Crosslinguistic Similarity and Structured Variation in Cantonese-English Bilingual Speech Production

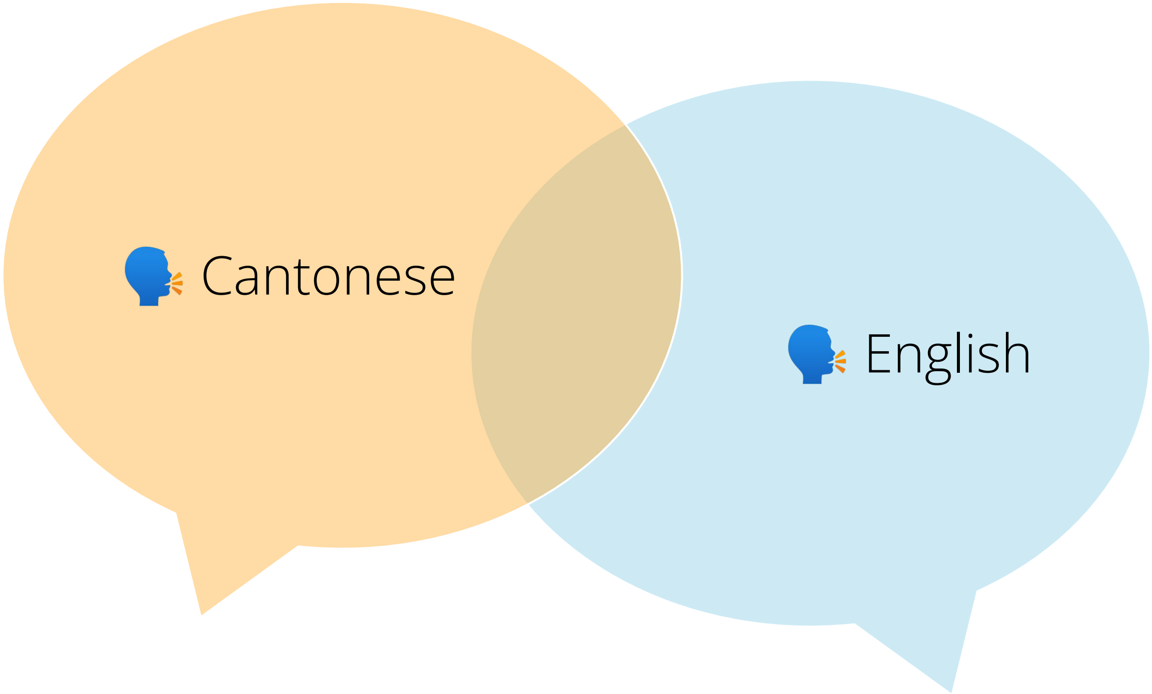
Khia A. Johnson

University of British Columbia, Linguistics
Final Oral Defense | December 10, 2021

A "bilingual is NOT the sum of two complete or incomplete monolinguals; rather, [they have] a unique and specific linguistic configuration…a different but complete linguistic entity"

