Alignment of beat gestures and prosodic prominence in German

Sophie Repp[1], *Lara Muhtz*[1], *Johannes Heim*[2]

[1]University of Cologne
[2]University of Aberdeen
sophie.repp@uni-koeln.de, lara.muhtz@gmx.de, johannes.heim@abdn.ac.uk

Abstract

We present evidence on the alignment of beat gestures and prosodic prominence from a video corpus consisting of six German educational videos for students from six presenters. Our analysis of 120 beat gestures (with a substantial variety of hand shapes) shows that beat gestures almost always align with prosodically prominent syllables, i.e., syllables carrying a pitch accent. Specifically, the stroke always starts before, or – more often – on, a pitch-accented syllable; the apex mostly falls on the accented syllable (74%) but may also occur in subsequent syllables. The degree of prosodic prominence of the accented syllable (in terms of DIMA-prominence levels) is predictive for the position of the apex, which occurs within rather than after the accented syllable more often for higher degrees of prominence. These findings provide new insights into the alignment of prominence-lending features of prosody and gesture, thereby broadening the empirical landscape for beat gestures.

Index Terms: beat gestures, prosodic prominence, alignment, prominence level

1. Introduction

Speech prosody and gesture show similarities regarding phrasing and prominence, which has led to recent analogous analyses for the two modalities [1], [2]. Intuitively, beat gestures, which are a subtype of co-speech gestures, lend themselves particularly well to a comparison with prosody because they are synchronized with the rhythm of speech [3]–[5], although referential gestures have equally shown to be aligned with speech [6]. Still, with one exception [7], phrasing and alignment analyses in German have focused on referential gestures [8], which tend to be associated with semantic meaning more often than beat gestures [9]. In this paper, we address this gap by presenting novel data on alignment and prominence of German beat gestures based on a small-scale video corpus.

1.1. Beat gestures

Beat gestures have long been of only marginal interest to gestural researchers. The reasons for this disregard are twofold: beat gestures appear to be less associated with semantico-pragmatic meaning [6] and their form and distribution seem easier to account for than is the case for referential gestures [10]. However, these assumptions fall short of accounting for the observed complexity of beat gestures: Beat gestures can correlate with new and contrastive information [11] and come in different forms and complex distributional patterns [12]. Thus, just like referential gestures, beat gestures show significant variation in form, function, and distribution and therefore require detailed analysis. With recent attempts to treat all gestures uniformly [2], the traditional distinctions between 'meaningful' vs. 'meaningless' gestures and between 'simplex' vs. 'complex' distribution need revising.

1.2. Prominence in gesture and speech

Prominence has a form dimension and a function/meaning dimension. A gesture or syllable can be regarded as perceptually prominent (i.e., in form) because of greater articulatory effort, which is prominence-lending. As a result of the perceptual prominence, the function/meaning associated with the form can also be regarded as prominent because it stands out in comparison to the surrounding function/meaning units. Typically, both formal and functional prominence are considered to be gradient [13], [14].

In speech prosody, prominence is expressed through a variety of parameters like the categorical parameters accentuation and pitch accent type, and gradient parameters like later alignment and greater excursion of pitch movements, longer duration, higher intensity, segmental changes resulting in hyperarticulation or a changed spectral balance, e.g., [15]. None of these acoustic measures are strictly necessary for signaling prominence [16], but pitch changes affect perceptual prominence the most [15]. With greater perceptual prominence comes, for instance, a stronger sense of information to be new or focused rather than given [14]. Contextual cues that induce expectations may amplify perception of prosodic prominence in English [13], more so than in German [16].

Gestural prominence is more difficult to define, which, however, may simply be a matter of detail in description. The gestural equivalent of an acoustically prominent pitch accent (peak) is the gestural apex [17], the endpoint of a hand stroke/movement, i.e., of the gestural trajectory. Just as with pitch accents, the apex is embedded into larger gestural units. We might even consider the apex to be the head of these units, just as pitch accents head prosodic phrases. Beat gestures share these accent-like characteristics [18], [19], but their form is traditionally regarded as simple, staccato-like movements. Still, perceptual prominence is the product of both intonational and gestural forms [19] and recent approaches argue for a multidimensional analysis of all gestures [3].

