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

Individual strategies in multimodal hyperarticulation

Alina Gregori

Goethe University Frankfurt
gregori@lingua.uni-frankfurt.de

Abstract

Hyperarticulation refers to speakers adapting their articulation to the communicative needs of their interlocutor. Thus, speakers enhance specific features to highlight elements or hypoarticulate and reduce them if the context allows. This was found for prosody, and recently also in prosody-gesture interaction [1], licensed by focus and background contexts. These findings show that hyperarticulation is not unimodal but can be found in multiple modalities. Further investigating the connection between modalities, this study addresses whether individual speakers use hyperarticulation strategies similarly to mark focus, whether they combine strategies or whether they prefer one to highlight information in focus. Drawing on data from [1], individual hyperarticulation strategies are examined in a German conversational corpus.

Results reveal that multimodal hyperarticulation is more stable across speakers than prosodic hyperarticulation is, as most speakers use the multimodal strategy, while less than 50% use the prosodic strategy. Importantly however, all speakers use at least one strategy to hyperarticulate in focus, with the modality strategies making up for each other if not both show hyperarticulation. This reinforces the conclusion that hyperarticulation is a multimodal phenomenon and that both unimodal and multimodal hyperarticulation strategies are needed to establish a comprehensive view on hyperarticulation in focus including individual strategies.

Index Terms: hyperarticulation, focus, individual strategies, prosody, gesture, multimodality

1. Introduction

Variation in the prosodic signal arises from the communicative context, but it also depends on the individual speaker [2]. Research on individual differences in German prosody has shown that speakers generally differ in prosodic prominence marking [2]. More specifically, they show differences in marking both information status [3] and focus [4], [5]. Focus is a domain of pragmatic prominence that involves choosing a referent from a set of alternatives based on the linguistic context [6]. This distinguishes focused constituents from unfocused background constituents and provides a setting to investigate individual strategies in contexts with different communicative goals. Overall, prosody has been shown to vary in focus and background contexts: focused referents are highlighted acoustically [7], [8]. In German, pitch accentuation and enhanced acoustic cues are known to be correlated with focus [9], [8]. As such, the slope of f0-falling accents is steeper in (contrastive) focus than it is in background contexts, see [1]. Other acoustic cues like intensity or duration are also involved in focus marking [10], [11]. Individual variation in this domain concerns the prosodic cues that speakers use to mark focus and to which extent they use each cue [2].

Communication is fundamentally multimodal, and it is expressed not only in text form or by the acoustic modality, but also by the visual modality and other channels [12]. Importantly, it is the combination and interplay of multiple modalities that makes for a comprehensive utterance. It is crucial to consider comparable rhythmic structures of the acoustic and visual modality and to relate them in their temporal properties since communication goes beyond the words that are said. Theories working on this domain of multimodality by investigating prosody-gesture interaction assume a close connection of prosody and gesture, while highlighting their independence [13], [14]. Nevertheless, both modalities have been shown to employ similar communicative functions, for which focus marking is one example. Therefore, investigating both prosody and gestures is important to explore individual strategies of focus marking.

Generally, gestures have been shown to occur more often [15], [16], [17] and to be produced more saliently [18] in focus contexts. Moreover, gesture apexes and pitch accents are temporally closer aligned in focus than in background, which points to a more precise alignment in the peaks of the two modalities [1], [19]. In terms of individual variation, gestures are used differently by speakers in terms of their occurrence frequency and saliency (see [20], [21] for a review). These variations are often discussed in relation to working memory or to characteristics in the discourse context. Research on individual variability in gestures' temporal properties and its relationship with prosody has not been conducted yet. However, given that individual speakers differ in the separate modalities, their interaction is likely to be affected by individual variation as well. It is important to investigate individual differences between speakers for a more comprehensive picture of the multimodal strategies involved in focus marking.