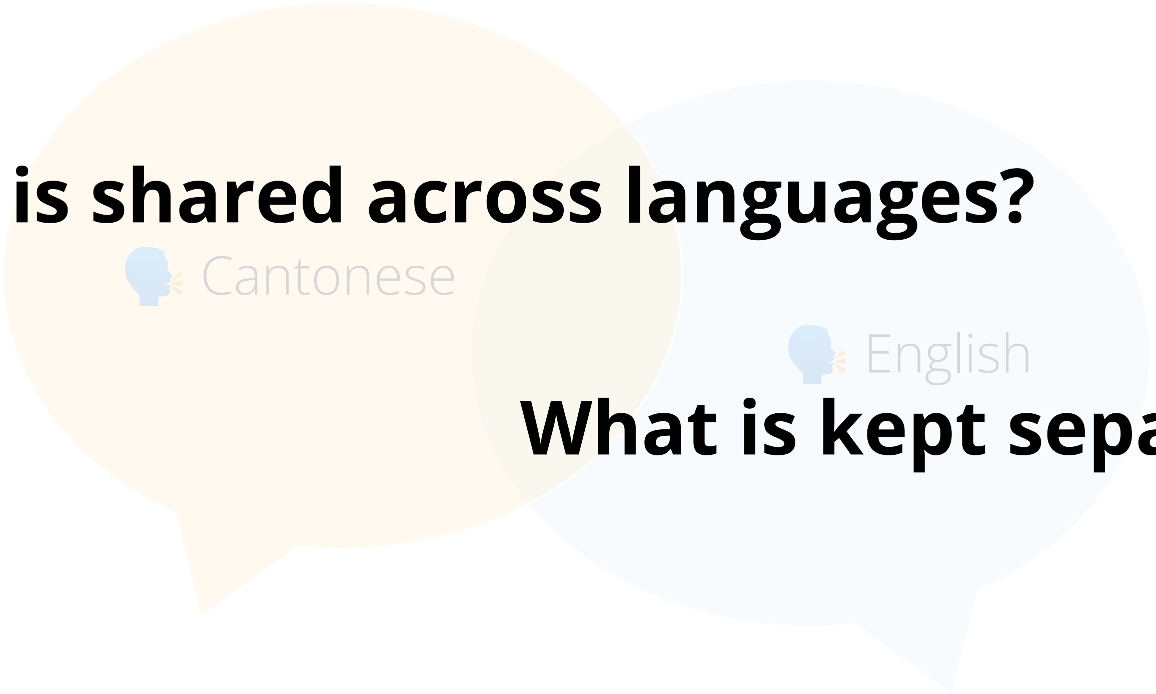
Grosjean, 1989: p. 6

Bilingual sound systems overlap



Bilingual sound systems overlap

What is shared across languages?



What is kept separate?

The SpiCE Corpus

(**Sp**eech **i**n **C**antonese and **E**nglish)

What is SpiCE?

UBC Research Data Collection (University of British Columbia)

Scholars Portal Dataverse > University of British Columbia >
UBC Research Data Collection >

SpiCE: Speech in Cantonese and English

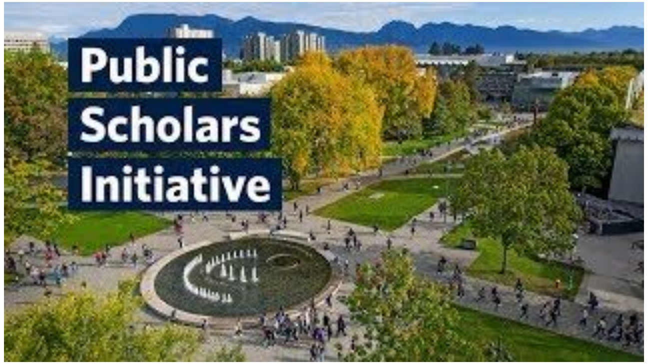
Version 1.0

Johnson, Khia A., 2021, "SpiCE: Speech in Cantonese and English", https://doi.org/10.5683/SP2/MJOXP3, Scholars Portal Dataverse, V1, UNF:6:c6HNIwwpBuQOA349cyCu7w== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

OPEN ACCESS



SpiCE by the numbers

🗣 **Cantonese**
🗣 **English**

34 early bilinguals × **2** languages × **3** tasks

18-34 years old
50% female
50% male

📕 sentence reading
🖼 storyboard narration
👥 conversational interview

SpiCE by the numbers

32.8 total hours

219,000 words

Cantonese
English

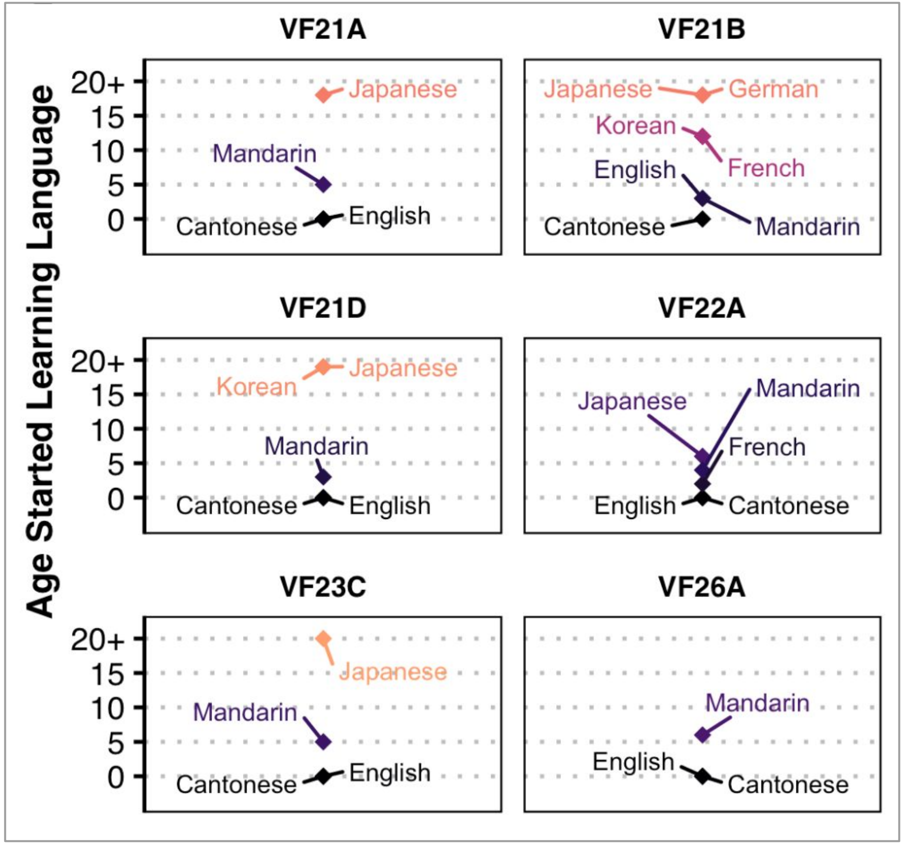
34 early bilinguals × **2** languages × **3** tasks

