

Gestural and accentual prominence are cumulative – An exploratory analysis of a German TED Talk

Klymentii Myslyvyi¹, Janne Lorenzen² and Stefan Baumann¹

¹*IfL-Phonetik, University of Cologne, Germany*, ²*Johannes Gutenberg University Mainz, Germany*
k.myslyvyi@uni-koeln.de, janne.lorenzen@uni-mainz.de, stefan.baumann@uni-koeln.de

Gestures and speech have often been claimed to represent two dimensions of the same idea unit (e.g., McNeill 1992). In fact, it is widely accepted that not only speech but also speech-accompanying gestures are characterized by prosodic features (e.g., Shattuck-Hufnagel & Ren 2018), accounting for the rhythm and phrasing of gestural movements. We are interested in the interplay of the two modalities in cueing multimodal prominence.

Our study examines the relationship between the prosodic components of gesture and speech in a Standard German TED Talk, a text genre which offers a complete discourse context and features a clear, engaging speech style. We investigate whether specific speech-prosodic properties (such as accentual prominence, type and status of pitch accents as well as their temporal alignment with segmental and visual landmarks) can predict hand gesture occurrence and the level of perceived gestural prominence. Our overarching goal is to ascertain whether the contribution of gestural strokes and pitch accents to the encoding of prominence is additive or compensatory in nature.

Gestures and speech were annotated independently to avoid circularity. Five annotators labelled gestural units and strokes. Consensus labels were developed on the basis of joint discussions, during which 103 G-units and 378 strokes were identified. At the level of the stroke, gestural prominence was annotated using a scale of 1 (weak), 2 (typical) and 3 (strong). In the analysis, we use the median prominence score of each stroke. In the spoken modality, accentual prominences were labelled using a similar scale of 1 (weak), 2 (strong) and 3 (extra strong). Non-prominent syllables were automatically assigned the label 0 (n = 2257). In addition, we identified accent contour shapes (low, falling, high, rising) and accent status (pre-nuclear, nuclear).

Results show that accentual prominence increases the probability of *gesture occurrence*, with more prominent syllables being more likely to attract gestures than less prominent syllables (see Fig. 1). While pitch accent type has no robust effect on gesture presence, our findings reveal that the occurrence of a gesture is more probable with a prenuclear rather than a nuclear accent.

Looking at the correlation between *perceived prominence levels* in the spoken and visual modalities, we find that the presence of a strongly highlighted syllable increases the probability of a strongly prominent gestural stroke, while minimally prominent syllables tend to co-occur with rather weak manual gestures (see Fig. 2). Additionally, we observe that pitch accent contours affect gestural prominence. Specifically, rising and high accents are more likely to co-occur with more prominent strokes (see Fig. 3).

Finally, we investigate gestural prominence as a function of the temporal alignment (measured as syllable distances) of apices with accented syllables. We observe that apices are more likely to be closely synchronized with accented syllables in more prominent gestures and more likely to occur outside of accented syllables in weakly prominent gestures (see Fig. 4).

In sum, we interpret all of these findings as evidence for a tight link between gestural and accentual prominence, in line with the *Cumulative-Cue Hypothesis* (Ambrazaitis & House 2023).

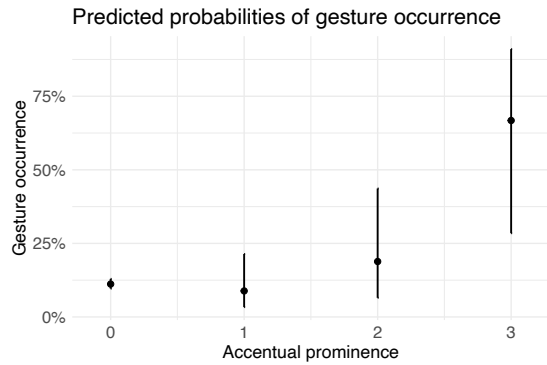


Figure 1: Predicted probabilities of gesture occurrence as a function of accentual prominence.

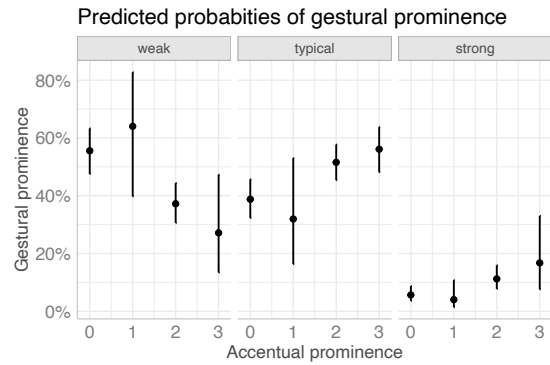


Figure 2: Predicted probabilities of gestural prominence as a function of accentual prominence.

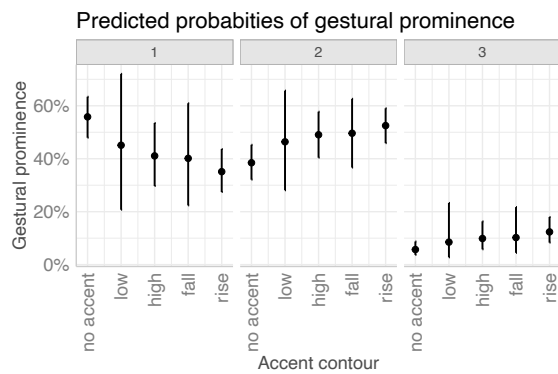


Figure 3: Predicted probabilities of gestural prominence as a function of pitch accent contour.

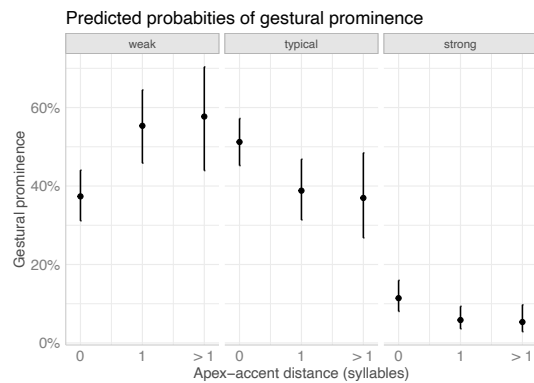


Figure 4: Predicted probabilities of gestural prominence by apex-accent distance (in syllables).

References:

[1] McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

[2] Shattuck-Hufnagel, S. & A. Ren (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.01514>

[3] Ambrazaitis, G. & D. House (2023). The multimodal nature of prominence: Some directions for the study of the relation between gestures and pitch accents. *Proceedings of the 13th International Conference of Nordic Prosody*, 262–273, doi: 10.2478/9788366675728-024.