1.3. Prosody-gesture alignment

In prosody-gesture alignment analyses, different gestural units have been taken as reference points for alignment with speech: apex [17], stroke [20], or the starting point of a gesture [21]. In general, gestural strokes have been observed to be aligned with either focused [22], lexically stressed [17] or prosodically accented syllables [23]. The alignment trend is captured by McNeill's phonological synchronicity rule, which assumes that the stroke precedes or coincides with the acoustic high point of the syllable [3]. If gesture and speech prosody are not aligned,

gesture lag is perceived as less natural than a gesture preceding the acoustically prominent syllable [24].

Corresponding with these suggestions for gesture in general, beat gestures have been observed to show a significant degree of systematicity in where they occur: they tend to be aligned with stressed syllables [3], [25]. Most beat gesture strokes have been found to overlap with accented syllables [19].

1.4. Aims of this study

The current study aims to address the paucity of research on German beat gestures by quantifying their alignment with syllables of different levels of prosodic prominence. As already mentioned, prosodic prominence results from changes in various parameters, which can be gradient or categorical. Based on the previous findings for referential and beat gestures discussed above, we expect to find a large degree of alignment of German beat gestures and accented syllables as reported for other languages. We explored both the position of the stroke and the position of the apex. We also looked at a potential role of the hand shape for alignment but had no specific hypotheses.

To explore the relation of beat gestures and prosodic prominence on a more fine-grained scale than just the presence of a pitch accent, we used the measure of (perceived) prominence level from the DIMA annotation system (*Deutsche Intonation, Modellierung und Annotation* [26]). This system distinguishes three levels of prominence: (1) weak, (2) strong, and (3) extra strong, the latter occurring for instance on narrowly or contrastively focused words. We considered more fine-grained annotations, for instance in terms of accent types, to be unsuitable for the small size of our video corpus because more categories would have made it more difficult or impossible to recognize emergent patterns.

We hypothesized that the alignment of beat gestures might be impacted by the prominence level such that highly prominent syllables show a tighter, or more precise, alignment with the gestures, for instance regarding the apex position being within rather than near the accented syllable. Such a precise alignment might increase the perceived level of prominence of the accented syllable due to the combination of both articulatory channels, which might have communicative advantages for the function/meaning aspect of prominence. Lower prosodic prominence, in contrast, might correlate with greater variance of gesture placement.

2. Method

2.1. Materials

We created a corpus consisting of six pop-educational videos hosted on YouTube, which cover topics in German grammar [27], English grammar [28], chemistry [29], maths [30], politics [31], and finance [32], targeting a young audience. Three of these videos were presented by men, the other three by women; all presenters were trained speakers; the videos were scripted. The video frame was large enough to capture both the speaker's faces and hands (upper body; see Figure 1). Duration varied between 3:02 minutes and 17:53 minutes. We extracted two minutes of continuous speech per video with full visibility of the gestural space. Beat gestures were used as a reference point for the selection of relevant stimuli. We chose 120 gestures (20 per speaker) contained within 68 gestural units with a total duration of 2:42 minutes. 80 utterances were aligned with these gestural units, with an overall speech duration of 3:16 minutes.

Figure 1. *Three of six speakers with full visibility of the gestural space and upper body.*

2.2. Annotation and data preparation

Data annotation was performed in ELAN (Version 6.4 [33]). Beat gestures were annotated with muted videos. We annotated gestural unit, hand shape, trajectory/gesture phase, and position of the apex. The overwhelming majority of gestures was carried out with both hands, so we do not report on differences between the hands. Hand shape annotation followed the M3D annotation system [3] with some additions, see Section 3.3 for details. The annotation on the trajectory/gesture phase tier was purely descriptive and included (i) a vertical (up / down) trajectory; (ii) a horizontal trajectory describing the relation of the two hands (moving apart / approaching / moving together), and (iii) a 'hold' annotation. Forward and backward movement did not occur on the stroke or showed no variation. In some of the strokes, the apex did not coincide with the end of the trajectory because there was a very small, relaxed continuation at the very end of the trajectory after the visually most prominent point of the trajectory. Thus, our annotation did not follow the standard definition of the stroke [2], [12]. The reason is that we wished to analyze the relation between trajectory and apex in greater detail. We will nevertheless use the word *stroke* in our analyses for all main trajectories containing a clearly identifiable apex.

