How to achieve a prominence GOAL! in different speaking styles

Lena Pagel, Simon Roessig and Doris Mücke – University of Cologne
lena.pagel@uni-koeln.de, simon.roessig@uni-koeln.de and doris.muecke@uni-koeln.de

1 Introduction

Speakers vary the degree of vocal effort in speech production to successfully convey a message to listeners. Vocal effort can be globally increased over entire utterances, as in loud speech. Furthermore, it can vary within an utterance to locally highlight important information, which is then referred to as prosodic prominence. The global and local modulations can be interpreted in light of the H&H continuum, which assumes that speech varies on a scale between hypoarticulation and hyperarticulation (Lindblom 1990). A wide range of studies show that vocal effort has a number of acoustic and articulatory correlates (e.g. Geumann 2001, de Jong 1995, Cho 2005). Particularly, for highlighting related to prosodic prominence, two strategies have been carved out. Through a greater opening of the vocal tract, more acoustic energy radiates from the mouth, yielding a sonority expansion (Beckman et al. 1992). An enhancement of vocalic place features, i.e. localised hyperarticulation, yields more distinct vowel productions (de Jong 1995).

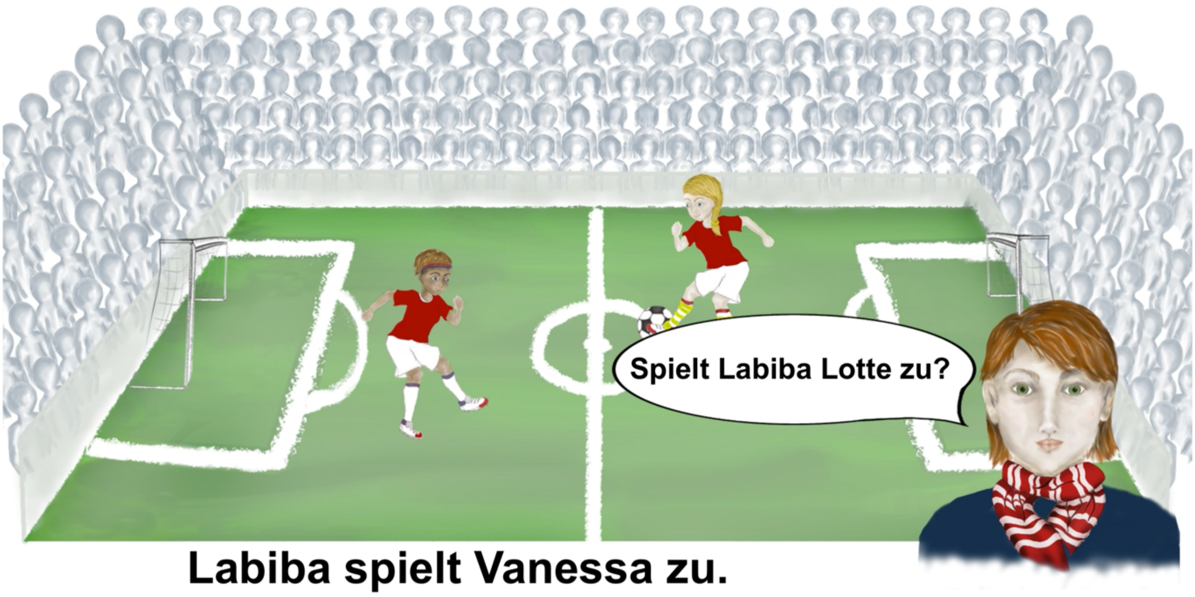
We can assume that there is an interaction between the increase in vocal effort on the global level (i.e. loud speech) and on the local level (i.e. to mark prosodic prominence). However, these interactions have not yet been investigated in the domain of articulation. The present study examines supralaryngeal signatures of prominence in loud speech as opposed to habitual speech. We ask: If speakers increase the vocal effort of an utterance globally in loud speech, can they still enhance vocal effort locally to mark prominence?

2 Methods

We recorded 20 native German speakers acoustically and articulatorily using 3D Electromagnetic Articulography. The subjects were engaged in an interactive experimental game that was set in a football stadium. It involved a question-answer task, in which participants described scenes on the field to a virtual avatar. This avatar, Marie, was introduced as their friend who accompanied them to the match but forgot her glasses and consequently asked for descriptions of the situations on the field (cf. Fig. 1).

