A cross-linguistic investigation of acoustic voice spaces

Yoonjeong Lee[1], Marc Garellek[2], Christina Esposito[43] and Jody Kreiman[1]

[1]Department of Head and Neck Surgery, David Geffen School of Medicine at UCLA; Department of Linguistics, University of California, Los Angeles
[2]Department of Linguistics, University of California, San Diego; [3]Department of Linguistics, Macalester College
yoonjeonglee@ucla.edu | linguistylee.com

3aSC16

1. Variability in Voice Quality

The acoustic signal

- input to the perceptual system
- highly variable
- critical for formulating models of voice quality and talker recognition

Prototype-based models for voice identity perception

- **Population prototype**: A context-dependent "average-sounding" voice residing at the center of a multidimensional acoustical voice space
- **Reference pattern**: Each voice's unique deviations from the group prototype

To what extent does the phonological structure of a language impact acoustic variation in voice spaces for individual and populations of speakers?

HYPOTHESIS: A few biologically relevant measures will emerge commonly across languages, while some variance will depend on the structure of the language.

2. Language Datasets

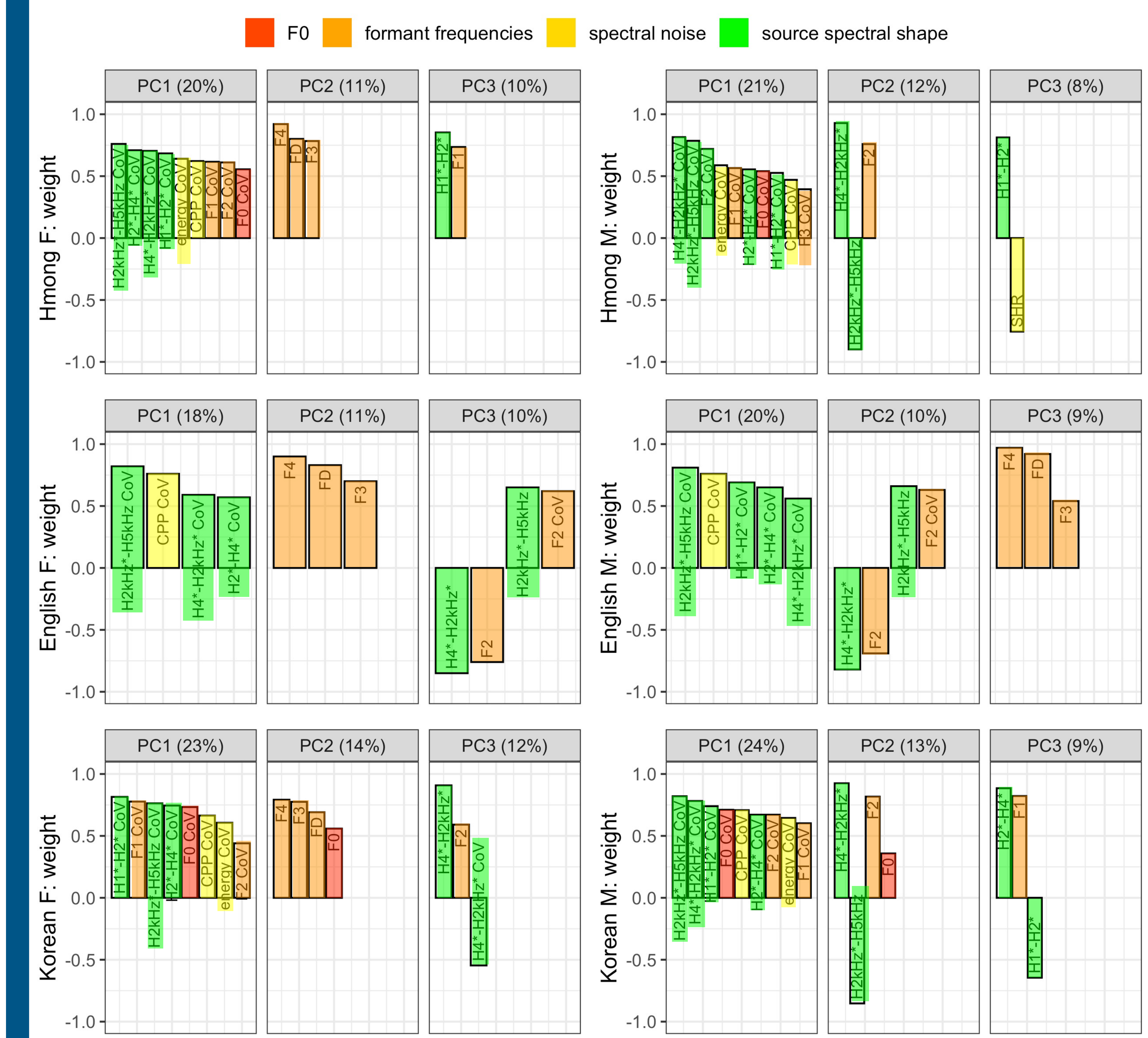
Language	Tone	Phonation	Speaker	Speech task
Hmong	Y	Y	F: 5 M: 3	Story reading
English	N	N	F: 50 M: 50	Sentence reading, spontaneous phone conversation
Seoul Korean	N; specific phrase intonation patterns	N	F: 5 M: 5	Sentence reading

3. Method

Variable categories	Acoustic variables
Pitch	F0
Formant frequencies	F1, F2, F3, F4, formant dispersion (FD; average interval between formants)
Harmonic source spectral shape	H1*–H2*, H2*–H4*, H4*–H2kHz*, H2kHz*–H5kHz
Source/spectral noise	CPP, energy, subharmonics ratio (SHR)
Variability	coefficients of variation (CoV) for all measures

