Compounds in interaction: patterns of synchronization between manual gestures and lexically stressed syllables in spontaneous Swedish

Gilbert Ambrazaitis[1]*, Margaret Zellers*[2]*, & David House*[3]

[1] Linnaeus University, Växjö, Sweden
[2]Kiel University, Germany
[3]KTH (Royal Institute of Technology), Stockholm, Sweden

gilbert.ambrazaitis@lnu.se, mzellers@isfas.uni-kiel.de, davidh@speech.kth.se

Abstract

Prosody and gesture share common functions in communication, one of which is the signaling of prominence. However, we still know relatively little about how these two domains are coordinated with one another. The current study explores timing relationships between hand gestures and stressed syllables in Swedish compounds. Compounds in Swedish usually have two stressed syllables, enabling us to investigate possible differences in alignment between gestures and stressed syllables as a function of their phonological status (primary, secondary stress). We find a tendency for stressed syllables to be accompanied by movement phases of gestures, with the primary stressed syllable somewhat more likely to arise with a movement phase than the secondary stressed syllable. The stressed syllables also show tendencies for close temporal alignment with transitions between one gesture phase to another, though not all gesture phase types may start in alignment with the second stressed syllable. Our data thus provide additional support for the common function of prosody and gesture in signaling prominence in spoken communication and indicate that transitions between gesture phases as well as the gestures themselves may contribute to prominence marking.

Index Terms: prosody, hand gesture, prominence, temporal alignment, Swedish, compounds

1. Introduction

Speech prosody and gesture share a number of common functions [1], and it has been argued that the two should be treated as parts of the same system in terms of their linguistic functions [2]. One of these common functions is the signaling of prominence. In fact, gestures have been found to be synchronized with elements of prosodic prominence (e.g., stressed syllables or pitch accents) [3][4][5][6][7][8][9] and it has been argued that gestures in the visual modality can add to perceived prominence [10][11][12][13][14][15]. However, gestures are multifunctional, just as speech prosody is, and we still know too little about the constraints that prominence structures (both phonetic and phonological) may put on patterns of gestural movement or alignment. For instance, how well are different types of gestures or gesture phases aligned with units such as stressed syllables? And does the lexical phonological level of prominence play a role in the planning and execution of gestures? The existence of such a role would be a possible prediction under the assumption that speech and gesture production arise from a common generation process [16][17].

This paper aims to approach these questions by means of studying the synchronization of manual gestures and stressed syllables in compound words in spontaneous Swedish dialogue. Compounds make a particularly good test case, since in Swedish, a compound comprises two lexical stresses – a primary and a secondary one – which are clearly distinguished phonologically with respect to how they associate with tones (see 1.1). The question we ask in this paper is whether both stresses also display evidence of association with gestures, and whether and how they differ in this respect: Are different gesture types or gesture phases more regularly synchronized with the primary vs. the secondary stress? We also discuss potential functions of such alignment. For instance, is alignment *per se* prominence-lending?

1.1. Compounds in Swedish

Compounds can be defined as words that are formed from other lexical elements, either full words or stems [18]. In Swedish, compounds usually have two lexically specified stressed syllables (a primary and a secondary one), while simplex words have only one. The two stresses in a compound are located in the same syllables as in the underlying simple words or morphemes, and the primary stress is usually found in the first element. For instance, the word *konferensartikel* 'conference paper', derived from *konfe'rens* and *ar'tikel*, receives a compound stress pattern according to *konfe'rensar¸tikel*.

In addition to lexical stress, Swedish exhibits a binary tonal word-accent distinction (Accent I, Accent II). Phonologically, the distinction can be conceived of as a timing distinction, where Accent I is modelled as an early fall in pitch (H+L*) and Accent II as a late fall (H*+L) in the Stockholm variety [19][20]. Swedish word accents are assigned according to phonological and morphological rules, where compound words receive Accent II. Sentence-level prominence, in Stockholm Swedish, is typically encoded by means of an additional rise in pitch (modeled as a high tonal target, H), following the word-accent fall. In Accent II, this results in a two-peaked pitch pattern, with the accentual fall aligned with the primary stress (H*+L H). In compounds, the double-peak pattern of Accent II also requires a double association with the two stressed syllables [21].