18-34 years old
50% female
50% male

sentence reading
storyboard narration
conversational interview

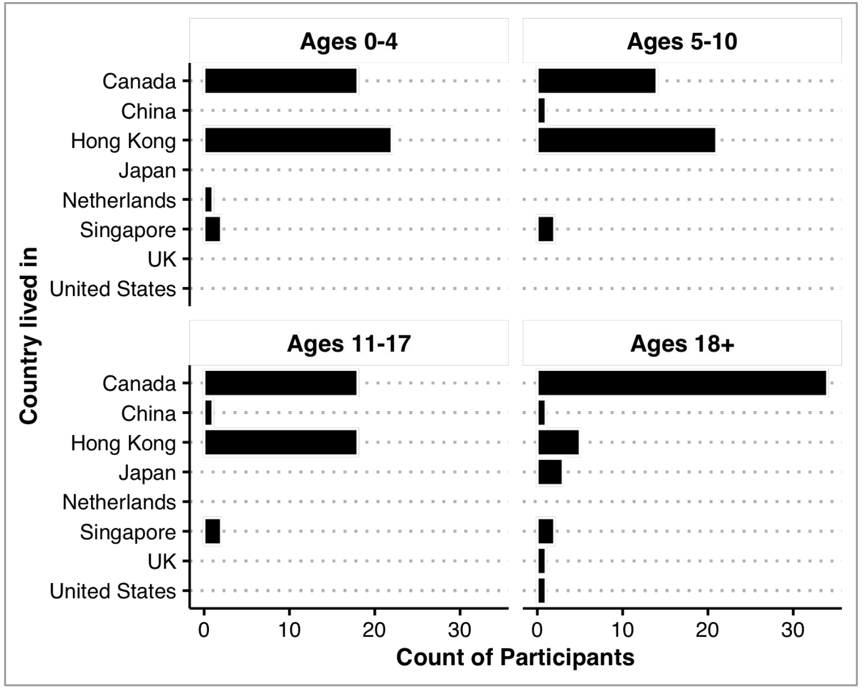
The SpiCE bilinguals

- English and Cantonese learned very early
- **Multi**lingualism is widespread
- For example, 6 talkers →



The SpiCE bilinguals

- Have roots from all over the diaspora
- Don't necessarily speak the same varieties



SpiCE is well-suited for...

- Corpus phonetics
- Within-talker designs
- Studying variation

Study 1
Shared structure in voices

Voices are highly variable

- Voice variability is largely idiosyncratic (Lee, Keating, & Kreiman, 2019)
- To know a voice is to know how it varies across environments, physical states, and emotions
- Is this variation influenced by **language**?

The role of language in voice variability

- Segmental, suprasegmental, & prosodic aspects of languages vary
- Few Cantonese-English voice quality comparisons (Ng et al., 2012)
    - 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 (Orena, Polka, & Theodore, 2019)

Methods overview

- Identify all voiced speech with Praat algorithm (Boersma & Weenink, 2020)
  - Vowels
  - Voiced consonants
- Make acoustic measurements every 5 ms in VoiceSauce (Shue et al., 2011)

Acoustic measures

Pitch F0	**Source spectral shape** H1*-H2* H2*-H4* H4*-H2kHz* H2kHz*-H5kHz
Formants F1 F2 F3 F4	**Source spectral noise** CPP Energy SHR

(Kreiman et al. 2014)

