BILINGUAL ACOUSTIC VOICE VARIATION IS SIMILARLY STRUCTURED ACROSS LANGUAGES

INTERSPEECH 2020, Paper ID: 3095

KHIA A. JOHNSON, MOLLY BABEL, ROBERT A. FUHRMAN
DEPARTMENT OF LINGUISTICS, UNIVERSITY OF BRITISH COLUMBIA, CANADA

VOICES ARE HIGHLY VARIABLE

- Voice variability seems to be largely idiosyncratic[1]
- To know a voice is to know how it varies across environments, physical states, and emotions
- Is this variation influenced by language?

[1] 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.

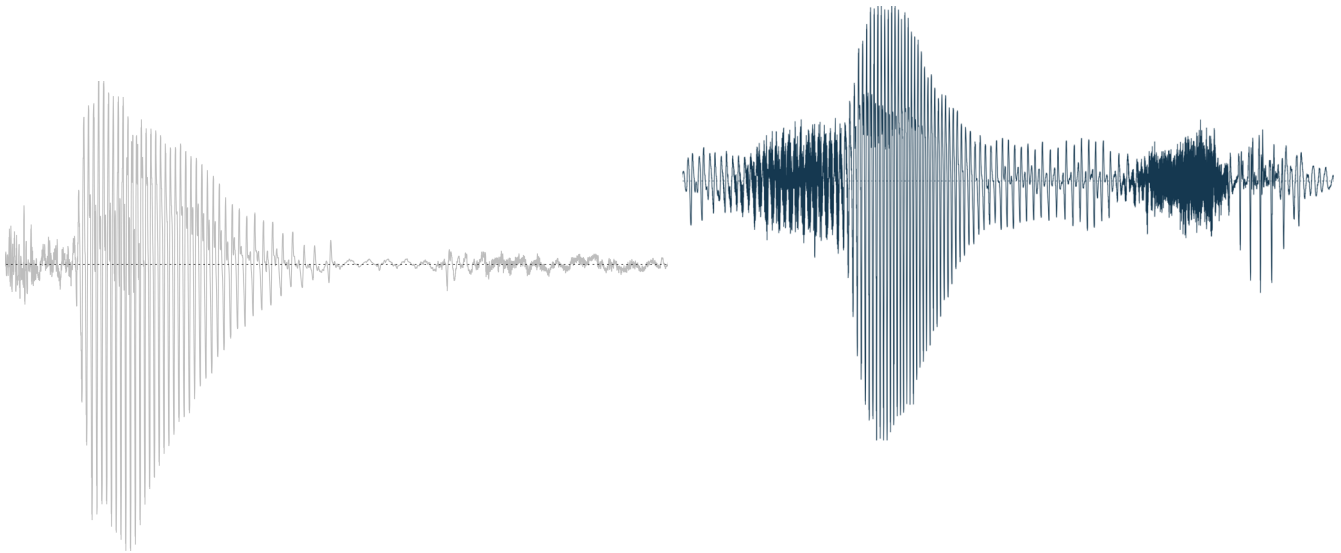
THE ROLE OF LANGUAGE IN VOICE VARIABILITY

- Segmental, suprasegmental, & prosodic aspects of languages vary
- Few Cantonese-English voice quality comparisons[2] :
  - ➞ English tends to be creakier (or less breathy)
  - ➞ Cantonese tends to have lower, more variable pitch
- Perceptual evidence that bilingual talkers can be identified after a language switch, especially by other bilinguals[3]

[2] Ng, M. L., Chen, Y., & Chan, E. Y. K. (2012). 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.

[3] Orena, A. J., Polka, L., & 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.

DO BILINGUAL TALKERS HAVE THE SAME VOICE IN EACH OF THEIR LANGUAGES?



