

Towards a Unified Model of Word Pronunciation

T. Florian Jaeger (fjaeger@bcs.rochester.edu)

Celeste Kidd (ckidd@bcs.rochester.edu)

Department of Brain & Cognitive Sciences, University of Rochester
Rochester, NY 14627 USA

Abstract

Speakers pronounce predictable words shorter and with less articulatory detail (e.g. Bell et al., 2002). Other studies have found that speakers also modulate the duration of a word, if the next words are not readily available for pronunciation to buy more time (e.g. Clark & Fox Tree, 2002). We present the first studies ever that directly compare and integrate these two principles (redundancy avoidance and strategic lengthening) into a unified model. We model the duration of 65,000 determiners from a time-aligned corpus of spontaneous American English speech. We present cases studies on individual determiner types (e.g. 15,000 cases of "a"; 20,000 cases of "the") as well as combined a study on all determiner types (including controls for differences in phonological complexity). Our studies provide evidence that speakers' modulation of speech rate is due to three underlying causes: redundancy avoidance (predictable words are reduced), strategic lengthening before unavailable material, and joint retrieval of collocations.

References

- Bell, A., Jurafsky, D., Fosier-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2002). Effects of disfluencies, predictability, and utterance position of word form variation in English conversation. *Journal of the Acoustical Society of America*, 113(2), 1001-1024.
- Clark, H. & Fox Tree, J. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84, 73-111.