



Head beats as pitch-accompanying visual correlates of primary and secondary lexical stress: evidence from Stockholm Swedish compounds

Anneliese Kelterer¹, Gilbert Ambrazaitis^{1,2}, David House³

¹Lund University, Sweden; ²Linnaeus University, Växjö, Sweden

³KTH (Royal Institute of Technology), Stockholm, Sweden

anneliese.kelterer.722@student.lu.se, Gilbert.Ambrazaitis@lnu.se, davidh@speech.kth.se

Abstract

This study examines the interplay of (verbal) prosody with (visual) head and eyebrow movements in a 24-minute corpus of Swedish television news readings. The paper focuses on ‘double’ beat gestures, asking whether their occurrence relates to a word’s lexical prominence structure (simplex; compound), to lexical tonal prosody (Accent 1; Accent 2), or rather to prominence levels (+/- focal accent; +/- nuclear position).

The results suggest that double eyebrow beats are a marginal phenomenon. Double head beats are also rare (only 28 of the 688 words annotated for head beats in our 4088-word corpus), but their usage follows a clear pattern: There is no preference for the nuclear position, but a strong preference to occur on a focally-accented compound (Accent 2), which is usually realized with two pitch peaks. In conjunction with previous findings on (single) head beats, the present results suggest that a head beat in this type of data can associate with lexical (primary or secondary) stress in case the stressed syllable is also marked by a (tonal or intonational) pitch peak.

Index Terms: audio-visual prosody, pitch accent, gesture, multimodality, prominence, nuclear accent

1. Introduction and background

A growing body of evidence on audio-visual, or multimodal prosody suggests that elements of verbal prominence, such as pitch accents, are frequently accompanied by visual beat gestures, i.e. rapid movements of a hand, the head, or an eyebrow, e.g. [1]–[15]. Occasional ‘double’ beats by the head were observed in a recent, preliminary study on head and eyebrow movements in Swedish television news readings [13]. The results of that study suggest that double head beats might be restricted to Accent 2 words, and more specifically to compounds, which contain two lexical stresses in Swedish. If corroborated, such findings might suggest an association of head beats to lexical stress.

Using a crucially enlarged corpus, the present study further explores the occurrence of such complex beats and whether it relates to a word’s basic prominence structure (one vs. two lexical stresses), to lexical tonal prosody (Accent 1 vs. Accent 2), or rather to prominence level (\pm sentence accent; \pm nuclear position).

1.1. Stress and tonal prosody in Stockholm Swedish

Swedish is a pitch accent language, where each prosodic word carries one of two so-called *word accents*, Accent 1 or Accent 2 [16]. Phonologically, the distinction can be conceived of as a timing distinction, where Accent 1 is modelled as an early fall (H+L*) and Accent 2 as a late fall

(H*+L) in the stressed syllable in the Stockholm variety [17]. Monosyllabic words can only have Accent 1. That is, the distinction between Accent 1 and Accent 2 only applies to words with at least one post-tonic syllable; i.e. the minimal structure required for Accent 2 is a disyllabic word with initial stress. The word accents are assigned according to phonological and morphological rules and do not reflect different prominence levels [18]. In many varieties including Stockholm, however, we can distinguish between two phonological tonal prominence levels which are orthogonal to the word accent contrast: a lower level which has been referred to as ‘non-focal’ [19] or recently as ‘small’ accent [20], and a higher level, referred to as sentence accent [17], focal accent [19], or recently as ‘big’ accent [20]. The focal or big accents arise from an additional H-tone that is realized in sequence with the basic H+L pattern of the small accents. Focal Accent 1 (H+L* H) typically surfaces as a single peak in the stressed syllable. Focal Accent 2 (H*+L H) retains its first peak and surfaces with two peaks (Figure 1), where the second peak falls into the post-tonic syllable in simplex words, and in the secondary stressed syllable in compounds.