However, there are a few lexicalized compounds that exhibit a simplex-word stress pattern (only one stress) along with Accent I, such as *'blåbär* 'blueberry', *'trädgård* 'garden' (literally 'tree-yard') and the days of the week, e.g. *'fredag* 'Friday' [22]. Finally, 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. *ˈunderˌbar* 'wonderful'. In addition, there are some prefixes that have a similar effect, as they take the primary stress and leave a secondary stress on the word stem, likewise resulting in a prosodic compound, e.g. *or-* in *ˈorˌsak* 'reason'.

1.2. Alignment between prosody and gesture

Many studies have reported patterns of alignment or overlap between pitch accents or stressed syllables on the one hand, and gestures – manual, as well as head or eyebrow movements – on the other [3][4][5][6][7][8][9][23]. For instance, eyebrow raises have been found to precede pitch accents by on average 60 ms in a corpus of English face-to-face dialogue [23]. Furthermore, the apices of gesture strokes have been shown to align with pitch accents [5][6]. Temporal alignment between gesture and prosody can thus probably be regarded as a basic mechanism of spoken language production, which indeed has been shown to be acquired as early as at the age of 16-18 months [24].

A question that suggests itself in the light of such findings is whether the temporal alignment between gesture and prosody is a mere effect of a common generation process, or whether patterns of synchrony and alignment themselves can also fulfil communicative functions, such as signaling prominence. As mentioned above, prominence is an established common function of gesture and prosody, and is in fact often produced by simultaneous audio-visual cues, as for instance shown in [4] and [9]. However, it is still an open question how exactly gestures contribute to prominence, or in other words, which characteristics of gestures are prominence lending. It could, for instance, be the mere co-occurrence of certain types of gestures that adds to prominence, but possibly also the alignment patterns between gestures and stressed or accented syllables. Traditionally, the prominence function has been associated with so-called "beat gestures" [17][25]. However, it is now widely agreed that, rather than distinguishing between gesture types, gestures should be understood in terms of dimensions. That is, a gesture can be, say, iconic, and have a prominence-lending function at the same time [25]. In this paper, we do not explicitly address the communicative functions of temporal alignment, but we discuss the potential relevance of gestural phase transitions and their alignment with stressed syllables for the signaling of prominence.

So far, previous studies have only implicitly suggested a link between gestures and prosodic phonology, as gesture strokes have been shown to associate and temporally align with lexically stressed syllables, for instance [8]. However, in our recent studies on head movements in read speech (television news), we have observed occasional instances of compound words that were produced with two distinguished head movements, with, presumably, a prominence-lending function (i.e., beats) [9][26]. With very few exceptions, such 'double head beats' only occurred in connection with words with two lexical stresses (i.e. formal compounds or certain derivations, see 1.1), and only if associated with a two-peaked pitch accent (1.1). A conclusion drawn from these results was that head beats seem to require a lexical stress, either primary or secondary, to associate with, as well as an F0-peak realized on that syllable. These findings suggest that even syllables with secondary stress are units with which gestures may be associated. However, this does not necessarily indicate that primary and secondary stresses would behave differently in this respect. Furthermore, to the best of our knowledge, we still lack data on the temporal alignment between gestures (be it manual or head gestures) and stresses, comparing primary and secondary stressed syllables.

1.3. Research question

This is an exploratory study asking whether and how manual gestures are coordinated with stressed syllables in Swedish compounds. More specifically, we compare primary with secondary stressed syllables and ask whether both types of stressed syllables are equally accessible as sites of synchronization with gestures.

