



# Prosodic and gestural marking of focus types in Catalan and German

Alina Gregori<sup>1\*</sup>, Paula G. Sánchez-Ramón<sup>1,2\*</sup>, Pilar Prieto<sup>3,2</sup> & Frank Kügler<sup>1</sup>

<sup>1</sup>Goethe University Frankfurt, <sup>2</sup>Universitat Pompeu Fabra, <sup>3</sup>ICREA

\*Shared first authorship

gregori@lingua.uni-frankfurt.de, paulaginesa.sanchez@upf.edu

## Abstract

Prosody and gesture coordinate with each other in speech. This study investigates the contribution of prosody and gesture to the marking of focus types comparing Catalan and German. We hypothesize that multimodal prominence increases along increasing layers of pragmatic meaning in focus, namely background (no focus) < information < contrastive < corrective. To test this, we conducted a semi-spontaneous production experiment in Catalan and German, that systematically varied eliciting contexts for each focus type, in which participants (n per language = 15) interacted with a digital character. Target focused adjectives (791 items) were annotated for pitch accentuation and gesture presence, as well as degrees of perceived prosodic and gestural prominence.

Results suggest that while target adjectives are systematically produced with pitch accentuation across focus types in both languages, the number of head gestures tends to significantly increase with focus pragmatic strength in both languages. Crucially, a significant positive correlation was found between focus types and degrees of perceived prosodic and gestural prominence in both languages. In short, an increasing multimodal marking was observed across focus types, in terms of number of gestures produced and perceived prosodic and gestural prominence, which indicates an integrated behavior of prosody and gesture in speech.

**Index Terms:** Multimodality, Focus, Gesture-Prosody link, Prominence, semi-controlled production

## 1. Introduction

Multimodal research builds on the premise that speech and gestures constitute an integrated system in communication (e.g., [1], [2] among others). This can be observed in the temporal coordination between speech and gesture (cf. [3]) or through prominence patterns (e.g., [4]). The present cross-linguistic study examines the influence of different focus types on the prosody-gesture link in Catalan and German, two languages with distinct prosodic systems.

An assessment of the multimodal marking of focus types will need to target both acoustic and visual cues to prominence. In linguistic terms, prominence can be interpreted as a relational property that structures an element within its context ([5]). In general, greater linguistic prominence is expected to be emphasized in speech e.g., by the use of increasing values of acoustic cues that signal prominence ([6], [7], [8], [9]). A few studies have started to explore methods to pin down the cues that contribute to perceived prosodic prominence (e.g., [10]), showing that intonation (pitch accentuation) is an important reflection of prosodic prominence. For gesture, the term of gestural prominence refers to the saliency of a communicative bodily movement and is associated with its kinematic cues ([11]). In addition, [12] found that alignment patterns between pitch

accentuation and gestures were used to mark prominence, but that the relation between prosodic and gestural prominence goes beyond pitch accentuation. In the present study, we investigate the multimodal marking of linguistic prominence in the form of focus constituents cross-linguistically in two languages with distinct prosodic systems.

We understand focus as a cognitive domain of information structure that “indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions” ([13]:247). While certain constituents in a sentence are focused, others are so-called “background” constituents ([13]). Focused constituents can display different layers of pragmatic meaning, which can be categorized in different *focus types*. “Information focus” commonly applies to the most important information in a sentence ([13]). If an additional layer of pragmatic meaning is added, constituents can evoke a notion of contrast, which is usually called “contrastive focus”. Adding another layer of disagreement to a previous statement in contrastive focus, results in the expression of “corrective focus” ([14]). The more layers of pragmatic meaning (conceived here as a scale of pragmatic strength) a focus constituent carries, the more communicative effort is expected to be taken to express this constituent.

Research on the relationship between information structure and prosody has shown that focused information is marked by higher levels of prosodic prominence ([15]), while background constituents tend to be prosodically reduced or unaccented (e.g., [6], [7] for German). Among focus types, contrastive or corrective focus constituents have been suggested to be marked by higher levels of prosodic prominence (e.g., [6], [16], [17], [18]) compared to information focus. For instance, [19] suggest that higher *f0* is perceived as more prominent and developed a “pitch accent prominence scale” for German. Additionally, in Romance languages, focused information has been found to be located in the right periphery of the sentence (e.g., [20], [21], [22]). Typologically, Catalan and German differ in their rhythm class and their accentuation patterns ([23], [24]).

