Beyond the Lab: Acoustic Analysis of Speech Data from Smartphones and Traditional Recordings

Ke Zhang (kezhang@uwm.edu) & Hanyong Park (park27@uwm.edu)
Department of Linguistics, University of Wisconsin-Milwaukee

MidPhon 28

October 20-21, 2023
Purdue University
West Lafayette, IN

Research Question

- **Do Smartphone recordings provide quality speech data good for acoustic analysis?**

Backgrounds

- Lab recording is becoming more challenging due to various factors (e.g., COVID-19, participant recruitment, etc.).
- With the advancement of technology, smartphone could be an alternative way for collecting speech data.
- Few smartphone studies have examined acoustic properties relevant to linguistic research.
  - e.g., monophthongs [1][2]; diphthongs only at midpoint [3]; fricatives only with COG [4]
- Thus, we examined acoustic properties of fricatives and vowels in depth.

Methods

- Recordings with studio recording equipment (*iMac, Earthworks M30 microphone, Sound Devices USBPre 2 Audio interface*) vs. Recordings with participants' own smartphones (*iPhone/Galaxy, internal microphone*) in a sound booth
  - iMac: Praat (6.2.23) – 44,100 kHz, Mono
  - Phone: *Awesome Voice Recorder* app. Setting: WAV/OGG, High Encode Quality, 44,100 kHz, 256kbps, Mono (20 iPhone, 3 Android)
- Participants: 23 speakers (F = 16, Age M = 22 yrs.) with diverse L1 backgrounds (English = 11, Arabic = 6, Chinese = 1, Persian=1, Japanese=1, Spanish=1, Czech=1, & Dutch=1)
- Reading *the North Wind and the Sun* first with phone and then through the microphone

Analysis

- 8 fricatives: [f] [v] [θ] [ð] [s] [z] [ʃ] [h]
  - spectral moments (center of gravity: COG, variance, skewness, & kurtosis), fricative duration, & global intensity. Praat script [6]
- 9 monophthongs: [a] [æ] [ʌ] [ɔ] [ɛ] [ɪ] [i] [ʊ] [u] in various stressed location (0 = no, 1 = primary, 2 = secondary)
  - F1, F2, and F3 at one third, mid, & two thirds points, F0, & duration
- 3 diphthongs: [aɪ] [eɪ] [oʊ]
  - F1, F2, F3 at one third, mid, & two thirds points, F0, duration
- *The Online Forced Aligner (the Penn Phonetics Lab Forced Aligner for English)*

Results

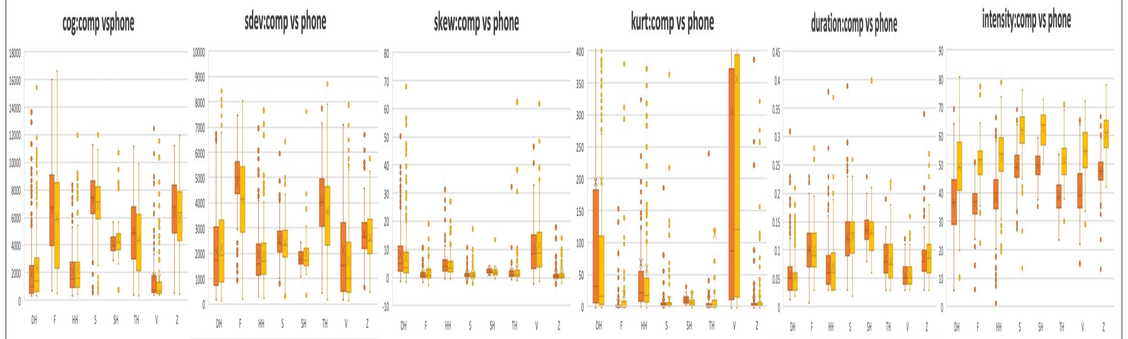

Figure 1: Boxplots for cog(Hz), sdev(Hz), skewness, kurtosis, duration (sec.), & intensity(dB) for each fricative from studio recording (com) and smartphone recording (phone)