The target sentences were the answers to the avatar's questions. They were of the form "X spielt Y zu" ("X passes to Y"). The questions were constructed so that the focus structure of the target sentences was manipulated in two different ways: Either the whole sentence was in broad focus or the object (position Y) of the sentence was in contrastive focus and the subject (position X) was in the background.

Figure 1: Experiment screen for one trial



The target words occurred in either the subject (X) or the object (Y) position. We call the subject position initial and the object position medial. When the target word is medial, it is either in broad focus or in contrastive focus; when it is initial, it is either in broad focus or in the background. Accordingly, the focus conditions are called broad-broad (henceforth BR-BR) and background-contrastive (henceforth BA-CO). It is important to note that when going from the BR-BR to the BA-CO condition, we expect a prominence increase in the medial position, as opposed to a prominence decrease in the initial position.

The three pseudo words Labiba /la'bi:ba/, Sabima /za'mi:ma/ and Nabima /na'bi:ma/ were selected as target words. The target syllable was the penultimate syllable, which carried the lexical stress and included the vowel /i/. Exemplary question-answer pairs for the target word Nabima are presented in Table 1. During the first half of the virtual football game, participants spoke in a habitual style. For the second half, they were told that the atmosphere in the stadium heated up such that they needed to speak very loudly to be understood by their friend.

Table 1: Exemplary question-answer pairs for the target word Nabima

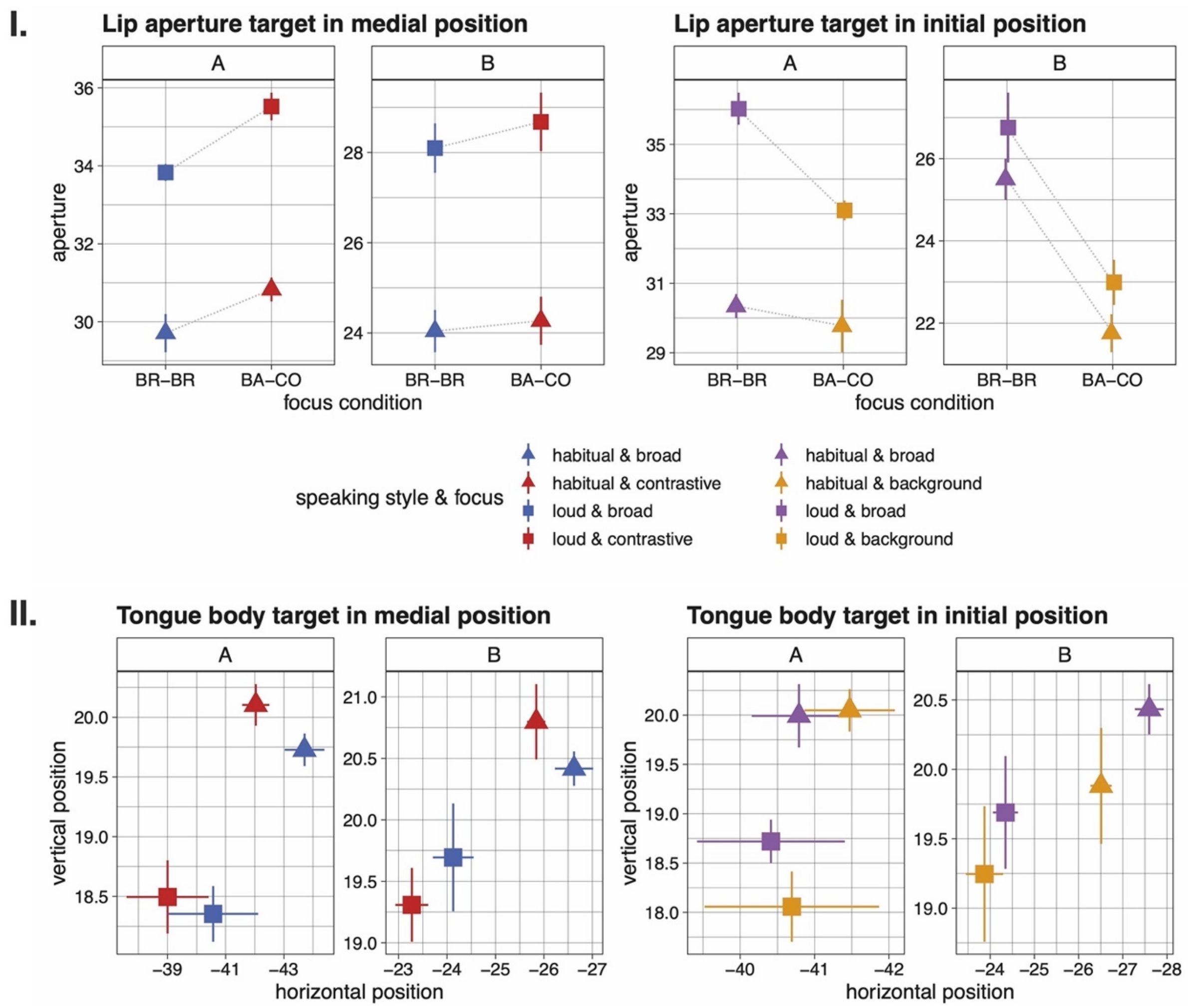
Position	**Focus**	**Question**	**Answer**
initial	BR-BR	Was passiert gerade?	[Nabima spielt Vanessa zu]F.
initial	BA-CO	Spielt Nabima Holly zu?	Nabima spielt [Rebecca]F
medial	BR-BR	Was passiert gerade?	[Carlotta spielt Nabima zu]F.
medial	BA-CO	Spielt Annette Lotte zu?	Annette spielt [Nabima]F zu.

