Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

Examining the role of prosody and information structure in Hungarian speech and co-speech gestural coordination: An EMA study

Csilla Tatár[1], Ezra Keshet[1], Jelena Krivokapić[1]
[1]University of Michigan
cstatar@umich.edu

Abstract

This study investigates the temporal coordination of speech and co-speech gestures, focusing on how linguistic structure (particularly pragmatic and prosodic prominence) affects their alignment in Hungarian, a typologically distinct language relative to prior studies. Research has shown that gestures frequently align with prosodic prominence (mainly pitch accents) and discourse functions (especially focus). However, in discourse configurational languages such as Hungarian, where pragmatic prominence is primarily expressed through word-order modulation rather than obligatory pitch accentuation, the mechanisms of speech–gesture coordination remain unclear. Electromagnetic articulography data were collected along with acoustic and video recordings, testing the effect of pragmatic and prosodic prominence on gestural coordination. We examine which landmarks of consonant, vowel, and f0 gestures coordinate with which landmarks of pointing gestures and how the coordination is affected by varied discourse conditions. Preliminary results from one speaker show that the onset of the pointing gesture slightly precedes the onset of the target word, and the shortest and least variable temporal lag is between the targets of the pointing gesture and the vowel of the first syllable of the target word, regardless of prosodic prominence and information structure, indicating that it is these two landmarks that are coordinated.

Index Terms: co-speech gestures, multimodality, information structure, f0, prominence, articulatory kinematics, Hungarian

1. Introduction

This study examines the coordination of speech and co-speech gestures in Hungarian. Previous studies have found co-speech gestures and prosodic prominence to co-occur with some systematicity [1-11]. Peak-to-peak alignment has often been reported with the target (maximum displacement, also called apex) of the co-speech gesture aligning with the pitch accented syllable [1-2, 6-9, 11]. In the absence of a pitch accent, co-speech gestures may align with other landmarks (such as an accentless tone on the left edge of prosodic words in Turkish [2] and the focused word in Cantonese [12]) or show a tendency to start and end before and after the corresponding speech unit, referred to as *containment* [13-14]. If co-speech gestures are aligned with prosodically prominent units, the manner in which they are aligned is, conceivably, mediated through the prosodic system of the particular language (or even, as [15] argues, the gestural and prosodic systems are two realizations of the same multimodal system), possibly producing typological variation.

Another element in speech and co-speech gesture alignment is pragmatic prominence. Studies report on the role of *information structure*, which pertains to the organization of information within an utterance (see [16]). Three core areas of information structure relate to (1) *focus* (the identification of the contextually most important or contrastive piece of information in an utterance), (2) *topic* (the matter under discussion), and (3) *givenness* (whether the utterance provides *new* or *given* information within its context). Research highlights the frequent occurrence of co-speech gestures with pragmatically more prominent units of speech. In particular, items in focus, new (rather than given) information, and to some extent topic, have been found to be frequent anchors of co-speech gestures [1-2, 14, 17-22]. Further, [1] show gesture peak alignment to be temporally closer with focus-related pitch accents than with *background* (non-focus) pitch accents. These findings demonstrate the strong connection between information structure and gesture.

An open question concerns how prosodic and pragmatic prominences factor into alignment: since prosodic prominence is often elicited via focus, it remains difficult to evaluate if the close alignment observed occurs mainly due to information structural factors, prosodic factors, or a complex interplay of the two. Specifically, the questions are: (1) What speech and co-speech gestural landmarks coordinate in Hungarian? (2) (How) does pragmatic prominence related to information structure influence the coordination of speech and co-speech gestures? (3) (How) does prosodic prominence factor into coordination? (4) (How) do information structure and prosodic prominence co-influence coordination?