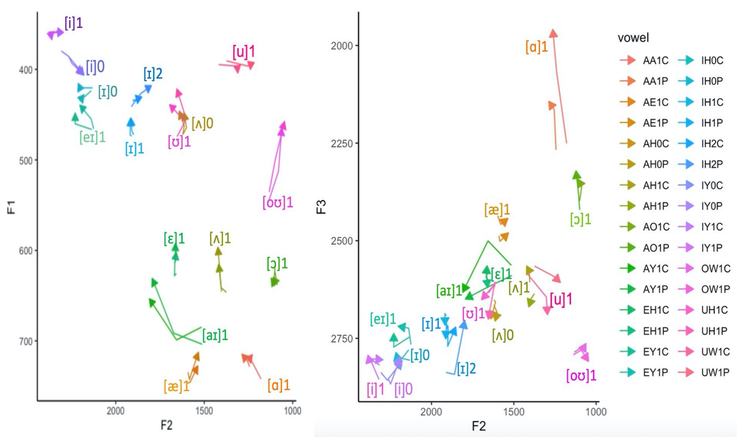

Figure 2: Comparison of F1 and F2, F2 and F3 (Hz) values yielded by two recording types

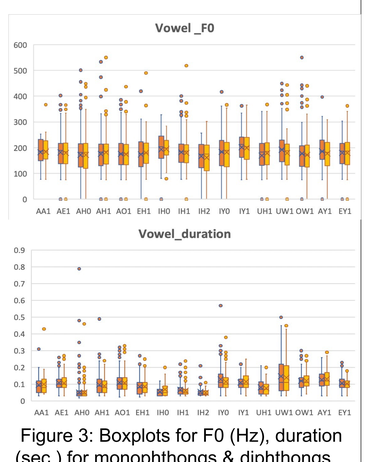

Figure 3: Boxplots for F0 (Hz), duration (sec.) for monophthongs & diphthongs from studio recording (com) and smartphone recording (phone)

Results & Discussion

- Acoustic measures were comparable across all recordings, except for global intensity.
- Fricatives result:

	Our results	Previous studies
Center of gravity (COG)	[s][z][f] > [θ][ʃ]>[ð][h][v]	[s][z] > [θ][ð][f][v] > [ʃ][ʒ] [5]
Variance (standard deviation)	[f][θ] > [s][z][ʃ][ð][h][v]	[v] > [f] > [θ][ð] > [ʃ][ʒ] > [s][z] [5][7]
Skewness	[v][ð] positive strong concentration	[ʃ][ʒ] positive-strong concentration of energy in lower frequencies [5]
Kurtosis	[v][ð] with clearly defined peaks	[s][z] with clearly defined peaks [5]
Duration	voiceless > voiced	[s][z][ʃ][ʒ] > [θ][ð][f][v] [5], voiceless > voiced [7]
Intensity	sibilants > non-sibilants	[s][z][ʃ][ʒ] > [θ][ð][f][v] [5], voiceless > voiced [7]

Table 1. Comparison of fricative results between the current and previous studies

- **Smartphone recordings yielded speech data suitable for acoustic analysis.**
  - With a lossless mobile phone application & in a quiet environment

Remaining issue

- Would the results differ when recordings are conducted outside a sound booth (e.g., in a quiet room), a more realistic recording environment?

Reference

[1] De Decker, P., & Nycz, J. (2011). For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics*, 17(2), 7.

[2] Zhang, C., Jepson, K., Lohfink, G., & Arvaniti, A. (2021). Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America*, 149(6), 3910-3916.

[3] Sanker, C., Babinski, S., Burns, R., Evans, M., Johns, J., Kim, J., Smith, S., Weber, N., & Bowern, C. (2021). (Don't) try this at home! The effects of recording devices and software on phonetic analysis: Supplementary material. *Language*, 97(4),

[4] Sanker, C., Babinski, S., Burns, R., Evans, M., Johns, J., Kim, J., Smith, S., Weber, N., & Bowern, C. (2021). (Don't) try this at home! The effects of recording devices and software on phonetic analysis: Supplementary material. *Language*, 97(4), e360-e382.

[5] Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252-1263.

[6] DiCanio, C. [2021]. Spectral moments of fricative spectra script in Praat. Version 4.0.

[7] Silbert, N., & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *The Journal of the Acoustical Society of America*, 123(5), 2769-2779.

Acknowledgements: We thank Preston Ruh and Amirah Alruwais for their help in collecting data. Travel supported by the UWM Linguistics Department.