Temporal alignment of manual gestures' phase transitions with lexical and post-lexical accentual F0 peaks in spontaneous Swedish interaction

Jule Nabrotzky[1], Gilbert Ambrazaitis[1], Margaret Zellers[2], David House[3]
Linnaeus University Växjö[1], Kiel University[2], KTH Royal Institute of Technology[3]
jule.nabrotzky@lnu.se, gilbert.ambrazaitis@lnu.se,
mzellers@isfas.uni-kiel.de, davidh@kth.se

Abstract

Many studies investigating the temporal alignment of co-speech gestures to acoustic units in the speech signal find a close coupling of the gestural landmarks and pitch accents or the stressed syllable of pitch-accented words. In English, a pitch accent is anchored in the lexically stressed syllable. Hence, it is unclear whether it is the lexical phonological dimension of stress, or the phrase-level prominence that determines the details of speech-gesture synchronization. This paper explores the relation between gestural phase transitions and accentual F0 peaks in Stockholm Swedish, which exhibits a lexical pitch accent distinction. When produced with phrase-level prominence, there are three different configurations of lexicality of F0 peaks and the status of the syllable it is aligned with. Through analyzing the alignment of the different F0 peaks with gestural onsets in spontaneous dyadic conversations, we aim to contribute to our understanding of the role of lexical prosodic phonology in the co-production of speech and gesture. The results, though limited by a small dataset, still suggest differences between the three types of peaks concerning which types of gesture phase onsets they tend to align with, and how well these landmarks align with each other, although these differences did not reach significance.

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

1. Introduction

1.1. The speech-gesture link

Evidence for the link between speech and co-speech gesture proposed by e.g. McNeill (1985) and Kendon (2004) has been readily found in the temporal alignment of gestural peaks and phonological as well as acoustic landmarks. The alignment of gestural apexes with pitch-accented syllables is especially well documented (De Ruiter, 1998; Esposito et al., 2007; Esteve-Gibert & Prieto, 2013; Jannedy & Mendoza-Denton, 2005; Karpiński et al., 2009; Leonard & Cummins, 2011; Loehr, 2007, 2012; McClave, 1994; Pouw & Dixon, 2019; Rusiewicz, 2010; Shattuck-Hufnagel & Ren, 2018; Yassinik et al., 2004).

However, in a language like English, a pitch accent is anchored in the lexically stressed syllable. Hence, it is unclear what determines the details of synchronization between pitch accents and gestures: the lexical-phonological dimension of stress, or the phrase-level prominence. There is less information about the synchronization of gestures with non-primary stressed syllables or non-stressed syllables, although there is some evidence that these are less likely to be aligned with gestural phase transitions and apices (Ambrazaitis et al., 2020; Rusiewicz, 2010, pp. 201–206). Additionally, the synchronization of gesture with acoustic landmarks related to larger linguistic units such as the boundaries of intonation phrases remains poorly understood (Karpiński et al., 2009).

For these reasons, it is not yet settled what determines the alignment of gestural and acoustic landmarks. Therefore, the primary motivation for this study was to explore the alignment of F0 peaks to gestural phase transitions both in non-stressed syllables and conversely, in stressed syllables which, however, represent a lexical rather than a phrase-level pitch accent. The Stockholm Swedish variety is well suited for this study, because it exhibits a lexical pitch accent distinction between Accent 1 (A1) and Accent 2 (A2), where words with A2 are realized with two consecutive F0 peaks when the word is given phrase-level prominence (a so-called big accent, BA, see Sec. 1.2). In this case, the first peak is associated with the stressed syllable, primarily representing a lexical pitch accent, while the second one occurs in the following, non-stressed syllable.

Together, phrase-level (PL) accented A1 and A2 words provide us with the following three possible configurations of F0 peak lexicality and syllable status: (i) A1: a phrase-level accentual peak aligned with the lexical stress (as in English), (ii) A2: a lexical accentual peak aligned with the lexical stress, and (iii) A2: a phrase-level accentual peak aligned with a non-lexically stressed syllable. This means that peak type (ii) occurs first in the A2 word and peak type (iii), the phrase level peak, afterwards.