A paradigm that concerns the variation of communication is hypo- and hyperarticulation theory (H&H-theory, [22]), which states that speakers adapt their communicative signals to the needs of their interlocutor to achieve a balance between conveying relevant information and efficient articulation. Thus, speakers articulate stronger or more precisely to highlight important information and to reduce other aspects that are not critical to the current discourse. Hypo- and hyperarticulation are thus gradient and need to be evaluated in relation to each other. H&H-theory was established in the acoustic modality, operating on the segmental [22] or on the prosodic level [23], [24]. Recently, visual discourse-directed adaptations in gesture [1] and in sign language [25] research provided evidence for hyperarticulation being a multimodal phenomenon, involving prosodic modulation, as well as modulation of the interaction of the acoustic and visual modalities. H&H-theory and focus marking can be related to each other as both are characterized by the variation of communicative signals to convey different degrees of relevance in a conversation. However, the two paradigms have different origins and purposes, which makes

10.21437/SpeechProsody.2026-119

them operate orthogonally. Their similar characteristics deem a correlation probable and hyperarticulation constitutes a bridge between strategies, allowing to relate two modality-specific measures with properties that are not easily comparable.

The aim of this study is to address individual strategies in multimodal hyperarticulation in German focus marking. This evaluates whether and how speakers make different use of the acoustic and visual modalities. The present study makes use of previous analyses and findings in a paper submitted by the author [1]. There, both unimodal prosodic and multimodal prosody-gesture alignment hyperarticulation in focus were observed and the complementing roles of both strategies in distinguishing focus contexts were found. Prosodic hyperarticulation expressed by the steepness of f0-slopes of falling accents distinguishes information focus contexts from contrastive focus contexts with steeper slopes in contrastive focus. A multimodal analysis that examined the temporal alignment of pitch accents and gesture apexes distinguishes between background and focus contexts with a tighter alignment in focus. Multimodal effects were stronger than prosodic effects overall, but both strategies generally point towards hyperarticulation in focus contexts. This suggests that H&H-theory is a multimodal paradigm, as speakers use hyperarticulation in unimodal as well as multimodal strategies. In addition, the domain of focus licenses more precise articulation in the form of hyperarticulation in a unimodal as well as a multimodal setting.

This study aims to address whether individual speakers use both hyperarticulation strategies for marking focus, whether they use one of them preferably or whether some speakers do not hyperarticulate in focus at all. Individual analyses of prosodic and multimodal hyperarticulation strategies are conducted on German spontaneous speech and gesture data to investigate the following research question: *Do individual speakers use both unimodal and multimodal hyperarticulation in focus marking or do they prefer one strategy?* As focus is a pragmatic context that licenses an increase of communicative effort in terms of prosodic and gestural marking [18], a general pattern of hyperarticulation in focus is hypothesized for the majority of participants. However, since individual variation has been found in prosodic [2] and gestural [20] measures, variation in hyperarticulation strategies is expected. Finally, because multimodal hyperarticulation was found to be stronger than prosodic hyperarticulation in [1], multimodal hyperarticulation should be more stable than prosodic hyperarticulation, which may show more individual variation.

2. Methods

2.1. Data and participants

The database used in this study is the SaGA corpus [26], [27], which contains audio- and video-recordings of German spontaneous conversations. In the recordings, one speaker describes a tour through a virtual reality town environment to a second speaker who is then supposed to find their way through the town. 18 dialogues with audio- and video-data are included in the SaGA corpus, which makes 36 participants eligible for the individual analysis. Only those participants who provided datapoints in background and focus for both analyses (i.e., at least one falling pitch accent in the prosodic analysis and at least one apex-accent pair in the temporal alignment analysis in each condition) were considered for the individual analysis. This leads to 24 (10 female, 14 male) participants in the final sample.