2. Method

We used selected dialogues from the Spontal Corpus to study the synchronization between gesture phases and stressed syllables in compound words. To this end, manual gestures were labelled, and all compounds in the selected material were identified and segmented acoustically. Synchronization was studied in two steps: First, we counted overlaps of gesture phases with the primary and the secondary syllable. Second, we measured the temporal alignment (in seconds) between the onset of a primary or secondary stressed syllable and the nearest gesture phase transition.

2.1. Corpus

Data were drawn from Spontal [27], which comprises audio, video, and motion capture data for spontaneous, unrestricted two-party conversations in Swedish; the current study makes use of the audio and video data. To place a minimum structure on the dialogues, the recordings were formally divided into three ten-minute blocks, although the conversation was allowed to continue seamlessly over the blocks, with the exception that participants were informed, briefly, about the time after each ten-minute block. After 20 minutes, they were also asked to open a wooden box which had been placed on the floor beneath them prior to the recording. The box contained objects whose identity or function was not immediately obvious. The subjects could hold, examine and discuss the objects taken from the box, but they could also choose to continue whatever discussion they were engaged in or talk about something entirely different. In total, Spontal contains more than 60 hours of unrestricted conversation in over 120 dialogues.

Six approximately five-minute chunks of conversation which had previously been annotated for hand gestures were used as the basis for the study. All of the chunks were taken from the first 20 minutes of each conversation and did not include passages in which the participants manipulated the box or its contents. Different pairs of participants occurred in the chunks, with the exception of two chunks which featured the same pair of participants. Thus, our corpus comprises speech from ten (rather than twelve) speakers (five female, five male). Two of the authors listened to the conversations without watching the accompanying video and identified compounds. Compounds were defined as any content word which could be transparently seen to have been built from two smaller simplex words, regardless of whether the compound was apparently constructed in the moment by the speaker (e.g. *ˈpocketˌdjup*, lit. 'paperback book deep', an adjective describing a shelf of the right depth to hold a paperback book) or was commonly in use (e.g. *ˈVasaˌstan*, the name of a neighborhood in Stockholm). In total, 122 compounds were identified across the five conversations.

2.2. Annotations

2.2.1. Gesture annotations

The gesture labelling was carried out by annotators in ELAN [28] using only the video, so that they could not be influenced in their interpretations by what was simultaneously being spoken. Hand gesture phrases and phases were identified in two passes, following McNeill [29] and Kendon [16]. A gesture phrase started when the hand(s) moved away from a rest position into the gesture space in front of the gesturer's torso, and lasted until the hand(s) returned to the rest position. Once the gesture phrases were in place, the second annotation pass was used to divide them into phases. The phases, following McNeill, were:

- Preparation: hand moves from rest position to start of gestural movement
- Stroke: the "meaningful" part of the gesture; normally a movement across or within the gesture space
- Hold: hand is held in one location preceding or following a stroke
- Retraction: hand moves from gesture space to rest position

Of the gesture phases, only Strokes are obligatory; all other phases are considered to be optional. Hand gestures were identified for each hand separately, but were partially merged in the analysis, so no distinction is made between which hand is used for a particular gesture.

Beat gestures can be difficult to classify using the traditional phases, since they sometimes only involve a biphasic movement. In these cases, the gestures are labelled based on the relationship of the movement direction to the location of the visually prominent beat:

- Toward: hand moves towards the location of the rhythmic beat; thus, the perceived timing of the beat is the end of the Toward phase
- Away: hand moves away from the location of the rhythmic beat; either before the Toward phase in preparation for it, or after the Toward phase as a retraction

2.2.2. Acoustic segmentation

The compounds were segmented acoustically in Praat [30], based primarily on visual inspection of the spectrogram and the waveform, and secondarily on the audio for the purpose of validation. Segmentation was performed at the phoneme level as far as possible (strongly coarticulated stretches without clear phoneme boundaries were subsumed in one segment), and at the syllable level on a separate tier. For the present study, only syllable and word onset boundaries were used. (Word onset boundaries are most often, but not always, identical to primary stress syllable onsets.) Acoustic syllable onset boundaries were usually clearly identifiable in the spectrogram. The few exceptions were, coincidently, mostly words that did not overlap with a gesture and were thus excluded from the alignment measures anyway; one of the problematic cases (a syllable starting with a voiceless stop after a pause) was among the words with a gestural overlap, and this boundary was excluded from the temporal alignment measurements.