Recent studies assessing the relationship between information structure and gesture have shown that gestures performed by different articulators (e.g., head nods, eyebrows movements or hand movements) act as discourse markers, such as focus (e.g., [1], [25], [26], [27]). Gestures have also been shown to play a role in the marking of information status, such that new referents are more likely to be accompanied by a gesture than given referents (e.g., [28], [29], [30], [31]). Studies that have jointly assessed prosody, gesture and information status have also found that new information is marked by a higher number of gestures and pitch accents than given information, while indicating an independent contribution of gesture to information status marking ([4], [32], [33]).

A few studies have considered focus and the prosody-gesture link. For example, [34] found that gestures align more closely with pitch accents on focused constituents than on non-focus constituents in German. In children’s narrative speech,

manual gestures were reported to associate more often with focus than with background constituents in Catalan ([35]). On the perception side, [36] found that head nods and eyebrow movements are essential in the marking of contrastive focus in Catalan (see also [23], [37] for production), pointing towards an inter-layered contribution of pitch accentuation prominence and gestural prominence to the perception of contrastive focus. Finally, the only study to our knowledge that has addressed the multimodal marking of focus types is [38] on 24 French-speaking children. The authors found that head gestures were significantly used to mark focus, with the highest number of gestures occurring in corrective focus, and the lowest number of gestures occurring in broad focus. They also found that focused words co-occurring with gestures had a longer syllable duration and a wider pitch range compared to focused words produced without gestures.

All in all, the previous studies provide evidence of a potential unique contribution of gesture in the marking of information structure. Given that this evidence is mostly based on the presence of gesture and pitch accentuation, the present study has the goal of broadening this description by investigating the role that prosody and gesture take in the marking of focus types by not only assessing prosodic and gestural cues, but also including the perceived levels of prosodic and gestural prominence. Thus, we address the following research question: Are focus types reflected in multimodal prominence in Catalan and German? We hypothesize that focus types do reflect gestural and prosodic prominence, meaning that multimodal prominence should increase along focus types with increasing layers of pragmatic meaning: Information focus < Contrastive focus < Corrective focus. This is expected for both Catalan and German since, despite their different prosodic patterns of focused NPs (cf. [23]), in this study we examine the focused element, which usually receives the strongest accent in both languages.

## 2. Method

### 2.1. Participants

In total, 30 healthy adults, 15 per language, were recorded in Barcelona and Frankfurt (Catalan-Spanish bilinguals, mean age = 21,06; 3m, 12f; German, mean age = 28,8; 6m, 9f). 93,3% of the Catalan speakers reported themselves as being Catalan-language dominant. None of the speakers reported any speech or hearing disorders. All were paid 10 € for participation.

### 2.2. Procedure

We adapted the focus elicitation task by [38] to adult speakers to elicit four conditions (background (no focus), information focus, contrastive focus or corrective focus). To encourage the use of target productions, participants are asked to naturally interact with a digital character that is a language learner (named Maria) and to speak using short instructions. In each trial, the speaker instructs Maria to take a certain colored item from her bag. All elicited sentences contain a noun (target object) and an adjective (color of the object), and the narrow focus is on the color adjective in each sentence. In order to elicit the different focus types, the target item and competitor items in Maria's bag change. See Figure 1 for an example of the contrastive and corrective focus conditions. Participants see a context picture in which the target object is introduced, and afterwards, they see the bag and are prompted to produce a sentence instructing Maria to take the correct item. Maria either takes the correct object (information and contrastive focus) or makes a mistake and

needs to be corrected (corrective focus) before moving on. In information focus, there is no competitor item in the bag. In the case of the background condition, the focus is shifted from the object to a certain action to make the object non-focused.

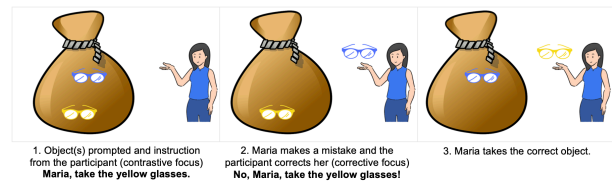


Figure 1: Example trial of the study – contrastive and corrective focus item, preceded by a context slide.