2.2. Annotations and measures

Focus annotations were done following the LISA guidelines [28]. ELAN [29] was used to annotate gestures following M3D [30], annotating manual gesture strokes and apexes with and without referential properties. In Praat [31], pitch accents were annotated following GToBI [32]. Reliabilities for all annotations were high with Cohen's kappa (focus: N = 1370, k = 0.946, z = 40.4, p < 0.001; gesture: N = 421, k = 0.805, z = 18.6, p < 0.001; accent: N = 327, k = 0.75, z = 14.7, p < 0.001). Two analyses were run to test hyperarticulation strategies across modalities. For a unimodal prosodic analysis of falling pitch accents (H+L*), f0 was extracted at ten points per syllable using the Praat script ProsodyPro [33] to receive time-normalized f0-contours of the accented and its preceding syllable. From these points, f0-maxima and f0-minima were identified to calculate the slope of falling contours as sketched in Figure 1 (left). Falling accents were selected as they have shown to vary in f0-slopes in different focus contexts in Germanic languages [24], [1]. For a multimodal prosody-gesture alignment analysis, starting from timepoints of gesture apexes, the temporally closest pitch accent within the same focus domain was identified to form an apex-accent pair. This allows to derive the absolute distance between the two landmarks as a measure, see Figure 1 (right). The multimodal analysis is not restricted to a pitch accent type as the type itself is not decisive for assessing temporal alignment, but it is necessary to annotate to identify the correct timestamp.

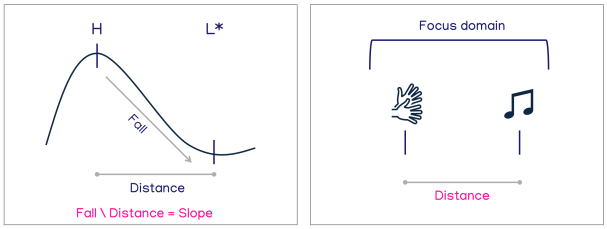

Figure 1: *Measures of the prosodic (left) and the multimodal (right) hyperarticulation strategy.*

2.3. Analysis procedure

Based on these annotations and measures, mean f0-slope and apex-accent distances were aggregated for each participant in background and focus to account for variation in the number of productions. For each strategy separately, the focus condition means were subtracted from one another to analyze whether speakers hypo- or hyperarticulated in focus and how big the difference in means between focus conditions was (section 3.1). This shows the strength of hyperarticulation per strategy. From that, the difference between condition means was z-transformed to 0 to enable comparability between strategies (which use different scales and units). Using the z-transformed differences, power relations between modality strategies were derived by subtracting the difference scores (section 3.2). These scores indicate which hyperarticulation strategy was used in focus marking and how big the power difference between strategies was. In the following, hyperarticulation refers to enhancement in focus (opposed to enhanced values in background). Bayesian binomial models with weakly informative priors were used for analysis. The models indicated no convergence issues. Visualization using *ggplot* [34] and inferential statistics using *brms* [35] were done in RStudio [36]. The results section first addresses each modality strategy separately, followed by comparisons between strategies.

3. Results

3.1. Individual hyperarticulation per modality

Figure 2 shows the means of each modality strategy per participant and focus condition. In the prosodic analysis (Figure 2, left plot), the f0-slope of falling accents is the measure of interest and hyperarticulation in focus is represented by steeper values of slope in focus than in background. Comparing background and focus means, no linear pattern of hyperarticulation in focus is found for all speakers. While most participants have slope means of up to 350 Hz/sec, few means range up to 450 Hz/sec, which are observed in background. Background means show a higher degree of variation than focus means, which in turn are more similar between speakers. All participants have different slopes in background and focus, as condition means differ. Therefore, participants either hypo- or hyperarticulate in focus rather than having a consistent slope in both contexts. Ten participants (41.7%) have steeper f0-slopes in focus, and 14 participants (58.3%) have steeper slopes in background. In short, some participants use prosodic hyperarticulation in focus, but more participants do not, which hints to speaker-dependent prosodic hyperarticulation.