By uncoupling the pitch accent movements from word-level prominence in this way, we aim to shed more light on the factors that cause acoustic landmarks to align with gestural phase transitions. Our primary research questions are therefore which gestural phase transitions the three different F0 peaks most closely align with, and, especially, whether peak type (i) behaves more like peak type (ii) or like peak type (iii) in this respect. If it is lexical stress that aligns to the gestural strokes, we would expect peak types (i) and (ii) to align more often with stroke onsets. We would also expect this alignment to be closer than the alignment of peak type (iii) to stroke onsets. Instead, one could expect more close alignments with retraction onsets for this peak type. If phrase-level prominence were the basis for the alignment to gestures, we would expect the peak types (i) and (iii) to be more similar and (ii) to be less closely aligned with stroke onsets, but rather more closely aligned with preparation onsets.

It is however important to clarify initially that because we want to compare F0 peaks directly to each other, this paper does not study the alignment of F0 peaks to gestural phase transitions, but vice versa. Before describing the analytical procedure adopted in this study, the next section gives some more background on Swedish word accents to illustrate why these are well suited for this task.

1.2. Swedish word accents

Swedish lexical prosody involves lexical stress as well as two basic types of pitch contours that can occur at the word level, termed Accent 1 and 2 or acute and grave respectively. Which contour is assigned to a word depends on both its morphology and some lexical or phonological factors, such as origin and syllabicity. It is not usually thought to be semantically distinctive, although there are a small number of minimal pairs where the word accent contour differentiates between two possible meanings of a word. Instead, its most probable function is the on-line prediction of morphological endings to facilitate speech processing (Roll, 2022). In Stockholm Swedish, the two contours differ in the alignment of the peaks with the stressed syllable, with Accent 1 occurring earlier than Accent 2 (Bruce, 1977), as shown in Fig. 1.

Additionally, each of the pitch accents can occur in either a one-peaked or two-peaked version, where the two-peaked version is also referred to as the big accent, typically used to signal sentence-level prominence. In A1, however, the first peak of a big accent is often elided (it is otherwise realized before the stressed syllable), resulting in a single peaked accent which is realized in the stressed syllable (L*H). In A2, both peaks are typically realized in a big accent (H*LH), where the first peak, aligned with the stressed syllable, represents the lexical pitch accent. The second peak, encoding the sentence-level prominence, is typically realized in the unstressed post-stress syllable. Only words that carry big accents, that is, only accents encoding higher-level prominence, are included in the present study. Additionally, compound words (generally A2) were excluded, because in these, the second peak is associated with a stressed syllable as well (Ambrazaitis et al., 2020).

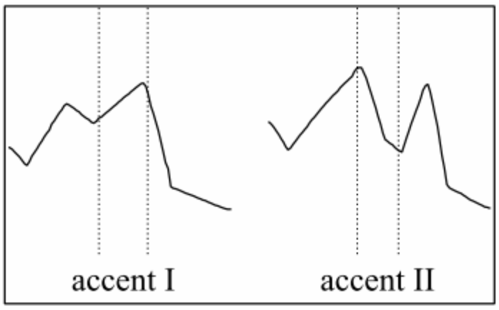

Figure 1: *Illustration of the lexical pitch-accent contrast of Swedish. Stylized F0-curves based on authentic productions of a short phrase; Accent 1: "(det var) med bilen"* '(it was) by car'; *Accent 2: "(det var) med bilar"* '(it was) with cars'*; dotted lines delimit the rhyme of the stressed syllable; it can be seen that the entire F0 pattern is timed earlier in A1 than in A2; these realizations represent so-called big accents (see text); in a small accent, the second F0 peak would be absent; notably, in this example, the first peak in a big-accent A1 is visible, although it is often heavily reduced (see text). Illustration from Ambrazaitis et al. (2012).*

2. The study

2.1. Materials

This study is based on materials taken from the Spontal corpus (Edlund et al., 2010). We analyzed four five-minute chunks of dialogues between two Swedish speakers from Stockholm for this study (eight speakers in total, four females, four males). The dialogues were recorded with separate microphones and cameras for each speaker so that a detailed segmentation with minimal interference from the other channel was possible.

2.2. Annotation procedure

The dialogues were annotated on the acoustic level as well as the gestural level, with several steps in both processes. This section explains the procedure for both annotations.