The present study explores these questions, investigating coordination in varied focus and topic structures to examine the potential effect of different prominence types on gestural alignment. Alignment has been defined in a number of ways, including temporal overlap between prominent words or phrases and co-speech gestures [4] and co-occurrence within a specified amount of time, such as the average duration of a syllable [14]. In the present paper, the questions are probed through the lens of Articulatory Phonology [23-25], employing the theoretically established and computationally modelled definition of alignment called *gestural coordination*, where *in-phase* coordination is considered the most stable (such as *onset-to-onset* and *target-to-target*). The definition relies on the notion of stability in gestural coordination, i.e. finding the least variable temporal lag, which indicates a systematic relationship between landmarks (see also [7, 27-28]).

The object language is Hungarian, a discourse configurational language in which pragmatic prominence relations are expressed primarily via word-order modulation [29-31]. Pragmatic prominence can additionally be expressed via pitch accentuation [32-33] and/or phrasing (i.e. the insertion of a pause before an expression to signal its prominence [34-35], but pragmatic and prosodic prominence may also be separated [33, 36]. The examination of a typologically new language (relative to prior studies on speech-gesture alignment)

10.21437/SpeechProsody.2026-100

offers further cross-linguistic insight into the universal and language-specific aspects of alignment.

2. Methods

2.1 Participants and materials

Seven native speakers of Hungarian participated in the study. Participants reported no speech impediments or hearing loss. They were compensated for their time and effort. The present paper reports preliminary results from one participant analyzed to date.

The study designed conditions to test gestural coordination under five different pragmatic prominence conditions (*focus*: broad focus, confirmation focus, corrective focus; *topic*: aboutness, contrastive). For the purpose of another study, a further prosodic boundary condition and a condition without gesturing were collected as well. A target sentence was constructed with a contextually given target word (*Mima* ['mimɒ]). The sentence functioned as the answer to context-setting questions, and, via varying the questions, it was placed in different pragmatic contexts to elicit the correct interpretation for the different focus and topic types. In all non-broad focus contexts, the target word was followed by the verb *bebetonozta* 'cemented' (1/a) and in broad focus and topic contexts, it was followed by *betonozta be* 'cemented' (1/b), the difference lying in the position of the verbal prefix *be*, reflecting the discourse-configurational nature of Hungarian word order [29-31]. In all conditions, the target word was preceded by the same sequence (*ma* [mɒ]). Lexical stress falls on the first syllable of the target word. Thus each sentence contained the sequence [mɒ mimɒ bɛ], allowing for a labelable sequence of consonants and vowels, and the sequence was voiced to allow for the labeling of f0.

Ma ***Mima*** *be- betonozta a hírnevét.* **(1/a)**
Today Mima.NOM vpx-cemented.3pSg the reputation.3pPoss.ACC
"Today, Mima cemented her reputation."

Ma ***Mima*** *betonozta be a hírnevét.* **(1/b)**
today Mima.NOM cemented.3pSg vpx the reputation.3pPoss.ACC
"Today, Mima cemented her reputation."

The question-answer pairs were presented in blocks. Each block consisted of two question-answer pairs, presented multiple times in pseudo-random order and interspersed with fillers. Participants first read the pair silently, then produced it aloud. Each pair was repeated ten times, for a total of 50 sentences (5 conditions *x* 10 repetitions).

2.2 Data collection procedure

Kinematic, visual, and acoustic data were collected at the EMA lab of the University of Michigan. Kinematic data was collected with a Carstens Medizinelektronik AG501 3D electromagnetic articulography (EMA) system. EMA sensors were placed on the tongue body and dorsum to track movement for vowels and on the upper and lower lips to track bilabial consonants. To provide reference points for the head and to correct for head movement, two sensors were placed on the left and right mastoid processes and one on the upper right incisor. A microphone was placed in front of the participant. Two cameras recorded each participant; these were placed to record frontal and profile views.

The question-answer pairs were displayed on a monitor in front of the participant at a comfortable reading distance. Around the monitor, images were placed of the characters that appear in the question-answer pairs. Participants were instructed to point at the image of the character they mentioned. Prior to recording, the natural resting position of participants' hands was marked with a sticker on their leg, to serve as a reference point. Participants were asked to rest their gesturing hand on this sticker when not gesturing (see [28, 37]).

