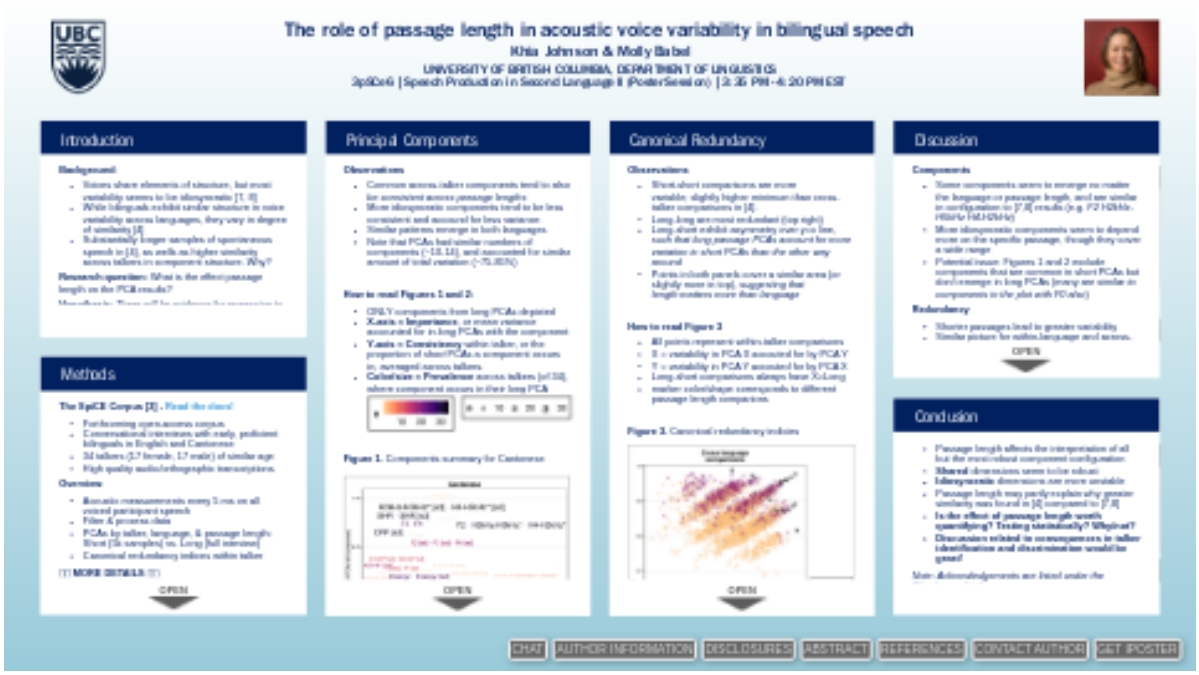
The role of passage length in acoustic voice variability in bilingual speech



Khia A. Johnson & Molly Babel

UNIVERSITY OF BRITISH COLUMBIA, DEPARTMENT OF LINGUISTICS

3pSCe6 | Speech Production in Second Language II (Poster Session) | 3:35 PM - 4:20 PM EST

PRESENTED AT:

ACOUSTICS VIRTUALLY EVERYWHERE

ASA ACOUSTICAL SOCIETY OF AMERICA

179th Meeting

7 -11 December 2020

INTRODUCTION

Background:

- Voices share elements of structure, but most variability seems to be idiosyncratic [7, 8]
- While bilinguals exhibit similar structure in voice variability across languages, they vary in degree of similarity [4]
- Substantially longer samples of spontaneous speech in [4], as wells as higher similarity across talkers in component structure. Why?

Research question: What is the effect passage length on the PCA results?

Hypothesis: There will be evidence for regression to the mean, such that shorter samples will be more variable.

METHODS

SpiCE Corpus [3] → spice-corpus.readthedocs.io (http://spice-corpus.readthedocs.io)

- Forthcoming open-access corpus
- Conversational interviews with early, proficient bilinguals in English and Cantonese
- 34 talkers (17 female, 17 male) of similar age
- High quality audio/orthographic transcriptions

Overview:

- Acoustic measurements every 5 ms on all voiced participant speech
- Filter & process data
- PCAs by talker, language, & passage length: Short (5k samples) vs. Long (full interview)
- Canonical redundancy indices within talker

👇 **MORE DETAILS** 👇

NOTE: The methods are nearly identical to [4] which were adapted from [7,8].

