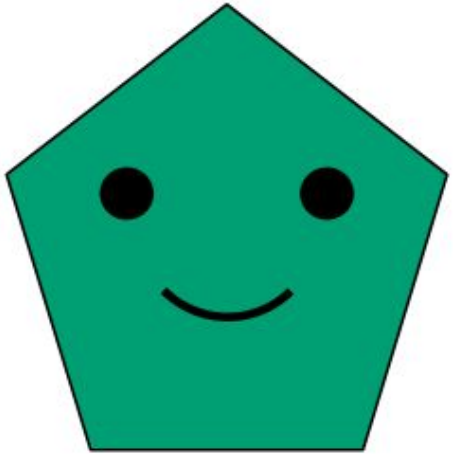
SpiCE: A new open-access corpus of conversational bilingual **Sp**eech **i**n **C**antonese & **E**nglish

NEW EMPIRICAL CONTRIBUTIONS TO CANTONESE LINGUISTICS & LANGUAGE PROCESSING

FEBRUARY 28, 2020

Khia A. Johnson • University of British Columbia • Linguistics

A pair of very important acknowledgements



Nancy Yiu

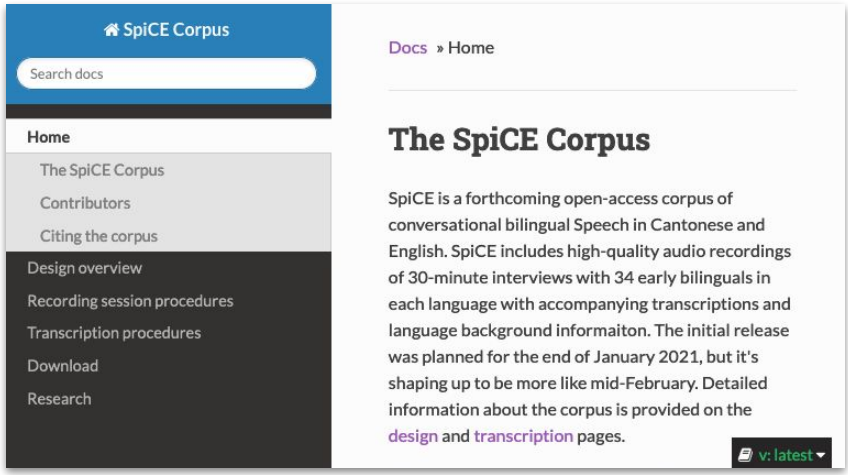
Undergraduate Research Assistant

Ivan Fong

Research Assistant

What is SpiCE?

- Open-access speech corpus
- 34 early Cantonese-English bilingual participants
- Transcribed, high-quality recordings
- Speech samples in two languages
- Language background summary

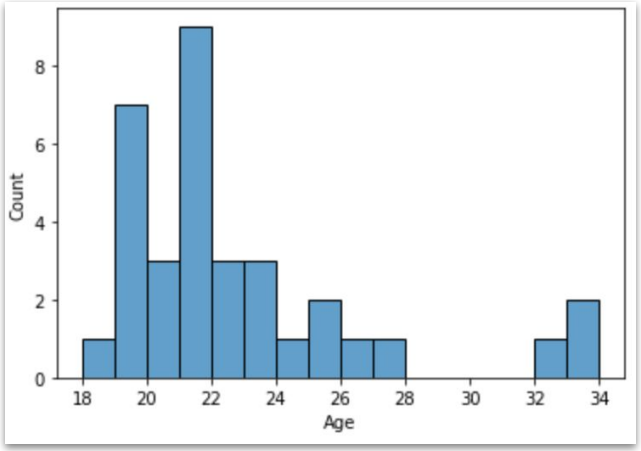


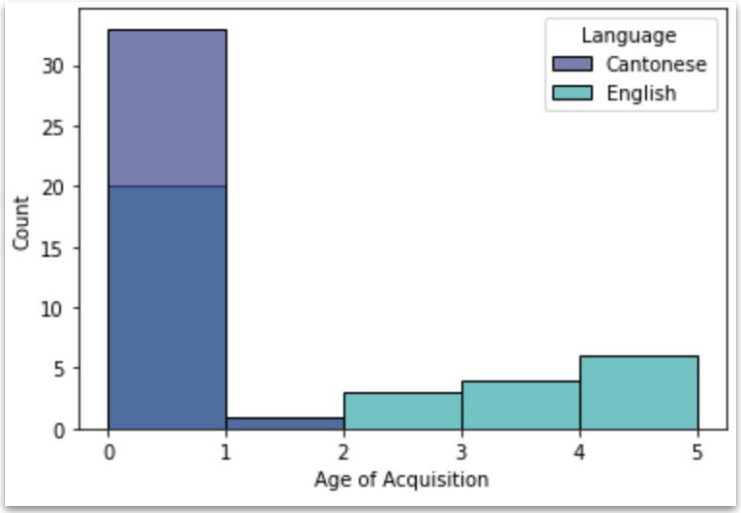
Why create this corpus?