2.3 Data processing and annotation

Utterances with disfluencies and speech errors were excluded. The utterances were examined in Praat; the sentences were produced with the expected prominence and boundary type. The f0 contours were further examined and are described in section 3.3. The video recordings were time-aligned with the acoustic signal and kinematic data using *Adobe Premiere Pro 2025*. Co-speech gestural data was extracted from the visual data using the full body tracking module of *EnvisionBox* [38]. The module takes a video as input and outputs kinematic time-series data, tracking the movement of points on the hands, face, and body in 3D. The co-speech gesture data stream was time-aligned with the kinematic data from EMA using a custom script by Yoonjeong Lee (University of Southern California).

The kinematic data from the two sources were annotated using *mview* (software by Mark Tiede at Haskins Laboratories, New Haven, CT, USA), see Figure 1.

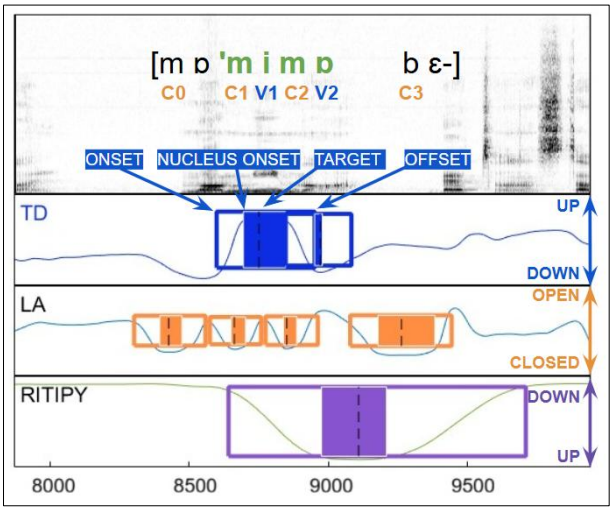

Figure 1: *Speech and co-speech gesture labeling in mview: spectrogram, tongue dorsum (TD, vertical) lip aperture (LA, vertical), right index fingertip (RITIPY, vertical). The target word* ['mimɒ] *is highlighted in green.*

Consonants were labeled on the lip aperture trajectory (LA, the Euclidean distance between upper and lower lip markers) in the target word and in the pre- and post-target bilabial consonants, as well as the vertical movement of the tongue dorsum (TD) in the target word. Timepoints extracted for speech gestures were onset, peak velocity, target (nucleus onset, maximum displacement) and offset. For co-speech gesture movement (RITIPY), the vertical movement of the pointing index finger was labeled, with the same landmarks. Maximum displacement was not reliably identified for the pointing gesture, as the gesture was often held for a longer period of time; nucleus onset was used as the finger displacement measure instead.

Time-aligned acoustic data was annotated in Praat [39], where pitch accent onsets and targets (maximum height)

associated with the target word were labeled ("PA" label). In cases where f0 was flat throughout the first (prominent) syllable of the target word, the target word onset was marked as the onset and the midpoint of the lexically stressed vowel was marked as the peak. Additionally, the f0 maxima and its onset were labeled ("f0" label). When the second (non-prominent) syllable exhibited a rise, the beginning and end points of the contour were additionally marked as an onset and a peak (f0). When the PA peak in the first syllable is the f0 maxima in the target word, the PA and f0 labels coincide. Amplitude maxima timepoints associated with the target word were extracted via parabolic interpolation to examine intensity peak and co-speech gesture target coordination [26].

2.4 Data analysis

To examine how pragmatic and prosodic prominence may modulate gestural coordination, we examine the variability and, when inconclusive, also the duration (in absolute values) of the temporal lag between articulators for onset-to-onset and target-to-target coordination. The intervals between the pointing gesture time points and those corresponding to the pre-target word, the stressed syllable of the target word, and prosodic events are listed in Table 1. The least variable interval is considered to be the most reliable for coordination [7, 27-28].