Data preparation steps:

1. Identify voiced participant speech using Praat algorithms [1] with Parselmouth [2]:
   - *Point Process (periodic, cc)*
   - *To TextGrid (vuv)*
2. Collect acoustic measurements every 5 ms with Voicesauce [9], based on psychoacoustic voice quality model [6]:
   - Pitch: F0
   - Formants: F1, F2, F3, and F4
   - Source spectral shape: H1*–H2*, H2*–H4*, H4*–H2kHz*, H2kHz*– H5kHz*
   - Spectral noise: CPP, Energy, SHR
3. Process data:
   - Filter impossible values
   - Compute rolling standard deviations
   - Identify full length (all samples from interview) and 5k contiguous samples for passage length comparisons (approximately matches spontaneous speech in [8]).

Analysis steps: Code available on OSF (https://osf.io/ybdkw/)

1. Conduct principal components analyses by talker, language, and passage length:
   - Adjusted Kaiser-Guttman rule [5]
   - Interpret PCA loadings > |0.32|
   - Output: Lower dimensional structure of voice variability
   - Interpret component structure with respect to consistency, importance, and prevalence.
2. Conduct canonical redundancy comparisons [see 5] across languages for same talker, same passage length:
   - Use all loadings, ignores component order
   - Output: proportion of variation in Cantonese PCA accounted for by English PCA, and vice versa

PRINCIPAL COMPONENTS

Observations

- Common across-talker components tend to *also be consistent across passage lengths*
- More idiosyncratic components tend to be less consistent and account for less variance
- Similar patterns emerge in both languages
- Note that PCAs had similar numbers of components (~10-14), and accounted for a similar amount of total variation (~75-85%)

How to read Figures 1 and 2:

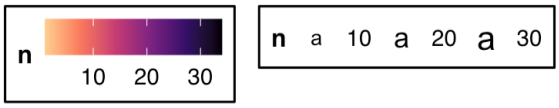
- ONLY components occurring in the long PCAs depicted
- **X-axis ≈ Importance**, or mean variance accounted for in *long* PCAs with the component
- **Y-axis ≈ Consistency** within talker, or the proportion of *short* PCAs a component occurs in, averaged across talkers
- **Color/size ≈ Prevalence** across talkers (of 34) 