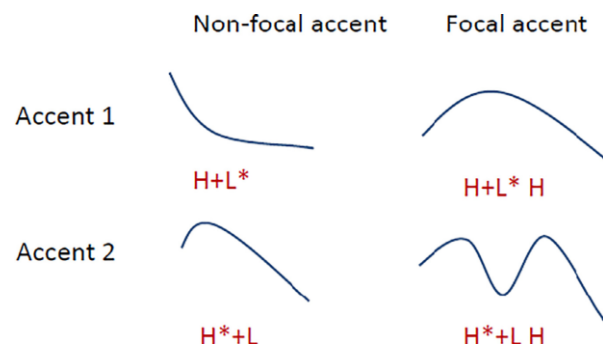


Figure 1. Schematic representation of the non-focal (or ‘small’) and focal (or ‘big’) realization of Accent 1 and Accent 2, respectively (taken from [14]).

In addition to the tonal word accent distinction, Swedish exhibits variable lexical stress. A morphologically simplex word has one lexically specified stressed syllable, while a compound word usually has two (irrespective of its number of elements), a primary and a secondary one. Word accent is phonologically associated with (primary) stress. Compounds are as a rule assigned Accent 2. However, there are a few lexicalized compounds that exhibit a simplex-word stress pattern (only one stress) along with Accent 1, such as *trädgård* ‘garden’ (literally ‘tree-yard’) or the days of the week, e.g. *fredag* ‘Friday’ [21]. In addition, there are some derivational suffixes such as *-bar* which introduce a secondary stress and thereby generate formal simplex words that behave,

prosodically, like compounds, e.g. *'under, bar* 'wonderful'. Finally, there are some prefixes that have a similar effect, taking primary stress and leaving secondary stress on the word stem, likewise resulting in a prosodic compound, e.g. *or-* in *'or, sak* 'reason'.

Finally, there are indications that the nuclear position is a relevant concept in Swedish, similar to intonation languages [22].

1.2. Audio-visual integration of prominence

During the last two decades, research on the connection of gestures with speech and prosody in particular has become increasingly vivid [1]-[15]. These studies have shown that prosodic prominence is frequently accompanied by so-called beat gestures, which are understood as rapid movements whose function is related to rhythm or prominence, rather than to encode a meaning. They are thus different from other types of gestures such as deictic gestures, iconic gestures, or emblems [23]. Beats can be produced by different parts of the body, e.g. hands, head or eyebrows.

Beats from different modalities (such as head and eyebrow) can either occur on their own or they can form a cluster of visual markers. In such cases, they seem to pattern in certain ways with respect to their temporal alignment with the pitch accent or the stressed syllable, see [2], [5], [7], [9], [24] and [25]. For instance, it has been shown that head beats are usually aligned with the stressed syllable [26], whereas eyebrow beats tend to align *before* the head beat [25].

There is also evidence that the perceptual strength of a verbal prominence is a good predictor for the occurrence of head and eyebrow beats, as well as their clustering: moderately prominent words have been shown to more likely coincide with a beat (head or eyebrow but seldom both) than non-prominent words, while only strongly prominent words were likely to coincide with a head-eyebrow cluster [6]. Such a cumulative interplay of beat gestures with verbal prominence has been partly corroborated by acoustic data for Swedish [15], although head and eyebrow beats also appear to exhibit a cue trading relation [12].

1.2.1. Differential usages of head and eyebrow movements

Previous findings on different languages and genres are, however, not always converging when it comes to the communicative functions fulfilled by head and eyebrow movements, respectively. The two visual modalities might fulfil equivalent functions in the sense that either the one or the other can be used as a visual cue to prominence [6]. Alternatively, they might assume crucially different roles in multimodal prominence coding, as has been suggested for Swedish news readings [14].

First, head beats have been found to occur very frequently in news readings (23% and 24% of all words in a Dutch [6] and a Swedish corpus [14], respectively), and in the Swedish case, much more frequently than eyebrow beats (7% of words). Although their distribution within a given read text has been argued to encode aspects of information structure, their density suggests that they do not only occur on individual words that are marked as important in the discourse [14]. It would appear that head beats cooperate with focal accents in Swedish news readings in the construction of a sentence's basic rhythmical structure.

