

# Multimodal levels of prominence: a preliminary analysis of head and eyebrow movements in Swedish news broadcasts

Gilbert Ambrazaitis<sup>1</sup>, Malin Svensson Lundmark<sup>1</sup> and David House<sup>2</sup>

<sup>1</sup>Centre for Languages and Literature, Lund University

<sup>2</sup>Department of Speech, Music and Hearing, KTH Stockholm

## Abstract

*This paper presents a first analysis of the distribution of head and eyebrow movements as a function of (a) phonological prominence levels (focal, non-focal) and (b) word accent (Accent 1, Accent 2) in Swedish news broadcasts. Our corpus consists of 31 brief news readings, comprising speech from four speakers and 986 words in total. A head movement was annotated for 229 (23.2%) of the words, while eyebrow movements occurred much more sparsely (67 cases or 6.8%). Results of  $\chi^2$ -tests revealed a dependency of the distribution of movements on the one hand and focal accents on the other, while no systematic effect of the word accent type was found. However, there was an effect of the word accent type on the annotation of ‘double’ head movements. These occurred very sparsely, and predominantly in connection with focally accented compounds (Accent 2), which are characterized by two lexical stresses. Overall, our results suggests that head beats might have a closer association with phonological prosodic structure, while eyebrow movements might be more restricted to higher-level prominence and information-structure coding. Hence, head and eyebrow movements can represent two quite different modalities of prominence cuing, both from a formal and functional point of view, rather than just being cumulative prominence markers.*

## Introduction

People gesture while they speak, using various parts of the body. Hands, the head and certain facial areas (such as the eyebrows) have so far received most attention in research on multimodal communication, i.e. the interaction of gestures and speech. While a single gesture often serves several functions at once, a basic typology of gestures would include emblems, iconic gestures, deictic gestures and beat gestures (Casasanto, 2013). In this typology, beat gestures are special in that they do not necessarily convey any semantic content. Beat gestures are generally understood as simple, rapid movements of a hand, a finger, the head, or the eyebrows, “often repeated, and timed with prosodic peaks in speech” (*Ibid.*, p. 373). Moreover, “the cognitive and communicative functions of beats are not well understood” (*Ibid.*, p. 373).

However, recent studies have begun solving a couple issues concerning beat gestures: For instance, it has been shown that beat gestures can facilitate both speech production (Lucero et al.,

2014) and speech processing (Biau and Soto-Faraco, 2013; Wang and Chu, 2013).

A growing body of evidence also suggests that hand, head and eyebrow movements are aligned with pitch accents in speech and in this way contribute to the production and perception of prosodic prominence (Yasinnik et al., 2004; Flecha-García, 2010). For instance, in a database of video recordings of two shorter (5-7.5 minute) academic lectures by two male speakers of English, Yasinnik et al. (2004) found that beat gestures by the hands, head, or eyebrows occurred in close alignment with ToBI-labelled pitch accents in about 65-90% of the cases. Similarly, Flecha-García (2010) found that eyebrow raises preceded pitch accents by on average 60 ms in a corpus of English face-to-face dialogue.

This study is part of the research project *Multimodal levels of prominence*, investigating the interaction of head movements, eyebrow movements and pitch accents at the sentence level (so-called *focal accents* in the Swedish tradition), in Stockholm Swedish. The main research question of the project is whether the

three modalities (pitch accent, head, eyebrows) can interact in various ways in order to produce different levels of prominence, and whether these prominence levels are used to encode different shades of information structure (such as *new* vs. *accessible* information).

This paper presents a first analysis of the distribution of head and eyebrow movements as a function of verbal prominence levels and word accent categories in a corpus of news readings from Swedish Television. Our approach is inspired by Swerts and Krahmer's (2010) study on Dutch newsreaders, which argues that news readings "represent natural data that are still sufficiently constrained to be able to explore specific functions of their expressive style" (p. 198). Swerts and Krahmer (2010) found that the more accented a word was on an auditory scale (no accent, weak accent, strong accent), the more likely the word was to also be accompanied by a head movement, an eyebrow movement or both (most common in the strongly-accented words). While our materials are largely comparable to theirs, the studies differ in the important respect that we – in this first step – did not establish a perceptual prominence rating. Instead, we are making use of the fact that Swedish has two phonological prosodic prominence levels, which can rather easily be distinguished when inspecting the fundamental frequency contour. Thus, our point of departure is the question whether *phonological* prominence levels – which often, but not always necessarily reflect perceptual prominence levels – have an effect on the distribution of head and eyebrow movements.