- Corpus phonetics for crosslinguistic influence
- Pre-existing bilingual corpora
  - Not spontaneous
  - Too small
  - Poor(er) recording quality
  - Focused on well-studied European language pairs
  - Recorded for other purposes
- Buckeye-like bilingual corpus (Pitt et al., 2007)
- Support from UBC Public Scholars Initiative

Participants

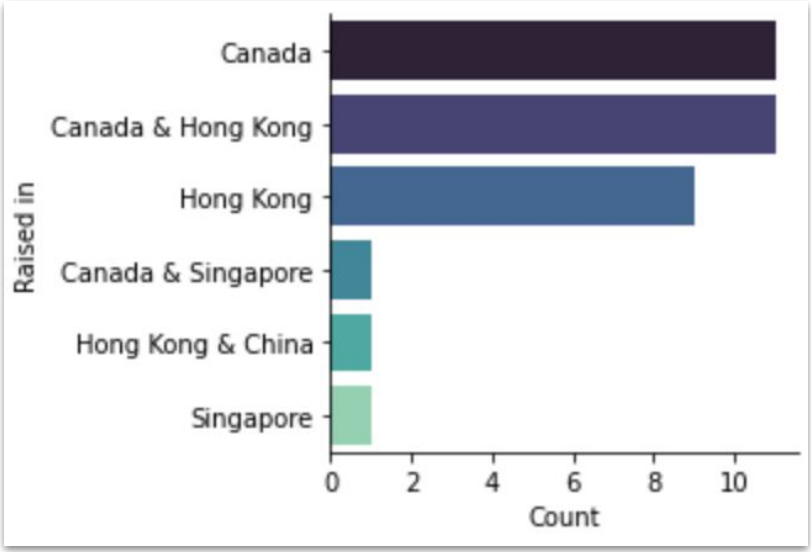
- 17 female, 17 male
- Age 19-34 at recording
- Early bilinguals