Acoustic measures

Pitch F0	**Source spectral shape** H1*-H2* H2*-H4* H4*-H2kHz* H2kHz*-H5kHz
Formants F1 F2 F3 F4	**Source spectral noise** CPP Energy SHR

(Kreiman et al. 2014)

Acoustic measures

Pitch F0	**Source spectral shape** H1*-H2* H2*-H4* H4*-H2kHz* H2kHz*-H5kHz
Formants F1 F2 F3 F4	**Source spectral noise** CPP Energy SHR

(Kreiman et al. 2014)

Acoustic measures

Pitch F0	**Source spectral shape** H1*-H2* H2*-H4* H4*-H2kHz* H2kHz*-H5kHz
Formants F1 F2 F3 F4	**Source spectral noise** CPP Energy SHR

(Kreiman et al. 2014)

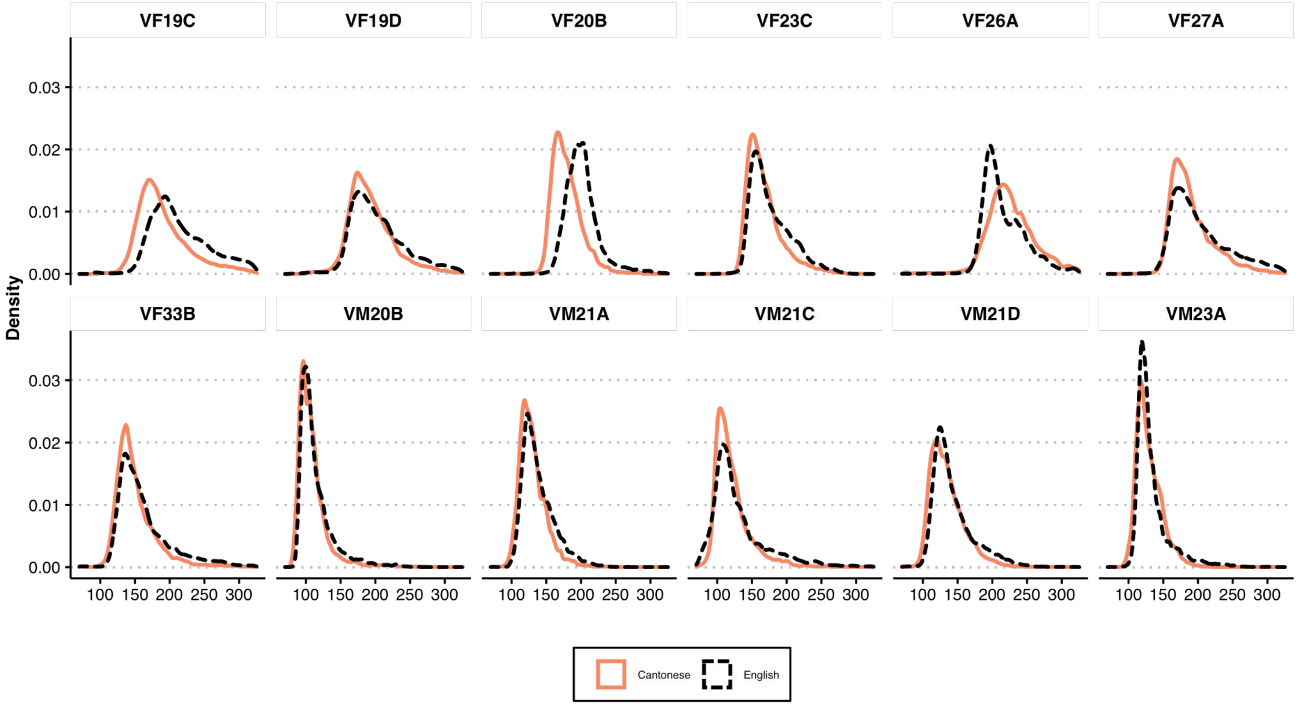
Methods overview, continued

- Remove impossible values
- Calculate rolling standard deviations
- 3-part analysis