Unlike so-called intonation languages like English and German, Swedish is a pitch-accent language, making use of pitch contrasts at the lexical level. In particular, Swedish has a binary distinction between two word accents (Accent 1 and Accent 2), two different pitch accents assigned to words by means of lexical/ morphological rules. In addition, words can be highlighted at the sentence level, just as in English or German. For Stockholm Swedish, a phonological distinction is generally assumed between the non-focal, accented realization of a word (tonal pattern in Stockholm Swedish: H[igh]-L[ow]; with a different timing of the HL for Accent 1 and 2, cf. Bruce, 1977), and a focal realization of a word (HLH, i.e. an additional High tone). Note: While the non-focal vs. focal accents represent two different phonological prominence levels, no difference in prominence

is generally assumed between the two word accents (Accent 1 vs. 2).

Therefore, the hypothesis was that focally accented words would coincide with head or eyebrow movements more often than non-focal words, while the word accent category (Accent 1 vs. 2) should have no effect of the distribution of head or eyebrow movements.

## **Method**

Audio and video data of 31 brief news readings from Swedish Television (SVT Rapport, 2013) were analyzed. The corpus included speech from four newsreaders: two female (Sofia Lindahl, Katarina Sandström) and two male (Pelle Edin, Alexander Norén). Each piece of news typically contained 1-3 sentences, amounting to 986 words in total. The recordings were retrieved on DVD from the National Library of Sweden (Kungliga Biblioteket).

The material was transcribed, segmented at the word level, and annotated using ELAN (Wittenburg et al., 2006). Word segmentations were adjusted using Praat (Boersma and Weenink, 2014) and re-imported in ELAN prior to doing the annotations. Head and eyebrow movements, as well as focal accents were labelled manually by three annotators. In the analysis, a word was counted as coinciding with one of the three events (focal accents, head, eyebrow movements) in the event of an agreement between at least two annotators. The annotation scheme was simple in that only the presence vs. absence of any of the three events was judged upon. That is, no time-aligned annotations were made for the purpose of this study, and hence, no decisions had to be made upon temporal onsets and offsets of the movements. A word was annotated for bearing a (head or eye-brow) movement in the event that the head or at least one eyebrow rapidly changed its position, roughly within the temporal domain of the word. That is, slower movements were ignored, which could occur, for instance, in connection with the re-setting of the head position, which often spanned several words. No distinctions were made between types of directions of movements. However, test annotations revealed that a word may contain either one or two clearly distinguishable beats within a single word and hence, we introduced a distinction between 'simple' and 'double' instances of head or eyebrow movements; we can anticipate that 'double beats' were only

recognized for head, and never for eyebrow movements.

A focal accent was annotated when a rising F0 movement corresponding to the focal H- tone in the Lund model of Swedish prosody (Bruce, 1977; 2007), or to the LH prominence tone in Riad (2006), was recognizable in the F0 contour; note that this F0 movement was expected in the stressed syllable for Accent 1 words, while later in the word, surfacing as a second peak, in Accent 2 words. Praat was used, again, for inspecting F0.

In addition, all 986 words were tagged for lexical pitch accent category. In this very first approach, all words were simply classified as either Accent 1 or Accent 2, according to the lexical or morphological rules of Swedish. That is, it was not actually judged whether a word was, phonetically (non-focally) accented or not. However, as words may be de-accented, which is frequently the case, e.g., in the case of function words, a more detailed analyses of the data will be an important future task.

The analysis in this study is restricted to studying the distribution of head (simple and double movements) and eyebrow movements as a function of word accent type (Accent 1 vs. 2) and prosodic prominence level (focal vs. non-focal). In a first step, a co-occurrence of a focal accent and a movement was counted as such only if an annotation of both events had been made for the same word. However, annotations for focal accents on the one hand, and head- or eyebrow movements on the other, often fell on adjacent words, where it appeared obvious in many of these cases that both events relate to the same word. Therefore, in a second analysis, even such annotations of focal accents and movements on adjacent words were counted as co-occurrences.

Chi-squared tests were used in order to determine whether prominence levels (non-focal, focal) and word accent categories (Accent 1, Accent 2) have a significant effect on the distribution of head and eyebrow movements.