Verbal information was annotated for utterance, sentence, word, syllable, and prominence level of the syllable (DIMA). Non-prominent syllables were not annotated.

The videos had a sampling rate of 25 frames per second. To accommodate this low resolution, we examined the position and length of gestural and verbal events (apex, stroke, prominent syllable) in frames rather than milliseconds. Furthermore, we applied normalization procedures for some analyses (see Section 3 for details). The statistical analyses were conducted with R (Version 4.2.2 [34]) using the packages lme4 [35] and lmerTest [36] for fitting (general) mixed linear models starting with maximal models (with by-subject random slopes) and reducing the random structure (with only by-subject intercepts) in case of convergence issues. We specify the various fixed factors in the Results section.

3. Results

3.1. Apex position

Apexes were strongly associated with accented syllables. 89 (74%) of apexes occurred within an accented syllable, one apex occurred in the syllable preceding an accented syllable, one occurred two syllables earlier; 20 apexes (17%) occurred in the syllable following an accented syllable, nine (8%) occurred two syllables later. Thus 92% of apexes occurred in very close association with a prominent syllable: within or adjacent to it.

Whether or not an apex occurred within or after an accented syllable interacted with the prominence level of the

accented syllable. Table 1 suggests that the frequency of occurrence within the accented syllable rather than after it increased with an increased prominence level (the two gestures occurring before a prominent syllable are ignored in Table 1).

Table 1. *Proportion of apexes within the accented syllable by prominence level*

Prominence level	Nb. of apexes	Nb. of apexes within accented syllable	Proportion
1	19	10	0.53
2	46	34	0.74
3	53	45	0.85

The statistical analysis of the apex position within vs. after the accented syllable (generalized mixed model with binomial logit function; treatment coding with prominence level 1 as baseline; subject intercept)) showed that apexes associated with syllables of prominence level 3 occurred more often within these syllables rather than in subsequent syllables in comparison to apexes associated with level 1 syllables (b = 1.75, SE = 0.64, z = 2.73, $p < 0.01$), and apexes associated with level 2 syllables occurred marginally more often within these syllables than apexes associated with level 1 syllables (b = 1.09, SE = 0.62, z = 1.75, p = 0.08).

To assess the relative apex position in the accented and in the subsequent syllable, we normalized the apex position by syllable duration separately for all syllables. The mean relative position of apexes in the stretch comprising the accented and the subsequent syllable was at 0.77 (SD = 0.46), i.e., after 77% of the duration of the accented syllable. (Apexes within the accented syllable were realized at 0.64 (SD = 0.26), i.e., after about two thirds of the duration of that syllable.) Figure 2 shows the distribution of the relative apex position in relation to the prominence level. The statistical analysis revealed no effect of the annotated prominence level on the relative position of the apex.

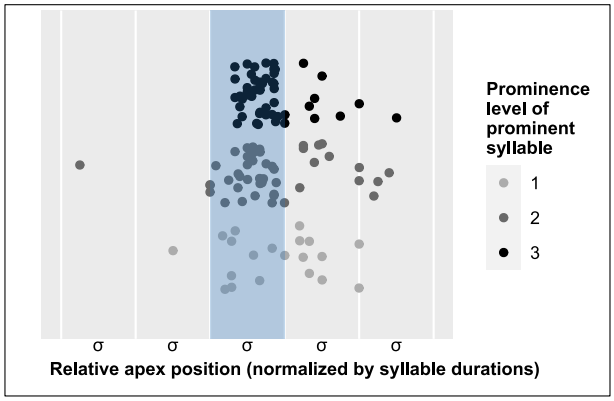

Figure 2. *Position of apex (with vertical jitter) in relation to prominence level. The accented syllable is marked in blue. Normalization by duration of syllables (i.e., relative position of apex within the syllable where it occurs).*