2.3. Analysis

Segmentations made in Praat [30] were imported into the gesture annotation file in ELAN [28], and this combined annotation file was exported for analysis. The data were analyzed in two steps. First, we explored patterns of overlap between stressed syllables and gesture phases. To this end, we first identified the compounds that displayed at least partial overlap with a gesture, according to our annotations (57 out of 122, where overlap was minimal in one case, only affecting the word-initial consonant in an unstressed syllable: *kera*ˈ*mikˌkurs* 'pottery course'). We then qualitatively inspected the configurations of gesture phase labels that overlapped with the word, and classified them into a smaller number of categories. This was done in order to identify general, recurrent patterns. The categorization focused on whether or not either the primary (henceforth, s_1), or the secondary stress (s_2) syllable was co-occurring with a movement phase, assuming that movement can potentially serve as a visual prominence cue. For this categorization, we counted Preparations, but not Retractions, as potentially relevant 'movements' (besides Strokes and the elements of beats, i.e. Away and Toward phases). This was motivated by informal inspections of the video, which revealed that Preparations might well be perceived as prominence-related in some cases. Note, however, that this analysis is exploratory in nature and does not make any a priori claims as to whether gesture phase types are prominence-lending or not.

Table 1: *Gesture phase patterns co-occurring with compounds (number of occurrence). s_1= first (and primary) stressed syllable; s_2=second (and secondary) stressed syllable.*

Gesture phases	**Count**
Single phase overlapping with both s1 and s2	
Retraction	6
Hold	3
Hold + stroke (different hands)	2
Stroke (variations*)	19
One phase of a beat	2
	32
	(57.1%)
Different phases overlapping with s1 and s2	
*Movement in **s_1** and **s_2***	
s_1: str. → ***s_2***: str.+hold (or hold+str. → str.)	4
s_1: prep. → ***s_2***: str. (+retr.)	2
Ongoing beats	4
Movement in s1	
s_1: (prep.+) stroke+hold → ***s_2***: hold (+retr.)	3
s_1: str. (+retr.) → ***s_2***: retr. or no gesture	5
s_1: beat phase → ***s_2***: hold or no gesture	3
s_1: prep. → ***s_2***: hold (+retr.)	1
*Movement in **s_2***	
s_1: hold → ***s_2***: str. (+retr.)	2
	24
	(42.9%)
Total	**56**

Second, we identified the nearest gesture phase transition for each s_1 and s_2. This was done for all syllables that were overlapping with a gesture: Among the 56 compounds with gestural overlap, 100% of the primary stressed and 89% (50 of 56) of the secondary stresses overlapped with a gesture. Based on the annotation files, we then calculated the temporal distance (in seconds) between syllable onset and nearest phase transition for each stressed syllable. The aim of this analysis was to explore whether it was possible to find evidence of synchronization between syllables and gesture onsets, and whether phase types, on the one hand, and primary and secondary stressed syllables, on the other hand, differ in this respect.

3. Results

Table 1 shows the distribution of movement phases of gestures arising with compounds. In total, 56 of the 122 compounds were accompanied by hand gestures. The gesture phases could be organized in several ways in terms of their coordination with the stressed syllables. First, a single gesture phase could be ongoing during both of the stressed syllables; this arose in 32 cases, as shown in the top portion of Table 1. Alternatively, the two stressed syllables could be accompanied by different gesture phases: either both syllables could be accompanied by movement phases (Preparation, Stroke, or beats), or one syllable was accompanied by a movement phase and the other by a non-movement phase (Hold) or the end of the gesture (Retraction or no gesture).