Since the big accent contour was the basis of the analysis, all dialogues were annotated for the occurrence of these contours. As explained above, the big accent manifests itself as an F0 peak on the first syllable for A1 and on the unstressed post-stress syllable for A2. In practice, this means that Accent 1 words were marked as carrying a big accent when they were realized with an F0 peak on the stressed syllable (as opposed to no peak), and words with an Accent 2 were marked when they showed two peaks (as opposed to only a peak on the stressed syllable). The annotations were done by the first author and the second author independently, and only the big accents that both annotators agreed on were included in the analysis. Inter-rater-reliability was calculated in R (RStudio Team, 2022) using the irr package (Gamer et al., 2019) and reached a good agreement of κ=.782 for all speakers pooled.

The gestural annotations were conducted by the third author and were divided into two phases. Firstly, the videos were manually "time-stamped" for the presence of gestures without precision regarding the on- and offset of the gestures. Afterwards, the gesture boundaries were manually refined through a frame-by -frame inspection. Additionally, the general annotation was specified with a labelling of the preparation, stroke, retraction and hold phases of gestures following the classification by Kendon (2004). According to this classification, not all phases had to be present in the gesture; only the stroke phase was obligatory. Less complex beat gestures were further divided into "towards" and "away" movements in relation to the apex of the gesture, with only the movement toward the apex counting as a stroke. Non-beat gestural apices were not annotated.

2.3. Analysis

First, a list of all words that carried a big accent and were accompanied by a gestural stroke were created. The rationale for the decision to only include stroke overlaps was that this is generally regarded as the "meaningful" phase of a gestural phrase as a whole (Esteve-Gibert & Prieto, 2013; Kendon, 2004, p. 112; Leonard & Cummins, 2011). Then, a new variable was created that included the type and timing of the relevant F0 peak. Since there are two relevant F0 peaks in every Accent 2 word, this means that the alignment of all Accent 2 words is considered twice in the analysis, once for every F0 peak in the word. After this, a list of all gestural phases and their onsets, offsets and duration was created. However, phases were not grouped into gestural phrases, which means that the data contained no information on which phases belonged together.

To match each F0 peak to the closest gesture phase transition, each F0 peak's time stamp was compared with all gesture phase onsets that had the same speaker ID and the closest match was determined. Then, the information on this closest matching phase was appended to the information about the specific F0 peak. This also meant that only one of the gesture phases was assigned to the word in case of parallel onset of the gesture with both hands (which was by default the left). While this procedure could cause problems in case both hands started different phase onsets at the exact time, this happened only four times in our data set and none of those onsets were aligned with a F0 peak.

The statistical analysis was conducted in R (RStudio Team, 2022) and will be detailed in the results section.

3. Results

3.1. Descriptive Statistics

Our 20 minutes of dialogue contained a total of 873 words with a big accent. Of these, 209 simplex (non-compound) big accent words in our data were accompanied by a gesture, 154 of which were assigned Accent 1 and 55 of which were assigned Accent 2. Due to irregularities in the speaker's voices or non-modal voice qualities, we could only annotate big accent peaks for 185 of those cases, resulting in 140 instances of Accent 1 and 45 instances of Accent 2.

Independent of whether they overlap with a BA-accented word, a total of 1013 individual gesture phase onsets are present in the data. However, this figure is slightly misleading, because every gesture phase that was executed by both hands constitutes two gesture phases in our dataset (since the data was annotated for each hand separately). Additionally, the number of gesture phrases was much lower, since several gesture phases can make up a gesture *phrase*. And finally, since some gesture phrases do not overlap with any BA words, not all of them are possible anchor points for BA words in the data. Keeping this in mind, the gesture data consists of 49 toward and 50 away phases, 205 preparations, 308 strokes, 195 holds and 205 retractions.

Table 1: Number of occurrences for every phase type and F0-peak

phase type	*(i) PL prominence + lexical stress*	*(ii) lexical stress*	*(iii) PL prominence*
away	5	0	0
hold	30	9	12
preparation	7	2	0
retraction	29	8	14
stroke	63	22	15
toward	6	4	4

Table 1 gives an overview of the number of gesture phase onsets that are most closely aligned with every type of F0 peak as described above. For better comparisons, this is graphically depicted in Figure 2. In all cases, the F0 peak was most often closest aligned with the stroke phase onset. However, there is a greater tendency for F0 peaks which occur earlier in the word ((i) *PL prominence + lexical stress* in Accent 1 and (ii) *lexical stress* in Accent 2) to align with stroke onsets. The hold and retraction phase onsets are tied for second most likely. Interestingly, the peak type (iii) (PL prominence) is more commonly aligned with hold and retraction onsets than the other two types of F0 peaks. Alignments with preparation onsets are quite uncommon in the data, in fact, none of the PL prominence peaks are aligned with them.