Figure 1. Components summary for Cantonese

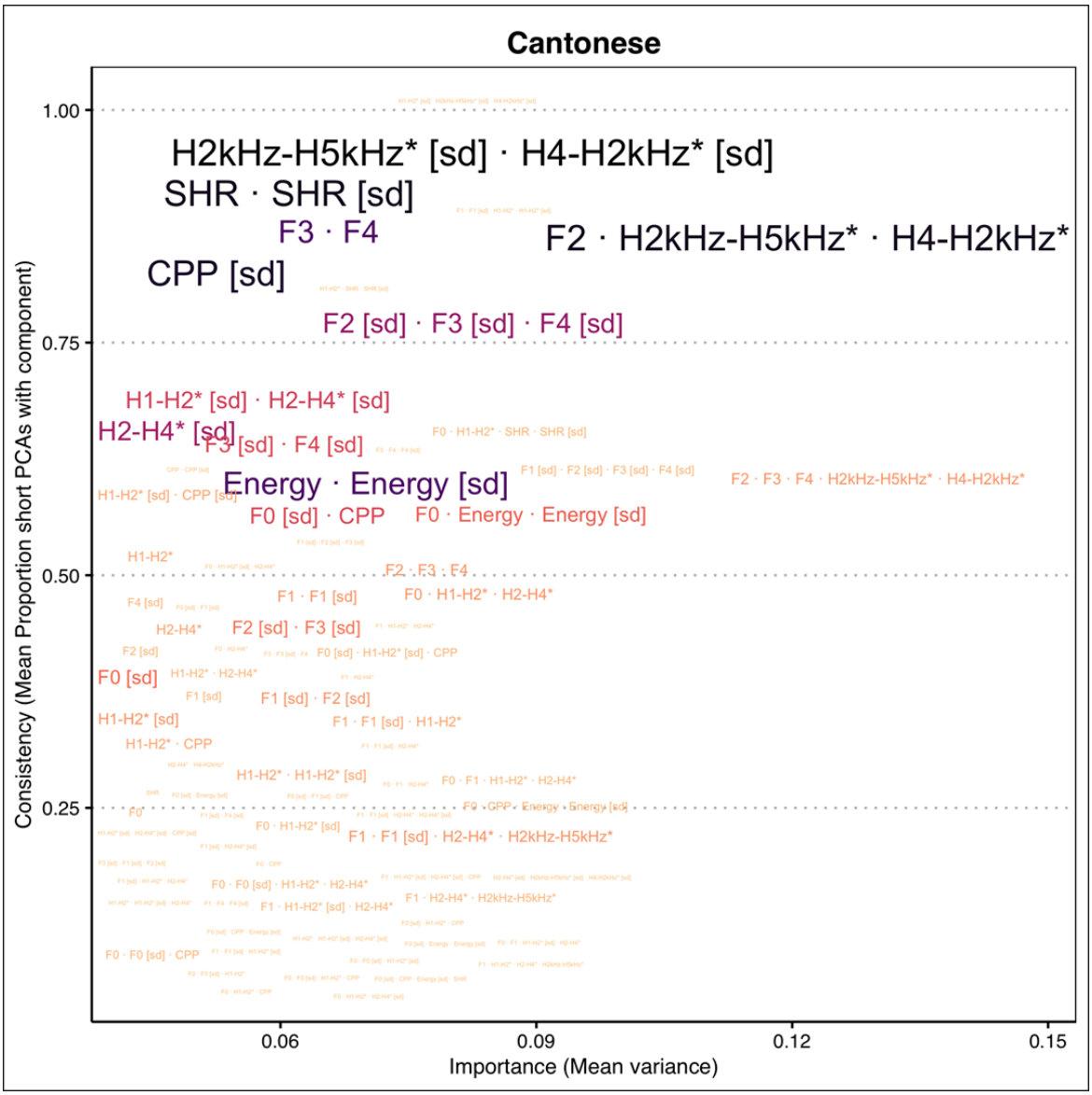
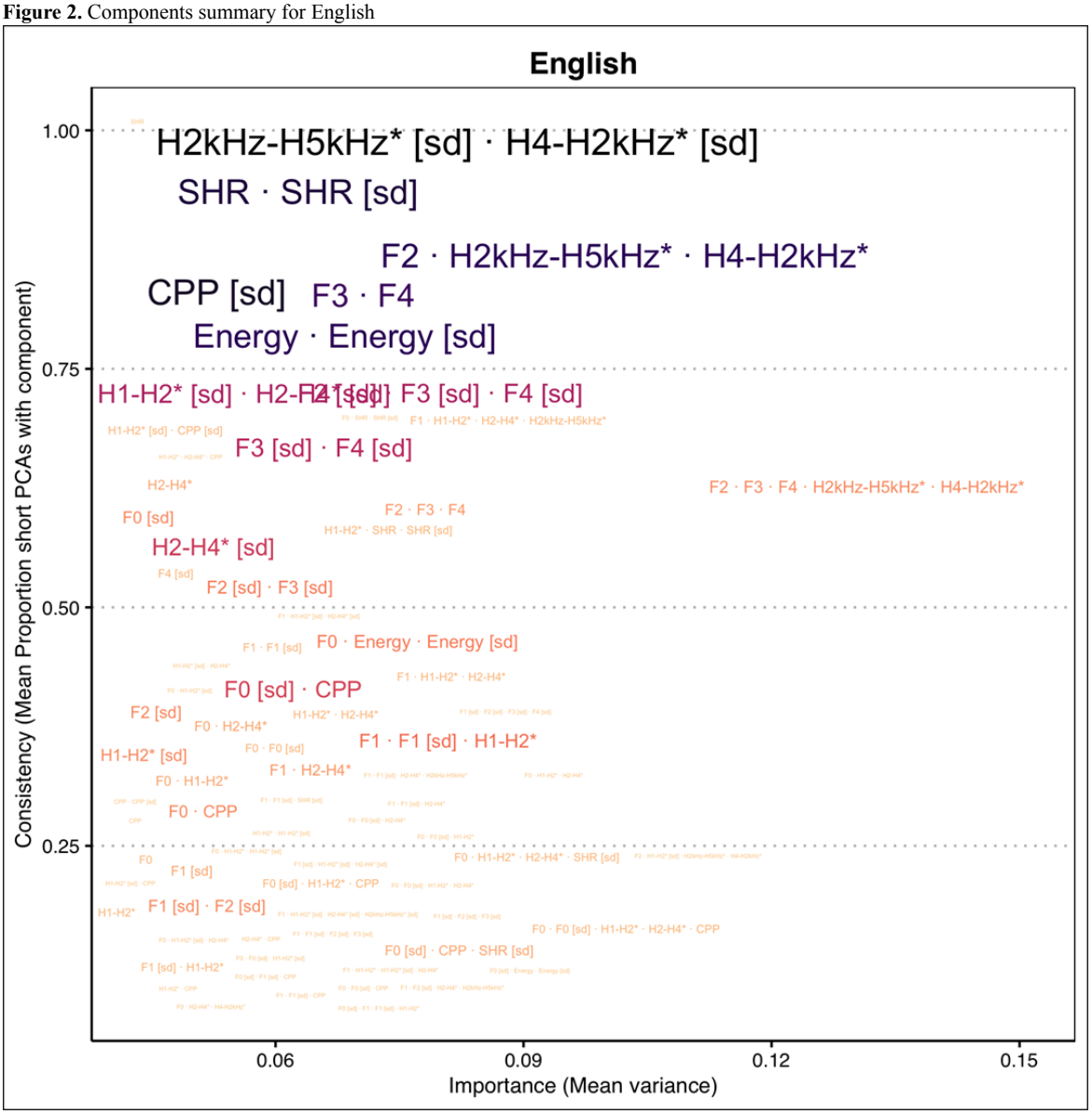