## Results

Results of the first and second analysis are displayed in Tables 1 and 2, respectively. Table 1 shows that more than half of the words in the corpus (514) were non-focal Accent 1 words; these include many, probably de-accented, function words. The remaining three accent

categories are about evenly distributed in the corpus (146-175 words in each category).

Table 1 further shows that eyebrow and (simple) head movements have been annotated on words of all accent categories, but movements of both types were much more frequent in focally accented words (see also Figure 1): Eyebrow movements were annotated for on average 3.6% of the non-focal words, as opposed to about 14% of the focal words; head movements were annotated for as much as half of the focally-accented words, and again, in far fewer cases for non-focal words.

This effect of the phonological prominence level (focal vs. non-focal) on the distribution of movements proved significant both for eyebrow ( $\chi^2=42.24$ ,  $p<.01$ ) and head beats ( $\chi^2=209.11$ ,  $p<.01$ ). In these chi-squared tests, samples for Accent 1 and Accent 2 were collapsed for each prominence category (non-focal, focal); a parallel set of tests was performed separately for Accent 1 and Accent 2 words, resulting in lower, but still significant  $\chi^2$  values.

Figure 2 suggests that the word accent type might have some effect on the distribution of,

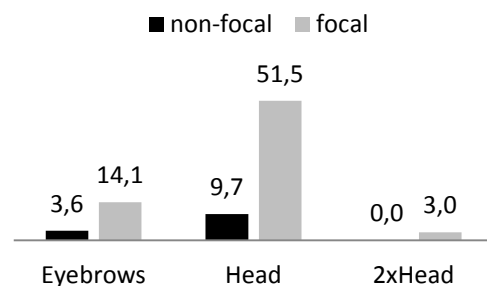


Figure 1. Distribution of eyebrow and head movements in % as a function of phonological prominence level (focal, non-focal), pooled across word accent types.

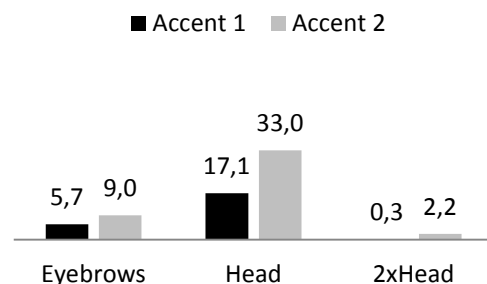


Figure 2. Distribution of eyebrow and head movements in % as a function of word accent category (Accent 1, Accent 2), pooled across prominence levels (focal, non-focal).

*Table 1. Distribution of eyebrow and head movements across a corpus of 986 words as a function of phonological prominence level (focal, non-focal) and word accent category (Accent 1, Accent 2). Co-occurrences of focal accent labels and movement labels on exactly the same word.*

Movement		Accent category				Total
		non-focal		focal		
		Accent 1	Accent 2	Accent 1	Accent 2	
Eyebrow	<b>Absent</b>	<b>498</b> (96.9%)	<b>166</b> (94.9%)	<b>129</b> (85.4%)	<b>126</b> (86.3%)	<b>919</b>
	<b>Present</b>	<b>16</b> (3.1%)	<b>9</b> (5.1%)	<b>22</b> (14.6%)	<b>20</b> (13.7%)	<b>67</b>
	Total	514	175	151	146	986
Head	<b>Absent</b>	<b>477</b> (92.8%)	<b>145</b> (82.9%)	<b>74</b> (49.0%)	<b>70</b> (47.9%)	<b>766</b>
	<b>Present</b>	<b>37</b> (7.2%)	<b>30</b> (17.1%)	<b>77</b> (51.0%)	<b>76</b> (52.1%)	<b>220</b>
	Total	514	175	151	146	986
2x Head	<b>Absent</b>	<b>514</b> (100%)	<b>175</b> (100%)	<b>149</b> (98.7%)	<b>139</b> (95.2%)	<b>977</b>
	<b>Present</b>	<b>0</b> (0.0%)	<b>0</b> (0.0%)	<b>2</b> (1.3%)	<b>7</b> (4.8%)	<b>9</b>
	Total	514	175	151	146	986

*Table 2. As Table 1, but: co-occurrences of focal accent labels and movement labels on the same word or adjacent words.*

