simple linear regression of *VOT∼ rate*, and accounts for differing default speech rates across languages [see 1]. As in [3], analyses with raw and residual VOT were comparable. Correlation coefficients and Holm-adjusted p-values are superimposed on each panel for: within-Cantonese comparisons (top/green), within-English comparisons (middle/blue), and across-language homorganic comparisons (bottom/orange). Correlations were also computed for across-language non-homorganic comparisons, but only the Cantonese /k/ ∼ English /p/ correlation was moderate and significant ($r = 0.56$; $p = 0.006$).

Mixed effects model

- **Formula**: VOT ∼ 1 + Place × Language + Rate + Pause + (Place × Language | Talker) + (1 | Word).
- Categorical variables were weighted effect coded
- Word ($SD = 11.45$) and Talker ($SD = 6.11$) intercepts accounted for the most random effects variation (rest < 2.8)
- Fixed effects results:

Parameter	Estimate	SE	*p*
Intercept	3.62	1.22	0.005
Place (T)	1.91	1.00	0.06
Place (K)	-1.09	0.65	0.095
Language (English)	2.81	0.59	<0.001
Rate (Average Phone Duration)	7.75	0.23	<0.001
Pause (Precedes: True)	2.96	0.38	<0.001
Place (T) × Lang. (En.)	0.08	0.72	0.91
Place (K) × Lang. (En.)	0.70	0.49	0.16

Discussion & Conclusion

- Some degree of structure in VOT variation was found, yet patterns are weaker compared to prior work, where strong within-language patterns were observed [e.g., 3]—**at best, a murky answer to the research question**
- Unexpected outcome for ordinal relationships potentially due to smaller token count, speech style, and/or multilingualism
- English VOT longer than Cantonese VOT [opposite of: 4, 10]; yet close proximity provides evidence that bilinguals can maintain contrast within long-lag zone [for similar result with vowels, see 7], possibly a composite category [5]
- By-word variability likely reflects prosodic position differences
- Implications for perception: where uniformity-flavored explanations have been proposed to account for perceptual adaptation [13] and multilingual talker identification [12]

References

[1] A. R. Bradlow, M. Kim, and M. Blasingame. 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.

[2] E. Chodroff and M. Baese-Berk. Constraints on variability in the voice onset time of L2 English stop consonants. In *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 661–665, 2019.

[3] E. Chodroff and C. Wilson. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61:30–47, 2017.

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

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

[6] M. Fricke, J. F. Kroll, and P. E. Dussias. Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language*, 89:110–137, 2016.

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

[8] K. A. Johnson. SpiCE: Speech in Cantonese and English [V1], 2021. Scholars Portal Dataverse.

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

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

[11] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger. Montreal Forced Aligner (1.0.1) [software], 2019.

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

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

Acknowledgements

I would like to thank Márton Sóskuthy, and members of the Speech-in-Context lab for their comments and work on the SpiCE corpus. This research was supported by a UBC Public Scholar's initiative award to Khia A. Johnson and a SSHRC grant to Molly Babel.