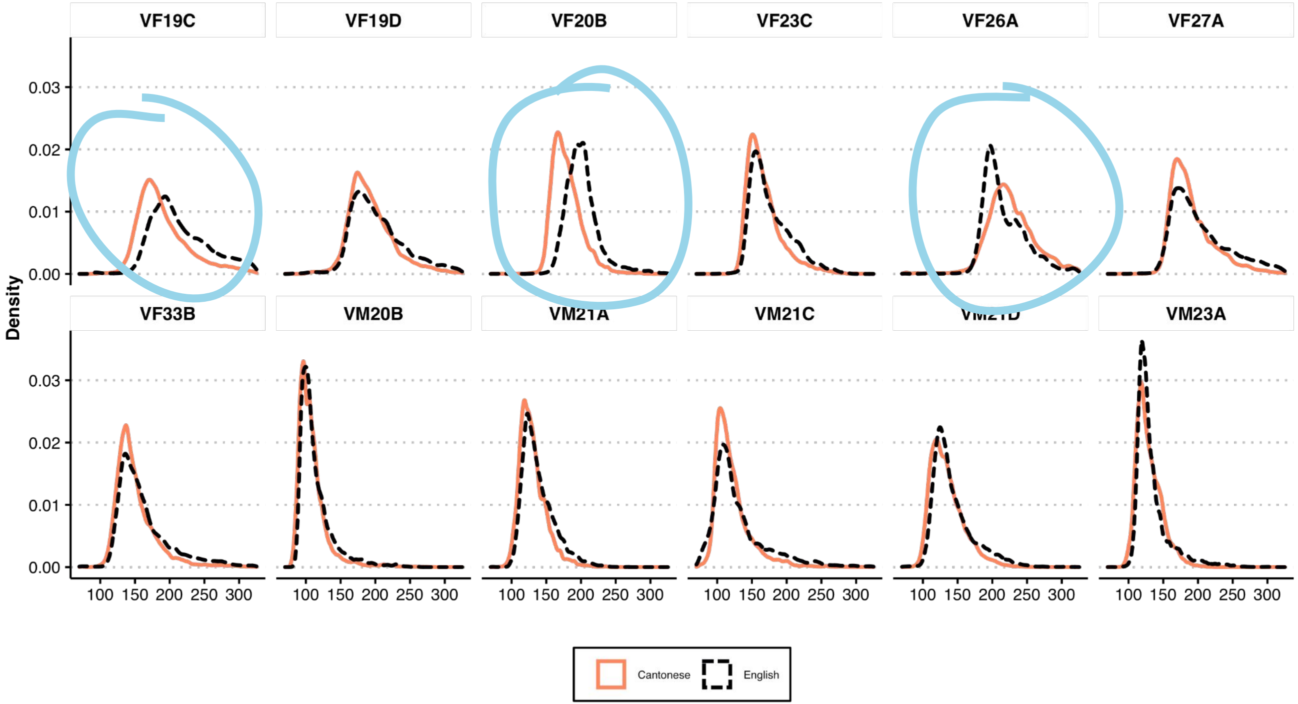
Comparing acoustic measure means

- 24 across-language comparisons per talker
- For example, do talkers differ F0 across languages?
- Cohen's *d* → is the difference...
  - Trivial
  - Small
  - **Medium**
  - **Large**