Movement		Accent category				Total
		non-focal		focal		
		Accent 1	Accent 2	Accent 1	Accent 2	
Eyebrow	<b>Absent</b>	<b>509</b> (99.0%)	<b>171</b> (97.7%)	<b>121</b> (80.1%)	<b>118</b> (80.8%)	<b>919</b>
	<b>Present</b>	<b>5</b> (1.0%)	<b>4</b> (2.3%)	<b>30</b> (19.9%)	<b>28</b> (19.2%)	<b>67</b>
	Total	514	175	151	146	986
Head	<b>Absent</b>	<b>488</b> (94.9%)	<b>154</b> (88.0%)	<b>64</b> (42.4%)	<b>61</b> (41.8%)	<b>767</b>
	<b>Present</b>	<b>26</b> (5.1%)	<b>21</b> (12.0%)	<b>87</b> (57.6%)	<b>85</b> (58.2%)	<b>219</b>
	Total	514	175	151	146	986
2x Head	<b>Absent</b>	<b>514</b> (100%)	<b>175</b> (100%)	<b>149</b> (98.7%)	<b>139</b> (95.2%)	<b>977</b>
	<b>Present</b>	<b>0</b> (0.0%)	<b>0</b> (0.0%)	<b>2</b> (1.3%)	<b>7</b> (4.8%)	<b>9</b>
	Total	514	175	151	146	986

first and foremost, head movements. However, Table 1 shows that this effect is not very consistent, as it is mainly observed for (simple) head movements on non-focal words (and also 2xHead annotations on focal words, see below), and hardly for eyebrow movements.

Accordingly, Chi-squared tests did not reveal any effects of word accent category on eyebrow movements, neither when testing separately for non-focal and focal accents, nor for both prominence levels collapsed ( $\chi^2=3.7$ ,  $p=.052$ ). However, the last mentioned result is marginally significant. An even clearer, significant effect of word accent was revealed for head movements, both when testing for both prominence levels collapsed ( $\chi^2=31.49$ ,  $p<.01$ ), and for non-focal accents alone; however, no significant effect of word accent was found for focal accents alone.

This tendency towards fewer (head, and to some degree eyebrow) movements in non-focal Accent 1 than in non-focal Accent 2 words

(Table 1, Figure 2), can probably be explained as an artefact of the composition of the corpus: we do find some movements on both non-focal Accent 1 and non-focal Accent 2 words, which might indicate that certain non-focally, but still accented, words indeed attract movements. However, the non-focal Accent 1 sample probably contains many unaccented (function) words, which do not attract as much movement. Therefore, we find relatively fewer movements among non-focal Accent 1 than in non-focal Accent 2 words.

Turning to the ‘2x Head’ annotations, Table 1 and Figures 2-3 show that double head movements were annotated very sparsely, and only for focally accented words. Of the nine annotated items, seven are Accent 2 words. This effect of the word accent category proved significant ( $\chi^2=8.46$ ,  $p<.01$ ) with non-focal and focal words collapsed. As discussed above for simple head movements, this result could

likewise be explained by the somewhat deviating non-focal Accent 1 sample. However, traces of this effect are still seen when only focal words are included, although not reaching significance ( $\chi^2=3.04$ ,  $p=.08$ ).

Table 2 displays the results of the second analysis. As mentioned above, in a number of cases, movement annotations did not exactly coincide with words labelled as focally accented, but rather with words directly preceding or following the focal word. In the second analysis, such adjacent movements were also ascribed the focally accented word. Accordingly, as compared to Table 1, Table 2 reveals a certain shift of tokens from the ‘Absent’ to the ‘Present’ rows for the focal words, and vice versa for the non-focal words. Note that the ‘2x Head’ annotations are not affected, since these always coincided with focally accented words.

Overall, Chi-squared tests for analysis 2 did not provide any surprisingly different results than those performed for analysis 1.

## **Discussion**

The results have revealed a dependency of the distribution of eyebrow and head movements on the one hand and focal accents on the other, confirming one part of our hypothesis. Furthermore, no (strong) effect of the word accent type on eyebrow and head movements was found; the effect found can probably be explained as an artifact of the composition of the database (see above), which means that the second part of the hypothesis is largely confirmed.