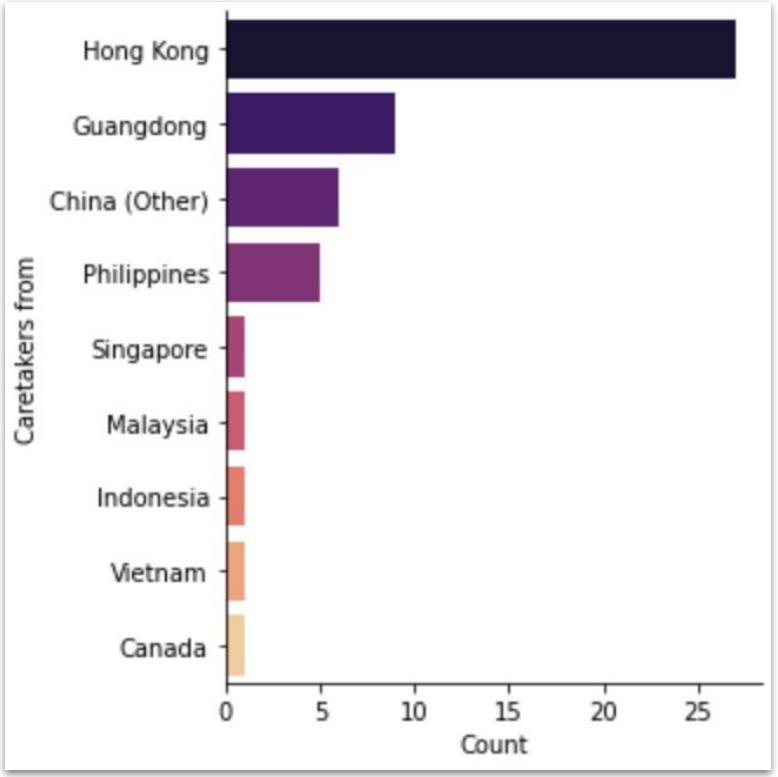




Participants

- Heterogeneous backgrounds

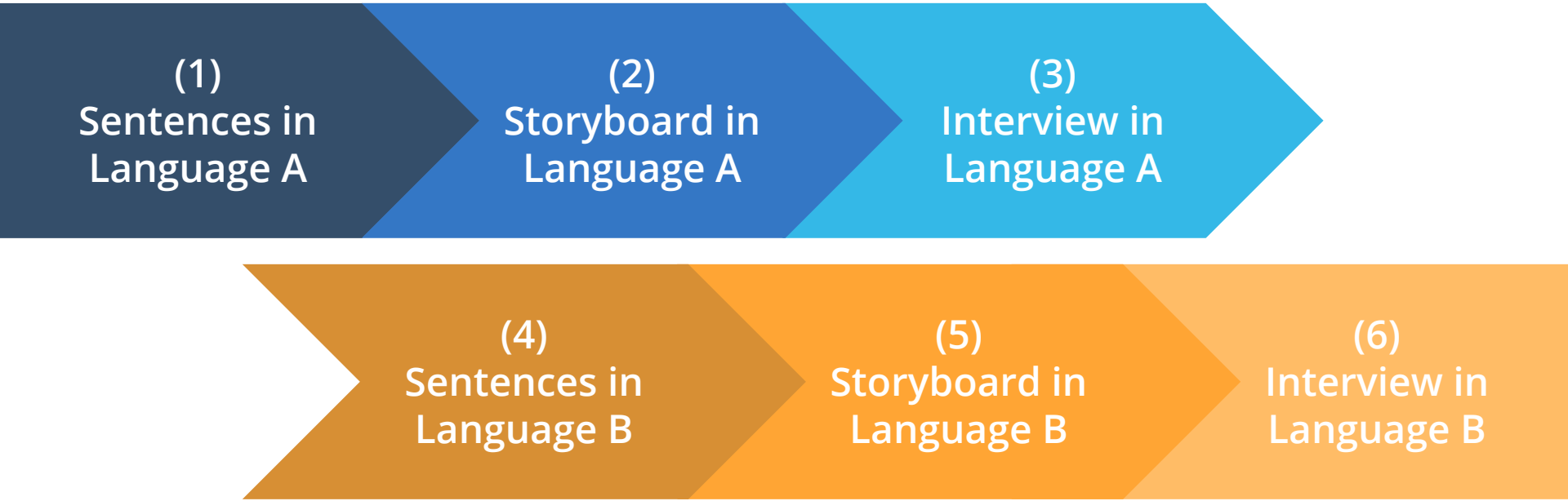




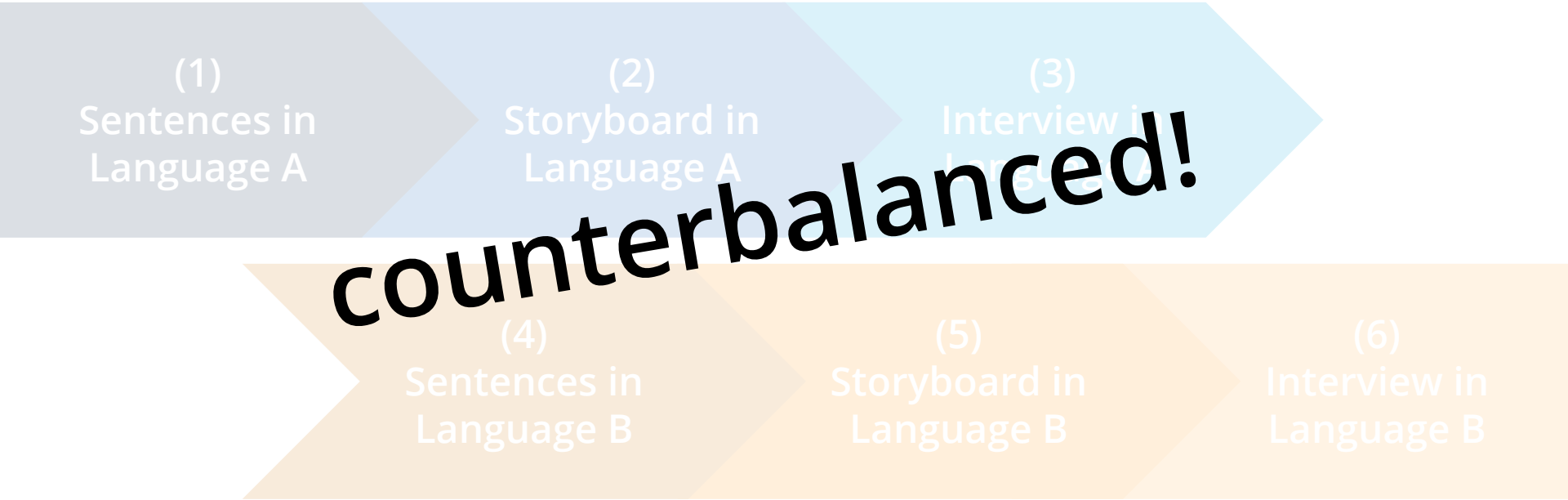
Recording setup

- Quiet room, seated around table
- IF managed recording, NY conducted interviews
- Equipment:
  - Head-mounted mics → isolates participant speech when overlap
  - Sound Devices USBPre 2 audio interface
  - PC laptop running Audacity
- Settings:
  - Stereo
  - 44.1 kHz sampling rate
  - 24-bit depth

Recording session



Recording session



1. Sentence reading

	Characters	Jyutping	English translations
1.	新年快樂	san1 lin4 faai3 lok6	Happy New Year
2.	恭喜發財	gung1 hei2 faat3 choi4	Congratulations on happiness and prosperity
3.	身體健康	san1 tai2 gin6 hong1	May your health be well
4.	快高長大	faai3 gou1 zoeng2 dai6	Grow quickly
5.	龍馬精神	lung4 ma5 zing1 san4	Have the spirit of the horse and dragon
6.	學業進步	hok6 yip6 zeon3 bou6	Progress in your education
7.	年年有餘	lin4 lin4 yau5 yue4	Excess in each year
8.	出入平安	cut1 yap6 ping4 on1	Leave and enter in safety
9.	心想事成	sam1 soeng2 si6 sing4	Accomplish that which is in your heart
10.	生意興隆	saang1 yi3 hing1 lung4	Have a prosperous business
11.	萬事如意	maan6 si6 yu4 yi3	A thousand things according to your will
12.	天天向上	tin1 tin1 hoeng3 soeng6	Upwards and onwards every day
13.	笑口常開	siu3 hau2 soeng4 hoi1	Laugh with an open mouth frequently
14.	大吉大利	daai6 gat1 daai6 lei6	Much luck and much prosperity
15.	五福臨門	mm5 fuk1 lam4 mun4	Five blessings for your household
16.	招財進寶	ziu1 coi4 zeon3 bou2	Seek wealth welcome in the precious
17.	盤滿缽滿	pun4 mun5 but3 mun5	Basins full of wealth