In the majority of cases (42 of 56), both stressed syllables were accompanied by movement phases in the hand gestures. When only one syllable was accompanied by a movement phase, however, this was almost always the first stressed syllable; in only two cases was the second stressed syllable accompanied by a movement phase when the first stressed syllable was not.

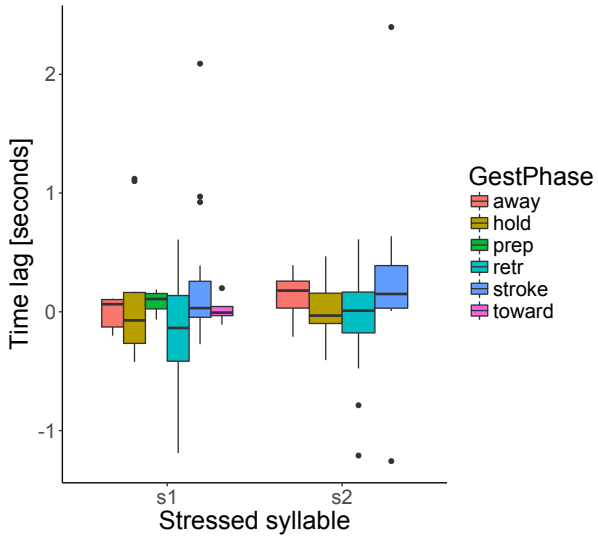

Figure 1: *Boxplots displaying results of alignment measurements between onsets of stressed syllables (primary = s1, secondary = s2) and the nearest gesture phase transition, per gesture phase type. GestPhase= Type of gesture phase (that starts nearest to syllable onset); Time lag = Time from gesture phase onset to syllable onset (negative value = syllable starts before gesture phase).*

Figure 1 shows the temporal relationship between the onset of the first and the second stressed syllable, respectively, compared to the onset of the gesture phase that was the nearest to the syllable onset, whether that gesture phase began before or after the stressed syllable.

There appears to be a relatively tight temporal alignment of gesture phase onsets to the onset of both stressed syllables. However, there are differences across the two stressed syllables as well as differences in the gesture phases. While all six phase types may arise in the first stressed syllable, we do not observe any Preparations or Towards movements closely aligned with the second stressed syllable. For the primary stress, Stroke phases and the Toward phase of beat gestures tend to be more tightly aligned to syllable onset than other phase types. Preparations, for instance, display a rather constant alignment slightly before syllable onset. Moreover, Toward phases and Preparations are less variable in their timing compared to other phrase types. The largest variation is clearly observed for Retractions. For the secondary stress, however, Retractions, as well as Holds, tend to align to syllable onset, and all four phrase types display a moderate level of timing variation.

Due to the small size of the dataset, we have not carried out any inferential statistics, and thus our results should be regarded as exploratory. However, the data reported here clearly suggest both similarities as well as an asymmetry in gesture behaviour between the first and the second stressed syllable in compounds: both s_1 and s_2 seem to display close alignment to gesture phase transitions, but the distribution of available gesture phases is more constrained in s_2 than in s_1.

4. Discussion

In a set of 56 Swedish compounds accompanied by hand gestures, we find a difference in distribution of movement phases, with these somewhat less likely to arise in the second stressed syllable than the first. Additionally, we find evidence of close alignment of gesture phase onsets with both primary and secondary stressed syllable onsets.

4.1. Distribution of gesture phases in relation to stressed syllables

The Stroke phase is considered to be the only obligatory gesture phase, and is often the focus of attention of studies investigating prominence relationships between gestures and speech, see, for instance, [31]. However, in our investigation of Swedish compounds, we find that it is not necessarily only the stroke phase which is closely affiliated with the prominent syllables. While many of our stressed syllables were accompanied by gesture strokes, they could also be accompanied by other gesture phases. We find a preference for the stressed syllables to be accompanied by movement phases (Preparations, Strokes, or Toward/Away), but there is no prohibition on the syllables being accompanied by a Hold, or with the gesture being retracted during the stressed syllable.