However, there were traces of an effect of the word accent type on the annotations of ‘double’ head movements. This effect might be explained as follows: Only nine words in the entire corpus were annotated with a double head movement (to be compared with 220 annotations of simple head movements), of which seven were Accent 2 words, all of which were compounds. Compounds are characterized by a complex lexical stress pattern, comprising a main and a secondary stress. In addition, focal Accent 2 words are produced with two pitch peaks. If this effect were corroborated by additional data in future studies, it could imply an association of a head movement, if it is used to add prosodic prominence to a word, with a linguistic/phonetic prominence. Possibly, it requires a lexically stressed syllable to associate with, in a similar manner as is known for accentual tones. A similarly “linguistic” behavior does

not seem to be evidenced for eyebrow movements.

A way of interpreting the results by Swerts and Krahmer (2010) is that head movements and eyebrow movements have quite equivalent, cumulative functions as building blocks of prominence, as each of them seems to mirror a minor degree of prominence, while their combination adds up to a higher degree of prominence. In our study, about equally many head movements were annotated (229 of 986 words, Head and 2xHead annotations collapsed) as in the Dutch data in Swerts and Krahmer (228 of 985 words). However, we annotated far fewer eyebrow movements (67 vs. 303). That is, we annotated about two eyebrow movements, on average, per piece of news. This suggests that eyebrow movements were used rather sparsely by the speakers, presumably mostly restricted to words representing the (absolutely) most important information.

This (tentative) conclusion on eyebrow movements, in combination with our (tentative) conclusion drawn above on the “linguistic behavior” of head movements, as well as their relatively frequent occurrence, might suggest that head and eyebrow movements can represent two quite different modalities of prominence cuing, both from a formal and a functional point of view. That is, head movements might have a closer association to low-level prominence and phonological prosodic structure, while eyebrow movements might be more restricted to higher-level prominence and information-structure coding.

The present analysis has provided a preliminary, but significant insight into the usage of head and eyebrow movements in Swedish news-readers. However, it will need to be further developed, for example by means of an additional classification of the ‘non-focal’ words into phonetically de-accented and accented words. What we have also neglected so far are co-occurrences of head and eyebrow movements. Future studies will also need to incorporate an analysis of the information-structural conditions underlying the distributions of the movements, as well as a phonetic analysis of the focal and non-focal accents.

A further question for future studies (using the present and additional materials) concerns the timing of movements and focal accents. As an informal note, we have observed a number of instances of very (phonetically and visually) prominent words, which also seem to represent

the informational focus of the message, and which were associated with both a head and an eyebrow movement. In these cases, eyebrow movements (usually raises) often seem to precede the head movement. This could indicate that eyebrow movements can function as a kind of upbeat for a (multimodal) prosodic prominence, which would imply further evidence for the claim that eyebrow and head movements do not simply represent cumulative functions.

## Acknowledgements

This work was supported by a grant to the first author from the Marcus and Amalia Wallenberg Foundation.

## References

- Boersma P, Weenink D (2014). Praat: doing phonetics by computer [Computer program]. <http://www.praat.org/>
- Biau E, Soto-Faraco S (2013). Beat gestures modulate auditory integration in speech perception. *Brain & Language*, 124: 143-152.
- Bruce G (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.
- Bruce G (2007). Components of a prosodic typology of Swedish intonation. In: Riad T & Gussenhoven C, eds, *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*. Berlin, 113-146.
- Casasanto D (2013). Gesture and language processing. In: Pashler H, ed, *Encyclopedia of the mind*. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage, 372-374.
- 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: 542-554.
- Lucero C, Zaharchuk H, Casasanto D (2014). Beat gestures facilitate speech production. *Proc. of the 36th Annual Conference of the Cognitive Science Society*, Austin, TX, 898-903.
- Riad T (2006). Scandinavian accent typology. In: Viberg Å, ed, *Special issue on Swedish. Sprachtypologie und Universalienforschung (STUF)*, 59: 36-55.
- Swerts M, Kraemer E (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, 38: 197-206.
- Wang L, Chu M (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, 51: 2847-2855.
- Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H (2006). ELAN: a professional framework for multimodality research. *Proc. of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. See also: <http://tla.mpi.nl/tools/tla-tools/elan/>
- Yasinnik Y, Renwick M, Shattuck-Hufnagel S (2004). The timing of speech-accompanied gestures with respect to prosody. *Proc. of From Sound to Sense*, MIT, Cambridge, MA, 97-102.