Cantonese
Consistency (Mean Proportion short PCAs with component)
1.00
0.75
0.50
0.25
H2kHz-H5kHz* [sd] · H4-H2kHz* [sd]
SHR · SHR [sd]
F3 · F4
F2 · H2kHz-H5kHz* · H4-H2kHz*
CPP [sd]
F2 [sd] · F3 [sd] · F4 [sd]
H1-H2* [sd] · H2-H4* [sd]
H2-H4* [sd]
F3 [sd] · F4 [sd]
Energy · Energy [sd]
F2 · F3 · F4 · H2kHz-H5kHz* · H4-H2kHz*
H1-H2* [sd] · CPP [sd]
F0 [sd] · CPP
F0 · Energy · Energy [sd]
H1-H2*
F2 · F3 · F4
F1 · F1 [sd]
F0 · H1-H2* · H2-H4*
F4 [sd]
H2-H4*
F2 [sd] · F3 [sd]
F2 [sd]
F0 [sd] · H1-H2* [sd] · CPP
F0 [sd]
H1-H2* · H2-H4*
F1 [sd]
F1 [sd] · F2 [sd]
H1-H2* [sd]
F1 · F1 [sd] · H1-H2*
H1-H2* · CPP
H1-H2* · H1-H2* [sd]
F0 · F1 · H1-H2* · H2-H4*
F0
F0 · GPP · Energy · Energy [sd]
F0 · H1-H2* [sd]
F1 · F1 [sd] · H2-H4* · H2kHz-H5kHz*
F0 · F0 [sd] · H1-H2* · H2-H4*
F1 · H1-H2* [sd] · H2-H4*
F1 · H2-H4* · H2kHz-H5kHz*
F0 · F0 [sd] · CPP
0.06
0.09
0.12
0.15
Importance (Mean variance)

