Towards understanding the multimodality of prominence production and perception

Gilbert Ambrazaitis[1], Johan Frid[2], and David House[3]
[1]*Linnaeus University, Växjö,* [2]*Lund University Humanities Lab,* [3]*KTH Stockholm, Sweden*

Previous research on the visual dimensions of speech suggests that prominent syllables or words are frequently co-produced with beat-like movements of various parts of the body, and the present work focuses on head movements (e.g., [1-2]). Several studies have shown for a number of settings that visually perceived head movements can contribute to the overall prominence impression (e.g., [3-4]), and our own ongoing research (in preparation) has confirmed this effect for head movements produced by Swedish television news presenters: We let two groups of naïve Swedish participants (44 in an audio-visual and 41 in an audio-only condition) perform prominence ratings in a selection of 16 news clips (218 words in total) using a three-level scale. The clips had previously been annotated for head movements. Words co-produced with a head movement were overall rated more prominent than words without, and this difference was significantly larger in the audio-visual than in the audio-only condition.

In another study, based on an extended set of Swedish news data, we have shown that pitch accents on words co-produced with head movements (or head and eyebrow movements) tend to be realized with larger accentual f_o movements than pitch accents on words without head movements [5]. Together, the two reported results suggest that, in order to make a word more prominent (as perceived audio-visually), pitch accents will be both strengthened acoustically and co-produced with a head movement. However, we still do not understand the details of the relationships between acoustic (such as f_o and durations) and kinematic parameters (such as head movements) and how these in turn relate to perceived prominence. A goal for our subsequent research is to scrutinize these relations by means of extending our perceptual ratings to the larger dataset used in [5].

Meanwhile, as a preliminary account to be presented at the conference, along with the overall rating results presented above (in preparation), we study the relations between f_o, head movements, and perceived prominence in two sets of carefully selected words (11 words in total) from the dataset rated so far (218 words). The five or six words in each set were selected in order to introduced a measure of experimental control over phonological-prosodic properties as well as f_o realizations: All five words in set A are di-syllabic words with initial stress and the same word-accent category (Accent 2) from the same female news presenter; all six words in set B are compounds (with Accent 2 and two lexical stresses), uttered by the same male news presenter. The words in both sets vary critically in terms of f_o realizations and head movements.

Results from the audio-visual condition suggest that the presence of head movements might play a crucial role for the perception of prominence, as, in set A, words with head movements were rated clearly higher than words without, given similar f_o-profiles (see Fig. 1 and Tab. 1); one word without a head movement (*inte* 'not') likewise received higher ratings, but that word was also produced with a larger f_o range. If the high ratings of the words *saknas* and *alla* in Fig. 1 are related to the presence of the visually perceived head movement, then we should expect considerably lower ratings for these two words in the audio-only condition. However, this was not the result obtained. Fig. 2 compares ratings for the two conditions for the word *saknas* as an example ($F(1,83)=1.317$, $p=0.254$). Turning to set B, however, we a find a difference between the conditions for at least one of the four words with head movements included in this set (Fig. 3, $F(1,83)= 4.22$, $p=0.043$*). These results show that, even if there is an overall effect of visually perceived head movements of the prominence rating (in preparation), this effect might be virtually absent in individual words, and other factors than f_o and head movements can have a rather strong influence on the prominence rating. This latter conclusion is well in line with our general knowledge on prominence, and not least on top-down effects (e.g., [6]). More detailed results will be presented and discussed at the conference.

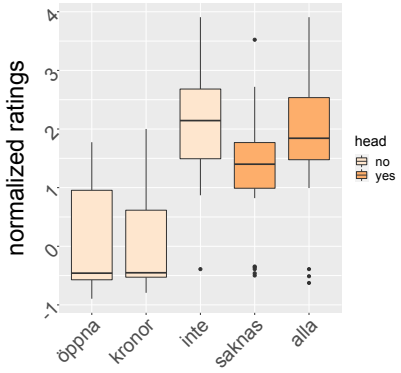

Figure 1. A*udio-visual prominence ratings (rater normalized) for five selected words by a female speaker (set A). See Tab. 1 for word semantics and f_0 characteristics*

Word	Meaning	Head	F0 fall *(st)*	F0 rise *(st)*
öppna	'to open'	no	0.57	6.83
kronor	Sw. currency	no	11.38	5.78
inte	'not'	no	9.90	14.29
saknas	'is missing'	yes	8.55	6.64
alla	'all'	yes	7.01	10.91
tågtrafiken	'railway traffic'	no	10.89	7.92
olyckstillbud	'incident'	no	9.21	11.50
fullmåne	'full moon'	yes	3.37	6.63
järnvägsnät	'rail network'	yes	8.93	12.12
skuldsatta	'indebted	yes	12.23	9.63
underhåll	'maintenance'	yes	14.34	14.58

Table 1. *Characteristics of 11 selected words (set A: upper part; set B: lower part – see text); f_0 fall and rise in semitones (st) refer to the fall and the rise of a two-peaked pitch accent; 'Head' indicates if a head movement is co-produced*

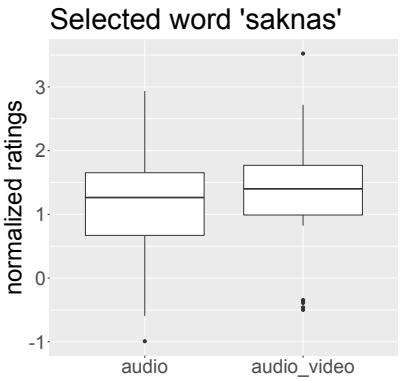

Figure 2. A*udio-visual vs. audio-only prominence ratings (rater normalized) for the word* saknas *'is missing' (set A).*

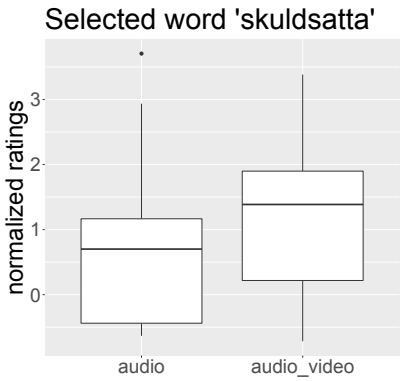

Figure 3. A*udio-visual vs. audio-only prominence ratings (rater normalized) for the word* skuldsatta *'indebted' (set B).*

[1] Swerts, M., & Krahmer, E. 2010. Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics* 38, 197-206.

[2] 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 Communication* 95, 100-113.

[3] Mixdorff, H., Hönemann, A., & Fagel, S. 2013. Integration of acoustic and visual cues in prominence perception. In *Proceedings of AVSP 2013*. Annecy, France.

[4] 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. *Journal of Phonetics* 49, 41-54.

[5] Ambrazaitis, G., & House, D., submitted. *Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters.*

[6] Wagner, P. 2005. Great expectations – Introspective vs. perceptual prominence ratings and their acoustic correlates. In *Proceedings of Interspeech 2005* (pp. 2381-2384). Lisbon, Portugal.