In the multimodal analysis, the measure of interest is the temporal distance between pitch accents and gesture apexes in temporally associated pairs in background and focus, shown in Figure 2 (right plot). Since tighter alignment is interpreted as hyperarticulation, a higher distance in background and smaller distance in focus represents the hyperarticulated pattern. The mean distances per participant and condition show this pattern for all but one participant. 23 participants (95.8%) have smaller distance means in focus than in background, indicating a stable pattern of multimodal hyperarticulation in focus. Note that five participants (20.8%) only have slightly deviating means in background and focus, indicating only weak hyperarticulation, with background and focus having a more neutral alignment. In addition to lower distance means in focus, the variability of means is much lower in focus for all participants, with most of them ranging between 150-250 ms, while background distance means mostly range between 100-600 ms. Hence, hyperarticulation in focus was applied across participants using the multimodal strategy.

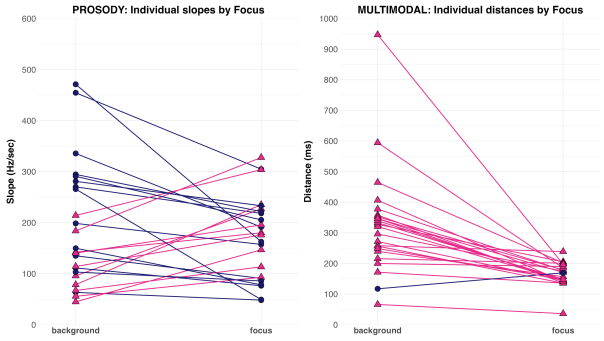

Figure 2: *Individual prosodic slope (left) and multimodal distance (right) means per focus condition and participant. Pink lines indicate hyperarticulation in focus, blue lines indicate hypoarticulation in focus.*

3.2. Modality strategy comparisons

To combine unimodal and multimodal hyperarticulation strategies for each participant, it is first assessed whether participants hypo- or hyperarticulated in focus across modality strategies. Figure 3 displays a matrix indicating whether or not a participant hyperarticulated in each modality in focus. The matrix shows that all participants hyperarticulate in focus in at least one modality, or in both, but no participant hypoarticulates in both modalities. This confirms the patterns already sketched in the individual strategy overviews in Figure 2. Nine speakers hyperarticulate using both the prosodic and the multimodal strategy (37.5%). 14 speakers only hyperarticulate multimodally (58.3%) and one speaker only hyperarticulates prosodically (4.2%). A Bayesian binomial model indicates a reliable difference between groups that use only one hyperarticulation strategy (coeff. = 2.37, 95%-CrI [0.95, 4.22]), with the multimodal strategy being more consistent. Importantly, nine speakers use both hyperarticulation strategies and no speaker uses none. If speakers do not hyperarticulate in one strategy, they do hyperarticulate in the other one.

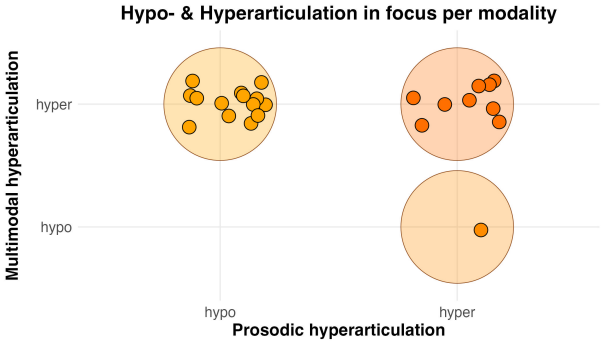

Figure 3: *Matrix of hyperarticulation strategy groups. 'hypo'/'hyper' indicate which strategy speakers chose in focus. X-axis refers to prosodic hyperarticulation; Y-axis refers to multimodal hyperarticulation.*