	English sentence
1.	Stop whistling and watch the boys march.
2.	Jerk the cord, and out tumbles the gold.
3.	Slide the tray across the glass top.
4.	The cloud moved in a stately way and was gone.
5.	Light maple makes for a swell room.
6.	Set the piece here and say nothing.
7.	Dull stories make her laugh.
8.	A stiff cord will do to fasten your shoe.
9.	Get the trust fund to the bank early.
10.	Choose between the high road and the low.
11.	Wish on every candle for your birthday.
12.	Deck the halls with boughs of holly.
13.	Ring in the new year with a kiss.
14.	Have a spooky Halloween.
15.	Enjoy the vacation with your loved ones.
16.	Be filled with joy and peace during this time.
17.	Relax on your holiday break.

2. Storyboard narration

3. Conversational interview

SpiCE Corpus Project: Cantonese Interview

Interview designed by Nancy Yiu, Ivan Fong, and Khia A. Johnson

These interview questions are a rough guide to what is covered in each of the SpiCE Cantonese interviews. Content was driven by what each participant was interested in talking about. Note that while many of these questions are designed to get participants talking about themselves, participants were instructed to refer to people and places in general terms (i.e., just say " my Aunt" rather than a name).

1. "Hometown" - you / parents / family
    a. Dialect?
    b. Identity? (Hong Kongese, Chinese, Chau Zhounese, Canadian, etc.)
    c. How is your parents' English?
        i. Weird quirks? Accent?
        ii. Do you correct them?
    d. Do they ever comment on your physicality? (you should eat more / less)
    e. Form of communication with parents / family
        i. Do you keep in contact?
        ii. Calling / whatsapp / wechat / skype?
    f. Hometown specialties - food, products, souvenirs?
        i. Can you cook them? does your family cook them?
        ii. Do you even like eating them? Were you ever forced to eat them due to good luck/superstition
        iii. Have you ever had a bad experience with eating traditional/hometown food and someone made a negative comment?
            1. Have you ever felt discrimination - called a 'banana' or had people say things like 'go back to your own country'?

2. Love life
    a. Are you single / dating?
    b. Dream partner

SpiCE Corpus Project: English Interview

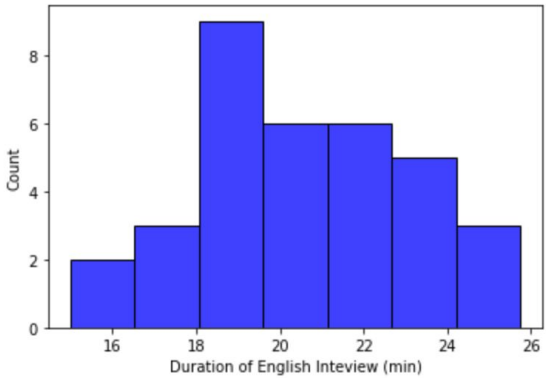
Interview designed by Nancy Yiu, Ivan Fong, and Khia A. Johnson

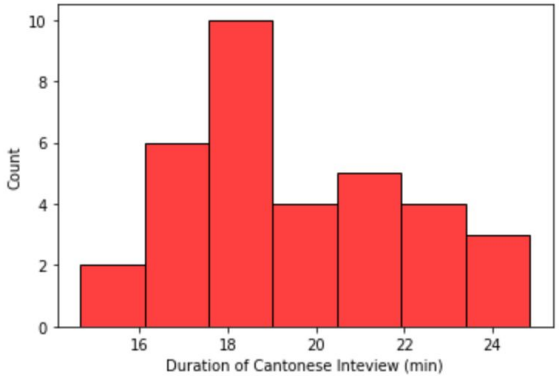
These interview questions are a rough guide to what is covered in each of the SpiCE English interviews. Content was driven by what each participant was interested in talking about. Note that while many of these questions are designed to get participants talking about themselves, participants were instructed to refer to people and places in general terms (i.e., just say " my Aunt" rather than a name).

1. Where & when were you born? where you grew up?
    a. Where
        i. From Vancouver – friends attend UBC?
        ii. From elsewhere – missing home/friends/what prompted to study here?
            1. Differences from home and Vancouver
                a. Favorite things about Vancouver or home
                b. Transit/Driving - potential differences in wheel/passenger side
            2. Culture shock in Vancouver
                a. Things you miss from your hometown
    b. When
        i. Lunar (Chinese) Calendar Birthday?
        ii. 90s or 00s kid? Differences you've noticed/been told?
    c. Childhood?
        i. Did you have a strict upbringing?
            1. Examples?
            2. What would change about your upbringing style
        ii. How does it compare to the upbringing of your peers/friends
            1. Do you like yours or theirs
        iii. Expectation of parents
            1. Finish high school/Uni
            2. Get a stable job? Become a lawyer/doctor/engineer?

How much speech?