Participants were recorded using a camera (Sony HDR-CX625 Camcorder) and an external microphone (RODE Wireless Go for Catalan, Sennheiser MD46 for German), sitting on a high chair and looking at an electronic visual display.

### 2.3. Data coding and analysis

Data were annotated in Praat ([39]) and ELAN ([40]). In a first step, target noun phrases (TP) were identified, segmented into adjective and noun as well as their accented syllables. For acoustic annotation, we labelled ToBI pitch accents types in the TPs for each language (using GToBI [41] and CatToBI [36]) as well as prosodic prominence of nouns and adjectives on a scale from 0 - 3 (following DIMA [42]). For gestural annotation, we identified manual gesture strokes, head movements and eyebrow movements overlapping with the TP following the M3D system ([11]). Gestural prominence was annotated on each adjective and noun on a scale from 0 (low prominence) to 3 (high prominence) ([11]). In case there were no gestures occurring in the word, this was annotated as “no gesture”.

For the current analysis, we investigated the multimodal cues occurring on the adjective (focused item) of the TP. We first extracted the number of acoustic and visual prominence cues - pitch accents, manual gestures, head movements and eyebrow movements occurring on the adjectives of each focus type. Then, we extracted the perceived prosodic prominence and perceived gestural prominence ratings per focus type. We analyzed the data inferentially with R ([43]). To determine the relation between focus types and prosodic and gestural variables, 12 Spearman correlation tests were applied. Spearman tests were set to analyze the relation between the focus type ordinal variable (4 levels) and each of the ordinal variables, namely “pitch accent type” (6 levels), “hand gesture” (2 levels), “head movement” (2 levels), “eyebrow movement” (2 levels), “prosodic prominence” (4 levels) and “gestural prominence” (5 levels). Correlation coefficients and their significance are reported below.

## 3. Results

### 3.1. Presence of prosodic and gestural cues

The results presented in this section comprise the analysis of 791 target adjectives within the focused noun phrases (395 in Catalan and 396 in German) across the four conditions, which are ordered from weakest to strongest multi-layered focus type (background < information focus < contrastive focus < corrective focus). We assess the counts of prosodic cues (number of pitch accents, pitch accent types) and gestural cues (hand gestures, head movements, eyebrow movements) that occur on the target adjectives in the four focus conditions.

Regarding pitch accent presence, in both languages every target adjective in all focus types was accented (with the exception of two data points in German). Figure 2 shows the distribution of pitch accent types across focus conditions.

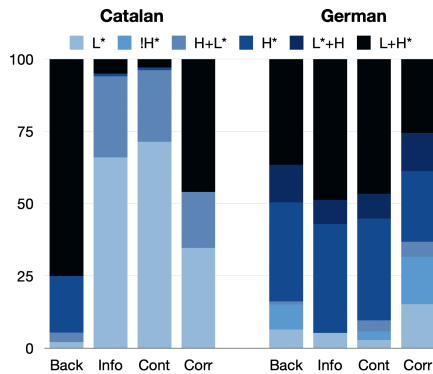


Figure 2: Distribution of pitch accent types per condition, split by language.

An inspection of the data plotted in Figure 2 reveals no clear pattern of pitch accents across focus types, neither for Catalan nor for German. For example, the most prominent pitch accent types (e.g., L+H\* and H\* cf. [19]) do not seem to represent the most frequent types in the pragmatically strongest focus types. In Catalan, the most prominent pitch accents primarily occur in Background, followed by Corrective focus, while Information and Contrastive focus are mainly marked by less prominent accents like L\*. In German, the pitch accent distribution across conditions is similar, although Information and Contrastive focus more often receive more prominent pitch accents than Background and Corrective focus. Statistically, the two Spearman correlation tests show a weak significant negative correlation between pitch accent types and focus types for the two languages (Catalan:  $\rho = -0.24$ ,  $p < 0.0001$ ; German:  $\rho = -0.13$ ,  $p < 0.05$ ).

