Bilingual sibilant acoustics in conversational Cantonese-English speech

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

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

Presentation Number: 2aSC2

Introduction

- Sibilants in different languages differ in their trajectories, even when other cues (e.g., center of gravity at midpoint) fail to differentiate them [6].
- Descriptions of Cantonese /s/ vary between [s], [ɕ], and [ʃ], though recent acoustic evidence suggests a more [s̪]-like configuration subject to much anticipatory coarticulation [8].
- Phonological convergence in bilingual speech: "Similar enough" speech sounds tend to converge (at least partially) in bilingual speech production [7].

Research Questions

Do Cantonese-English bilinguals distinguish Cantonese /s/ from the voiceless English sibilants in conversational speech?

PROOF OF CONCEPT QUESTION: How well do spectro-temporal measures work with conversational speech?

Methods

- **Corpus:** ≈ 30 minute conversational interviews with early Cantonese-English bilinguals ($n = 34$), from a new corpus currently being developed by the authors [3].
- **Participants:** Subset ($n = 5$) have been transcribed and force-aligned [5] in both languages. See Table 1.
- **Automatic delineation:** Prevocalic sibilants were filtered ($n = 2244$, pre-emphasis 2000 Hz, high-pass 2500 Hz), and the max intensity between force-aligned edges identified. A smoothing spline was fit to intensity, and on/offsets were set at 60% of peak intensity velocity [method adapted from 2]. Clear errors were discarded ($n = 301$).

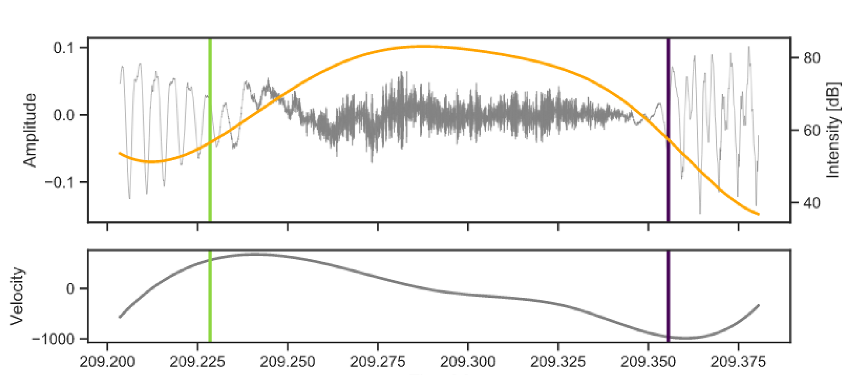

Figure 1: The sibilant delineated here is [s] in *listen*, produced by VF21B. Waveform and intensity (orange) are plotted in the upper panel, and intensity velocity in the lower panel. Onset and offset are marked in both panels.

- **Spectro-temporal measures** were made at 15 points during the sibilant from 6.25% to 93.75% [6, 8].
  - Peak ERB$_N$ trajectories A psychoacoustic measure based on peak frequency (higher → more anterior constriction) from recordings high-pass filtered at 300 Hz [4].
  - Spectral moments Center of gravity, standard deviation, skewness, and kurtosis capture spectral shape.
- **Duration** acts as control measure, as spectro-temporal measures are evenly spaced throughout sibilant.

Participants

Table 1: Summary of participants in this pilot study. All participants were female and ranged in age from 19 to 27 (encoded in Participant ID). Token counts indicate the number of prevocalic sibilants that were included in the final analysis.

