Time-varying spectral characteristics of Danish stop releases

Rasmus Puggaard-Rode, Leiden University Centre for Linguistics
r.p.hansen@hum.leidenuniv.nl

The aspirated alveolar stop /t/ in Modern Standard Danish is affricated quite dramatically, but does not pattern phonologically as an affricate. This has led to a variety of different descriptions and transcription strategies. /p k/ are usually taken to be ‘regular’ aspirated stops. No systematic phonetic studies have ever been made of these stops’ release characteristics, and it is not clear if the affrication in /t/ is simply mentioned more often because sibilant frication is most salient. There are no obvious acoustic phonetic heuristics for determining whether a sound is an affricated stop or an actual affricate, but with the right tools, examining the spectrum can go some way towards answering whether affrication is environmentally determined, speaker-specific, etc.

There are a number of methods in use for analyzing frication on the basis of the spectrum. Calculating spectral moments – i.e. treating the complex spectrum as a probability mass function – can be used to summarize the spectrum (Forrest et al., 1988). Particularly the spectral mean (or center of gravity) is a popular measurement in this regard. The mid-frequency spectral peak can pinpoint relatively minor differences in place of articulation quite precisely (Koenig et al., 2013), as can the four highest cepstral coefficients of a discrete cosine transform (Bunnell et al., 2004). These measurements are usually taken at static or normalized points in time (‘magic moments’; see Mücke et al., 2014). There are advantages and disadvantages to all these approaches, but perhaps the main advantage of all of them is that they reduce the complex, multi-dimensional information in the spectrum into a manageable number of discrete values, which can then serve as dependent variables in statistical models.

In this poster, I propose function-on-scalar regression (FOSR; Bauer et al., 2018) as a method to model spectra and their variation in time holistically and without the need to reduce dimensionality. In functional data analysis, variables do not need to be discrete values, but can instead be (smoothed) curves, such as spectra. An FOSR model is similar to a linear mixed-effects model or a generalized additive mixed model, except with a functional response instead of a discrete continuous response. The output of FOSR models give clear and easily interpretable overviews of the influence of various factors on the functional response. Examples are given in Figures 1–2. In Figure 1, we see the influence of time on the energy distribution in the spectrum; the structure is similar to spectrograms, with normalized time along the x-axis, frequency along the y-axis, and grey-scale shading indicating energy. Figure 1 shows significant energy at frequencies above 5 kHz early on in /t/ releases, and significant energy below 5 kHz towards the end of the release. This suggests that the stop releases are initially affricated, but that this affrication is gradually dominated by aspiration. Figure 2 shows how this pattern differs for male and female speakers.

For this study, I used the spontaneous monologues in the DanPASS corpus (Grønnum, 2009). For each aspirated stop (n = 2,334), multitaper spectra were extracted from 20 normalized time steps. Separate FOSR models were fitted for each phoneme /p t k/, and the influence of each dependent variable subsequently visualized. The results show that /t/ is indeed invariably affricated, although the spectrum is very dynamic throughout /t/ releases, with affrication gradually being replaced by aspiration. The acoustic characteristics of /p/ releases show a lot of inter-speaker variation, but also coarticulatory context effects, primarily during the first half of the release. Coarticulatory context effects greatly influence the spectra of /k/ releases throughout. These findings can all be explained with reference to general acoustic and articulatory principles, and with reference to previous studies of Danish stop articulation, in particular muscular and glottal activity (Fischer-Jørgensen and Hirose, 1974; Hutters, 1985). I believe that using FOSR models to analyze time-varying spectra could potentially be fruitful in the analysis of many problems in acoustic phonetics beyond this one.

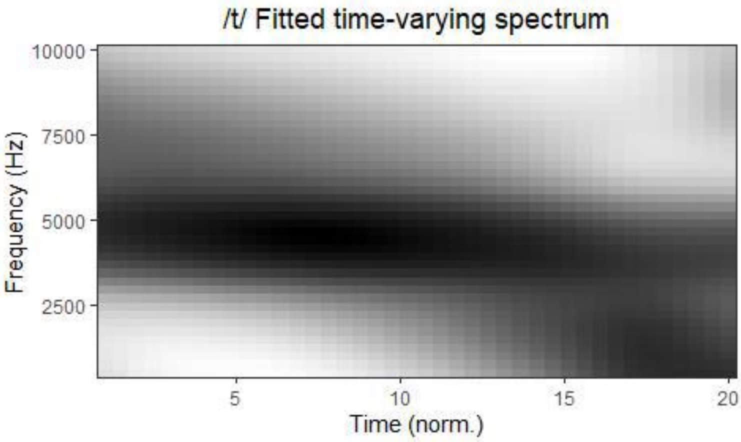

Figure 1: Fitted time-varying spectrum of /t/ releases (main effect of time).

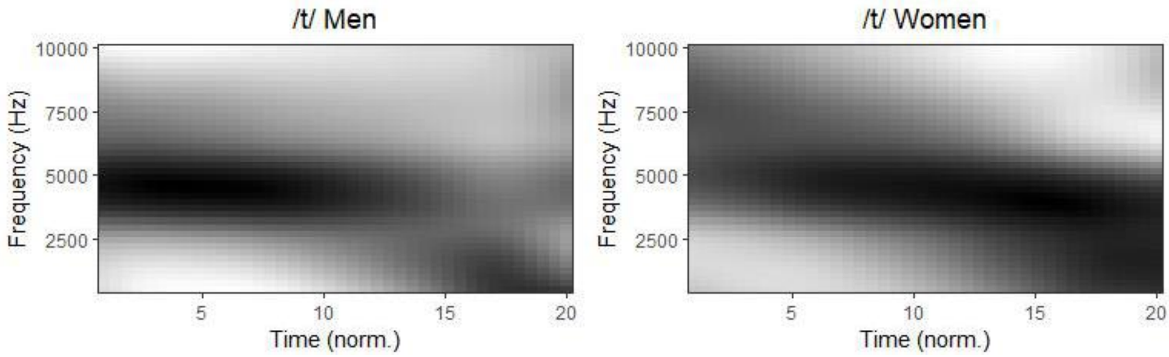

Figure 1: Fitted time-varying spectrum of /t/ releases for each direction of the sex variable.

References

Bauer, A., F. Scheipl, H. Küchenhoff & A.-A. Gabriel (2018). An introduction to semiparametric function-on-scalar regression. *Statstical Modelling, 18*, 346–364.

Bunnell, H.T., J. Polikoff & J. McNicholas. (2004). Spectral moment vs. bark cepstral analysis of children's word-initial voiceless stops. *International Conference on Spoken Language Processing, 8*.

Fischer-Jørgensen, E. & H. Hirose. (1974). A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production. *Status Report on Speech Research, 39/40*, 231–253.

Forrest, K., G. Weismer, P. Milenkovic & R.N. Dougall. (1988). Statistical analysis of word-initial voiceless obstruents. Preliminary data. *Journal of the Acoustical Society of America*, *84*, 115–123.

Grønnum, N. (2009). A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication, 51,* 594–603.

Hutters, B. (1985). Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica, 42,* 1–24.

Koening, L.L., C.H. Shadle, J.L. Preston & C.R. Mooshammer. (2013). Toward improved spectral measures of /s/. Results from adolescents. *Journal of Speech, Language, and Hearing Research, 56,* 1175–1189.

Mücke, D., M. Grice & T. Cho. (2014). More than a magic moment. Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics, 44*, 1–7.