Figure 2. Components summary for English



CANONICAL REDUNDANCY

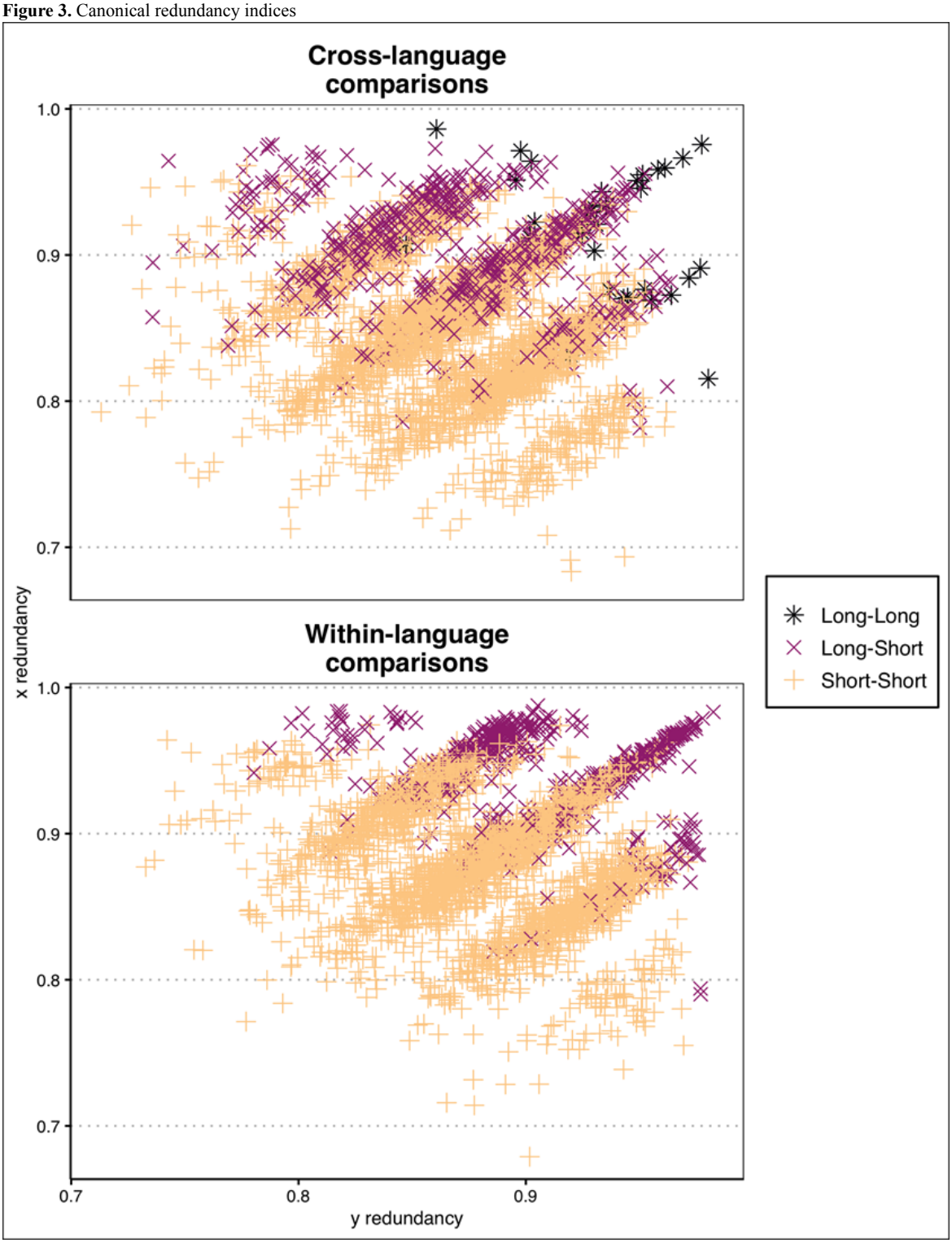
Observations

- Short-short comparisons are more variable; slightly higher minimum than cross-talker comparisons in [4].
- Long-long are most redundant
- Long-short exhibit asymmetry over y = x line, *such that long PCAs account for more variation in short PCAs than vice versa*
- Points in both panels cover a similar area (or slightly more in top), *suggesting that passage length might matter more than language*

How to read Figure 3

- All points represent canonical redundancy indices for within-talker comparisons
- X = variability in PCA X accouted for by PCA Y
- Y = variability in PCA Y accouted for by PCA X
- Long-short comparisons *always* have X = long
- Marker color/shape corresponds to different passage length comparions
- Top = across language comparisons; bottom = within-language comparisons

Figure 3. Canonical redundancy indices



DISCUSSION

Components

- Some components seem to emerge no matter the language or passage length, and are broadly similar in configuration to [7,8] results (e.g. *F2 H2kHz-H5kHz H4-H2kHz*)
- More idiosyncratic components seem to depend more on the specific passage, though they cover a wide range
- Potential issue: Figures 1 and 2 exclude components from short PCAs but don't emerge in long PCAs regardless of how common they are—*note that many involve F0.*