DATA

www.spice-corpus.rtfd.io

- SpiCE Corpus[4]
  - 34 high-proficiency, early Cantonese-English bilinguals
  - 30-minute conversational interviews in Cantonese & English
  - High-quality audio
- Pre-processing:
  - Select all voiced participant speech with Praat[5] algorithm
  - Includes vowels, approximants, & some voiced obstruents

[4] 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.
[5] Praat

ACOUSTIC MEASUREMENTS

- Drawn from psychoacoustic voice quality model,[6] measurements every 5 ms with VoiceSauce[7]

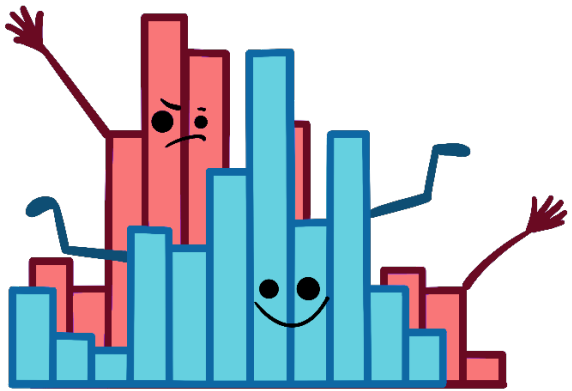
Pitch F0	**Source spectral shape** H1*-H2*, H2*-H4*, H4*-2kHz*, H2kHz*-5kHz*
Formants F1, F2, F3, F4	**Spectral noise** CPP, Energy, SHR

- Post-processing
  - Remove impossible values
  - Calculate moving s.d. for each measure

[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.
[7] 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.

METHODS 1/3

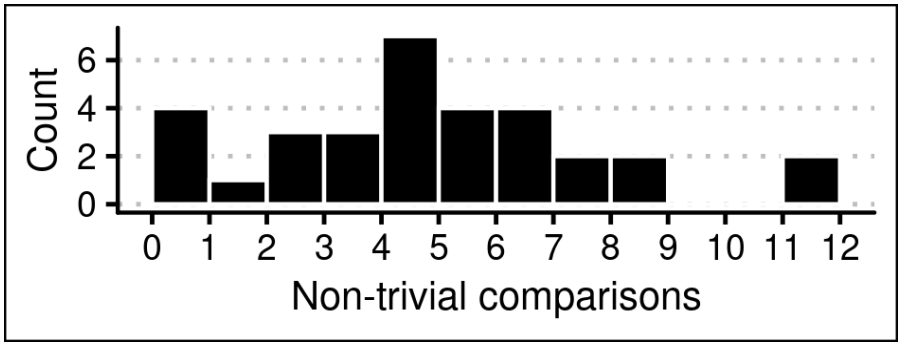
- Crosslinguistic comparison of acoustic measurements
  - ➞ Do bilingual talkers have the same mean values for each measure?
- Within- talker Principal components analyses (PCAs)
  - ➞ How is voice variability structured? How much of it is Idiosyncratic?
- Canonical redundancy analysis
  - ➞ How similar are talkers across languages?



artwork by @allison_horst

COMPARISON OF ACOUSTIC MEASUREMENTS

- Cohen's *d* for t-tests within-talker, across language
- Most talkers have relatively few non-trivial comparisons



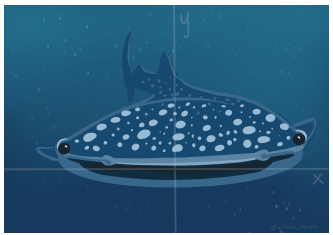
COMPARISON OF ACOUSTIC MEASUREMENTS

- Non-trivial differences tend to…
  - ➞ be small
  - ➞ lack a consistent direction
- When there is a consistent direction, it mirrors prior work
  - ➞ F0 tends to be lower in Cantonese
  - ➞ H1*-H2* consistently puts English on creakier end of spectrum

	Cohen's *d*		
Variable	**Trivial** ***0.0–0.2***	**Small** ***0.2–0.5***	**Medium** ***0.5–0.8***
F0	21	10	3
F0 s.d.	34	0	0
F1	24	9	1
F1 s.d.	29	5	0
F2	26	8	0
F2 s.d.	32	2	0
F3	24	9	1
F3 s.d.	29	5	0
F4	30	3	1
F4 s.d.	28	6	0
H1*–H2*	18	15	1
H1*–H2* s.d.	32	2	0
H2*–H4*	25	9	0
H2*–H4* s.d.	31	3	0
H4*–2kHz*	25	8	1
H4*–2kHz* s.d.	34	0	0
H2kHz*–5kHz*	23	10	1
H2kHz*–5kHz* s.d.	31	3	0
CPP	21	10	3
CPP s.d.	32	2	0
Energy	17	14	3
Energy s.d.	18	16	0
SHR	31	3	0
SHR s.d.	29	5	0

