Cross-linguistic differences in the phonetic implementation of /s/

Massimo Lipari, Morgan Sonderegger, and Meghan Clayards

McGill University

Background: Although there exists a long history of research and a rich body of literature on the acoustics of fricatives, the extent of variation across languages is not well documented. Only a small number of languages have been studied (with work on English being especially overrepresented), and most individual studies have focused on a single language. However, differences in the kinds of data used (e.g., laboratory speech vs naturalistic data, recording conditions and equipment) and in methodological choices (e.g., recording sampling rate, acoustic measures used) can make comparisons across studies challenging. The few existing large-scale cross-linguistic studies of fricatives (e.g., Nartey, 1982, 14 languages; Gordon et al., 2002, 7 langs.), moreover, are beginning to show their age. For one, whereas these papers rely on regular discrete Fourier transforms to obtain spectral estimates for fricatives, more robust multitaper analysis (Thomson, 1982) has become the standard in the field lately. For another, while most older work considers only static acoustic measures (e.g., spectral properties at the midpoint of each fricative), it has been shown in recent years that fricative dynamics can also differ across languages—cf. Reidy (2016), who finds differences in the timing of the point of maximal sibilance between English and Japanese /s/.

This exploratory study seeks to help update our understanding of the typology of fricative acoustics across languages, with a particular view to capturing time-dynamic information. As a first step, we focus on /s/ (due to its near-universality cross-linguistically), conducting a large-scale corpus study incorporating comparable data from several languages.

Data & methods: All data are from *GlobalPhone* (Schultz et al., 2013), a large multilingual (22 langs.) corpus of laboratory speech with comparable recording conditions—notably, similar microphones were used and all recordings were produced with a sampling rate of 16 kHz. So far, more than 80,000 tokens of word-initial, pre-vocalic /s/ produced by 591 speakers across 6 languages—Czech, French, Korean, Swedish, Thai, and Turkish—have been extracted. For each token, multitaper spectra with parameters K = 8, nW = 4 are computed over 20 ms windows at 17 equidistant points (following Reidy, 2016) using a modified version of the *R* script from Sonderegger et al. (2023). More languages are currently being analyzed.

Preliminary results: Modelled results are not yet available: tentatively, summaries of empirical distributions are given here instead. **Figure 1** shows the average (first within, then across speakers) /s/ center of gravity (COG) value at each measurement point for female (left) and male (right) speakers of each language. Four dimensions of cross-linguistic variation are suggested here: (1) the overall height of the trajectory/average COG value (e.g., higher in Thai, lower in French), (2) the peakedness of the trajectory (e.g., relatively flat curve in Turkish/Korean, a more prominent peak in Czech), (3) the timing of the COG maximum (e.g., relatively early in Swedish/Korean, relatively late in French), and (4) the size of the difference between gender groups (e.g., smaller in Turkish/Thai/French, greater in Czech). In **Figure 2**, the distribution of speakers' average (first within, then across, tokens) COG values is shown. Although a lot of overlap between the distributions for each language can be seen, systematic differences between languages emerge—both in terms of central tendency and of variance (for example, French, Thai, and Czech seem to show greater interspeaker variability, while Korean exhibits somewhat less). The amount of variability does not clearly seem to differ between the male and female speaker groups in the languages considered here.

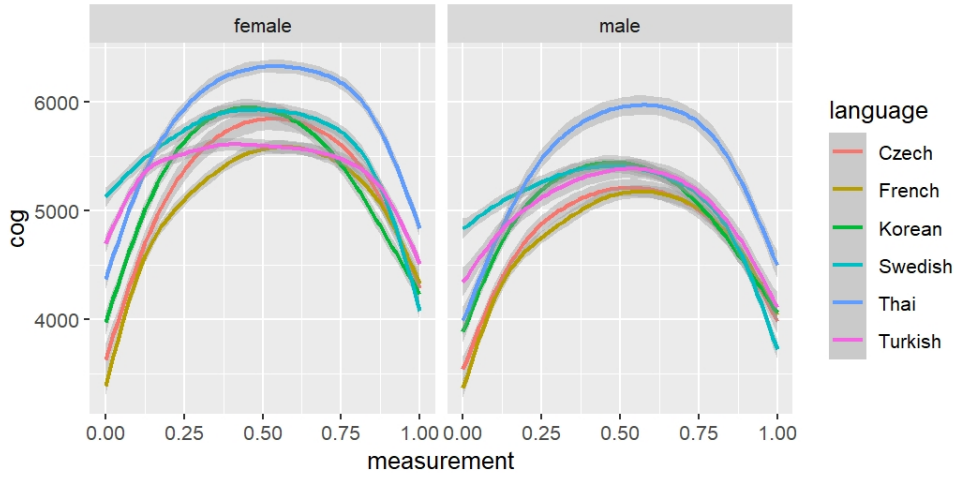

Figure 1: Change in COG over the average /s/ token across 5 languages, faceted by gender.

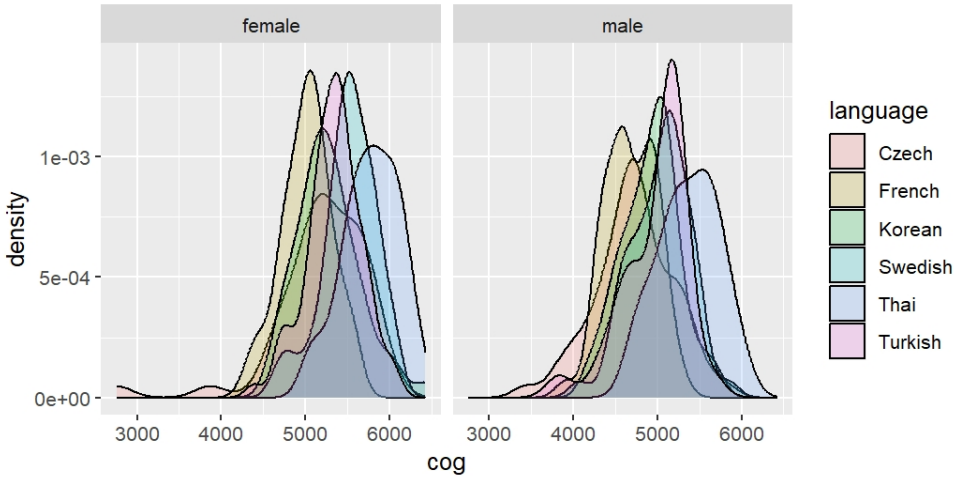

Figure 2: Distribution of speakers' average COG in /s/ across 5 languages, faceted by gender.

References

Gordon, Barthmaier, & Sands. (2002). A cross-linguistic acoustic study of voiceless fricatives. JIPA, 32(2), 141–174.

Nartey. (1982). On fricative phones and phonemes: Measuring the phonetic differences within and between languages. UCLA.

Reidy. (2016). Spectral dynamics of sibilant fricatives are contrastive and language specific. JASA, 140(4), 2518–2529.

Schultz, Vu, & Schlippe. (2013). GlobalPhone: A multilingual text & speech database in 20 languages. 2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 8126–8130.

Sonderegger, Stuart-Smith, Mielke, & SPADE Consortium. (2023). How variable are English sibilants? Proc. 20th ICPhS, 3196–3200.

Thomson (1982). Spectrum estimation and harmonic analysis. Proc. IEEE, 70(9), 1055–1096.