Table 1. *Pairs of landmarks considered for coordination.*

	Speech gesture	Co-speech gesture
ONSET TO ONSET	pre-boundary consonant onset (*C0*)	pointing gesture onset (*FING*)
	target word consonant onset (*C1*)	
	f0 onset (*PA*; *F0*)	
	target word vowel onset (*V1*)	
TARGET TO TARGET	pre-boundary consonant target (*C0*)	pointing gesture target (*FING*)
	target word consonant target (*C1*)	
	f0 target (*PA*, *F0*)	
	intensity peak (*INT*)	
	target word vowel target (*V1*)	

During prosodic verification, different types of prosodic contours on the target word were identified. To explore how these different contours might affect coordination patterns, f0 contour clustering was performed on the target word using the clustering method described in [40], with default settings.

Statistical analyses were carried out in R [41]. The following linear regression models were tested with a 0.05 threshold of significance:

$$I = \beta_0 + \beta_1(InformationStructure) + \varepsilon \quad (1)$$
$$I = \beta_0 + \beta_1(f0contour) + \varepsilon \quad (2)$$
$$I = \beta_0 + \beta_1(IS) + \beta_2(f0contour) + \beta_3(IS \times f0contour) + \varepsilon \quad (3)$$

The dependent variable is the interval ("I"), or temporal lag, between two given landmarks. The predictors are *Information Structure* condition (topic: aboutness or contrastive, focus: broad, confirmation, correction), *f0 contour* type (the contours that emerge from the contour clustering analysis), and the interaction of these.

3. Results

3.1. Stable coordination points

Figure 2/A illustrates the intervals with all information structure conditions conflated. Visual examination shows that, with some variability, the pointing gesture on average appears to begin before the onset of the target word (*ONSET_C1_FING*, *ONSET_V1_FING*), and the target of the pointing gesture follows the f0, intensity, and pitch accent targets (*TARGET_F0_FING, TARGET_INT_FING, TARGET_PA_FING*).

Prior literature [1] has shown the closest alignment between the pitch accent peak and co-speech gestural target to occur in narrow focus. Accordingly, closest alignment is expected in the corrective focus (a type of narrow focus) condition in our data. Figure 2/B illustrates the intervals in this condition. In general, the least variable intervals are target-to-target, with the lags being shortest in the f0, pitch accent, intensity, and vowel target to pointing target intervals (*TARGET_F0_FING, TARGET_PA_FING, TARGET_INT_FING, TARGET_V1_FING*). This indicates target-to-target coordination.

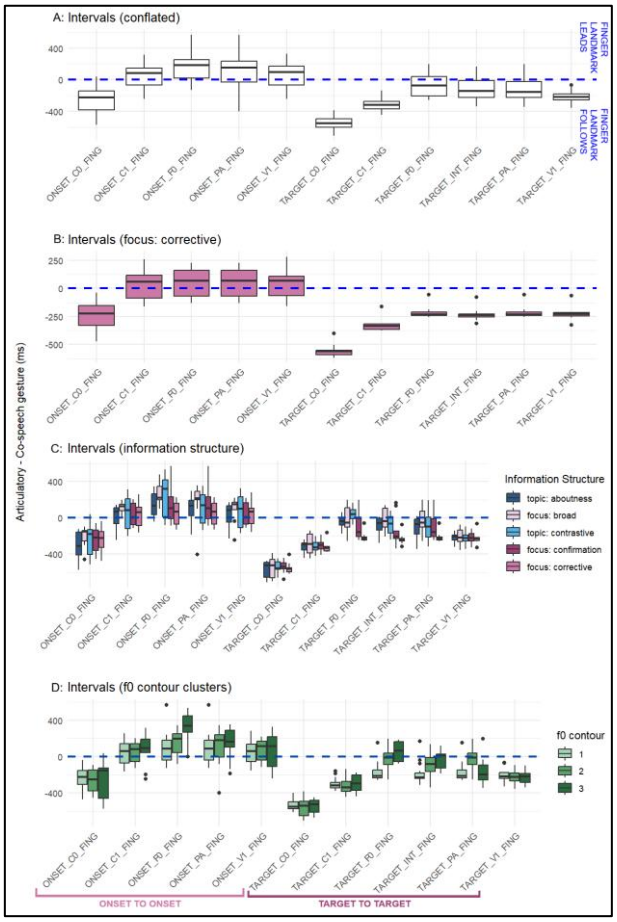

Figure 2: *Coordination configurations. The co-speech gestural landmark's timepoint is subtracted from the articulatory gestural landmark's timepoint (negative range indicates that the co-speech gestural landmark follows the speech one).*

