Dimensions of segmental variability: relationships between information density and prosodic structure

Zofia Malisz[1], Erika Schulz[1], Yoon-Mi Oh[2], Bistra Andreeva[1], Bernd Möbius[1]
[1]Universität des Saarlandes, Germany
[2]Université de Lyon, France
sfb-c1@coli.uni-saarland.de

Variation in information density may affect phonological and phonetic structure (Aylett and Turk, 2004; Jaeger and Levy, 2006; Aylett and Turk, 2006; Jaeger, 2010). For instance, if a word is predictable, phonemes are elided or acoustic-phonetic features are reduced in turn decreasing acoustic signal redundancy. Conversely, segments or syllables forming a less predictable word are lengthened and spectrally expanded. Reduction and expansion of acoustic-phonetic features is also characteristic of prosodic variability. Aylett and Turk (2004) in fact posit that prosodic structure mediates between signal variability and relative predictability.

In this study, we aim to assess the impact of information density and prosodic structure on phonetic encoding, both independently of each other and in interaction. We model segmental duration and vocalic quality as a function of information density measures (e.g. surprisal, estimated as the inverse of the unit's *log* probability), as well as syllable prominence, phrase boundary and speech rate. We also control for contextual factors known to affect segmental variability, e.g. the identity of preceding and following segments. Correlates of phonetic encoding density are extracted from a subset of the BonnTempo corpus (Dellwo et al., 2004) in which six speakers (3 male, 3 female) per language (Czech, German, English, Finnish, French, Polish) read a set text in three different speech rates. Surprisal is estimated from n-gram language models (LMs) trained on equivalent text corpora in the studied languages (WebCELEX, Lexique 3.8, etc.). Statistical analysis is based on linear mixed models.

First results for German, English and French show expected independent effects of prosodic factors on duration and a significant interaction between surprisal (via tri-gram LM) and syllabic prominence, supporting the dependence of both dimensions in their effect on duration. In addition, correlation analysis shows a positive relationship between surprisal (via bi-gram LM) and distinctiveness of vocalic quality: the higher the surprisal value of a respective bi-gram, the larger the distance between the vowel space centroid and formant values for each vowel.

We discuss the results in light of the present information theoretic models of spoken phenomena. We also reflect upon the usefulness of different n-gram LMs and measures of information density in modelling phonetic and acoustic variability.

References

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.

Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119:3048–3058.

Dellwo, V., Steiner, I., Aschenberner, B., Dankovicova, J., and Wagner, P. (2004). BonnTempo-Corpus & BonnTempo-Tools: A database for the study of speech rhythm and rate. In *Proceedings of Interspeech 2004*, pages 777–780.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1):23–62.

Jaeger, T. F. and Levy, R. P. (2006). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems*, pages 849–856.