

Multidimensional and multimodal cues to prosodic prominence: Can we compare monologic and dialogic scenarios?

Lena Pagel¹, Simon Roessig¹ and Doris Mücke¹

¹University of Cologne

{lena.pagel, simon.roessig, doris.muecke}@uni-koeln.de

Prosodic prominence can be cued multidimensionally through modulations of various phonetic parameters and speech-accompanying body movements. Prominent entities in an utterance are often accompanied by modulations in acoustic-prosodic cues (such as pitch accent placement, longer durations; [1], [2]), prosodically strengthened articulation (such as greater lip aperture, more extreme tongue targets; [3], [4], [5], [6], [7]), as well as co-speech body motion as visual prosody (such as head nods, eyebrow raises, manual gestures; [8], [9], [10], [11]). However, many existing studies are limited to a small number of parameters within a specific modality or domain. In this contribution, we aim to provide insights into the inherently multidimensional and multimodal nature of prosodic prominence cueing by integrating multiple parameters into a single experimental design. We also test whether prominence encoding is comparable between monologic and more natural dialogic scenarios.

Thirty native speakers of German were recorded for speech acoustics, supra-laryngeal articulation, and co-speech body motion, using 3D electromagnetic articulography. They played a game that elicited the production of controlled utterances and was completed in two communicative modes: first solo (by each speaker individually, in front of a screen), then dialogue (by two speakers paired into a dyad, in a cooperative setting). The speech material included four target words that were elicited in the experimental game scenario: Medina (/me'di:na/), Manila (/ma'ni:la/), Benali (/be'na:li/), and Milano (/mi'la:no/). These target words were embedded in consistent carrier phrases and were produced by speakers as responses to elicitation questions. They occurred in two focus conditions: background and corrective focus (see Example 1). Five parameters associated with the production of the target words were examined: word duration, F0 excursion, maximum vocalic lip aperture, vertical vocalic tongue body displacement, and maximum head velocity. Using Bayesian hierarchical regression, each parameter was modelled as a function of focus type and communicative mode as well as their interaction.

(1)

focus type	elicitation question	target utterance
background	Habe ich die Bohne aus Milano auf der Hand? 'Am I holding the bean from Milano?'	Du hast die [Vase] _F aus Milano auf der Hand. 'You are holding the [vase] _F from Milano.'
corr. focus	Habe ich die Bohne aus Manila auf der Hand? 'Am I holding the bean from Manila?'	Du hast die Bohne aus [Milano] _F auf der Hand. 'You are holding the bean from [Milano] _F .'

Model results are illustrated for all parameters in Figure 1, in terms of estimated conditional means by focus type and communicative mode (left panel per parameter) and posterior distributions with means and 95% credible intervals of the effect of focus per mode (right panel per parameter). Additionally, posterior probabilities of values being higher in corrective focus than in the background are included in the figure. We find compelling evidence for a difference between focus conditions in both modes, with words in corrective focus being produced with longer word durations, larger F0 excursions, greater vocalic lip aperture and tongue body displacements, as well as faster co-speech head movements. In each case, the credible interval of the posterior distribution for the focus effect does not include zero, and the posterior probability that values are higher in corrective focus is 1.00. Comparing the two modes, we observe that the magnitudes of these between-focus differences are consistently greater in the dialogue than in the solo mode (i.e., posterior distributions further away from zero). This is true for each parameter, although the extent of this mode effect varies, with the difference between modes being greater, e.g., in lip aperture and head velocity than in tongue body displacement. A visualisation of each speaker's parameter values over the course of the experiment sheds light on the temporal dynamics of the effect. It reveals that the mode effect emerges abruptly at the onset of the dialogue, rather than developing gradually over time (cf. Figure 2 for an example). This pattern suggests that the effect is not simply a by-product of task progression but instead reflects a genuine consequence of the shift in communicative mode, potentially driven by changing communicative demands.

Taken together, the study shows that speakers of German differentiate focus types through modulations of a range of phonetic cues and therefore emphasises the multidimensional and multimodal nature of prosodic prominence. Crucially, this differentiation is comparable between the monologic and dialogic scenario but tends to be more pronounced when speakers talk to an interlocutor.