3.2 The role of pragmatic prominence

Having established the possible patterns of coordination in corrective focus, we broaden our examination of these patterns in Figure 2/C, which illustrates the same intervals with all information structure conditions separated. The figure shows that it is the first vowel target and pointing gesture target intervals (*TARGET_V1_FING*) that are the least variable in their coordination across all conditions. Although coordination appears to be less variable under the pragmatically more

prominent conditions (confirmation and corrective focus, aboutness and contrastive topic) relative to the least prominent broad focus condition, the main effect of *Information Structure* is not statistically significant in predicting temporal lag.

3.3 The role of prosodic prominence

Contour clustering identified three characteristic f0 contours of the target word (Figure 3): H* pitch accent on the first syllable (focus only, predominantly corrective and confirmation); late peak or early rise mid-target word (mainly broad focus, confirmation focus, aboutness topic); flat first syllable followed by a high rise (mainly topic, predominantly contrastive). Table 2 lists the number and type of tokens.

To examine how contour type may factor into coordination, Figure 2/D examines coordination by contour type. This allows us to further investigate the hypothesis that co-speech gesture peaks coordinate with prosodic prominence when there is no pitch accent peak (see [2]). Although in general, the *TARGET_PA_FING* interval is expected to be the shortest and least variable across conditions (as pitch accents bear the main prominence), the figure shows that this interval is variable across contour types and is centered around 0ms only in *contour 2* (the difference between contours is statistically significant, with the main effect of *f0 contour* for this interval at β=154.47, p=0.001). In *contour 1*, the pitch accent peak is the only likely f0 landmark, coinciding with the f0 maxima of the contour. In *contour 3*, however, there are two potential landmarks: the mid-point of the vowel in the lexically stressed first syllable (on which prominence is normally realized), and the f0 peak at the end of the target word. Figure 2/D shows that the two intervals, *TARGET_PA_FING* and *TARGET_F0_FING* are identical in *contour 1* and *2* (as expected), but different in contour 3, where, although variable, the interval between the f0 peak and finger target is around 0ms. Crucially, the intervals between the f0 peak and finger target do not appear to be less variable than those between the pitch accent peak and finger target. Figure 2/C and D additionally show that the interval between the intensity peak and the pointing gesture target (*TARGET_INT_FING*) are variable across conditions as well.

As Figure 2/D shows, across the contours, the interval between the first vowel and pointing gesture target (*TARGET_V1_FING*) remains the least variable, as suggested in 3.2. There is no statistically significant difference across contour types for this interval.

3.4 Pragmatic and prosodic prominence in coordination

To explore if contour type and information structure interact in coordination, Figure 4 illustrates the two f0-related target-to-target intervals, along with the target word onset-gesture onset and the stable *TARGET_V1_FING* intervals across the three contours and all information structure conditions.

Regarding the f0-related intervals, Figure 4 shows that taking the f0 peak (rather than strictly the pitch accent peak) as a measure of coordination results in a difference in the temporal lag in the two topic conditions only (as these, especially contrastive topic, cluster into *contour 3* with a rising contour). The variability of the lag is similar across interval types.

TARGET_V1_FING remains the least variable across all conditions, and linear regression confirms that neither contour type, nor information structural condition, nor their interaction reliably affects the lag of this interval.