To explore if the relative apex position in the accented syllable and the subsequent syllable correlates with the duration of the accented syllable independently of perceived prominence levels, we calculated the repeated measurements correlation [37],[38] for the respective subset of the data. It was negative with r = -0.20 ($p < 0.05$), suggesting a shorter duration of accented syllables correlating with later apex positions.

3.2. Stroke position

109 (91%) of stroke trajectories were hands moving downwards, seven were hands moving apart, three upwards, and one moving towards each other and joining. 42 (35%) of the strokes started before the beginning of the accented syllable, the remaining 78 starting within the accented syllable. The distribution by prominence level was as follows: For accented syllables of prominence level 1, the stroke started before the accented syllable in 45% of the cases. For prominence-level-2 syllables, it did in 36% of the cases, and for prominence-level-3 syllables, it did in 30% of the cases. Descriptively, there is a continuous decrease with an increased prominence level: the higher the prominence level, the more often the stroke starts within the prominent syllable instead of before it. The statistical analysis showed no effect of prominence level.

The mean beginning of the stroke was 0.03 frames after the beginning of the prominent syllable (SD = 3.8). There were no prominence-level related gradient differences (level 1: -0.6 frames (SD = 4.0), i.e., 0.6 frames before the beginning of the prominent syllable; level 2: -0.4 frames (SD = 3.8); level 3: 0.6 frames (SD = 3.8)). The overall mean duration of the stroke was 5.4 frames (SD = 2.8), which corresponds to just above 200 ms. There were no prominence-level related differences regarding the duration of the stroke (level 1: 5.2 frames (SD = 2.6), level 2: 5.3 frames (SD = 2.9), level 3: 5.5 frames (SD = 2.7).

Figure 3 illustrates the relative position of the stroke (normalized by syllable durations) in relation to the accented syllable by prominence level. The Figure excludes 9 gestures for which the stroke started in a silence before the prominent syllable.

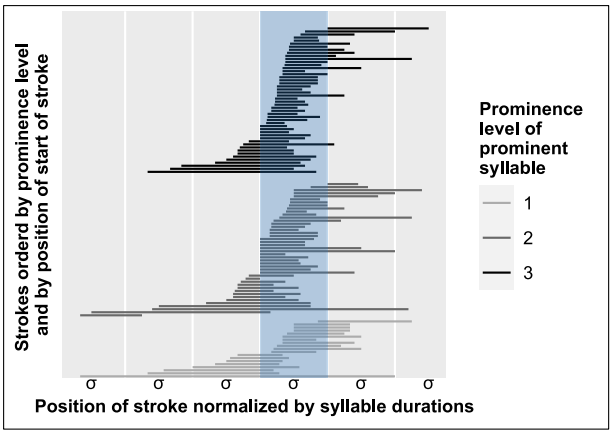

Figure 3. *Position of stroke in relation to the accented syllable (blue) by prominence level. Normalization by duration of syllables (see Figure 2).*

3.3. Hand shapes

We found 10 different hand shapes. Figure 4 shows the distribution of the seven hand shapes with more than three occurrences by speaker. Vertically parallel open hands were most frequent (n = 36), followed by palm-up-open hands (PUOH, n = 34) and vertically parallel knife hands (10). As Figure 4 shows, the distribution varied greatly between speakers. Some speakers show preferences for horizontal hand

shapes (e.g., PUOH), others for vertical forms (e.g., open), yet others mix. Speaker M3 shows no variation in form whereas all other speakers relied on at least two, if not more hand shapes. Given this variety of use, we did not investigate a potential influence of hand shapes on the alignment of beat gestures and prosodic prominence and leave this issue for future research.