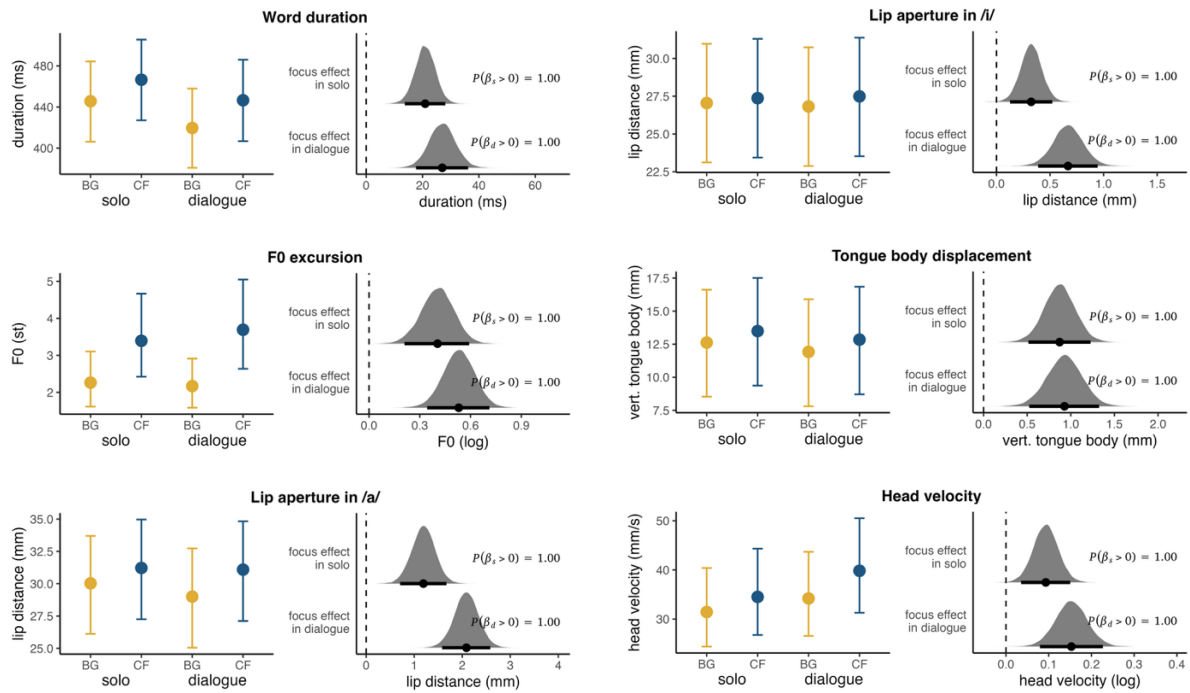


Figure 1: Model results with two panels per parameter; left: conditional means per focus type (BG = background, CF = corrective focus) and communicative mode; right: posterior distributions (including means and 95% credible intervals) of the effect of focus per mode, and posterior distributions of parameter values being larger in corrective focus than in the background

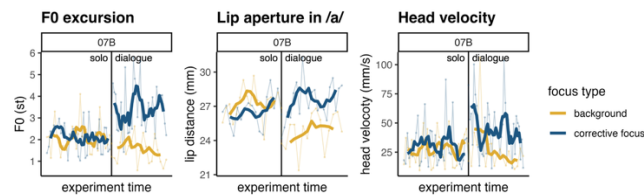


Figure 2: Visualisation of parameter values for both focus conditions over the course of the experiment, for one example speaker in three example parameters; a larger difference between the yellow and blue lines indicates greater focus differentiation

References:

- [1] R. D. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 2008.
- [2] S. Roessig, B. Winter, and D. Mücke, ‘Tracing the phonetic space of prosodic focus marking’, *Front. Artif. Intell.*, vol. 5, pp. 1–24, 2022, doi: 10.3389/fraci.2022.842546.
- [3] M. Beckman, J. Edwards, and J. Fletcher, ‘Prosodic structure and tempo in a sonority model of articulatory dynamics’, in *Gesture, Segment, Prosody*, G. J. Docherty and R. D. Ladd, Eds, Cambridge: Cambridge University Press, 1992, pp. 68–89. doi: 10.1017/cbo9780511519918.004.
- [4] T. Cho, *The effects of prosody on articulation in English*. in Outstanding dissertations in linguistics. New York/London: Routledge, 2002.
- [5] K. 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, 1995, doi: 10.1121/1.412275.
- [6] D. Mücke and M. Grice, ‘The effect of focus marking on supralaryngeal articulation - Is it mediated by accentuation?’, *J. Phon.*, vol. 44, pp. 47–61, 2014, doi: 10.1016/j.wocn.2014.02.003.
- [7] L. Pagel, S. Roessig, and D. Mücke, ‘The encoding of prominence relations in supra-laryngeal articulation across speaking styles’, *Lab. Phonol.*, vol. 15, no. 1, pp. 1–55, Sept. 2024, doi: 10.16995/labphon.10900.
- [8] G. Ambrazaitis and D. House, ‘Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings’, *Speech Commun.*, vol. 95, pp. 100–113, Dec. 2017, doi: 10.1016/j.specom.2017.08.008.
- [9] N. Esteve-Gibert, J. Borràs-Comes, E. Asor, M. Swerts, and P. Prieto, ‘The timing of head movements: The role of prosodic heads and edges’, *J. Acoust. Soc. Am.*, vol. 141, no. 6, pp. 4727–4739, 2017, doi: 10.1121/1.4986649.
- [10] 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, July 2024, pp. 891–895. doi: 10.21437/SpeechProsody.2024-180.
- [11] J. Krivokapić, M. K. Tiede, and M. E. Tyrone, ‘A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection’, *Lab. Phonol.*, vol. 8, no. 1, preprint preprint 3, 1–26, 2017, doi: 10.5334/labphon.75.