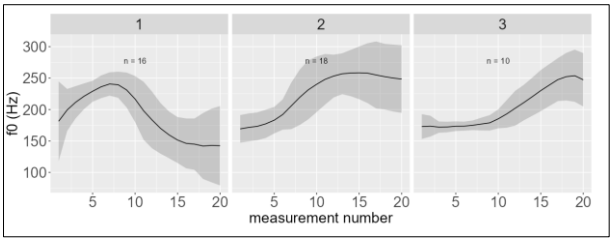

Figure 3: *f0 contour clusters.*

Table 2. *f0 contour tokens by information structure*

Information structure		*f0 contour type*			Total
		#1	#2	#3	
Focus	broad	1	6	1	8
	confirmation	5	4	0	9
	corrective	10	0	0	10
Topic	aboutness	0	6	2	8
	contrastive	0	2	7	9

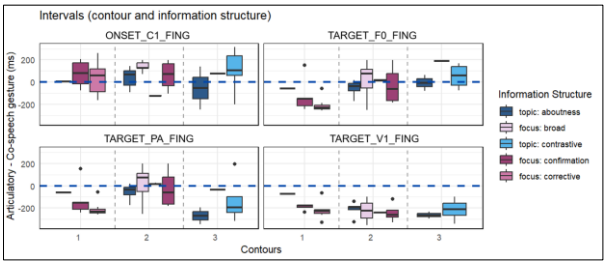

Figure 4: *Coordination configurations across contours and information structure conditions.*

4. Discussion

The present study examined how speech and co-speech gestures are coordinated in Hungarian, a discourse configurational language in which pragmatic prominence is primarily encoded through word order. By combining electromagnetic articulography (EMA) data with acoustic and visual data, we investigated how articulatory and prosodic landmarks align with manual pointing gestures under different information structural conditions.

Results from the participant analyzed to date show that the pointing gesture generally begins before the target word, and that the most stable alignment is a target-to-target coordination between the vowel of the stressed syllable and the pointing gesture (*TARGET_V1_FING)*. No stable coordination is observed with the f0 target (pitch accent or f0 maxima).

While the remaining speakers await analysis, these early results show a lack of evidence for the organizing role of f0 in speech-gesture coordination in Hungarian. Additionally, information structure does not appear to modulate coordination in this speaker's production. Although preliminary, the present results underscore the importance of examining gestural coordination in typologically distinct languages with diverse expressions of prominence.

5. Acknowledgements

The authors express their gratitude to Jungyun Seo, Ruaridh Purse, Yoonjeong Lee, Savithry Namboodiripad, Kendall Lowe, Frank Kügler, Natasha Abner, and Eric Swanson for their help and advice at various stages of the study.

6. References

[1] Kügler, Frank & Alina Gregori. (2023). "Iconic Gestures in Focus – Synchronization of Prosody and Gestures in Prominence." In: Skarnitzl, R. & Volín, J. (eds) *Proceedings of 20th ICPhS 2023*, Prague, Czech Republic August 2023. 4125-4129 (ID 232). Guarant International.

[2] Türk, O. & Calhoun, S., (2023). Multimodal cues to intonational categories: Gesture apex coordination with tonal events. *Laboratory Phonology*, 14(1).

[3] Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2023). Visualizing prosodic structure: Manual gestures as highlighters of prosodic heads and edges in English academic discourses. *Lingua*, 293, 103583.

[4] 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.

[5] 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), 850-865.

[6] Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory phonology,* 3(1), 71-89.

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

[8] Mendoza-Denton, N., & Jannedy, S. (2011). Semiotic layering through gesture and intonation: A case study of complementary and supplementary multimodality in political speech. *Journal of English Linguistics*, 39(3), 265-299.

[9] Roustan, B., & Dohen, M. (2010, May). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. In *Speech Prosody 2010-5th International Conference on Speech Prosody* (pp. 100110-1).

[10] Swerts, M., & Krahmer, E. (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, 38(2), 197-206.

[11] Yassinik, Y., Renwick, M. & Shattuck-Hufnagel, S. (2004). The timing of speech- accompanying gestures with respect to prosody. *Proceedings of the International Conference: From Sound to Sense: +50 Years of Discoveries in Speech Communication*, MIT, Cambridge, 10-13 June, C97 – C102.

[12] Fung, H. S. H., & Mok, P. P. K. (2018). Temporal coordination between focus prosody and pointing gestures in Cantonese. *Journal of Phonetics,* 71, 113-125.