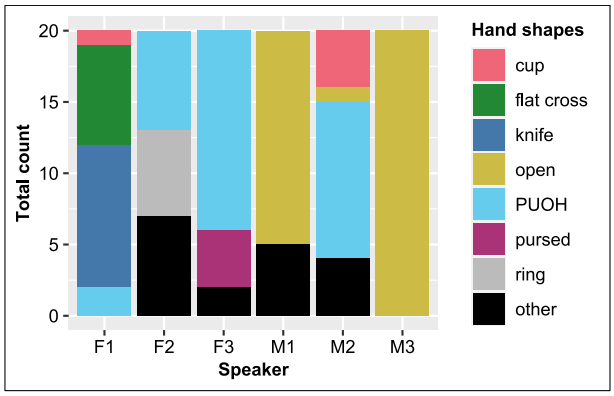

Figure 4: *Hand shape (n >3) by speaker.*

4. Discussion

Our investigation of the alignment of beat gestures with prosodic prominences in educational videos has shown that beat gestures overwhelmingly align with pitch-accented, i.e., prominent syllables in German, thus corroborating previous findings for other languages. Beyond confirming an overlap of strokes with the pitch-accented syllable, our investigation provides new observations regarding the specifics of the alignment. Strokes start before, or – more often – within the accented syllable. Furthermore, apexes occur within the accented syllable – about two thirds into that syllable – and less often in the subsequent syllable, although they also occur in the post-subsequent syllable. Apex position seems to weakly correlate with the duration of the accented syllable: in shorter syllables the apex tends to occur later. This might be because there is less time for the gesture trajectory. Alternatively, the pitch peak position of the accented syllable might play a role: it might be the peak that the apex aligns with. This issue must be investigated in future research with a larger database. Another novel finding is that the alignment between spoken and gestural prominence is tighter, or more precise, for higher levels of prosodic prominence: apexes occur more frequently within the accented syllable if that syllable is prosodically highly prominent, and descriptively, strokes start within the accented syllable more often if that syllable is highly prominent.

We take our data to lend further support to the phonological synchronicity rule of aligning prominence [2] but also suggest that it might benefit from additional specifications regarding apex placement and a potential sensitivity to prominence level. Overall, we propose that both stroke and apex are important reference points for speech-gesture alignment. For the beat gestures under investigation, stroke begin and apex often occur within the accented syllable, especially for highly prominent accented syllables. This tight association might be a consequence of the fairly short trajectories of this gesture type (5.4 frames, roughly 200 ms; the average length of the accented syllable was 6.7 frames, roughly 260 ms). Previous studies have emphasized the precedence or synchronicity of gestures [17], [22] but our results suggests that the speaker aims for maximal synchronicity especially for highly prominent syllables.

In the Introduction we discussed prominence as having a form and a function dimension. For spoken prominence, this is well established [18], [39]; yet for gestural prominence, particularly for beat gestures, the label of 'meaningless' gestures [6] insinuates the opposite. The correlation that we found between apex position and prominence level points to an analogous link between form and function for beat gestures. If we take greater spoken prominence as a proxy for the function of prominence (due to the established link between pitch accents and pragmatic meaning), the larger proportion of apexes within syllables with a greater prominence level than those with a lower prominence level is indicative of a greater effort in the speaker to emphasize content across domains. While DIMA prominence levels can only serve as an approximation for the degree of perceived prominence, the fact that we found differences in gesture alignment for the different prosodic prominence levels clearly suggests that beat gestures go beyond embodying sentence rhythm. They can be used to signal prominence just as much as pitch accents can. Thus, beat gestures may have both a rhythmic and a pragmatic function.