There are far fewer cases of alignment with beat gesture phase onsets in general, which is also related to the fact that they are much less common in the data. Both "away" and "towards" phase onsets have a similar likelihood for peak type (i) (Accent 1 words), while for the peak types (ii) and (iii) (Accent 2 words), there are no cases of alignment with the onset of the "away" phase, only with the "towards" phase.

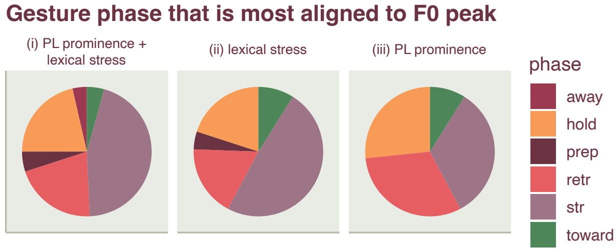

Figure 2: *Distribution of best-aligning gestural phase onset types for three different types of accentual pitch peaks. Example: For the first peak in a two-peaked Accent 2 word (type (ii) lexical stress), in about half of the cases, the phase onset that occurs closest to the accentual peak is a stroke onset.*

Figure 3 summarizes the alignment of the F0 peaks with gesture phase onsets in three boxplots. It should be noted that this figure does not reveal, generally, how well a given phase type onset tends to align with a given F0 peak type, because it only displays onset values for the phase onsets that happen to be closest to the corresponding F0 peak. The figure therefore does not show, for example, how closely "away" phase onsets in general are aligned to every F0 peak because it is operating with an incomplete data set of away phase onsets. Additionally, it is important to keep in mind, as mentioned earlier, that peak types (ii) and (iii) occurred in the same words.

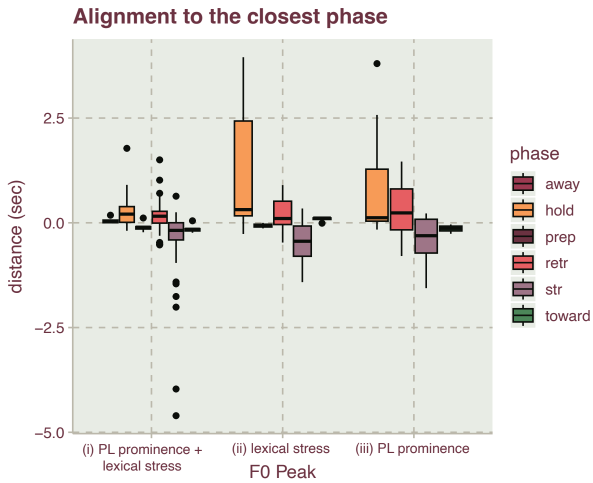

Figure 3: *Alignment (in seconds) between F0 and the closest occurring gestural phase onset (according to Tab. 1, Fig. 2), plotted per phase type.*

In general, the distribution of the data for Accent 1 appears to show a smaller dispersion than the data for Accent 2 words. Across F0 peak types, the onsets of hold and retraction seem mostly to occur after the F0 peak, while the stroke onset occurs before the F0 peaks in most cases. Retraction onsets are also more commonly found after the F0 peak. Preparations, toward and away phase onsets are all closely aligned with the F0 peaks with minimal dispersions, which may however be related to their limited occurrence (see Table 1). Overall, the hold phase onsets' alignment shows the greatest dispersions for Accent 2 words. Although they are skewed towards 0, a hold onset (although being the nearest phase transition) can be as far away as 2.5 seconds or more. In this case, it is clear that the two landmarks are not really aligned in a meaningful way. Instead, these values are likely a result of an F0 peak occurring at the end of a long gestural stroke.

3.2. Inferential statistics

To explore if the distributions of stroke, retraction and hold phase onsets differed significantly for the three types of F0 peak, we conducted a chi-square test of independence. The other three types of gesture phase onsets occurred less than five times in some conditions, which is why they were classed into a fourth "other" category. The values for three types of F0 peak did not significantly differ from each other (X^2 (1, N = 230) = 4.63333, p = .591623.