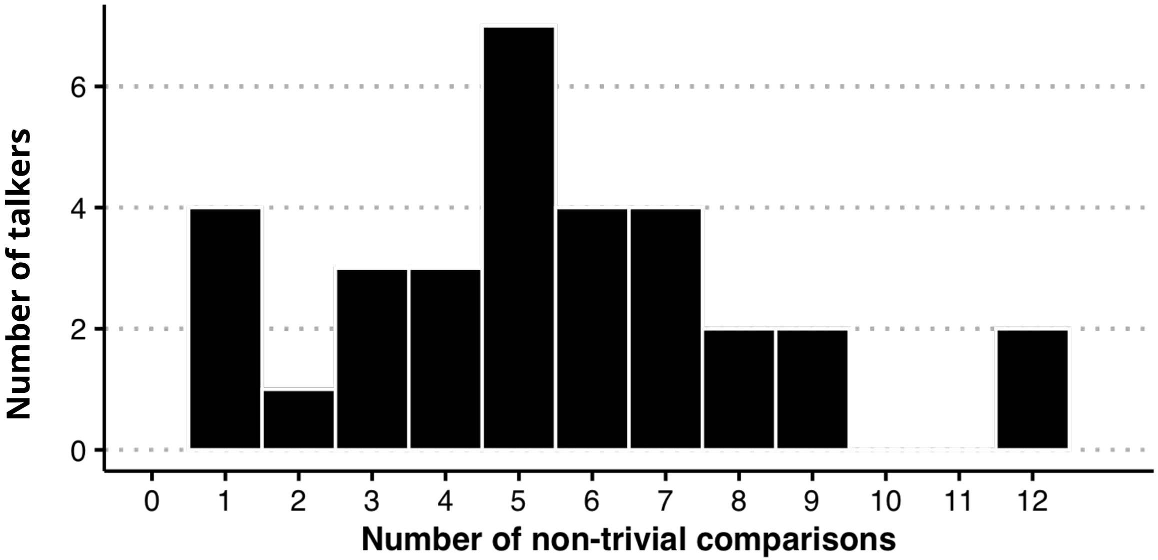
Non-trivial F0 differences



Non-trivial F0 differences



Counting non-trivial comparisons

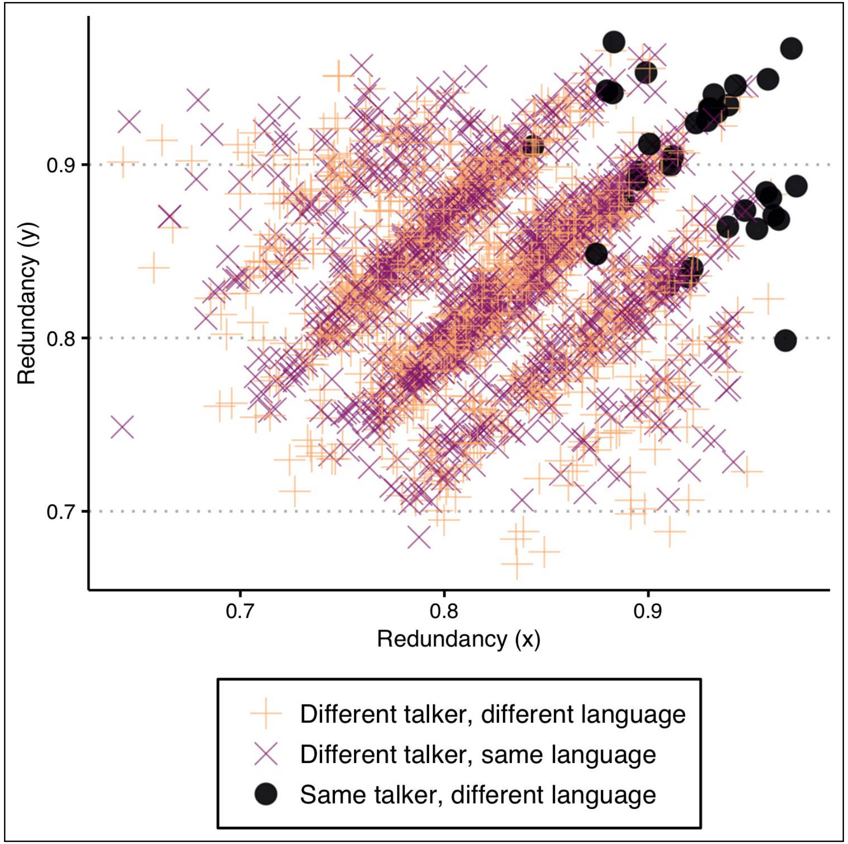


Principal Components Analyses

- Dimensionality reduction
  - Many measured dimensions → Fewer important components
  - Example: number of layers + daylight hours + temperature → “coldness” component
- 1 PCA each talker+language
- Main takeaways?
  - Lots of shared components across PCAs
  - Examples: Brightness/creakiness component, Energy component

Canonical redundancy analysis

- Or, ignoring component order, **how similar are two PCAs?**
- Asymmetrical metric → variation in PCA #1 accounted for by #2 and vice versa
- Redundancy is overall very high
- Redundancy is especially high within talker across languages



Takeaways

- Voices, broadly speaking, share lower dimensional structure (Lee et al., 2019)
- Much of idiosyncratic voice structure is retained across languages
- Voices *are* like auditory faces (Belin et al., 2004)

Study 2
Shared structure in consonants

Long-lag stops in Cantonese and English

- Both languages have long-lag /p/, /t/, and /k/ word-initially
- Typical long-lag voice-onset time in isolated speech
  - Cantonese: ~91 ms (Clumeck et al., 1981)
  - English: ~80 ms (Lisker & Abhramson, 1964)
- Often analyzed as “linked” sounds
- Classic methods in crosslinguistic depend on detecting/modulating *differences*

Crosslinguistic *uniformity*

- **Uniformity**: a constraint on within-talker phonetic variation, in which articulatory gestures or phonological primitives are implemented systematically in speech production (Chodroff & Wilson, 2017; Faytak, 2018; Ménard et al., 2008)
- A framework for identifying overlap among segments through the structure of variation
- *Study 2 extends this set of methods to crosslinguistic comparisons*

Methods overview

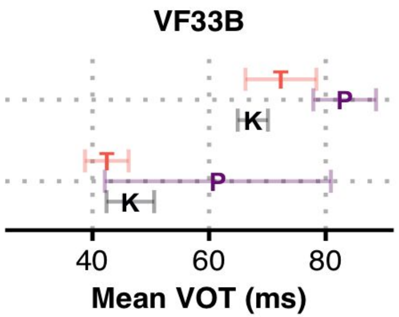
- Prevocalic word initial / p t k / with following stress
- VOT identified, refined, and measured

Table 4.1: The number of stop tokens (overall and range across talkers) and word types for each language and sound category.