Figure 4 shows the direction of hyperarticulation per participant and strategy as well as strategy power differences. To get the power of each strategy, hypoarticulated values were subtracted from hyperarticulated ones. These were then z-transformed to a reference value of 0 to enable comparability between strategies. The normalized values were plotted in relation to each other per participant. The zero-line marks the turning-point between hypo- and hyperarticulation: points above zero indicate hyperarticulation in focus. The resulting plot in Figure 4 confirms that if a participant hypoarticulates in one strategy, the other strategy shows hyperarticulation. Seven pasrticipants (29.2%) show stronger hyperarticulation in the unimodal prosodic strategy, while 17 participants (70.8%) hyperarticulate stronger in the multimodal alignment setting. For reasons of plot readability, modality preferences are plotted in separate plots, but the data are to be interpreted together.

In addition, the extent of a difference between prosodic and multimodal hyperarticulation differs between participants, represented by the line colors in Figure 4. Eight speakers have a small difference in power between hyperarticulation strategies (diff. < 0.5; 33.3%), meaning that they use both strategies to a similar extent. Six speakers have a medium preference for one strategy (0.5 < diff. < 1; 25.0%), and ten speakers have a big preference for one hyperarticulation strategy (diff > 1; 41.7%). Notably, all ten participants with a big difference in strategy power hyperarticulate more using the multimodal strategy (right plot), making up for hypoarticulation in the prosodic strategy. This is visible as all strong difference preferences have hypoarticulation values in the prosodic modality and

hyperarticulation values in the multimodal one. In the small and medium difference groups, the number of participants preferring each strategy is more balanced. Notably, participants that use both strategies have a smaller difference in strategy power and generally less extreme hyperarticulation. A Bayesian binomial model confirms these observations: in the group with a strong preference, the multimodal strategy is preferred compared to the other two groups (small – big: coeff. = 5.68, 95%-CrI [1.42, 12.19]; medium – big: coeff. = 5.73, 95%-CrI [1.28. 12.48]). The small and medium preference groups do not indicate a unified preference for one strategy (small – medium: coeff. = 0.00, 95%-CrI [-2.26, 2.26]).

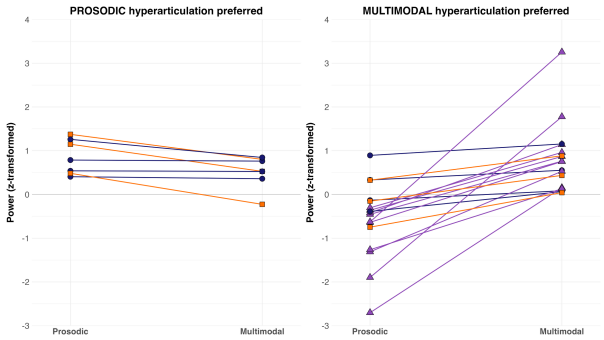

Figure 4: *Power relations between strategies per speaker. Left: prosodic strategy preferred; right: multimodal strategy preferred. Values above 0 indicate hyperarticulation in focus, values below 0 indicate hypoarticulation. Blue = small, orange = medium, purple = big preference.*

4. Discussion

This study analyzed individual strategies of unimodal and multimodal hyperarticulation in focus contexts of German spontaneous speech. Unimodal prosodic hyperarticulation in the steepness of f0-slopes in falling accents showed substantial inter-speaker variation. Multimodal hyperarticulation in terms of the temporal alignment of prosody and gesture showed a more stable pattern of hyperarticulation across speakers. All speakers used at least one, if not both, strategies to hyperarticulate in focus. This means that speakers used the strategies to make up for each other in case one was used only weakly or not in a contrasting fashion.