To understand if the alignment of the F0 peaks to the gesture phase onsets differed between our three types of peaks, we fitted a linear mixed model to the data using the lme4 package (Bates et al., 2015). We evaluated models that took into account the predictors "F0 peak type" and "phase type" as well as "speaker" as a random factor to predict the alignment of the gesture phase to the F0 peak. Table 2 shows the model fit of three models that were tested. The R^2-values indicate that the best model fit is the one that considers a possible interaction of Peak Type and Phase Type to explain the alignment measures. Additionally, the values indicate that the random factor of speaker seems to have no explanatory power for the data (because the results display no differences between R^2_m and R^2_c). However, when we conducted likelihood ratio tests on the data, the differences between the models were not significant (M1 vs. M2: $\chi2=10.548$, $df = 7$, $p=.1596$, M2 vs. M3: $\chi2=3.6379$, $df = 2$, $p=.1622$, M1 vs M3: $\chi2=14.186$, $df = 9$, $p=.1159$).

Table 2: Model fits

Name	Model	R^2_m	R^2_c
M1	Peak Type * Phase Type	.2773	.2773
M2	Peak Type + Phase Type	.2433	.2433
M3	Phase Type	.2311	.2311

4. Discussion

Our analysis of the temporal alignment of the types of F0 peak to gesture phase transitions was done in two steps. First, we compared the distribution of phase onset types that were most closely aligned with each peak. Then, we analyzed the degree of alignment for those most closely aligned phase onsets. While the descriptive statistics showed some interesting tendencies for these two measures that are worth discussing, none of the inferential tests indicated significant differences between our peak types. This discussion is therefore primarily intended to highlight further possibilities for research and analytical difficulties that come with quantifying this kind of data.

First and foremost, comparing the pie charts for the three types of F0 peaks we see a greater similarity between the two peak types that are associated with lexical stress. The third type of peak, which is only associated with sentence prominence, seems to more often align with phase transitions to retractions, a phase that occurs late in a gesture phrase. This might simply be due to the fact that this peak occurs later in the word. On the other hand, if it was the sentence-level prominence that generally aligned with gestural stroke onsets, we might have seen more of a similarity between the peak types (i) and (iii). These results back the hypothesis that it is the lexical stress feature that determines the alignment with gestural stroke onsets, rather than sentence-level prominence.

Additionally, since there are almost no F0 peaks that are aligned with preparation onsets, it seems that it is unusual for the gestural peak to follow the stressed syllable, further supporting the hypothesis that lexical stress determines alignment. These results also confirm the observations of previous studies that also found a tendency for early onsets of gestural peaks (De Ruiter, 1998; Pouw & Dixon, 2019).

Our data on the closeness of alignment pose some analytical challenges. The higher dispersion of Accent 2 words seems interesting but might be caused by several factors. In general, some stroke phases included in the data were quite long due to their complexity, which added to the high dispersions present in the data. However, the higher number of Accent 1 words in the dataset might have contributed to the higher number of outliers that are visible in the box plot. It could also be caused by the fact that Accent 1 words are on average shorter than Accent 2 words. This is simply owed to the fact that all monosyllabic words are assigned Accent 1 in Swedish. Of course, when a stroke is significantly longer than the accented word, this does not play a role. However, it might be the case that longer words more often partially overlap with a stroke phase, which could contribute to a higher distance between F0 peak and gestural phase transition.

Therefore, we are hesitant to claim that the higher dispersions for Accent 2 words can be interpreted as weaker alignment due to two possible alignment points. The fact that models which include the predictor "F0 peak type" to explain alignment did not perform significantly better than models without this predictor indicates that there might be other reasons for the wider dispersion. Additionally, an analysis that includes more data points might add clarity to these preliminary results. The number of words included in the analysis was comparatively low, owing to the time-consuming nature of annotating naturally occurring gesture.

A limitation of the method adopted in this study was that we measured the alignment to the onset of the gestural phases rather than, for example, their apex or point of maximal extension as in other studies. This is because the apex is a measure that is unique to gestural strokes and therefore would not be relevant or even measurable for the other types of gestural phases. Following from the research question that aimed to compare the three F0 peaks, we used the F0 peaks as reference points when measuring the most closely aligned phase transition as opposed to using the phase transitions as reference points to determine the most closely aligned F0 peak. An alternative analysis that used, for example, the gestural stroke onsets as the starting point could give more information on what peak they are most often aligned with and if there are differences in the degree of alignment depending on the F0 peak.