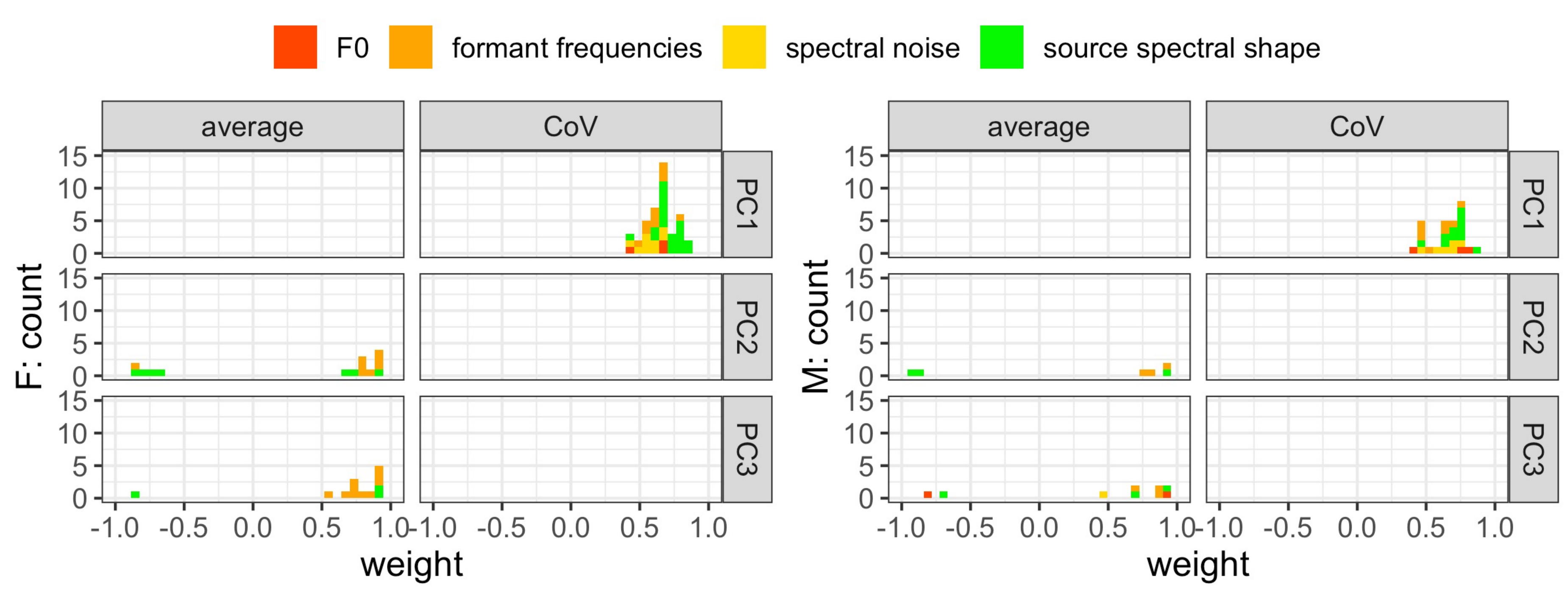
- Acoustic variables were measured every 5 ms on vowels and approximants using VoiceSauce.
- PCA was performed on the acoustic data (values of *moving averages & moving coefficients of variation*).

4. Between-Speaker PCA: Group Voice Spaces



(N%) = variance explained; 'CoV' = coefficient of variation

5. Within-Speaker PCA: Individual Voice Spaces for Hmong Voices



'CoV' = coefficient of variation

6. The Structure of Acoustic Voice Spaces

Results revealed both substantial similarities and differences across languages:

- Across languages and speaker groups 7-9 PCs are extracted for F groups and M groups, accounting for 70% (HF), 68% (HM), 71%(EF), 64% (EM), 67% (KF), and 68% (KM) of the cumulative variance.
- Reference patterns for speakers are mainly computed over the balance between higher harmonic amplitudes and inharmonic energy (**degree of perceived breathiness or brightness**) and over formant dispersion (**speaker identity & vocal tract length**), regardless of language spoken.
- The first three PCs are largely shared across speakers, together accounting for ~50% of the explained variance in the underlying acoustic data.
- The remaining PCs, cumulatively explaining ~20% of the variance, differ widely across speakers.

Language effect

- **F0 variability** commonly emerged for Hmong and Korean voices.
  - Hmong: tonal contrast in the phonology
  - Korean: Seoul speakers' systematic use of F0 for phrasal/accentual information.
  - Difference from English: **F0 variability** only emerged in English speakers' spontaneous speech, not read speech
- **H1-H2** (**correlated with phonation**) accounted for substantial variation only in Hmong voices.
- Unlike English, for Hmong and Korean voices, **lower formant frequencies** (i.e., **vowel quality**) account for the most acoustic variance within and across talkers.

7. Conclusion

- These results further replicate our findings that the same small set of acoustic variables characterizes acoustic variability across virtually all voices, regardless of language spoken.
- Patterns of acoustic variability in multi-talker spaces are largely similar to the patterns found within speakers.
- However, this shared structure accounts for about a half of acoustic variability in the individual and group data, with remaining variability being idiosyncratic.
- Our findings suggest that acoustic voice spaces are shaped by both biologically and phonologically (language-specifically) relevant factors.
- This might be a mechanism for the "own language" advantage in speaker perception.
- Prototypes may not be "average tokens" but may instead be specified by a very small number of acoustic attributes.

Supported by **NIH DC01797** & **NSF IIS-1704167**