In contrast, eyebrow movements were found to be used more in connection with specific, emotionally loaded words, or connected to information structure, e.g. marking contrastive focus ([5], [14]), relating more to higher-level, i.e. sentence- or discourse-level, highlighting.

These interpretations of the distribution or usage of head vs. eyebrow movements appear to be well in line with the findings on temporal alignment reported in 1.2 above: head beats align in time with a stressed syllable, and they are used in a similar manner as sentence-level pitch accents. Together, these findings suggest that head beats, but not generally eyebrow beats, associate with the stressed syllable in a similar manner as pitch accents do.

Another piece of evidence in favour of this tentative conclusion comes from a previous study of a smaller corpus of Swedish news readings [13]. This study reported occasional 'double' head beats – but no double eyebrow beats – with reference to a word. These double head beats predominantly occurred in compounds, i.e. words that have two stresses (cf. 1.1), which again suggests that head beats might associate with lexical (although either primary or secondary) stress.

1.3. Research question

The observations concerning double head beats reported in 1.2.1 are rather preliminary since only 9 were detected (of which 7 were found in compounds) in a corpus of about 1000 words of news speech [13]. Therefore, the present study further investigates the differential nature of head and eyebrow beats (cf. 1.2.1) by examining the occurrence of 'double' beats in an enlarged corpus of Swedish news readings.

Given the cumulative prominence function of head and eyebrow movements (cf. 1.2), an alternative hypothesis for the usage of double beats could be that an increased complexity of the visual cue is a correlate of a higher level of prominence.

In order to solve this issue, we examine patterns of occurrence of complex (i.e. 'double') beats asking whether they are conditioned by the lexical stress pattern of a word (simplex; compound), its lexical tonal category (Accent 1; Accent 2), by a focal accent or the nuclear position.

2. Method

2.1. Materials

The corpus used in this study consists of 142 short news-readings from Swedish TV (SVT Rapport), each one to three sentences long, comprising about 24 minutes of speech by five speakers (see Table 1). The recordings were retrieved on DVD from the National Library of Sweden (Kungliga Biblioteket). A subset of this material was used in [13] and [14].

Table 1. Number of recordings, words and total duration for each speaker.

Speaker (sex)	Recordings	Words	Duration (min:sec)
Alexander (m)	28	981	05:49
Filip (m)	34	904	04:55
Katarina (f)	18	508	03:02
Pelle (m)	26	964	06:08
Sofia (f)	36	731	04:45
Total	142	4088	24:39

2.2. Annotations

The recordings were transcribed in ELAN [27], manually segmented into words using Praat [28], and then annotated for head movements (*HB* for ‘head beat’), eyebrow movements (*EB* for ‘eyebrow beat’) and ‘focal’ accents (*FA*) in ELAN. Each word was annotated for the presence vs. absence of *HB*, *EB* and *FA*. A word was annotated for *HB* or *EB* if the head or an eyebrow moved rapidly in one direction, roughly within the time span of a word. Temporal onsets and offsets of movements were not labelled and slower movements spanning several words were ignored. In the event of two separate rapid movements within a word, such a ‘double movement’ was annotated as *2xHB* or *2xEB*, respectively. This label was introduced after a pilot annotation had revealed occasional double head movements [13].

Focal accents were annotated manually by listening and with the help of the F0 contour display in Praat, but without access to the video channel. A focal accent was annotated according to phonological criteria, i.e. if there was a H-tone recognizable in the F0 contour (cf. 1.1 and Fig. 1).

Two slightly differing annotation procedures were applied to a newer part of the corpus (111 files) and an older part (31 files that were used in previous studies, see 2.1.). For the older part, full annotations were performed by three labellers each. These three annotations were then converted to a single consensus annotation, where an annotation of e.g. *HB* was counted as such in the event of an agreement between at least two labellers. For the newer part of the corpus, two new labellers annotated either *HB* or *EB* in one portion of the files, and vice versa in the remaining files. The *EB*-labeller of a given file also annotated *FA* in the same file during a second round of annotations.