In over half of the compounds in our study, both stressed syllables were accompanied by the same gesture phase. However, in the cases where the two stressed syllables were accompanied by different gesture phases, an asymmetry between the two syllables arose. It was much more frequent for s_1 to arise accompanied by a movement phase and s_2 with a non-movement phase than the reverse. If we take movement phases to be more visually prominent than non-movement phases, then it seems that the first stressed syllable may have some kind of

privilege in terms of carrying prominence compared to the second syllable. This seems to contrast with the acoustic prominence features, which are sometimes observed to be stronger on the second stressed syllable than the first: Recall that sentence-level prominence is encoded through an additional pitch peak that associates, in compounds, with the secondary stress (see 1.1); high levels of prominence are typically encoded through an extension of the pitch peaks, but especially the second one. A closer investigation of the relationship between the acoustic prominence features and the gesture features is needed to further clarify this relationship.

However, the observed asymmetry suggests that gesture is a part of the prosodic marking of a compound as a single, coherent word, in that it supports the distinction between a main prominence (the first stressed syllable) and a secondary prominence (the second stressed syllable). It has been found that listeners can identify congruence between even unintelligible speech and gesture and use gesture to disambiguate when prosodic information is ambiguous or conflicting [32]. Thus, a co-occurrence of movement phases and stressed syllables in compounds could help support listeners in the ongoing process of speech perception.

The annotations in the current study only gave information about the gesture phases, without specifying any detail as to the form of the gesture (e.g. hand shape, direction or type of movement, apex of non-beat gestures). A classification of the forms of the gestures would likely give us additional information about the relationships between movements and prominences, but is beyond the scope of the current study.

4.2. Temporal relationship between gesture phase onsets and stressed syllables in compounds

Our exploration of the temporal relationship between the onset of a gesture phase and the onset of the stressed syllables supports the proposal of a close temporal alignment [3][5][6][8], while also providing additional evidence of an asymmetry between the first and second stressed syllables in the compound. Across the movement phases, there appears to be a closer alignment of the gesture onset with the syllable onset in the first stressed syllable than in the second stressed syllable. The first stressed syllable can also be accompanied by any gesture phase, while we found no Preparations or Towards movements accompanying the second stressed syllable.

The second stressed syllable did, however, show a relatively narrower alignment with the onset of Retraction phases than the first stressed syllable. The onset of the Retraction may be considered the end of the gesture, since it involves the release of any hand shape and the return of the hand(s) to the neutral position. Retractions in sign language also show similar timing patterns with the offset of speech in spoken language [33]. Thus, the alignment of the onset of the retraction with the second stressed syllable could be related to the final prominence of the compound signaling the upcoming end of the word.

As discussed in section 4.1, it was not always the Stroke phase of a gesture which accompanied the stressed syllables. Similarly, in the second portion of the study, it was not always the Stroke onset which was the closest phase onset to the onset of the stressed syllable. While movement phases appeared to be more closely aligned, especially to the first stressed syllable, no phase type was in principle excluded from alignment with the first stressed syllable.

It is possible that specific kinematic features of gestures might be prominence lending. In addition, as we observe patterns of temporal alignment between stressed syllables and various types of gesture phase onset, our results raise the possibility that the transition from one gesture phase to another also contributes to the perception of prominence. The perception of phonetic segments has been shown to focus primarily on the transition from one phone to the next [34], and similarly, across many languages, phrase boundary locations attract prominence [35] or prosodic boundaries are used to set off prominent elements [36]. A boundary between gestural phases could function similarly in terms of contributing to or at least being aligned with prominent portions of speech.

