Is phonetic prominence underlyingly multimodal?

Gilbert Ambrazaitis[1] and David House[2]
[1]*Linnæus University, Växjö*, [2]*KTH (Royal Institute of Technology), Stockholm, Sweden*

Stating that spoken language is multimodal – involving an auditory and a visual mode – can today almost be considered trivial [1-6], and the same holds for the phonetic – or better: the physical – dimensions of prosodic prominence, involving both acoustic and kinetic parameters, in production and in perception [7-14]. The purpose of this contribution is to discuss the nature of this multimodality, based on previous and recent empirical evidence. For, despite a growing consensus on the multimodality of speech and prominence, we still lack a common understanding of the relation between the modes: Are kinetic/visual correlates of prominence optional? On the one hand, we still lack sufficient empirical evidence. On the other hand, we believe that we can, on the basis of the evidence we have today, already argue for a hypothesis of an underlying multimodality of prominence: That is, even if not always surfacing, we argue that it is reasonable to assume that in the execution of prominence, both kinetic and acoustic expressions are generally targeted.

To this end, we will review some previous and recent work on the co-occurrence of auditory and visual cues to prominence. For instance, it has been shown that visible articulatory movements such as jaw openings and lip articulations are subject to variation due to prominence, besides their primary function of encoding phonemes (e.g., [7-8]). In addition, a growing line of research focuses on the role of gestures in prominence production and perception. For instance, pitch accents are frequently accompanied by so-called beat gestures, typically produced by the hands, the head or the eyebrows (e.g., [9-11], [15-19]).

Furthermore, some attempts have been made to test whether the acoustic and the kinetic dimensions of prominence would "influence each other" in the sense that accompanying beat gestures would "affect" the phonetic realization of pitch accents [20-21]. Our own recent study [22] does in fact provide evidence for slightly larger pitch ranges found in connection with head movements, and even larger ones when also eyebrow movements are added.

These results, we argue, suggest a cumulative relation between acoustic and kinetic prominence expressions in speech production: the more of the one, the more of the other we tend to get. However, we will argue that the assumption that gesture "affects" pitch (or why not vice versa) only describes the relation (in an inappropriate manner), rather than explaining it. It is more reasonable to assume that a relatively large pitch range on the one hand, and a relatively strong involvement of gestures on the other hand, are two parallel responses to the communicative need to produce a high prominence level.

To make this explanation work out, we hypothesize that prominence is underlyingly multimodal, just as speech in general is. After all, spoken language is produced through bodily (incl. articulatory) movements, where some are both visible and audible (certain articulatory movements), some are audible only (many articulatory movements), while some are only visible (gestures). All three types of movements have been shown to take part in the production of prominence [7-19]. The simplest way to model the involvement of these different channels would be to assume that they all are controlled by the same force, i.e. the need to make a unit of speech more prominent. However, prominence need not always be multimodal at the surface, which is explainable by the fact that all prominence-related channels, both acoustic and kinetic, are multifunctional: Hence, the degree to which they are exploited for prominence in a given situation may depend on the balance between the communicative needs to signal prominence vs. other functions; it may also be language-specific and possibly individual.

At the workshop, we will further discuss the hypothesis of an underlying multimodality in the implementation of prosodic prominence, and some implications of this hypothesis.

[1] McGurk, H., & MacDonald, J.1976. Hearing lips and seeing voices. *Nature* 264, 746-748.
[2] Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.
[3] McNeill, D. 2005. Gesture and Thought. University of Chicago Press.
[4] Biau, E., & Soto-Faraco, S. 2013. Beat gestures modulate auditory integration in speech perception. *Brain Lang.* 124, 143–152.
[5] Wagner, P., Malisz, Z., & Kopp, S. 2014. Gesture and speech in interaction: An overview, *Speech Commun.* 57, 209-232.
[6] Graziano, M., & Gullberg, M. 2018. When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Front. Psychol.,* 9:879.
[7] Beskow, J., Granström, B., & House, D. 2006. Visual correlates to prominence in several expressive modes. In: *Proc. of Interspeech 2006*, (Pittsburg, PA), 1272–1275.
[8] Hermes, A., Becker, J., Mücke, D., Baumann, S., & Grice, M. 2008. Articulatory Gestures and Focus Marking in German. In: *Proc. of Speech Prosody 2008*, (Campinas, Brazil), 457–460.
[9] Swerts, M., & Krahmer, E. 2010. Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phonet.* 38, 197–206.
[10] Esteve-Gibert, N., & Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech Lang. Hear. Res.* 56 (3), 850–864.
[11] Ambrazaitis, G., & House, D. 2017. Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Commun.* 95, 100-113.
[12] Dohen, M., & Loevenbruck, H. 2009. Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206.
[13] Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. 2015. Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *J. Phonet.* 49 (1), 41–54.
[14] Scarborough, R., Keating, P., Mattys, S. L., Cho, T., Alwan, A., 2009. Optical phonetics and visual perception of lexical and phrasal stress in English. *Lang. Speech* 52, 135–175.
[15] Jannedy, S., & Mendoza-Denton, N. 2005. Structuring information through gesture and intonation. *Interdiscip. Stud. Inf. Struct.* 3, 199–244.
[16] Leonard, T., & Cummins, F. 2011. The temporal relation between beat gestures and speech. *Lang. Cogn. Processes* 26, 1457–1471.
[17] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *J. Assoc. Lab. Phonol.* 3, 71–89.
[18] Rusiewicz, H.L. 2010. *The Role of Prosodic Stress and Speech Perturbation on the Temporal Synchronization of Speech and Deictic Gestures* (Doctoral dissertation). University of Pittsburgh.
[19] Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. 2004. The timing of speech-accompanied gestures with respect to prosody. In: *Proc. of From Sound to Sense,* (MIT, Cambridge, MA), 97–102.
[20] 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, 396–414.
[21] Roustan, B., & Dohen, M. 2010. Co-Production of Contrastive Prosodic Focus and Manual Gestures: Temporal Coordination and Effects on the Acoustic and Articulatory correlates of Focus. In: *Proc. of Speech Prosody 2010*, (Chicago).
[22] Ambrazaitis, G., & House, D. 2019. Multimodality in prominence production and its sensitivity for lexical prosody. Paper at *3rd Phonetics and Phonology in Europe conference (PaPE 2019)*, Lecce, Italy.