A subset of 13 files (12% of the newer part) was fully annotated by both labellers for reliability testing. Inter-rater reliability was tested by means of Cohen’s kappa [30], confirming good reliability (*FA*: $\kappa=0.77$; *HB*: $\kappa=0.69$; *EB*: $\kappa=0.72$), and by means of Fleiss’ kappa [29] for the older part of the corpus (3 labellers), again, confirming fair to good reliability (*FA*: $\kappa=0.88$; *HB*: $\kappa=0.77$; *EB*: $\kappa=0.77$).

Additionally, all words with a *2xHB* annotation were tagged for lexical pitch accent category (Accent 1; Accent 2), lexical stress pattern (simplex; compound), and position in the intonational phrase (nuclear; non-nuclear).

3. Results

In this data, 688 words were annotated for head movements (16.8%). Among these words, 28 *2xHB* annotations were found (4%), compared to 660 single *HB* annotations (96%).

Figure 2 displays an analysis of the 28 *2xHB* annotated words according to the factors word accent (Accent 1 vs. Accent 2), simplex (usually 1 stress) vs. compound word (2 stresses), focal- vs. non-focal accent, and nuclear vs. non-nuclear position. Only one of the 28 words is not focally accented and only three are Accent 1 words; *fjrtiosex* ‘forty-six’, *kriminaliseras* ‘to be outlawed’ and *förbud* ‘ban’. The one word without a focal accent is a compound (*dödsorsaken* ‘cause of death’). Of the remaining 25 Accent 2 words, 22 are compounds (79% of all words, 88% of Accent 2 words). The three simplex Accent 2 words are *berättade* ‘told’, *tredje* ‘third’ and *åtalas* ‘to be prosecuted’. The last of these, *åtalas*, is composed of *tala* ‘to tell’ (the -s suffix marks the passive) and the prefix *å-*, thereby containing two stresses (cf. 1.1).

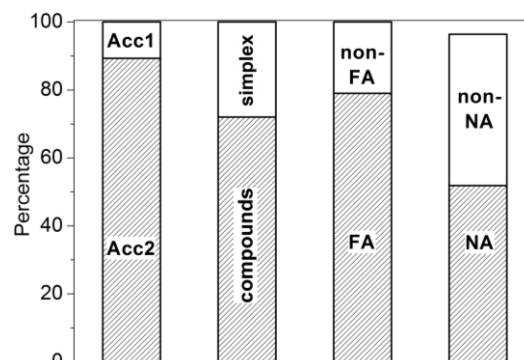


Figure 2. Percentage of Accent 2 vs. Accent 1 words, compounds vs. simplex words, focal vs. non-focal words and words in nuclear vs. non-nuclear position. $n = 28$, except for the right-most bar ($n=27$).

There is no apparent preference for nuclear accent, i.e. the last *FA* in a phrase. One of the 28 words was excluded from this categorization since it is not focally accented. Of the other 27 words, 14 are in nuclear position (52%) and 13 are not (48%). However, speakers show different preferences. Nine of Pelle’s 11 *2xHB* annotations coincide with a nuclear accent. None of Sofia’s nuclear accents carry a *2xHB* annotation. Alexander does not show any strong preference for nuclear (4 words) or non-nuclear accents (6 words). For the other speakers, there was only one instance of *2xHB* (Katarina: nuclear, Filip: non-nuclear). The nuclear accents of each speaker with the respective percentage are presented in Table 2.

Table 2. Number of focal accents and nuclear accents produced in combination with *2xHB* for each speaker.

Speaker	Focal Accents	Nuclear Accents
Alexander	10	4 (40%)
Filip	1	0 (0%)
Katarina	1	1 (100%)
Pelle	11	9 (82%)
Sofia	4	0 (0%)
Total	27	14 (52%)

Table 3. Number of tokens of the combinations of variables found with a *2xHB* annotation, excluding nuclear accent. Percentages are given in brackets.