5. Conclusions and outlook

We investigated hand gestures associated with the two stressed syllables in Swedish compounds, finding that movement phases are often associated with these syllables, with a slight bias towards the first syllable, and that the onsets of both stressed syllables often share a close temporal alignment with transitions between gesture phases (i.e. the end of one phase and the onset of the next). The current study was intended as exploratory and used only a single landmark, syllable onset, for measuring gestural alignment. Future work could include and compare alternative landmarks, such as the syllable nucleus, which is often considered to be the domain for prosodic alignment. The study was nonetheless promising in terms of providing insight into alignment between gestures and acoustic prominences. Further work should extend this analysis into additional types of body movement, such as head movements, as well as considering additional kinematic detail in order to be able to more precisely measure the temporal alignment between gestures and speech. Similarly, a wider range of acoustic measures, including F0 and duration, should be considered, to further elucidate relationships between different elements of the prosodic system. A comparison between compounds and non-compounds will also give more insight into the different prosodic structures and feature constructions that are available.

6. Acknowledgements

This work was supported by the Swedish Research Council (VR-2017-02140) and the research network GEHM (Gesture and Head Movements in Language), which is funded by the Independent Research Fund Denmark (grant 9055-00004B). We are grateful to Simone Löhndorf for performing the acoustic segmentation.

7. References

[1] P. Wagner, Z. Malisz, and S. Kopp, "Gesture and speech in interaction: an overview," *Speech Communication*, 57, pp. 209-232, 2014.

[2] D. Gibbon, "Gesture theory is linguistics: on modelling multimodality as prosody," *Proceedings of PACLIC 23*, Hong Kong, pp. 9-18, 2009.

[3] Y. Yasinnik, M. Renwick, and S. Shattuck-Hufnagel, "The timing of speech-accompanied gestures with respect to prosody," in *Proceedings: From Sound to Sense*, MIT, Cambridge, MA, pp. 97-102, 2004.

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

[5] T. Leonard and F. Cummins, "The temporal relation between beat gestures and speech," *Lang. Cogn. Processes,* 26, pp. 1457-1471, 2011.

[6] D. Loehr, "Temporal, structural, and pragmatic synchrony between intonation and gesture," *Lab. Phonol.: J. Assoc. Lab. Phonol.,* 3, pp. 71-89, 2012.

[7] N. Esteve-Gibert and P. Prieto, "Prosodic structure shapes the temporal realization of intonation and manual gesture movements," *J. Speech Lang. Hear. Res.,* 56 (3), pp. 850-864, 2013.

[8] S. Alexanderson, D. House, and J. Beskow, "Aspects of co-occurring syllables and head nods in spontaneous dialogue," *Proceedings of the 12th International Conference on Auditory-Visual Speech Processing (AVSP2013),* Annecy, France, 2013.

[9] 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*, 95, pp. 100-113, 2017.

[10] E. Krahmer and M. Swerts, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *Journal of Memory and Language,* 57 (3), pp. 396–414, 2007.

[11] M. Swerts, and E. Krahmer, "Facial expressions and prosodic prominence: Effects of modality and facial area," *Journal of Phonetics,* 36 (2), pp. 219-238, 2008.

[12] D. House, J. Beskow, and B. Granström, "Timing and interaction of visual cues for prominence in audiovisual speech perception," In *Proc of Eurospeech 2001*, Aalborg, Denmark, pp. 387-390, 2001.

[13] S. Al Moubayed, J. Beskow, B. Granström, and D. House, "Audio-Visual Prosody: Perception, Detection, and Synthesis of Prominence," in A. Esposito, A. M. Esposito, R. Martone, V. C. Müller, and G. Scarpetta (eds), *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues*, Lecture Notes in Computer Science, vol 6456, Berlin, Heidelberg: Springer, pp. 55-71, 2011.

[14] P. Prieto, C. Puglesi, 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* 49 (1), pp. 41-54, 2015.

