Identifying speaker specific properties in Russian fricatives

Natalja Ulrich[1], Marc Allassonnière-Tang[2], François Pellegrino[3]
Laboratoire Dynamique Du Langage (DDL) UMR 5596, CNRS/Université Lyon 2, Lyon, France
Laboratoire Eco-Anthropologie (EA) UMR 7206, CNRS/MNHN/Université Paris Cité, Paris, France

natalja.ulrich@univ-lyon2.fr

Abstract (max 300 words)

Phonetic research on speaker-specific information assumes that idiosyncratic characteristics are reflected in the physical properties of speech sounds (He & Dellwo, 2014) and can be exploited by listeners, and through the extraction of acoustic cues (Dellwo et al., 2007). Investigating fricative sounds, studies found considerable inter speaker variation, suggesting their further exploration (Gordon et al., 2002; C. Kavanagh, 2011; C. M. Kavanagh, 2012; Narayanan et al., 1995; Schindler & Draxler, 2013; Silbert & de Jong, 2008).

On the example of the Russian fricatives /f/, /s/, /S/, /v/, /z/, /Z/, /sj/, /Sj/, produced by 58 native speakers of Russian, the current study explores the question of how much speaker-specific information is carried by noise sounds. First, we identify gender-specific acoustic properties. Second, we zoom into the individual level. To do so, acoustic speech features, such as temporal, spectral, and harmonicity measurements and 13 Mel Frequency Cepstral Coefficients (MFCCs) are extracted. The statistical mean, median, and standard deviation are also computed.

Applying various statistical and machine learning methods (such as decision-tree based algorithms), the following overall patterns are detected: i) significant gender and individual differences are to a certain degree sound and cue specific, with alveolar fricatives carrying the most idiosyncratic differences; ii) speakers differ significantly in their productions and variability, with female speaker producing higher spectral energy as well as more variation across the production of the same token; iii) females produce less distance between, for instance fricative categories /f/ and /s/, but a greater distance between /s/ and /S/; iv) applying machine learning models, speakers gender can be identified by acoustic cues with high accuracy for acoustic speech feature and MFCCs.

Besides its academic contribution, the investigation enhances our understanding of the distribution of linguistic and speaker idiosyncratic information, which contributes to the development of Automatic Speech Recognition (ASR) systems and the assessment of the impact of individual variation in linguistic research.

[Abstract Word count: 310]

References

Dellwo, V., Huckvale, M., & Ashby, M. (2007). How Is Individuality Expressed in Voice? An Introduction to Speech Production and Description for Speaker Classification. *Speaker Classification I, 4343*, 1–20. https://doi.org/10.1007/978-3-540-74200-5_1

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174. https://doi.org/10.1017/S0025100302001020

He, L., & Dellwo, V. (2014). Speaker idiosyncratic variability of intensity across syllables. *Interspeech 2014*, 233–237. https://doi.org/10.21437/Interspeech.2014-59

Kavanagh, C. (2011). *Intra- and inter-speaker variability in acoustic properties of English /s/*. The *Journal of the Acoustical Society of America*, *130*, 2519. https://doi.org/10.1121/1.3655046

Kavanagh, C. M. (2012). *New consonantal acoustic parameters for forensic speaker comparison*. Unpublished doctoral dissertation: University of York.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *98*(3), 1325–1347. https://doi.org/10.1121/1.413469

Schindler, C., & Draxler, C. (2013). Using spectral moments as a speaker specific feature in nasals and fricatives. *Interspeech 2013*, 2793–2796. https://doi.org/10.21437/Interspeech.2013-639

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. https://doi.org/10.1121/1.2890736