Participant ID	Dominant language	Raised in	Age of acquisition Ca.	Age of acquisition En.	Speaking rating Ca.	Speaking rating En.	Token counts Ca. /s/	Token counts En. /s/	Token counts En. /ʃ/
VF19A	English	Canada	0	0	Good	Excellent	216	145	73
VF20A	English	Canada	0	4	Fair	Excellent	188	159	36
VF21A	English	Canada/HK	0	0	Excellent	Excellent	193	307	44
VF21B	Cantonese	HK/Canada	0	3	Excellent	Excellent	172	99	37
VF27A	English	Singapore	0	0	Good	Excellent	123	134	17

Results

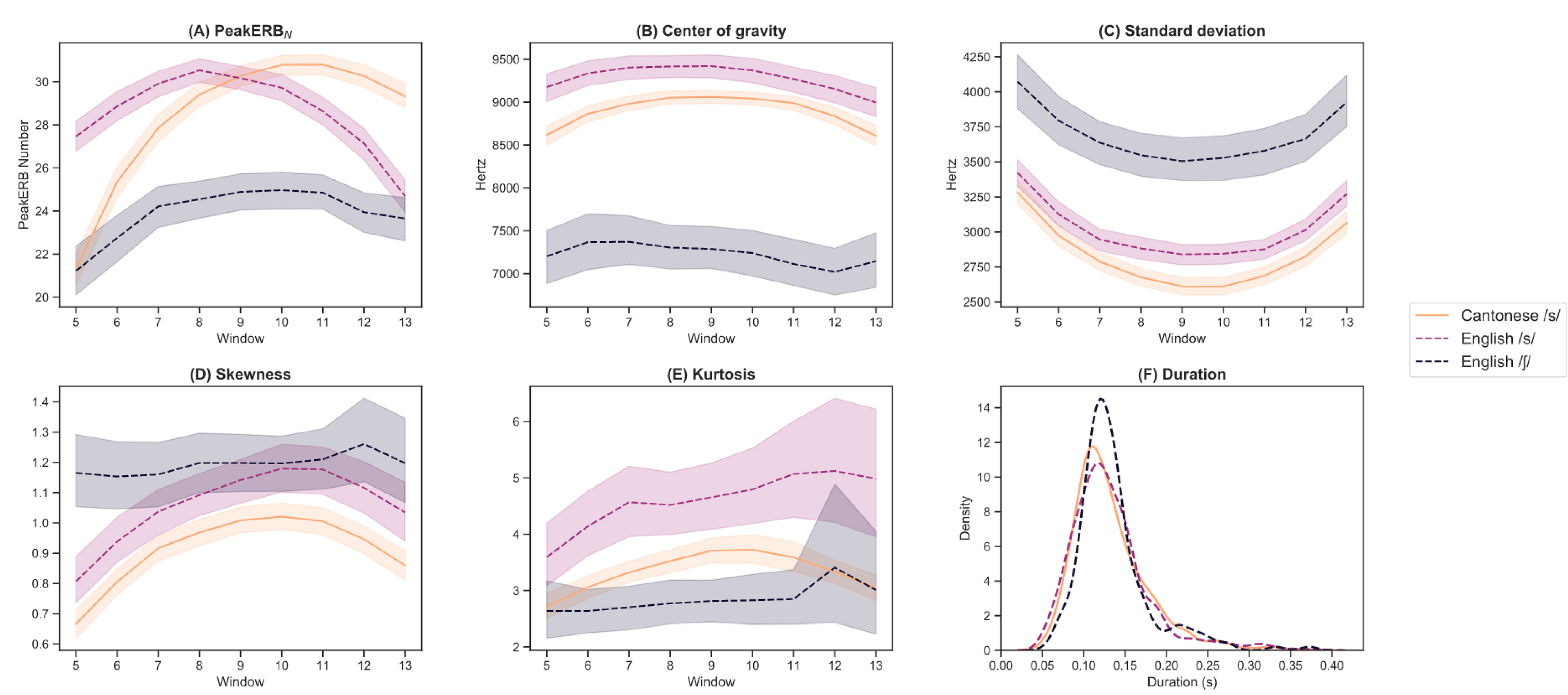

Figure 2: Panels (A)-(E) depict prevocalic sibilant trajectories for the five spectro-temporal measures, across windows centered at the middle nine points in order to capture the sibilants supposed "steady state" [as in 6, 8]. Panel (F) shows the duration distributions for each sibilant.