[15] G. Ambrazaitis, J. Frid, and D. House, "Word prominence ratings in Swedish television news readings: effects of pitch accents and head movements," in *10th International Conference on Speech Prosody, May 25-28, Tokyo, Japan, Proceedings,* pp. 314-318, 2020.

[16] A. Kendon, *Gesture: Visible action as utterance,* Cambridge: Cambridge University Press, 2004.

[17] D. McNeill, *Gesture and Thought*, Chicago: University of Chicago Press, 2005.

[18] P. ten Hacken, "Compounding in Morphology," *Oxford Research Encyclopedia of Linguistics,* Retrieved 27 May, 2020, from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-251, 2017.

[19] G. Bruce, *Swedish Word Accents in Sentence Perspective,* Lund: Gleerups, 1977.

[20] G. Bruce, "Intonational Prominence in Varieties of Swedish Revisited," in S. Jun (ed), *The Phonology of Intonation and Phrasing,* Oxford: Oxford University Press, pp. 410 429, 2005.

[21] T. Riad, "The origin of Scandinavian tone accents," *Diachronica* 15, pp. 63-98, 1998.

[22] T. Riad, *Prosodi i svenskans ordbildning och ordböjning,* Manuscript, 2009 (retrieved from http://www.su.se/polopoly_fs/1.29929.1320939953!/ProsodiOrdbildningRiadNov12.pdf)

[23] M. L. Flecha-García, "Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English," *Speech Communication,* 52, pp. 542-554, 2010.

[24] J. M. Iverson and E. Thelen, "Hand, mouth and brain. The dynamic emergence of speech and gesture," *Journal of Consciousness Studies*, vol. 6, nos. 11-12, pp. 19-40, 1999.

[25] P. Prieto, A. Cravotta, O. Kushch, P. Rohrer, and I. Vilà-Giménez, "Deconstructing beat gestures: a labelling proposal," In *9th International Conference on Speech Prosody, June 13-16, Poznań Poland, Proceedings,* pp. 201-205, 2020.

[26] A. Kelterer, G. Ambrazaitis, and D. House, "Head beats as pitch-accompanying visual correlates of primary and secondary lexical stress: evidence from Stockholm Swedish compounds," *Proc. TAL2018, Sixth International Symposium on Tonal Aspects of Languages,* Berlin, Germany, pp.124-128, 2018.

[27] J. Edlund, J. Beskow, K. Elenius, K. Hellmer, S. Strömbergsson, and D. House, "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), Valetta, Malta,* pp. 2992-2995, 2010.

[28] ELAN (Version 5.4) [Computer software], Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive, Retrieved from https://archive.mpi.nl/tla/elan, 2019.

[29] D. McNeill, *Hand and mind: What gestures reveal about thought*, Chicago: University of Chicago Press, 1992.

[30] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.1.16, retrieved 6 June from http://www.praat.org/, 2020.

[31] S. Shattuck-Hufnagel and A. Ren, "The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech," *Frontiers in Psychology*, 9: 1514, 2018.

[32] B. Guellaï, A. Langus, and M. Nespor, "Prosody in the hands of the speaker," *Frontiers in Psychology,* 5:700, 2014.

[33] C. de Vos, F. Torreira, and S. C. Levinson, "Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries," *Frontiers in Psychology*, 6:268, 2015.

[34] S. Furui, "On the role of spectral transition for speech perception," The Journal of the Acoustical Society of America, 80, pp. 1016-25, 10.1121/1.393842, 1986.

[35] U. Reichel, K. Mády, and F. Kleber, "How prominence and prosodic phrasing interact," in O. Jokisch (ed), *Elektronische Sprachverarbeitung 2016*, vol. 81, Studientexte zur Sprachkommunikation. Dresden, Germany: TUDpress, pp. 153–159. 2016.

[36] M. Dohen and H. Loevenbruck, "Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French," in *INTERSPEECH 2004, Jeju Island, Korea, Proceedings*, pp. 785-788, 2004.