The distribution of gestural cues per focus condition (see Table 1) shows a tendency for all cues to increase in number with increasing layers of pragmatic meaning in focus types in both Catalan and German. Starting with hand gestures, Catalan speakers used a total of 101 hand gestures across conditions. They used roughly the same number of hand gestures in Background and Information focus, but the number increases from Information focus to Corrective focus, leading to a weak positive correlation between hand gestures and focus types ( $\rho = 0.13$ ,  $p < 0.05$ ). German speakers used 151 hand gestures in total, for which the proportion increased from Background to Contrastive focus, but not to Corrective focus and the correlation is not significant ( $\rho = 0.05$ ,  $p = 0.28$ ). Concerning head movements, Catalan speakers used 84 head movements in the data set, which are positively correlated with increasing pragmatic strength of the focus types ( $\rho = 0.19$ ,  $p < 0.0001$ ). German speakers used head movements more frequently with a total number of 339. The use of head movements in German increases from Background to Contrastive focus but decreases for Corrective focus, leading to a significant but very weak positive correlation between focus type and head movement in German ( $\rho = 0.01$ ,  $p < 0.05$ ). Finally, Catalan speakers used 25 eyebrow movements in total, which do not show a clear significant increasing pattern of occurrence across focus types, although they used the least eyebrow movements in Background and the most in Corrective focus ( $\rho = 0.03$ ,  $p = 0.54$ ). In the German data, 54 eyebrow movements were produced. The occurrence of

eyebrows increased from Background to Contrastive focus, but decreased for Corrective focus, as observed in the other gestural cues. The correlation is not significant ( $\rho = 0.00$ ,  $p = 0.99$ ).

Table 1: Proportions (%) of gestural cues per condition (B = Background, I = Information focus, C = Contrastive focus, R = Corrective focus), split by language.

Cue	Catalan				German			
	B	I	C	R	B	I	C	R
Hand	19,8	18,8	26,7	34,7	17,9	29,1	29,8	23,2
Head	9,5	25	27,4	38,1	15,6	28,9	31,6	23,9
Eyeb.	20	28	20	32	18,5	27,8	38,9	14,8

While in general the number of gestural cues show a more consistent tendency to increase along focus types than pitch accentuation presence or type, neither the presence of eyebrow movements in Catalan nor hand gestures or eyebrow movements in German seem to behave as primary predictors of focus type. Only the presence of head movements shows a positive correlation with focus types in both languages. Since the joint assessment of multimodal cues shows systematic pitch accent presence and inconclusive use of pitch accent types as well as tendencies for focus marking by gestural cue presence, this highlights the need for adding analyses of prosodic and gestural prominence measures that go beyond pitch accentuation (in line with [12]) and gesture presence, which are provided in sections 3.2 and 3.3.

### 3.2. Perceived Prosodic Prominence

The stacked bar chart in Figure 3 shows the distribution of perceived prosodic prominence levels (0-1-2-3) across focus types for Catalan (left) and German (right). A visual inspection of the data shows a positive correlation in the expected direction of increased prominence for pragmatically stronger focus types, with small differences across languages. In Catalan it shows similarity across Background, Information, and Contrastive focus, but a strong increase of high prominence ratings for the corrective focus compared to the other focus types. In German, an increase can be observed across Background, Information and Contrastive focus, while a more similar picture is shown between the Contrastive and Corrective conditions. Spearman correlation tests reveal a significant positive relation between levels of prosodic prominence ratings and increased layers of pragmatic meaning for both languages (Catalan:  $\rho = 0.29$ ,  $p < 0.0001$ ; German:  $\rho = 0.33$ ,  $p < 0.0001$ ).

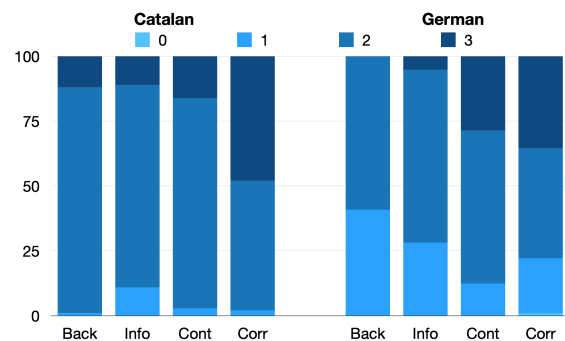


Figure 3: Perceived prosodic prominence ratings (0-3) per condition, split by language.

### 3.3. Perceived Gestural Prominence

The stacked bar chart in Figure 4 shows the distribution of perceived gestural prominence levels (0-1-2-3) across focus types. Again, a visual inspection of the data shows a positive correlation in the expected direction of increased prominence for pragmatically stronger focus conditions. A similar trend can be observed for both languages. Results of two Spearman correlation tests show a significant positive relation between levels of gestural prominence ratings and increased layers of pragmatic meaning for both languages (Catalan:  $\rho = 0.20$ ,  $p < 0.0001$ ; German:  $\rho = 0.12$ ,  $p < 0.05$ ).