Our annotation procedure also did not differentiate between referential and non-referential gestures. Although it has been argued that both types of gestures can relate to prominence (Prieto et al., 2018; Shattuck-Hufnagel & Prieto, 2019), non-referential gestures might be hypothesized to display stronger tendencies to align with phrase-level prominence than referential gestures. That said, results by Rohrer (2022, pp. 306–307) indicate that the two types of gestures behave similarly with respect to the association with pitch-accented syllables.

These preliminary results offer interesting ideas for further research. Firstly, it is clear that the decisions made in the analytical procedure have a large influence on the results of any study trying to understand gestural alignment. Choosing which type of F0 peak, which level of prominence, and which gestural landmark to prioritize influences the interpretations that can be made. For example, an argument could be made that only including words that overlap with gestural strokes limits the explanatory power of the results, especially regarding the other gestural phases. However, it is not clear if words that overlap with, say, only a preparation of an upcoming gesture, while the stroke occurs on another word entirely, should count as being co-occurring with gesture in any meaningful way.

Secondly, more data on this specific contrast, data on other languages that also dissociate between lexical stress and sentence-level prominence or data where other gestural landmarks (like stroke apices) are labelled could provide valuable insight into our research question and add nuance to the tendencies that we see in our data. While this can be seen as a starting point, further studies should consider the points we have raised when making decisions about the evaluation procedure.

5. 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).

6. References

Ambrazaitis, G., Frid, J., & Bruce, G. (2012). Revisiting Southern and Central Swedish intonation from a comparative and functional perspective. In O. Niebuhr (Ed.), *Understanding Prosody* (pp. 138–158). De Gruyter.

Ambrazaitis, G., Zellers, M., & House, D. (2020). *Compounds in interaction: Patterns of synchronization between manual gestures and lexically stressed syllables in spontaneous Swedish*. The 7th Gesture and Speech in Interaction (GESPIN2020), 7-9 September 2020, Stockholm, Sweden.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1).

Bruce, G. (1977). Swedish accents in sentence perspective. *Working Papers 12, Lund University, Department of Linguistics and Phonetics*.

De Ruiter, J. P. (1998). *Gesture and speech production.* Doctoral Dissertation, Radboud University Nijmegen.

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: A Swedish spontaneous dialogue corpus of audio, video and motion capture. *Proc. of the Seventh Conference on International Language Resources and Evaluation (LREC'10) :*, 2992–2995.

Esposito, A., Esposito, D., Refice, M., Savano, M., & Shattuck-Hufnagel, S. (2007). A Preliminary Investigation of the Relationship between Gestures and Prosody in Italian. In *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue* (pp. 65–74). IOS Press.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864.

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (0.84.1). https://CRAN.R-project.org/package=irr

Jannedy, S., & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure*, *3*, 199–244.

Karpiński, M., Jarmołowicz-Nowikow, E., & Malisz, Z. (2009). Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogues. *Speech and Language Technology*, *11*, 113–122.

Kendon, A. (2004). *Gesture Visible Action as Utterance* (1st ed.). Cambridge University Press.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, *7*(2), 179–214.

Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1).

McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, *23*(1), 45–66.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350–371.

Pouw, W., & Dixon, J. A. (2019). Quantifying gesture-speech synchrony. *Proceedings of the 6th Gesture and Speech in Interaction Conference*.

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: A labelling proposal. *Proceedings of the 9th International Conference on Speech Prosody 2018*, 201–205.

Rohrer, P. L. (2022). A temporal and pragmatic analysis of gestures-speech association: A corpus-based approach using the novel Multimodal MultiDimensional (M3D) labeling system. Ph.D. Thesis, Universitat Pompeu Fabra.

Roll, M. (2022). The predictive function of Swedish word accents. *Frontiers in Psychology*, *13*(910787), 1–11.

RStudio Team. (2022). *RStudio: Integrated Development Environment for R*. PBC. http://www.rstudio.com/

Rusiewicz, H. L. (2010). *The role of prosodic stress and speech perturbation on the temporal synchronization of speech and deictic gestures* [Doctoral Dissertation, University of Pittsburgh].

Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. *Proc. of the 19th ICPhS*, 1490–1494.

Shattuck-Hufnagel, S., & Ren, A. (2018). The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech. *Frontiers in Psychology*, *9*, 1514.

Yassinik, Y., Renwick, M., & Shattuck-Hufnagel, S. (2004). The Timing of Speech-Accompanying Gestures with Respect to Prosody. *Proceedings of From Sound to Sense*, *5*, 97–102.