Redundancy

- Shorter passages lead to greater variability
- Similar picture for within-language and across-languages suggests that length is more important for this measure than language
- Potential issue: Not all long samples have precisely the same length

CONCLUSION

- Passage length affects the interpretation of all but the most robust component configurations
- **Shared** dimensions seem to be robust
- **Idiosyncratic** dimensions seem more unstable
- Passage length may partly explain why greater similarity was found in [4] compared to [7,8]
- **Is the effect of passage length worth quantifying? Testing statistically? Why/not?**
- **Discussion related to consequences in talker identification and discrimination would be great!**

Note: Acknowledgements are listed under the "Disclosures" tab.

Code available on OSF → ***osf.io/ybdkw/*** ***(https://osf.io/ybdkw/)***

DISCLOSURES

Acknowledgements

This project draws on research supported by the Social Sciences and Humanities Research Council of Canada (SSHRC), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the University of British Columbia Public Scholars Initiative.

This work builds on our Interspeech 2020 paper with Robert Fuhrman (https://www.isca-speech.org/archive/Interspeech_2020/abstracts/3095.html).

We are grateful to members of the Speech-in-Context Lab (https://speechincontext.arts.ubc.ca/) for feedback on this work, and to everyone who has worked on the SpiCE Corpus, including: Nancy Yiu, Ivan Fong, Kristy Chan, Katherine Lee, Rachel Wong, Christina Sen, Ariana Hernandez, Natalia Oliveira, Michelle To, and Rachel Soo.

AUTHOR INFORMATION

Khia A. Johnson

khia.johnson@ubc.ca

Twitter: @khia_johnson

Molly Babel

molly.babel@ubc.ca

ABSTRACT

An individual's voice is determined in part by the limitations of their anatomy and physiology, in addition to language-specific phonological and phonetic structure. When a bilingual switches between languages, how much do they change their voice? Previous work using a corpus of spontaneous speech from early Cantonese-English bilinguals found surprisingly little variability across individuals' languages [Johnson, Babel, & Fuhrman, Proc. of Interspeech (2020)] compared to earlier research on across-talker acoustic voice variability [Lee, Keating, & Kreiman, JASA (2019)]. A crucial difference between these two studies, however, is passage length. A longer passage (e.g., 30 minutes) potentially allows for a more stable structure to emerge in a principal components analysis, while a shorter sample (e.g. 2 minutes or less) may instead be subject to ephemeral variation, and potentially misrepresent the overall variability of a voice. Building on Johnson et al. (2020), the present study asks: to what extent does passage length impact the results of principal components and canonical redundancy analyses designed to elucidate within-talker (across languages) and across-talker (within language) idiosyncratic variation? These results are important for theories of talker recognition, identification, and discrimination, in addition to improving understanding of talker-specific acoustic-phonetic variation.

REFERENCES

1. Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (.) [Computer software]. http://www.praat.org/
2. Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics, 71, 1–15. https://doi.org/10/ggc7w8
3. Johnson, K. A., Babel, M., Fong, I., & Yiu, N. (2020). SpiCE: A New Open-Access Corpus of Conversational Bilingual Speech in Cantonese and English. Proceedings of The 12th Language Resources and Evaluation Conference, 4089–4095. https://www.aclweb.org/anthology/2020.lrec-1.503
4. Johnson, K. A., Babel, M., & Fuhrman, R. A. (2020). Bilingual Acoustic Voice Variation is Similarly Structured Across Languages. Proc. Interspeech 2020, 2387–2391. https://doi.org/10.21437/Interspeech.2020-3095
5. Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer-Verlag. https://doi.org/10.1007/b98835
6. Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., & Zhang, Z. (2014). Toward a unified theory of voice production and perception. Loquens, 1(1), 009. https://doi.org/10/ggntxv
7. Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. The Journal of the Acoustical Society of America, 146(3), 1568–1579. https://doi.org/10/ggnsgh
8. Lee, Y., & Kreiman, J. (2019, December 5). Within- and between-speaker acoustic variability: Spontaneous versus read speech [Poster]. ASA 178, San Diego, CA.
9. Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. Proceedings of the 17th International Congress of Phonetic Sciences, 3, 1846--1849. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Shue/Shue.pdf