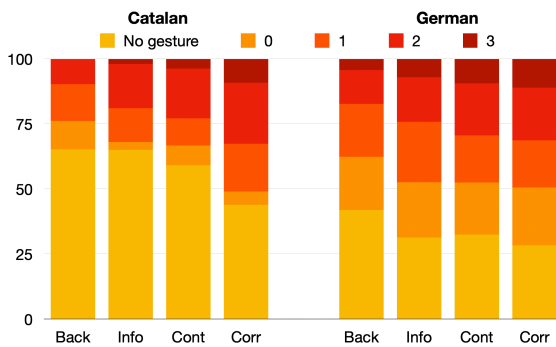


Figure 4: Perceived gestural prominence ratings (0-3) per condition, split by language.

## 4. Discussion

This multimodal production study aimed to assess the prosodic and gestural marking of focus types cross-linguistically in Catalan and German. We developed a study design and elicitation method that allow for a semi-controlled production of sentences according to the expected focus types while leaving room for spontaneous behavior by mimicking an interaction as natural but controlled as possible. Importantly, the interactive setting of teaching a digital character led to a high number of spontaneous target adjective productions (94,2%) while at the same time eliciting a frequent use of gestures.

Our results on multimodal cues showed that both pitch accent presence and pitch accent types did not indicate increasing layers of pragmatic meaning in any of the languages. With regard to gestural cues, only the occurrence of head movements showed a positive correlation with strength of the focus type in both languages (in line with [38]), in addition to hand gestures in Catalan. However, the mere presence of pitch accentuation and gestural cues do not seem to be a consistent marker of focus, which is why an assessment of the patterns of perceived prosodic and gestural prominence was conducted. Focusing on those results, crucially, we found a positive correlation between focus types and an increase in the ratings of perceived prosodic and gestural prominence in both languages (with prosodic prominence showing a stronger correlation). This result is directly relevant to our main research question, namely, whether focus types are reflected in multimodal prominence in Catalan and German. We can thus say that perceived multimodal prominence (assessed separately for gesture and prosody) increases along layers of pragmatic meaning in focus in both languages.

The dataset of this study has allowed us to assess the main research question of the study and offers many possibilities for further studies on the prosody-gesture link in semi-spontaneous production. Nevertheless, the elicitation method also has its limitations. The teaching scenario might have an influence on

the expressiveness of prosody and gesture, leading to higher accentuation rates. An example of that could be the background sentences in our data, which had a higher-than-expected accentuation rate and a similar prominence to information focus due to its prenuclear position, especially in Catalan. Further, politeness as a factor might have influenced the degree of accentuation of some of the focus conditions (especially corrective focus). That is, speakers might have expressed correction in a restrained way due to wanting to be polite to a language learner, even though they were instructed to use short and precise sentences. A second limitation of the study is the novel use of the perceived prominence scales in both languages, which would need to be further explored. Regarding prosody, we applied a prominence rating system (DIMA [42]) that has been used for German, but not for Catalan before. Thus, the rating system may need more fine-grained graduation in prominence levels for Catalan. The gestural prominence rating system ([11]) is a rather new system which may need more empirical testing in order to offer specific guidelines. We are currently conducting inter-rater reliability analyses for both scales in both languages.

While this study provides an insightful first assessment on prosodic and gestural marking of focus types in Catalan and German, more research is needed to get a comprehensive overview of multimodal focus marking. Assessing the joint contribution of prosody and gesture is important not only for focus marking but also for a more detailed examination of the prosody-gesture link. In addition, further research will be needed to assess precise acoustic features (e.g.,  $f_0$ , intensity, duration) and kinematic features (e.g., amplitude, velocity, position in space, combination of gestural articulators) that contribute to the perception of multimodal prominence in general. Further, from a language-comparison point of view, it would be relevant to assess prosodic and gestural behavior not only on the focused adjective, but also across the whole NP or even bigger domains, where language-specific differences are expected to be found with regard to rhythm and de-accentuation patterns ([23]).