Language	**Frequency**	**/p/**	**/t/**	**/k/**
Cantonese	Token (overall)	374	1373	1688
	Token (range)	0–32	17-79	19–116
	Type (overall)	60	157	68
English	Token (overall)	1035	1336	3155
	Range (tokens)	4–96	15–150	52–294
	Type (overall)	158	143	208

Mean VOT analysis

- Calculate mean VOT for each **talker × language × consonant**
- Ordinal relationships
  - Expected: /p/ < /t/ < /k/
  - Actual: Wildly inconsistent ordering across the board
- Pairwise correlations
  - Expected: Strong clear correlations
  - Actual: Moderate correlations at best, both within and across languages



Ordinal mean VOT relationships

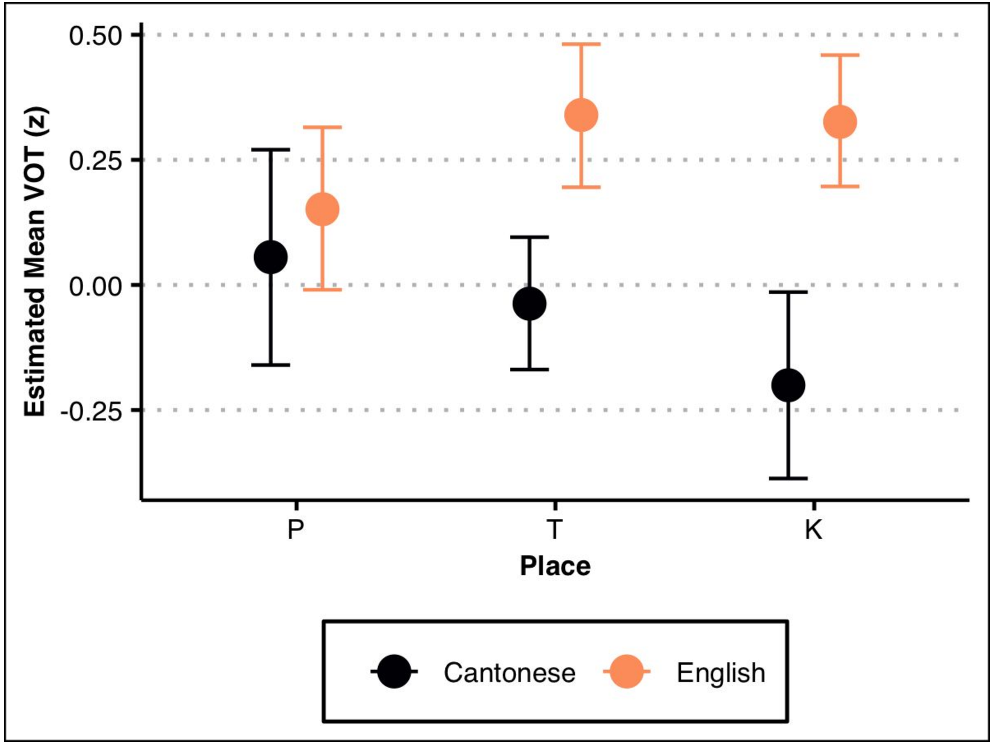
- **Expected**: p < t < k
- **Actual**: Wildly inconsistent ordering across the board
- An illustrative subset

Bayesian multilevel linear model

- Model designed to *figure out what's going on*
- Takes **speech rate**, **pauses**, and **word variability** into account
    - Note: these variables behaved *as expected*
- Code & model details → https://github.com/khiajohnson/dissertation

Model predictions

- Model-predicted means for Place and Language
- English has *slightly* longer VOT for /t/ and /k/



Sources of variation

- Words do the most heavy lifting
  - Example: 2 initial /k/ in →
- Removing Word from the model drastically alters the model's story
- High between-talker variability, too