METHODS 2/3

- Crosslinguistic comparison of acoustic measurements
  - → Do bilingual talkers have the same mean values for each measure?
- Principal components analyses (PCAs)
  - → How is voice variability structured? How much of it is idiosyncratic?
- Canonical redundancy analysis
  - → How similar are talkers across languages?



artwork by @allison_horst

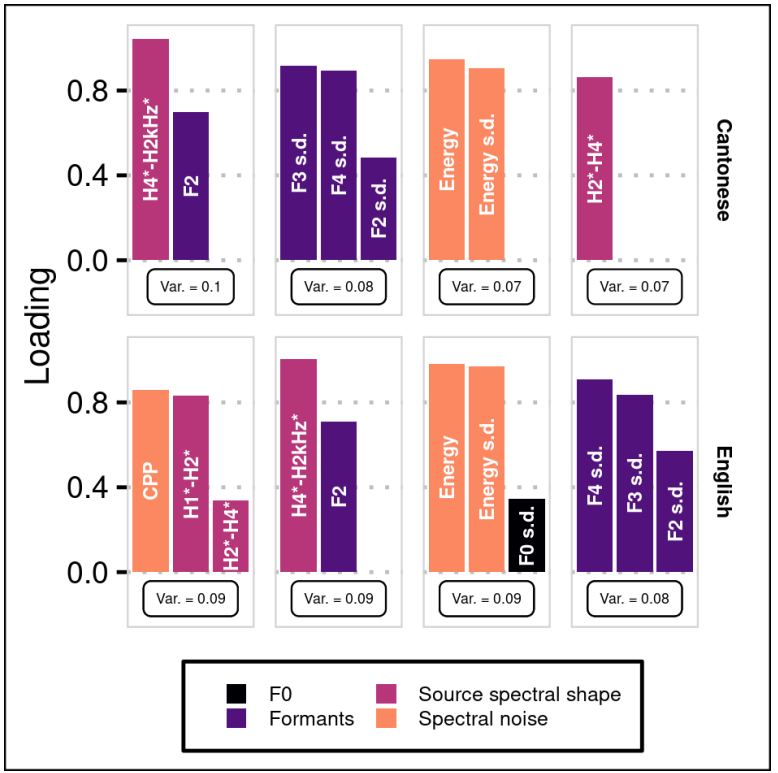
PCA DETAILS

- PCAs by talker and language
- All 24 measures (standardized)
- Oblique promax rotation
- Components retained if eigenvalues were > 0.7x the mean eigenvalue[8]
- Only |loadings| > 0.32 were interpreted

[8] Jolliffe, I. T. Principal Component Analysis, 2nd ed. New York: Springer-Verlag, 2002.

COMPONENT VARIABILITY

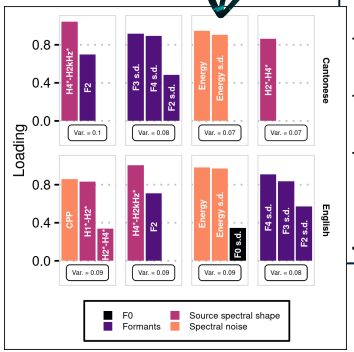
- 10–15 components accounted for 74.6–85.8% of the total variation
- Similar component structure across languages, but variable order



COMPONENT STRUCTURE

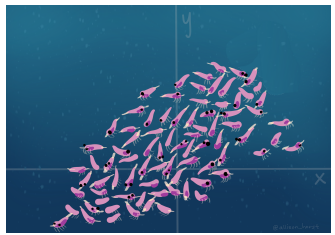
- Similar component composition across talkers and languages
- F0 is a less consistent variable
- Plenty of idiosyncratic variation