[13] Ferré, G. (2010). Timing relationships between speech and co-verbal gestures in spontaneous French. In *Language Resources and Evaluation, Workshop on Multimodal Corpora* (Vol. 6, pp. 86-91).

[14] Türk, O., & Calhoun, S. (2024). Phrasal synchronization of gesture with prosody and information structure. *Language and Speech*, 67(3), 702-743.

[15] Prieto, P., Esteve-Gibert, N., & Shattuck-Hufnagel, S. (2025). Towards a novel conceptualization of prosody that accounts for spoken and visual signals: The modality-neutral prosodic framework hypothesis. *Gesture*.

[16] Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica* (Since 2017 *Acta Linguistica Academica*), 55(3-4), 243-276.

[17] Rohrer, P. (2022). *A temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel MultiModal MultiDimensional (M3D) labeling system* (Doctoral dissertation, Nantes Université; Universitat Pompeu Fabra (Barcelone, Espagne)).

[18] Lehečková, E., Jehlička, J., & Králová Zíková, M. (2022). Multimodal marking of information structure: gesture-prosody alignment across languages. *Linguistica Pragensia*, 32(1), 19-38.

[19] Ambrazaitis, G., & House, D. (2016). Multimodal levels of prominence: the use of eyebrows and head beats to convey information structure in Swedish news reading. In *ISGS Conference 2016. 7th Conference of the International Society for Gesture Studies*. Paris, France, July 18-22, 2016 (pp. 319-319). New Sorbonne University Paris 3.

[20] Ambrazaitis, G., & House, D. (2017). Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication*, 95, 100-113.

[21] Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural viewpoint signals referent accessibility. *Discourse Processes*, 50(7), 431-456.

[22] Ebert, C., Evert, S., & Wilmes, K. (2011). Focus marking via gestures. In *Proceedings of Sinn und Bedeutung* (Vol. 15, pp. 193-208).

[23] Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180.

[24] Browman, C. P., & Goldstein, L. (1995). Dynamics and articulatory phonology. *Mind as motion: Explorations in the dynamics of cognition*, 175, 194.

[25] Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, 215-249.

[26] Pouw, W., Harrison, S. J., & Dixon, J. A. (2020). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, 149(2), 391.

[27] Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, 44, 62-82.

[28] Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory phonology*, 8(1).

[29] É. Kiss, K. (1998). Identificational focus versus information focus. *Language*, 74(2), 245-273.

[30] É. Kiss, K. (2016). Discourse functions: The case of Hungarian. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 663-685). Oxford University Press.

[31] Gyuris, B. (2012). The information structure of Hungarian. *The expression of information structure*, 159-186.

[32] Genzel, S., Ishihara, S., & Surányi, B. (2015). The prosodic expression of focus, contrast and givenness: A production study of Hungarian. *Lingua*, 165, 183-204.

[33] Mády, K. (2015). Prosodic (non-)realisation of broad, narrow and contrastive focus in Hungarian: A production and a perception study. *Proceedings of the International Conference on Speech Prosody* (ISCA).

[34] Mády, K., Kleber, F., Reichel, U. D., & Szalontai, A. (2016). *The interplay of prominence and boundary strength: a comparative study*.

[35] Mády, K., & Kleber, F. (2010). *Variation of pitch accent patterns in Hungarian.*

[36] Mycock, L. (2010). Prominence in Hungarian: the prosody–syntax connection. *Transactions of the Philological Society*, 108(3), 265-297.

[37] Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research,* 51(6), 1507-1521.

[38] Owoyele, B., Trujillo, J., De Melo, G., & Pouw, W. (2022). Masked-Piper: Masking personal identities in visual recordings while preserving multimodal information. *SoftwareX*, 20, 101236.

[39] Boersma, Paul & Weenink, David (2024*). Praat: doing phonetics by computer* [Computer program]. Version 6.4.06, retrieved from https://praat.org

[40] Kaland, Constantijn. (2021). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*. doi:10.1017/S0025100321000049

[41] R Core Team (2022*). R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.