Takeaways

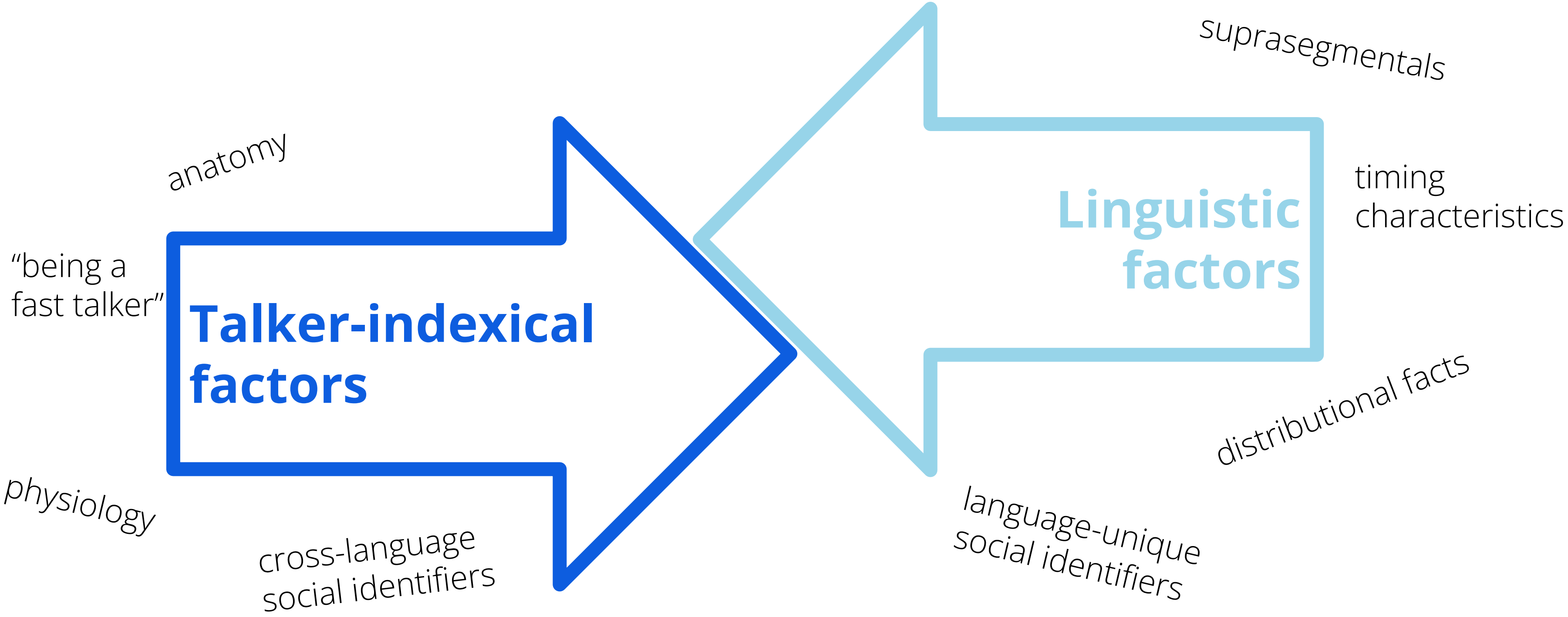
- Some evidence for within- and across-language uniformity, especially if a more “macro” approach taken
- Talkers can maintain *very* fine-grained crosslinguistic differences, though reason is unclear
- Spontaneous speech leads to drastically different results
    - Uncontrolled lexical content
    - Style differences

Discussion & Conclusion

What’s shared?

- Voices share mean values for acoustic dimensions and lower-dimensional structure
- Long-lag stops share a general target for VOT
- A lot, but not everything

Variation influencers



Key contributions

- The SpiCE corpus
- Reproducible analyses
- Improved understanding of *how* languages share sound structure
- Production-based groundwork for perception research
    - Talker identification
    - Processing uniform VOT

thank you!

Molly Babel, Kathleen Currie Hall, Márton Sóskuthy, Ivan Fong, Nancy Yiu, Katherine Lee, Kristy Chan, Natália Oliveira Ferreira, Michell To, Rachel Ching Fung Wong, Christina Sen, Ariana Zattera, Rachel Soo, Nikolai Schwarz, Robert Fuhrman, Carla Hudson Kam, Valter Ciocca, Matt Goldrick, Gloria Mellesmoen, John Alderete, Henny Yeung, everyone in the Speech in Context Lab, members of the linguistics department, my cohort, friends, family, dog!, and so many others I'm probably forgetting… the UBC Public Scholars Initiative, Molly's SSHRC grant, and my UBC fellowships. I couldn't have done this without all of you.

References available at:
https://github.com/khiajohnson/dissertation/blob/main/text/references.bib