The first observation to be taken from the data is that hyperarticulation generally persists in the individual marking of focus across modalities. Speakers use hyperarticulation to set apart focus and background constituents, by increasing the precision of their prosodic and multimodal productions in focus. The degree of variation between background and focus means differs, as speakers show more variation in background than in focus means, which are more similar across participants and strategies. This is a sign for hyperarticulation in both strategies, and it also highlights the presence of prosodic hyperarticulation. However, less than 40% of speakers hyperarticulate in both strategies, meaning that more than 60% only use one strategy for hyperarticulation. In turn, 60% of speakers hypoarticulate in one modality strategy (mostly prosody), indicating that hyperarticulation is not a fully straightforward way of marking focus, but is subject to individual preference. This is in line with research on individual prosodic variation, which showed that speakers can differ in their preference for cues in prominence marking, see [2].

The prosodic strategy shows a much higher degree of inter-speaker variation than the multimodal strategy does. While the multimodal strategy is consistent across speakers, the prosodic strategy shows considerable variation, with less than half of the participants (41.7%) using prosodic hyperarticulation in terms of falling f0-slopes. However, seven participants (29.2%) preferred the prosodic strategy over the multimodal one, indicating that the prosodic strategy is relevant for focus marking. This individual analysis revealed that generally weaker prosodic hyperarticulation as found in [1] may be due to the individual variation. Further, prosodic hyperarticulation was found to be primarily used in contrastive contexts in [1], which were not distinguished from other focus types in this study. It is possible that more stable prosodic effects would emerge from the data if focus were split for contrast instead. Note as well that this study took falling f0-contours as the acoustic measure for prosodic hyperarticulation. It did not include other accent types and their occurrence, or prosodic cues like duration or intensity, which are involved in focus marking and may be subject to hyperarticulation in this context.

In this dataset, there are no speakers that use no hyperarticulation in focus contexts. 15 speakers only use one strategy, with most of them preferring the multimodal one. Importantly, hypoarticulation in one modality is made up for by the hyperarticulated one. Nine participants use both strategies, and they show a reliable application of hyperarticulation in focus, using both strategies to a similar extent. These speakers had a smaller power difference between both strategies. Note, that this study only considered speakers that used accentuation and gestures in both focus types, which means that the dataset does not contain speakers who avoided a strategy in any of the focus contexts. Not using a strategy in a focus context could be a sign of hypoarticulation that is not visible in this dataset.

Both the stability in speakers using both strategies, and the consistent use of at least one strategy highlight the importance to view hyperarticulation as a multimodal phenomenon. The analysis of multimodal hyperarticulation in [1] showed that both strategies are needed to get a comprehensive picture of hyperarticulation in German focus marking. Adding to that, the analysis of individual variation in the present study highlights the complementary role that prosodic and multimodal hyperarticulation strategies take.

5. Conclusions

This analysis of individual modality-specific strategies on hyperarticulation in focus marking reinforces that hyperarticulation is multimodal. The multimodal strategy not only proved to be stronger, but also more stable across speakers than the unimodal prosodic strategy. Still, both strategies need to be considered jointly, as the individual analyses showed that both make up for each other if hyperarticulation is not used in one strategy, such that focus is marked by hyperarticulation in at least one modality for every participant. Focus licenses both unimodal and multimodal hyperarticulation, and both strategies are needed to establish a comprehensive view on hyperarticulation including individual variation.

6. Acknowledgements

This research was funded by the German Research Foundation (DFG; SPP 2392 (Visual Communication); KU 2323/5-1/2). Thanks to Andy Lücking for providing the SaGA corpus and to Hans Rutger Bosker for sparking the idea.

7. References

[1] A. Gregori and F. Kügler, "Temporal alignment of prosody and gesture in focus: Unimodal and Multimodal Hyperarticulation," submitted.

[2] J. Lorenzen, "Individual variability in the encoding and decoding of prosodic prominence relations," PhD Thesis, Universität zu Köln, Köln, 2024.

[3] J. Lorenzen and S. Baumann, "Does communicative skill predict individual variability in the prosodic encoding of lexical and referential givenness?," in *Speech Prosody 2024*, ISCA, Jul. 2024, pp. 802–806. doi: 10.21437/SpeechProsody.2024-162.