- 11.6 hours English *participant* speech
  - Shortest: 15 min
  - Median: 20 min
  - Longest: 25 min
- 11.1 hours Cantonese *participant* speech
  - Shortest: 15 min
  - Median: 19 min
  - Longest: 25 min





Speech-to-text

Use Google Cloud Speech-to-text software to expedite the transcription process

Hand-correct

Hand-correct the orthographic transcripts for accuracy and consistency

Force-align

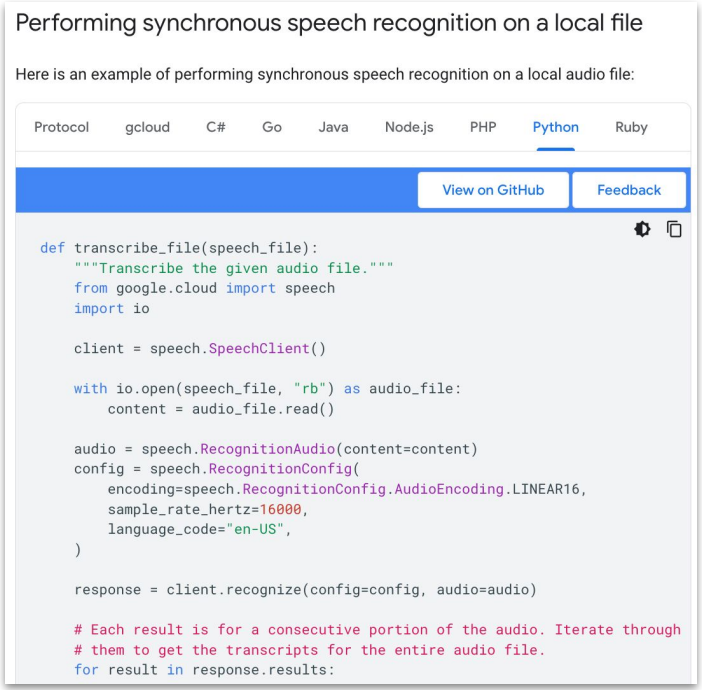
Use the Montreal Forced Aligner 2.0 to generate phone level annotations

The SpiCE transcription pipeline

Automating the first step

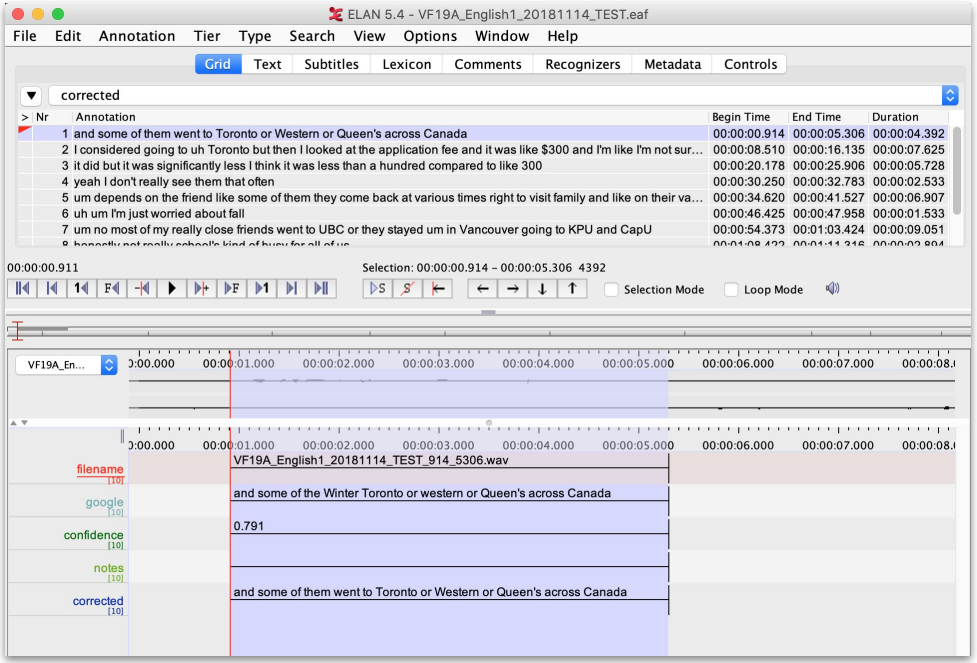
- Participant channel only
- Segment files by pauses
- Short synchronous speech recognition
  - Canadian English: **en-CA**
  - Hong Kong Cantonese: **yue-hant-HK**

https://cloud.google.com/speech-to-text/docs/sync-recognize#speech-sync-recognize-python →



Correcting the transcripts

- Bilingual RAs
- ELAN (Sloetjes & Wittenberg, 2008)
- Cantonese
  - Characters > Jyutping
  - [fused syllables]
- Conventions detailed in documentation & paper