Table 2: Significant main and interaction effects ($p < 0.05$) for each mixed effects growth curve model fit with *lmer* in *R* [1]. Models were fit to the middle nine points measured, as plotted above. Each column header indicates the dependent variable for the model with following specification, where *Measure* is a placeholder: *Measure* $\sim 1 + Duration + Phone \times (Time^1 + Time^2) + (1 + Time^1 + Time^2 \mid Speaker)$. Time^1 represents linear time, where positive values correspond to an overall increase across windows. Time^2 represents quadratic time, for which a negative value indicates concavity across windows. A model for Standard deviation (C) was run, but did not converge.

	(A) Peak ERB number	(B) Center of gravity	(D) Skewness, $\log_{10}$	(F) Kurtosis, $\log_{10}$
Intercept	24.1	8806	0.63	0.70
Duration	30.9	—	0.12	0.26
Phone: Eng. /s/	—	418	0.004	0.02
Phone: Eng. /ʃ/	−4.6	−1595	0.02	−0.06
Time^1	7.1	—	0.02	—
Time^2	−5.4	−499	−0.002	−0.07
Time^1-Phone: Eng. /s/	−10.1	—	—	—
Time^1-Phone: Eng. /ʃ/	−4.7	—	—	—
Time^2-Phone: Eng. /s/	—	167	—	—
Time^2-Phone: Eng. /ʃ/	2.9	404	0.02	0.04

Discussion & Conclusion

- Duration (F) distributions are comparable across all three target segments, despite differences in timing for the two languages (syllable- vs. stress-timing).
- Cantonese /s/ is more similar to English /s/ than /ʃ/, on a variety of measures, but not all.
  - Peak ERB$_N$ (A) English /s/ and Cantonese /s/ reach a similar front cavity size (see main effect of *Phone: Eng. /s/*), and have similar curves, given the non-significant interaction between *Time*^{2} and *Phone: Eng. /s/*. English /ʃ/ exhibits a shallower curve and is overall much lower than Cantonese /s/. These differences are not captured by center of gravity (B).
  - Standard deviation (C) The model did not converge; visually, English /ʃ/ appears more diffuse than either English /s/ or Cantonese /s/.
  - Skewness (D) The main effect for *Phone: Eng. /s/* is small, and the interactions with Time^1 and Time^2 are not significant.
  - Kurtosis (E) Model results for English /s/ and /ʃ/ follow from prior work [4], in which higher values correspond to narrower constrictions. Here, they are rather variable. Interestingly, Cantonese /s/ falls in the middle, with a narrower confidence interval. This compliments the plot for Standard deviation (C).

Take Home Points

1. While bilinguals produce /s/ similarly in both languages along most spectro-temporal measures, English /ʃ/ is consistently different from Cantonese /s/.
2. Spectro-temporal measures were successfully used with automatic methods and conversational speech.

References

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

[2] T. Ennever, F. Meakins, and E. R. Round. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1):1–32, 2017.

[3] K. A. Johnson, M. Babel, I. Fong, and N. Yiu. SpiCE: A new open-access corpus of conversational bilingual speech in Cantonese and English. Manuscript in preparation, 2019.

[4] K. Maniwa, A. Jongman, and T. Wade. Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, 125(6):3962–3973, 2009.

[5] M. McAuliffe, M. Socolof, E. Stengel-Eskin, S. Mihuc, M. Wagner, and M. Sonderegger. Montreal Forced Aligner, 2017.

[6] P. F. Reidy. Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, 140(4):2518–2529, Oct. 2016.

[7] M. Simonet. The Phonetics and Phonology of Bilingualism. *Oxford Handbooks Online*, Oct. 2016.

[8] A. C. L. Yu. Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *The Journal of the Acoustical Society of America*, 139(4):1672–1690, Apr. 2016.

Acknowledgements

Collection of the corpus used in this study was funded by the UBC Public Scholars Initiative, and a SSHRC grant awarded to Molly Babel. We would like to thank the Speech-in-Context Lab members for their support with the corpus project, in particular Nancy Yiu, Ivan Fong, Kristy Chan, Ariana Hernandez, and Christina Sen.