Form-wise, the beat gestures in our corpus were restrictive in the use of space: predominantly, the trajectory was up and down, rather than sidewards (or forwards/backwards). Yet, variation in hand shape was present across most speakers with ten different hand shapes overall, although there was a concentration on two hand shapes (open, PUOH). The variation in form, (assumed) function, and distribution of beat gestures in our sample unambiguously show that beat gestures deserve detailed description and analysis. They exhibit a complex alignment relation to acoustically prominent syllables and a complex variation in hand shape. The fact that no two speakers had the same shape distribution suggest that hand shapes of beat gestures show a similarly rich variation as those of other types of gestures. Speakers seem to have idiosyncratic hand shape profiles that predominantly exhibit an up-down trajectory. What contributes to the choice of hand shape of the individual beat gesture is still unknown. Overall, integrating speech prosody and semantico-pragmatic functions in descriptions of beat gestures seems strongly recommended [2].

5. Conclusions

Our findings from the annotation of spoken and gestural prominence in German educational videos provide evidence for a close alignment of speech prosody and beat gestures, and an interrelation with degrees of prosodic prominence. These results shed light on cross-linguistic parallels in beat gesture alignment with interesting avenues for future research. Gestural apexes served as a useful measure for investigating the alignment of gestural and spoken prominence because they allow identifying alignment within the syllable. Future research must show if the apex aligns with the pitch peak of the accented syllable. Furthermore, future studies of the variation in shape of beat gestures must investigate potential factors influencing the choice. All in all, the present analysis suggests that beat gestures are similarly complex as referential gestures.

6. Acknowledgements

The study expands on the Master's thesis of the second author at the University of Cologne [40]. Funding for this study: DFG Deutsche Forschungsgemeinschaft (German Research Foundation), Project-ID 281511265, SFB 1252.

7. References

[1] S. Shattuck-Hufnagel, A. Ren, M. Mathew, I. Yuen, and K. Demuth, 'Non-referential gestures in adult and child speech: Are they prosodic', in *Proceedings from the 8th international conference on speech prosody*, Boston University Boston, 2016, pp. 836–839.

[2] P. L. Rohrer *et al.*, 'The MultiModal MultiDimensional (M3D) labeling system', Aug. 2020, doi: 10.17605/OSF.IO/ANKDX.

[3] D. McNeill, 'Hand and Mind: What Gestures Reveal about Thought Univ'. of Chicago Press, Chicago, 1992.

[4] R. L. Birdwhistell, *Kinesics and context: Essays on body motion communication*. U of Pennsylvania press, 2010.

[5] D. Bolinger, 'Intonation and gesture', *Am. Speech*, vol. 58, no. 2, pp. 156–174, 1983.

[6] J. Weisberg, A. L. Hubbard, and K. Emmorey, 'Multimodal integration of spontaneously produced representational co-speech gestures: an fMRI study', *Lang. Cogn. Neurosci.*, vol. 32, no. 2, pp. 158–174, 2017.

[7] W.-M. Roth, 'From action to discourse: The bridging function of gestures', *Cogn. Syst. Res.*, vol. 3, no. 3, pp. 535–554, 2002.

[8] E. Fricke, *Grammatik multimodal: wie Wörter und Gesten zusammenwirken*, vol. 40. Walter de Gruyter, 2012.

[9] C. Müller, 'Forms and uses of the Palm Up Open Hand: A case of a gesture family', *Semant. Pragmat. Everyday Gestures*, vol. 9, pp. 233–256, 2004.

[10] J. B. Bavelas, N. Chovil, D. A. Lawrie, and A. Wade, 'Interactive gestures', *Discourse Process.*, vol. 15, no. 4, pp. 469–489, 1992.

[11] 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.

[12] P. Prieto, A. Cravotta, O. Kushch, P. Rohrer, and I. Vilà-Giménez, 'Deconstructing beat gestures: a labelling proposal', in *Proceedings of the 9th international conference on speech prosody*, 2018, pp. 201–205.

[13] J. Bishop, 'Information structural expectations in the perception of prosodic prominence', *Prosody Mean.*, vol. 25, p. 239, 2012.

[14] S. Baumann, C. T. Röhr, and M. Grice, 'Prosodische (De-) kodierung des informationsstatus im Deutschen', *Z. Für Sprachwiss.*, vol. 34, no. 1, pp. 1–42, 2015.