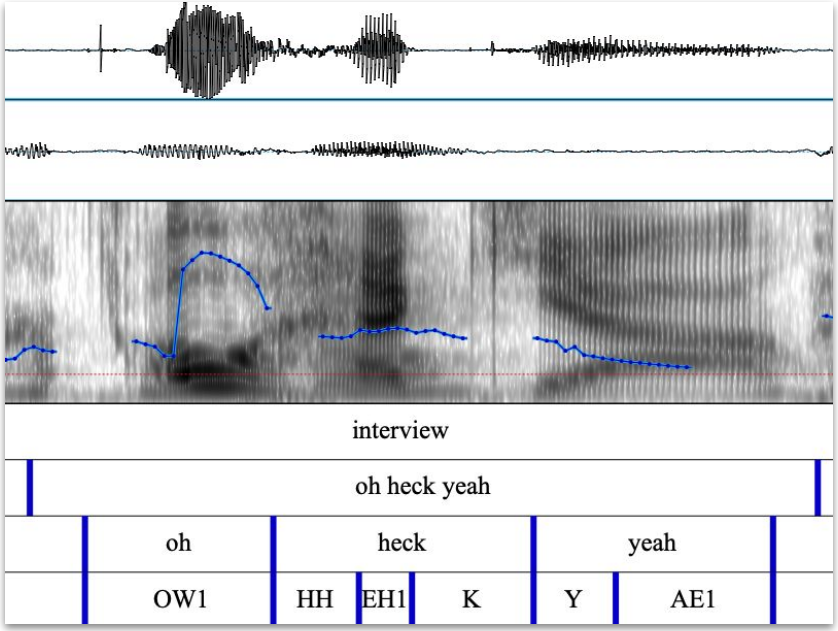


Annotating at the phone level

- Montreal Forced Aligner 2.0 (McAuliffe et al., 2021)
- Cantonese
  - Pronunciation dictionary built with `pycantonese` (Lee, 2021)
    - \+ manual additions
  - Align with `train`
- English
  - pronunciation dictionary provided with MFA
    - \+ manual additions
  - Align with provided acoustic model

...and the output!

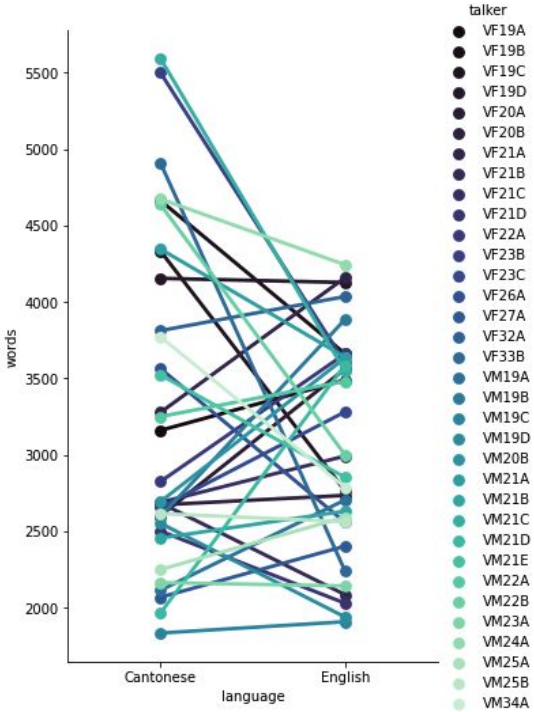
- 68 high quality, stereo WAV files
- 68 TextGrids with tiers for:
  - Task
  - Utterance (used in alignment)
  - Word
  - Phone
- Language background summary

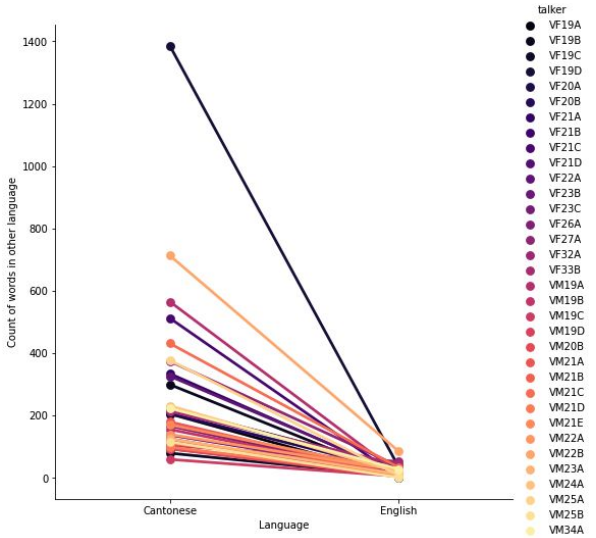


***not real output (:*

Counting words in SpiCE

- English: full word form
- Cantonese: longest word matching with **`pycantonese`** (Lee, 2021)
- More Cantonese → English switches



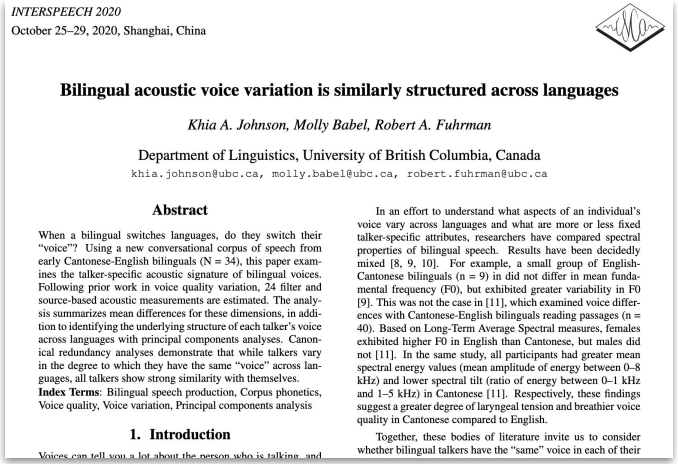


SpiCE is especially well suited for…

- Corpus phonetics
- Within-talker designs
- Cantonese-English bilingualism
- Understanding a speech community that we study *a lot*

Voice variability

- How does the structure voice quality compare across languages? (Johnson, Babel, & Fuhrman, 2020; Johnson & Babel, 2020)
- Exploratory data analysis methods (Lee et al., 2019)
- Within-talker design
- Takeaways:
  - Variation cross-language similarity
  - Identity > Language
- More to do here!



INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Bilingual acoustic voice variation is similarly structured across languages

Khia A. Johnson, Molly Babel, Robert A. Fuhrman

Department of Linguistics, University of British Columbia, Canada
khia.johnson@ubc.ca, molly.babel@ubc.ca, robert.fuhrman@ubc.ca

Abstract

When a bilingual switches languages, do they switch their "voice"? Using a new conversational corpus of speech from early Cantonese-English bilinguals (N = 34), this paper examines the talker-specific acoustic signature of bilingual voices. Following prior work in voice quality variation, 24 filter and source-based acoustic measurements are estimated. The analysis summarizes mean differences for these dimensions, in addition to identifying the underlying structure of each talker's voice across languages with principal components analyses. Canonical redundancy analyses demonstrate that while talkers vary in the degree to which they have the same "voice" across languages, all talkers show strong similarity with themselves.

Index Terms: Bilingual speech production, Corpus phonetics, Voice quality, Voice variation, Principal components analysis

1. Introduction

Voices can tell you a lot about the person who is talking, and

In an effort to understand what aspects of an individual's voice vary across languages and what are more or less fixed talker-specific attributes, researchers have compared spectral properties of bilingual speech. Results have been decidedly mixed [8, 9, 10]. For example, a small group of English-Cantonese bilinguals (n = 9) in did not differ in mean fundamental frequency (F0), but exhibited greater variability in F0 [9]. This was not the case in [11], which examined voice differences with Cantonese-English bilinguals reading passages (n = 40). Based on Long-Term Average Spectral measures, females exhibited higher F0 in English than Cantonese, but males did not [11]. In the same study, all participants had greater mean spectral energy values (mean amplitude of energy between 0–8 kHz) and lower spectral tilt (ratio of energy between 0–1 kHz and 1–5 kHz) in Cantonese [11]. Respectively, these findings suggest a greater degree of laryngeal tension and breathier voice quality in Cantonese compared to English.

Together, these bodies of literature invite us to consider whether bilingual talkers have the "same" voice in each of their

Extending uniformity across languages

- Presenting at ISBPAC 3 in June 2021
- Similarity at the level of speech sounds
- Extend articulatory uniformity framework (e.g., Chodroff & Wilson, 2017)
- Do bilinguals share a laryngeal gesture for long-lag stops?
- Work-in-progress

ISBPAC 2021

3rd International Symposium on Bilingual and L2 Processing in Adults and Children

Contact-induced sound change

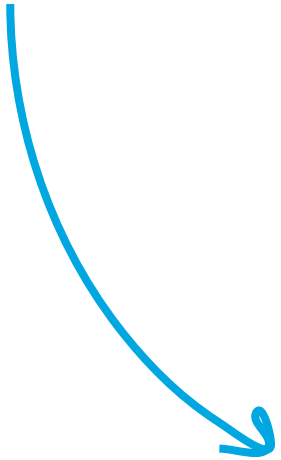
- Manuscript under review (🤞 ➡️ SOON 📸 📄)
- A mechanistic account for asymmetry in crosslinguistic influence, that adopts an approach from sound change (Harrington et al., 2018)
- Languages with categories that accept a wider range of variability are likely to be influenced by matched categories from other language that are less variable
  - English voiced stops ← Cantonese lack of voiced obstruents
  - English final stop releases ← Cantonese final stop categorically unreleased

What's next for us?

- Expanding on current work
- Perception studies in the pipeline
  - Talker identification (Lloy, Johnson, & Babel, 2020)
  - Talker discrimination
- Looking at the /n/-/l/ merger in spontaneous speech
  - **you'll hear more about this topic from Rachel Soo in a little bittion
- Sharing the corpus with you!

Where to get the SpiCE corpus

- Soon → doi.org/10.5683/SP2/MJOXP3
- Documentation → spice-corpus.rtfd.io
- Follow me on Twitter for updates