Combination of Variables	Tokens
FA + Acc2 + Comp	21 (75%)
FA + Acc2	3 (11%)
Acc2 + Comp	1 (3%)
FA + Acc1	3 (11%)
Total	28

Thus, excluding nuclear accent, Table 3 shows how complex head beats correlate with the combination of phonological tonal prominence level (focal vs. non-focal accent), lexical stress pattern (simplex vs. compound) and word accent category (Accent 1 vs. Accent 2). It is clear that focal accent,

compound and Accent 2 is the combination receiving the majority of double head beats (75%).

Most of the words accompanied by a *2xHB* annotation are focal Accent 2 words (86%), i.e. with a two-peaked F0 contour (cf. Fig. 1). Of these 24 focal Accent 2 words, 22 have two stresses (92%), either due to compounding or to affixation of the type '*ä, talas*' (cf. 1.1.). Of all 535 focal Accent 2 words in the corpus, 286 (53%) have two stresses, and of the 205 focal Accent 2 words with a *HB* in the corpus, 103 (49%) have two stresses. This shows that two-stressed words are grossly over-represented in focal Accent 2 words associated with a complex head beat, but not in words accompanied by a single head beat or without any head movement. Therefore, in comparison to other two-peaked words, two-stressed words are much more likely to be accompanied by double head beats.

In the 142 files, 19 words were annotated for eyebrow movements (0.47%) of which only three words were annotated for *2xEB* (< 0.1%). Two of the *2xEB* words were Accent 2 words, none of which was a compound. All of them were annotated for *FA* but none for *NA*. One of them also carried a *2xHB* annotation, one a simple *HB* and one no annotation for head beat at all.

4. Discussion and conclusion

In this study, we asked whether complex ('double') beat gestures produced by the head and the eyebrows in Swedish television news readings were conditioned by either of the following prosodic features of a word: lexical stress structure (simplex; compound), lexical tonal category (Accent 1; Accent 2), phonological prominence level (focal; non-focal), or position in the intonational phrase (nuclear; non-nuclear). Although only a relatively small number of *2xHB/2xEB* tokens were observed, the results are conclusive and reliable as they are taken from a relatively large corpus.

The results show that double eyebrow beats are a marginal phenomenon in the present type of data, which we take as strong evidence for our interpretation of previous research presented in 1.2 above. That is, eyebrow beats exhibit a crucially different pattern of occurrence in contrast to head beats. For the latter, our results provide further support for the assumption of an association with lexically stressed syllables.

No correlation was observed between the occurrence of *2xHB* and nuclear position, but a strong correlation was found with phonological tonal prominence level (focal vs. non-focal accent), with lexical stress pattern (simplex vs. compound) and with word accent category (Accent 1 vs. Accent 2). These results suggest that a double beat is not merely a signal of 'stronger prominence', since we would have expected *2xHB* annotations evenly distributed across both word accents in that case, as well as for both simplex and compound words, and possibly more frequently in connection with nuclear accents. However, there is some relation to prominence, as *2xHB* annotations almost exclusively co-occur with focal accents. At the same time, *2xHB* annotations also predominantly occur on Accent 2 compound words. One possible explanation for this is that double pitch peaks in focal Accent 2 words (see Fig. 1) attract double head beats. However, simplex Accent 2 words were by far less frequent among the *2xHB* words than compounds, indicating that the important prerequisite for a *2xHB* to occur seems to be two stresses, not Accent 2. This might indicate that a head beat requires a lexically stressed syllable, be it primary or secondary stress, to associate with.

Head beats are obviously not an obligatory feature of lexical stress, but if they are added to the multimodal prosodic make-up of a word, they might require a lexical stress position to dock onto. We could think of a stressed syllable as a *beat bearing unit*, in analogy to the concept of *tone bearing unit* in intonational and tonal phonology. This interpretation is well in line with previous findings on the temporal alignment of head beats with the stressed syllable (cf. 1.2), but future research will have to examine temporal head-stress alignment in more detail, also considering secondary stress in compounds.