[15] S. Baumann and B. Winter, 'What makes a word prominent? Predicting untrained German listeners' perceptual judgments', *J. Phon.*, vol. 70, pp. 20–38, 2018.

[16] C. T. Röhr, S. Baumann, P. B. Schumacher, and M. Grice, 'Perceptual prominence of accent types and the role of expectations', in *Proceedings of the 10th International Conference on Speech Prosody*, 2020, pp. 366–370.

[17] D. P. Loehr, *Gesture and intonation*. Citeseer, 2004.

[18] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[19] 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.

[20] D. Brentari, G. Marotta, I. Margherita, and A. Ott, 'The interaction of pitch accent and gesture production in Italian and English', *Studi E Saggi Linguist.*, vol. 51, no. 1, pp. 83–101, 2013.

[21] H. L. Rusiewicz, S. Shaiman, J. M. Iverson, and N. Szuminsky, 'Effects of prosody and position on the timing of deictic gestures', 2013.

[22] 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.

[23] S. Nobe, 'Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/threshold model of gesture production.', Doctoral dissertation, University of Chicago, 1997.

[24] T. Leonard and F. Cummins, 'The temporal relation between beat gestures and speech', *Lang. Cogn. Process.*, vol. 26, no. 10, pp. 1457–1471, 2011.

[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] 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.

[27] *Adjektiv, Substantiv, Subjekt - Wortarten und Satzglieder unterscheiden I musstewissen Deutsch*, (Oct. 24, 2017). Accessed: Mar. 05, 2023. [Online Video]. Available: https://youtu.be/I1rt3XNHyB4

[28] *past tenses: Wann verwendest du welche? - Englisch | Duden Learnattack*, (May 26, 2019). Accessed: Mar. 05, 2023. [Online Video]. https://youtu.be/NIMFBNiojCE

[29] *Die Masse von Atomen I musstewissen Chemie*, (Dec. 27, 2017). Accessed: Mar. 05, 2023. [Online Video]. Available: https://youtu.be/xQaafi2zLik

[30] *Was ist ein Term? I einfach erklärt mit Beispielen*, (Aug. 28, 2017). Accessed: Mar. 05, 2023. [Online Video]. Available: https://youtu.be/3RWgrC6kNSU

[31] *Das passiert, wenn die EU zerbricht*, (Apr. 08, 2020). Accessed: Mar. 05, 2023. [Online Video]. Available: https://www.youtube.com/watch?v=zOUQZvm7c7E

[32] *ETF & Steuern erklärt! Weniger Steuern auf deine ETFs zahlen | Finanzfluss*, (Feb. 23, 2020). Accessed: Mar. 05, 2023. [Online Video]. https://youtu.be/AyUiI8a_Rb8

[33] 'ELAN (Version 6.4)'. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. [Online]. Available: https://archive.mpi.nl/tla/elan

[34] R. C. Team, 'R: A language and environment for statistical computing', 2013.

[35] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-E ects Models Using lme', *J. Stat. Softw.*.

[36] A. Kunzetsova, P. Brockhoff, and R. Christensen, 'lmerTest package: Tests in linear mixed effect models', *J Stat Softw*, vol. 82, pp. 1–26, 2017.

[37] J. Martin Bland and D. G. Altman, 'Calculating correlation coefficients with repeated observations: Part 2–correlation between subjects', *BMJ*, vol. 310, no. 6980, p. 446, 1995.

[38] J. Z. Bakdash and L. R. Marusich, 'rmcorr: Repeated measures correlation. R package version 0.3. 0'. Avaliable at: https://CRAN. R-project. org/package= rmcorr (accessed June 24 2018.

[39] D. Bolinger, *Intonation and its uses: Melody in grammar and discourse*. Stanford university press, 1989.

[40] L. Muhtz, 'Zur zeitlichen Alignierung von Beat-Gesten und prosodischen Prominenzen. Eine Korpusstudie.', Cologne, 2020.