## 5. Conclusions

This study assessed the prosodic and gestural marking of focus types in Catalan and German. We developed a method that comprises a semi-controlled setting and the elicitation of multimodal data, and we established a database for further studies investigating focus and the prosody-gesture link. While the use of pitch accents and gestures were not stable markers of the different focus types in neither the Catalan nor the German data, perceived prominence ratings for both prosodic and gestural prominence showed positive correlations with increasing layers of pragmatic meaning in the form of focus types (information < contrastive < corrective) in the two languages. Concerning the prosody-gesture integration, our data shows that both prosodic prominence and gestural prominence increase across focus types which may suggest a related but independent behavior of prosody and gesture in speech.

## 6. Acknowledgements

We acknowledge the financial support of the German Research Foundation (DFG; SPP 2392 (ViCom); KU 2323/5-1 & PR 1780/1-1), of the Generalitat de Catalunya [2021 SGR 00922] and the Spanish Ministry of Science, Innovation and Universities [PID2021-123823NB-I00: “Multimodal Communication”, MICIU/AEI/10.13039/501100011033].

We thank this study’s participants and Leoni Hahn, Jonas Pfeffer and Carla Rufi for their help with data annotation.

## 7. References

- [1] D. McNeill, *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.
- [2] S. Kita, “How representational gestures help speaking,” *Lang. Gesture*, vol. 1, pp. 162–185, 2000.
- [3] S. Shattuck-Hufnagel and A. Ren, “The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech,” *Front. Psychol.*, vol. 9, p. 1514, 2018.
- [4] 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, 2020.
- [5] M. Grice and F. Kügler, “Prosodic Prominence – A Cross-Linguistic Perspective,” *Lang. Speech*, vol. 64, no. 2, pp. 253–260, Jun. 2021.
- [6] S. Baumann, M. Grice, and S. Steindamm, “Prosodic marking of focus domains-categorical or gradient,” in *Proceedings of speech prosody*, Dresden Germany, 2006, pp. 301–304.
- [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, 2008.
- [8] J. E. G. Hanssen, J. Peters, and C. Gussenhoven, “Prosodic effects of focus in Dutch declaratives,” 2008.
- [9] 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.
- [10] J. Cole and S. Shattuck-Hufnagel, “New methods for prosodic transcription: Capturing variability as a source of information,” *Lab. Phonol.*, vol. 7, no. 1, 2016.
- [11] P. L. Rohrer *et al.*, “The MultiModal MultiDimensional (M3D) labeling system,” 2023, doi: 10.17605/OSF.IO/ANKDX.
- [12] P. L. Rohrer, E. Delais-Roussarie, and P. Prieto, “Visualizing prosodic structure: Manual gestures as highlighters of prosodic heads and edges in English academic discourses,” *Lingua*, vol. 293, p. 103583, 2023.
- [13] M. Krifka, “Basic notions of information structure,” *Acta Linguist. Hung.*, vol. 55, no. 3–4, pp. 243–276, 2008.
- [14] S. Repp, “Contrast: Dissecting an elusive information-structural notion and its role in grammar,” 2014.
- [15] F. Kügler and S. Calhoun, “Prosodic encoding of information structure: A typological perspective,” 2020.
- [16] M. del Mar Vanrell, A. Stella, B. G. Fivela, and P. Prieto, “Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus,” *J. Int. Phon. Assoc.*, vol. 43, no. 2, pp. 195–220, 2013.
- [17] F. Kügler and A. Gollrad, “Production and perception of contrast: The case of the rise-fall contour in German,” *Front. Psychol.*, vol. 6, p. 1254, 2015.
- [18] M. Zimmermann, “Contrastive focus and emphasis,” *Acta Linguist. Hung. 2017 Acta Linguist. Acad.*, vol. 55, no. 3–4, pp. 347–360, 2008.
- [19] S. Baumann and C. T. Röhr, “The perceptual prominence of pitch accent types in German,” in *ICPhS*, 2015.
- [20] A. Dufter and C. Gabriel, “14 Information structure, prosody, and word order,” *Man. Gramm. Interfaces Roman.*, vol. 10, p. 419, 2016.