However, a head beat does not represent a visual cue to stress only, as almost all *2xHB* co-occurred with a focal accent. A possible explanation for this finding is that head beats align with a stressed syllable on the one hand, similar to a pitch accent, but that they generally also require a focal accent. However, results based on a subset of the present corpus clearly show that this is not a valid assumption, as single head beats frequently occur on words lacking a focal accent [14]. Nevertheless, this previous study also showed that head beats still prefer to coincide with a word that is at least non-focally accented ('small' accent, cf. 1.1), rather than with completely de-accented words [14]. This offers an alternative explanation suggesting that a head beat functions as a visual enhancement of a tonally-encoded prominence, where this prominence may well represent a low phonological level.

Our present results confirm this explanation while they also provide a refinement: A focal Accent 2 word is realized with two pitch peaks, and in the case of compounds, the two peaks are aligned with the primary and secondary stressed syllable, respectively. Accordingly, each head beat is associated with a tonally-encoded prominence, although the tones involved relate to different phonological prominence levels, i.e. that of stress and that of sentence accent (cf. 1.1). Thus, the phonological nature of the tone involved appears to be irrelevant; the crucial aspect is that it causes a pitch peak.

For Stockholm Swedish news readings, we can draw the following overall conclusion from the present findings in conjunction with previous results [14]: If a head beat is to be applied in the prosodic realization of a given word, it requires a lexically stressed syllable (primary or secondary) to associate with, which is phonetically made prominent by means of a pitch peak. We might therefore characterize head beats as optional pitch-accompanying visual correlates of primary or secondary lexical stress.

Previous studies have shown that beats from different modalities can have substantially different functions. However, since results from studies on different languages and genres have not always concurred in this respect, differential usages of head and eyebrow beats should be classified as a potential rather than a constraint. This study has refined this emerging understanding of the interplay of verbal and visual prominence cues, suggesting that head beats can be integrated in prosodic prominence at a basic phonological level.

5. Acknowledgments

We retrieved materials from the National Library of Sweden and received permissions from Swedish Television. We also thank our research assistants Malin Svensson Lundmark and Otto Ewald for assistance with data processing and annotations. This work was supported by the Marcus and Amalia Wallenberg Foundation [MAW 2012.01.03] and the Swedish Research Council [VR 2017-02140].