[4] F. Cangemi, M. Krüger, and M. Grice, "Listener-specific perception of speaker-specific productions in intonation," in *Individual Differences in Speech Production and Perception*, vol. 3, S. Fuchs, D. Pape, C. Petrone, and P. Perrier, Eds., Peter Lang International Academic Publishers, 2015, pp. 123–145.

[5] M. Grice, S. Ritter, H. Niemann, and T. B. Roettger, "Integrating the discreteness and continuity of intonational categories," *J. Phon.*, vol. 64, pp. 90–107, Sep. 2017, doi: 10.1016/j.wocn.2017.03.003.

[6] M. Krifka, "Basic notions of information structure," *Acta Linguist. Hung.*, vol. 55, no. 3–4, pp. 243–276, Dec. 2008, doi: 10.1556/ALing.55.2008.3-4.2.

[7] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *J. Phon.*, vol. 36, no. 4, pp. 680–703, Oct. 2008, doi: 10.1016/j.wocn.2008.05.001.

[8] F. Kügler and S. Calhoun, "Prosodic Encoding of Information Structure: A typological perspective," in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds., Oxford University Press, 2020, pp. 453–467. doi: 10.1093/oxfordhb/9780198832232.013.30.

[9] S. Baumann and C. T. Röhr, "The perceptual prominence of pitch accent types in German.," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, Scotland, 2015.

[10] G. Kochanski, E. Grabe, J. Coleman, and B. Rosner, "Loudness predicts prominence: Fundamental frequency lends little," *J. Acoust. Soc. Am.*, vol. 118, no. 2, pp. 1038–1054, Aug. 2005, doi: 10.1121/1.1923349.

[11] S. Baumann and J. Lorenzen, "Boosting or inhibiting - how semantic-pragmatic and syntactic cues affect prosodic prominence relations in German," *PLOS ONE*, vol. 19, no. 4, p. e0299746, Apr. 2024, doi: 10.1371/journal.pone.0299746.

[12] P. Perniss, "Why We Should Study Multimodal Language," *Front. Psychol.*, vol. 9, p. 1109, Jun. 2018, doi: 10.3389/fpsyg.2018.01109.

[13] G. Ambrazaitis and D. House, "Probing effects of lexical prosody on speech-gesture integration in prominence production by Swedish news presenters," *Lab. Phonol.*, vol. 24, no. 1, Aug. 2022, doi: 10.16995/labphon.6430.

[14] G. Ambrazaitis and D. House, "The multimodal nature of prominence: some directions for the study of the relation between gestures and pitch accents," in *Proceedings of the 13th International Conference of Nordic Prosody*, Sciendo, 2023, pp. 262–273. doi: 10.2478/9788366675728-024.

[15] C. Ebert, S. Evert, and K. Wilmes, "Focus marking via gestures," in *Proceedings of Sinn und Bedeutung 15*, Saarbrücken, Germany, 2011, pp. 193–208.

[16] S. Im and S. Baumann, "Probabilistic relation between co-speech gestures, pitch accents and information status," *Proc. Linguist. Soc. Am.*, vol. 5, no. 1, pp. 685–697, Mar. 2020, doi: 10.3765/plsa.v5i1.4755.

[17] N. Esteve-Gibert, H. Lœvenbruck, M. Dohen, and M. D'Imperio, "Pre-schoolers use head gestures rather than prosodic cues to highlight important information in speech," *Dev. Sci.*, vol. 25, no. 1, p. e13154, Jan. 2022, doi: 10.1111/desc.13154.

[18] A. Gregori, P. G. Sánchez-Ramón, P. Prieto, and F. Kügler, "Prosodic and gestural marking of focus types in Catalan and German," in *Speech Prosody 2024*, ISCA, Jul. 2024, pp. 891–895. doi: 10.21437/SpeechProsody.2024-180.