Variables	Cantonese N	Cantonese Var. %	English N	English Var. %
H4*–H2kHz*, H2kHz*–H5kHz*, F2, *F3*, *F4*	34	9.3–15.5	32	9.2–16.7
H4*–H2kHz* s.d., H2kHz*–H5kHz* s.d.	32	6.3–8.3	34	4.1–5.0
Energy, Energy s.d, *F0*	31	5.8–9.4	33	6.3–9.1
CPP s.d.	29	4.1–5.0	31	4.1–4.9
SHR, SHR s.d.	30	3.8–7.5	29	5.4–7.3
F3, F4, *F2*	26	6.0–8.5	29	5.8–8.5
F3 s.d., F4 s.d., *F2 s.d.*	26	5.3–8.6	29	4.7–8.6
H2*–H4* s.d., H1*–H2* s.d.	26	4.2–6.5	28	4.2–6.8



METHODS 3/3

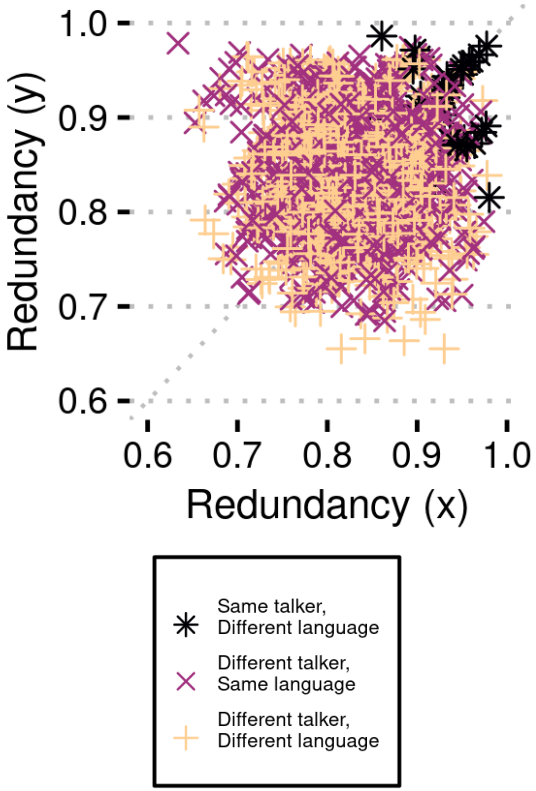
- Crosslinguistic comparison of acoustic measurements
  - ➞ Do bilingual talkers have the same mean values for each measure?
- Within- talker Principal components analyses (PCAs)
  - ➞ How is voice variability structured? How much of it is Idiosyncratic?
- Canonical redundancy analysis
  - ➞ How similar are talkers across languages?



artwork by @allison_horst

CANONICAL REDUNDANCY ANALYSIS

- Allows for comparison of two PCAs, and accounts for different component orders
- Asymmetrical ➞ variation in A accounted for by B and vice versa
- All loadings retained
- Within-talker comparisons are significantly more redundant: *Welch's t*(71.36) = –17.83, $p < 0.001$, d = 1.76



DISCUSSION

- Methodological differences from prior work[9] led us to find *more similarity across talkers* ➞ longer passages & decision to ignore component order
- Despite substantial segmental & suprasegmental differences across English & Cantonese, bilinguals exhibit similar spectral properties and structure in voice variability ➞ voices are like "auditory faces"
- Generates predictions related to bilingualism and cognitive organization of voices in speech perception

[9] 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.

THANK YOU!

SpiCE was developed with support from Nancy Yiu, Ivan Fong, Ariana Hernandez, Christina Sen, Kristy Chan, Katherine Lee, Rachel Wong, Rachel Soo, and members of the Speech-in-Context Lab: **www.speechincontext.arts.ubc.ca**

The SpiCE corpus will be available soon! Follow for updates: **@khia_johnson**

www.spice-corpus.rtfd.io

Natural Sciences and Engineering Research Council of Canada
Conseil de recherches en sciences naturelles et en génie du Canada
Canada

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Graduate and Postdoctoral Studies
WWW.GRAD.UBC.CA