

The Cumulative-Cue Hypotheses: An account of understanding the multimodal nature of prosodic prominence

Gilbert Ambrazaitis¹, David House²

¹ Department of Swedish, Linnaeus University, Växjö, Sweden

² Division of Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden

gilbert.ambrazaitis@lnu.se, davidh@kth.se

Abstract

Prosodic prominence can be argued to be multimodal, because co-speech gestures tend to temporally align with prominent units in speech (e.g., Flecha-García, 2010; Leonard and Cummins, 2011; Esteve-Gibert et al., 2017), in line with the phonological synchronization rule by McNeill (1992). In addition, speech and gesture may converge not only in the temporal, but also in the “spatial” domain, displaying correlations between the presence and “strength” of gestures (magnitude or complexity) and the strength of acoustic parameters in the production of prosodic prominence as reflected, for instance, in the accentual f_0 range (e.g., Krahmer and Swerts, 2007; Parrell et al., 2014; Pouw et al., 2021; Berger and Zellers, 2022; or see Ambrazaitis and House, 2023, for a review). This spatial convergence has been formulated in terms of the *Cumulative-Cue Hypothesis* (Ambrazaitis and House, 2022, 2023) and has been argued to result from an underlying compulsion to express prominence in both speech and gesture, all else being equal. This compulsion could be understood as part of a revised *Effort Code* (Gussenhoven, 2004): To signal prominence, we tend to produce vocal and gestural signals indicating an increased level of effort. In our talk, we will discuss this idea in more detail and present empirical evidence in favour of it, taken both from published research and our own previous and ongoing studies. For instance, in a data set comprising news readings from Swedish Television, we found a significant trend for larger f_0 rises in sentence-level pitch accents (so-called big accents, a.k.a. sentence accent or focal accent) as gestural complexity increased, measured in terms of number of accompanying head or eyebrow movements (no gesture vs. head vs. head plus eyebrows). In an ongoing study, we aim to replicate the previous analysis with spontaneous speech (Edlund et al., 2010), also including manual gesture strokes.

References

- Ambrazaitis, G., & House, D. (2022). Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters. *Laboratory Phonology*, 24(1). doi: 10.16995/labphon.6430
- Ambrazaitis, G., & House, D. (2023). The multimodal nature of prominence: some directions for the study of the relation between gestures and pitch accents. In *Proceedings of the 13th International Conference of Nordic Prosody* (pp. 262-273). *Sciendo*. doi: 10.2478/9788366675728-024
- Berger, S., & Zellers, M. (2022). Multimodal prominence marking in semi-spontaneous YouTube monologs: The interaction of intonation and eyebrow movements. *Frontiers in Communication*, 7. doi: 10.3389/fcomm.2022.903015
- Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. In

Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10) (pp. 2992-2995). Valetta, Malta.

- Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America*, 141(6), 4727–4739. doi: 10.1121/1.4986649
- Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication*, 52(6), 542–554. doi: 10.1016/j.specom.2009.12.003
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511616983
- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396–414. doi: 10.1016/j.jml.2007.06.005
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471. doi: 10.1080/01690965.2010.500218
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Parrell, B., Goldstein, L., Lee, S., & Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of Phonetics*, 42, 1–11. doi: 10.1016/j.wocn.2013.11.002
- Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., & Dixon, J. A. (2021). Gesture–speech physics in fluent speech and rhythmic upper limb movements. *Annals of the New York Academy of Sciences*, 1491(1), 89–105. doi: 10.1111/nyas.14532