6. References

- [1] D. House, J. Beskow and B. Granström, "Timing and Interaction of Visual Cues for Prominence in Audiovisual Speech Perception," *Proceedings of Eurospeech 2001*, Denmark, Aalborg, pp. 387-390, 2001
- [2] Y. Yasinik, M. Renwick and S. Shattuck-Hufnagel, "The timing of speech-accompanying gestures with respect to prosody," *Proceedings of From Sound to Sense*, MIT, Cambridge, MA, pp. 97-102, 2004
- [3] S. Jannedy and N. Mendoza-Denton, "Structuring information through gesture and intonation," *Interdisciplinary Studies on Information Structure*, vol. 3, pp. 199-244, 2005
- [4] D. McNeill, *Gesture and Thought*. Chicago: University of Chicago Press, 2005
- [5] Flecha-García, "Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English," *Speech Communication*, vol. 52, pp. 542-554, 2010
- [6] M. Swerts and E. Krahmer, "Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions," *Journal of Phonetics*, vol. 38, no. 2, pp. 197-206, 2010
- [7] T. Leonard and F. Cummins, "The temporal relation between beat gestures and speech," *Language and Cognitive Processes*, vol. 26, pp. 1457-1471, 2011.
- [8] D. Loehr, "Temporal, structural, and pragmatic synchrony between intonation and gesture," *Laboratory Phonology*, vol. 3, no. 1, pp. 71-89, 2012
- [9] N. Esteve-Gibert and P. Prieto, "Prosodic structure shapes the temporal realization of intonation and manual gesture movements," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 3, pp. 850-864, 2013
- [10] E. Biau and S. Soto-Faraco, "Beat gestures modulate auditory integration in speech perception," *Brain and Language*, vol. 124, no. 2, pp. 143-152, 2013
- [11] P. Wagner, Z. Malisz and S. Kopp, "Gesture and speech in interaction: An overview," *Speech Communication*, vol. 57, pp. 209-232, 2014
- [12] P. Prieto, C. Pugliesi, J. Borràs-Comes, E. Arroyo and J. Blat, "Exploring the contribution of prosody and gesture to the perception of focus using an animated agent," *Journal of Phonetics*, vol. 49, pp. 41-54, 2015
- [13] G. Ambrazaitis, M. Svensson Lundmark and D. House, "Multimodal levels of prominence: a preliminary analysis of head and eyebrow movements in Swedish news broadcasts," *Proceedings from Fonetik 2015, Lund, June 8-10*, pp. 11-16, 2015
- [14] G. Ambrazaitis and D. House, "Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings," *Speech Communication*, vol. 95, pp. 100-113, 2017
- [15] G. Ambrazaitis and D. House, "Acoustic features of multimodal prominences: Do visual beat gestures affect verbal pitch accent realization?" *Proceedings of The 14th International Conference on Auditory-Visual Speech Processing (AVSP2017)*, 2017
- [16] T. Riad, "Scandinavian accent typology," *Sprachtypologie und Universalienforschung (STUF)*, Berlin, vol. 59, pp. 36-55, 2006
- [17] G. Bruce, *Swedish Word Accents in Sentence Perspective*. Lund: Glerup, 1977
- [18] G. Bruce, "Intonational Prominence in Varieties of Swedish Revisited," S. Jun (ed.), *The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press, pp. 410-429, 2005
- [19] G. Bruce, "Components of a prosodic typology of Swedish intonation," T. Riad and C. Gussenhoven (eds.), *Tones and Tunes - Volume I: Typological Studies in Word and Sentence Prosody*, Berlin; New York: Mouton de Gruyter, pp. 113-146, 2007
- [20] S. Myrberg and T. Riad, "The prosodic hierarchy of Swedish," *Nordic Journal of Linguistics*, vol. 23, no. 2, pp. 115-147, 2015
- [21] T. Riad, *Prosodi i svenskans ordbildning och ordböjning*. Manuscript, 2009 (retrieved from http://www.su.se/polopoly_fs/1.29929.1320939953!/ProsodiOrd bildningRiadNov12.pdf)
- [22] G. Ambrazaitis, *Nuclear Intonation in Swedish - Evidence from Experimental-Phonetic Studies and a Comparison with German*. Travaux de l'institut de linguistique de Lund, vol. 49, 2009
- [23] D. McNeill, "So You Think Gestures Are Nonverbal?" *Psychological Review*, vol. 92, no. 3, pp. 350-371, 1985
- [24] K. Bergmann, V. Aksu and S. Kopp, "The Relation of Speech and Gestures: Temporal Synchrony Follows Semantic Synchrony," *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction*. Bielefeld, Germany, 2011
- [25] D. House, G. Ambrazaitis, S. Alexanderson, O. Ewald & A. Kelterer, "Temporal organization of eyebrow beats, head beats and syllables in multimodal signalling of prominence," *International Conference on Multimodal Communication: Developing New Theories and Methods*, Osnabrück, Germany, Accepted for publication
- [26] D. House, S. Alexanderson and J. Beskow, "On the temporal domain of co-speech gestures: syllable, phrase or talk spurt?" *Proceedings from Fonetik 2015, Lund, June 8-10*, pp. 63-68, 2015
- [27] H. Sloetjes and P. Wittenburg, "Annotation by category - ELAN and ISO DCR," *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*, 2008P.
- [28] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott International*, vol. 5, no. 9/10, pp. 341-345, 2001
- [29] J. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378-382, 1971
- [30] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37-46, 1960