- [21] E. Vallduví and E. Engdahl, “The linguistic realization of information packaging,” *ling*, vol. 34, no. 3, pp. 459–520, 1996.
- [22] M. L. Zubizarreta, “Prosody, focus, and word order,” *No Title*, 1998, Accessed: Jan. 10, 2024.
- [23] E. Krahmer and M. Swerts, “The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception,” *J. Mem. Lang.*, vol. 57, no. 3, pp. 396–414, 2007.
- [24] J. Cole, J. I. Hualde, C. L. Smith, C. Eager, T. Mahrt, and R. N. de Souza, “Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish,” *J. Phon.*, vol. 75, pp. 113–147, 2019.
- [25] A. Kendon, “Gesticulation and speech: Two aspects of the process of utterance,” *Relatsh. Verbal Nonverbal Commun.*, vol. 25, no. 1980, pp. 207–227, 1980.
- [26] C. Ebert, S. Evert, and K. Wilmes, “Focus marking via gestures,” in *Proceedings of Sinn und Bedeutung*, 2011, pp. 193–208.
- [27] D. P. Loehr, “Temporal, structural, and pragmatic synchrony between intonation and gesture,” *Lab. Phonol.*, vol. 3, no. 1, Jan. 2012.
- [28] S. Debreslioska, A. Özyürek, M. Gullberg, and P. Perniss, “Gestural Viewpoint Signals Referent Accessibility,” *Discourse Process.*, vol. 50, no. 7, pp. 431–456, Oct. 2013.
- [29] S. Debreslioska and M. Gullberg, “Discourse Reference Is Bimodal: How Information Status in Speech Interacts with Presence and Viewpoint of Gestures,” *Discourse Process.*, vol. 56, no. 1, pp. 41–60, Jan. 2019.
- [30] S. Debreslioska and M. Gullberg, “What’s New? Gestures Accompany Inferable Rather Than Brand-New Referents in Discourse,” *Front. Psychol.*, vol. 11, 2020.
- [31] P. Perniss and A. Özyürek, “Visible Cohesion: A Comparison of Reference Tracking in Sign, Speech, and Co-Speech Gesture,” *Top. Cogn. Sci.*, vol. 7, no. 1, pp. 36–60, 2015.
- [32] P. Rohrer, “A temporal and pragmatic analysis of gesture-speech association: a corpus-based approach using the novel MultiModal MultiDimensional (M3D) labeling system,” PhD Thesis, Nantes Université; Universitat Pompeu Fabra (Barcelona, Espagne), 2022.
- [33] S. Muñoz-Coego, J. Florit-Pons, P. L. Rohrer, I. Vilà-Giménez, and P. Prieto, “The prosodic and gestural marking of the information status of referents in children’s narrative speech: A longitudinal study,” *Proc Speech Prosody 2022*, pp. 401–405, 2022.
- [34] F. Kügler and A. Gregori, “ICONIC GESTURES IN FOCUS-SYNCHRONIZATION OF PROSODY AND GESTURES IN PROMINENCE”, 2023.
- [35] P. L. Rohrer, J. Florit-Pons, I. Vilà-Giménez, and P. Prieto, “Children use non-referential gestures in narrative speech to mark discourse elements which update common ground,” *Front. Psychol.*, vol. 12, p. 661339, 2022.
- [36] P. Prieto Vives, C. Puglesi, J. M. Borràs Comes, E. Arroyo, and J. Blat, “Exploring the contribution of prosody and gesture to the perception of focus using an animated agent,” *J. Phon. 2015 Mar 49 41-54*, 2015.
- [37] B. Roustan and M. Dohen, “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*, 2010, pp. 100110–1.
- [38] N. Esteve-Gibert, H. Løvenbruck, 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.
- [39] P. Boersma and D. Weenink, “Praat: doing phonetics by computer [Computer program].” 2024. [Online]. Available: <http://www.praat.org/>
- [40] “ELAN [Computer software].” Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive, 2023. [Online]. Available: <https://archive.mpi.nl/tla/elan>
- [41] M. Grice, S. Baumann, and R. Benzmüller, “German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic typology. The phonology of intonation and phrasing* (pp. 55–83).” Oxford University Press, 2005.
- [42] F. Kügler, S. Baumann, and C. T. Röhr, “Deutsche Intonation, Modellierung und Annotation (DIMA),” *Transkription Annot. Gespr. Sprache Multimodaler Interakt. Konzepte Probl. Lösungen*, p. 23, 2022.
- [43] R Core Team, “R: A Language and Environment for Statistical Computing. [Programming language].” Vienna, Austria., 2013.