The Dataverse® Project

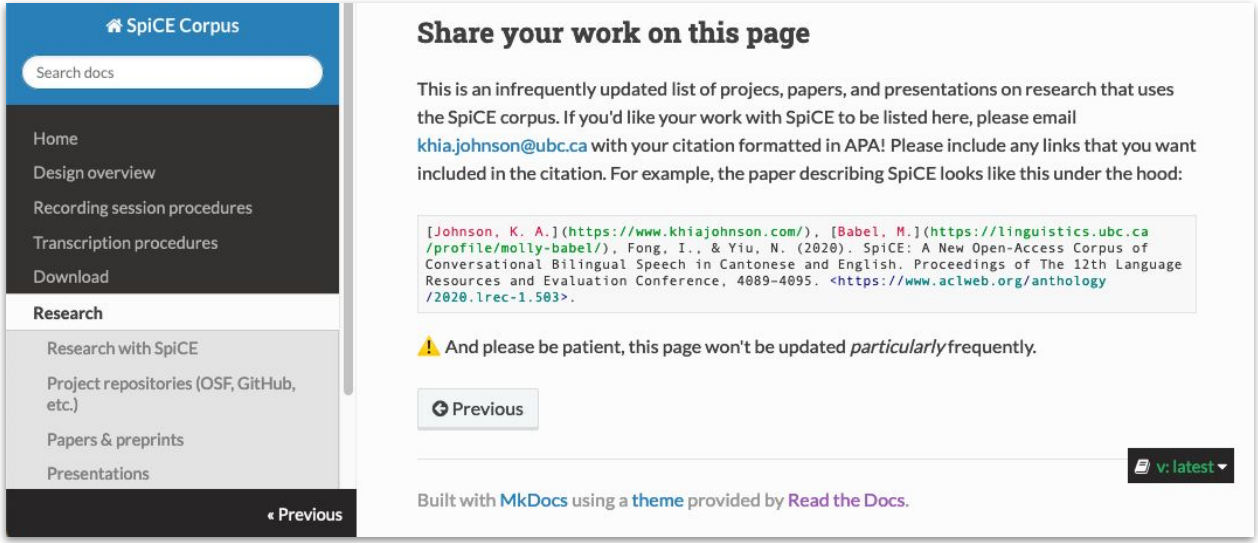
Read the Docs

How to cite SpiCE

- Creative Commons Attribution 4.0 International License
- Cite the data (preferred):
  - Johnson, Khia (2021). SpiCE: Speech in Cantonese and English, https://doi.org/10.5683/SP2/MJOXP3, Scholars Portal Dataverse.
- Cite the paper:
  - 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

If you use SpiCE in your research...

- ...we'd love to hear about it & feature your work here



Acknowledgements

- Molly Babel
- Nancy Yiu
- Ivan Fong
- Katherine Lee
- Kristy Chan
- Rachel Ching Fung Wong
- Rachel Soo
- Angelina Lloy
- Natália Oliveira Ferreira
- Michelle To
- Christina Sen
- Ariana Zattera
- Many members of the Speech-in Context lab!
- Our librarians - Susan Atkey & Doug Brigham

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

References (our research)

Johnson, K. A. & Babel, M. (2021). Language contact within the speaker: Phonetic variation and crosslinguistic influence. *Manuscript under review*.

Johnson, K. A. & Babel, M. (Accepted). Uniformity and crosslinguistic influence in Cantonese-English bilingual stops. The 3rd International Symposium on Bilingual and L2 Processing in Adults and Children, Virtual, June 3-4.

Johnson, K. (2021). SpiCE: Speech in Cantonese and English, https://doi.org/10.5683/SP2/MJOXP3, Scholars Portal Dataverse.

Johnson, K. A. & Babel, M. (2020). The role of passage length in acoustic voice variability in bilingual speech. The 179th Meeting of the Acoustical Society of America, Virtual, December 7–11.

Lloy, A., Johnson, K. A., & Babel, M. (2020). Bilingual talker identification with spontaneous speech in Cantonese and English: The role of language-specific knowledge. The 179th Meeting of the Acoustical Society of America, Virtual, December 7–11.

Johnson, K. A., Babel, M., & Fuhrman, R. A. (2020). Bilingual acoustic voice variation is similarly structured across languages. Proceedings of Interspeech 2020, 2387–2391. https://doi.org/10.21437/Interspeech.2020-3095

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

Johnson, K. A. & Babel, M. (2019). Bilingual sibilant acoustics in conversational Cantonese-English speech. The 178th Meeting of the Acoustical Society of America, San Diego, USA, December 2–6.

References (other)

Audacity Team. (2018). Audacity (R): Free audio editor and recorder. Retrieved from https://www.audacityteam.org

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. Journal of Phonetics, 61, 30–47. https://doi.org/10/ggktm8

Google. (2019). Cloud speech-to-text. Retrieved from https://cloud.google.com/speech-to-text/.

IEEE. (1969). IEEE recommended practice for speech quality measurements. IEEE Transactions on Audio and Electroacoustics, 17(3):225–246, September.

Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. Topics in Cognitive Science, 10(4), 707–728. https://doi.org/10/ggdfmt

Lee, J. L. (2018). PyCantonese (3.1.0) [Computer software]. https://pycantonese.org/

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

Littell, P. (2010). Thank-you notes [Version 1.0: Agent focus]. Retrieved from http://totemfieldstoryboards.org/ stories/thank_you_notes/

McAuliffe, M., Socolof, M., Stengel-Eskin, E., Mihuc, S., Wagner, M., and Sonderegger, M. (2021). Montreal Forced Aligner [Version 2.0]. Retrieved from https://montrealcorpustools.github.io/Montreal-Forced-Aligner/

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. Speech Communication, 45(1):89–95.

Sloetjes, H. and Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco, May. European Language Resources Association (ELRA).