[19] F. Kügler and A. Gregori, "Iconic Gestures in Focus – Synchronization of Prosody and Gestures in Prominence," in *Proceedings of ICPhS*, Prague, Czech Republic, 2023, pp. 4125–4129.

[20] M. Chu, A. Meyer, L. Foulkes, and S. Kita, "Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy.," *J. Exp. Psychol. Gen.*, vol. 143, no. 2, pp. 694–709, Apr. 2014, doi: 10.1037/a0033861.

[21] D. Özer and T. Göksun, "Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources," *Front. Psychol.*, vol. 11, p. 573555, Nov. 2020, doi: 10.3389/fpsyg.2020.573555.

[22] B. Lindblom, "Explaining Phonetic Variation: A Sketch of the H&H Theory," in *Speech Production and Speech Modelling*, W. J. Hardcastle and A. Marchal, Eds., Dordrecht: Springer Netherlands, 1990, pp. 403–439. doi: 10.1007/978-94-009-2037-8_16.

[23] K. J. De Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 491–504, Jan. 1995, doi: 10.1121/1.412275.

[24] J. Hanssen, J. Peters, and C. Gussenhoven, "Prosodic effects of focus in Dutch declaratives," in *Proceedings of Speech Prosody 2008*, ISCA, 2008, pp. 609–612. doi: 10.21437/SpeechProsody.2008-138.

[25] J. Mertz, L. Pagel, P. Perniss, G. Turco, and D. Mücke, "Coarticulation in sign language: A kinematic study on French Sign Language (LSF) using Electromagnetic Articulography (EMA)," in *ISSP 2024 - 13th International Seminar on Speech Production*, ISCA, May 2024, pp. 51–54. doi: 10.21437/issp.2024-14.

[26] A. Lücking, K. Bergmann, F. Hahn, S. Kopp, and H. Rieser, "The Bielefeld speech and gesture alignment corpus (SaGA)," in *LREC 2010 workshop: Multimodal corpora–advances in capturing, coding and analyzing multimodality*, M. Kipp, J.-P. Martin, P. Paggio, and D. Heylen, Eds., 2010, pp. 92–98.

[27] A. Lücking, K. Bergman, F. Hahn, S. Kopp, and H. Rieser, "Data-based analysis of speech and gesture: the Bielefeld Speech and Gesture Alignment corpus (SaGA) and its applications," *J. Multimodal User Interfaces*, vol. 7, no. 1–2, pp. 5–18, Mar. 2013, doi: 10.1007/s12193-012-0106-8.

[28] M. Götze *et al.*, "Information structure," *Interdiscip. Stud. Inf. Struct. ISIS*, no. 7, pp. 147–187, 2007.

[29] *ELAN [Computer software]*. (2024). The Language Archive, Nijmegen: Max Planck Institute for Psycholinguistics. [Online]. Available: https://archive.mpi.nl/tla/elan

[30] P. L. Rohrer *et al.*, "Multidimensional Labeling of Gesture in Communication: the M3D Proposal," *Corpus Pragmat.*, Jun. 2025, doi: 10.1007/s41701-025-00197-2.

[31] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*. (2024). [Online]. Available: http://www.praat.org/

[32] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in autosegmental-metrical phonology.," *Prosodic Typology Phonol. Intonation Phrasing*, pp. 55–83, 2005.

[33] Y. Xu, "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7–10.

[34] H. Wickham, *ggplot2: elegant graphics for data analysis*, Second edition. in Use R! Cham: Springer international publishing, 2016.

[35] P.-C. Bürkner, "brms: An R Package for Bayesian Multilevel Models Using Stan," *J. Stat. Softw.*, vol. 80, no. 1, 2017, doi: 10.18637/jss.v080.i01.

[36] R Core Team, *R: A Language and Environment for Statistical Computing. [Programming language]*. (2024). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/