The acoustic data was segmented using the Montreal Forced Aligner (McAuliffe et al. 2017) and hand-corrected in Praat (Boersma & Weenink 2021). Then, relevant articulatory landmarks were automatically extracted with emuR (Winkelmann et al., 2021) in R (R Core Team, 2020). Here, we examine target positions of labial and lingual movements during the production of syllables containing the vowel /i/. The Euclidean distance between the upper and lower lip was calculated to measure the target for the maximum lip aperture during /i/. In addition, the vertical (close-open) and the horizontal (front-back) movement dimensions are analysed to capture the maximum tongue raising and fronting during /i/.

3 Results

The results are presented for two exemplary speakers, labelled A and B. It is important to keep in mind that the BA-CO focus condition induces a prominence increase in the medial and a decrease in the initial position. Fig. 2 I (top) shows lip aperture targets, with higher values indicating a greater opening. Both speakers produce the vowel with an overall greater lip aperture in loud speech compared to habitual speech. Concerning the focus conditions, the lips are more open under increasing prominence (i.e. in BA-CO in the medial position) and less open under decreasing prominence (i.e. BA-CO in the initial position). This is the case in both speaking styles, but the degree of lip aperture modifications depends on the speaking style. The between-focus changes of the labial system are stronger in loud speech as compared to habitual speech.

Figure 2: Results for lip aperture (top) and tongue body position (bottom)



The results for the tongue body are shown in Fig. 2 II (bottom). The y-axis represents the vertical movement dimension, where higher values indicate a raising of the tongue in /i/. The x-axis is related to the horizontal movement dimension, with values to the left indicating a fronting of the tongue in /i/. Note that the figures can be read similarly to an acoustic vowel chart. The data reveal that the overall tongue target is lower and fronted in loud speech as compared to habitual speech, but we find differences due to focus structure and speaker-specific strategies. In habitual speech, both speakers modify the tongue target of /i/ in a way that the features [+high, +front] are enhanced under increasing prominence in the medial position. In the initial position, speaker A retracts the tongue under decreasing prominence, weakening the [+front] feature. Speaker B, on the contrary, exhibits a fronting of the tongue under decreasing prominence, strengthening the [+front] feature. Additionally, speaker B lowers the tongue, weakening the [+high] feature. Most of these patterns are mirrored in loud speech, except that speaker B produces a fronted but lower target in the medial position and speaker A additionally weakens the [+high] feature in the initial position.

4 Discussion and conclusion

The research aim of this contribution was to examine if vocal effort can be locally increased in loud speech – a speaking style that requires a globally high level of vocal effort. Despite speaker-specific strategies, kinematic lip and tongue body data reveal that prominence relations found in habitual speech can indeed be preserved in loud speech. In both speaking styles, we observe similar highlighting strategies: sonority expansion and localised hyperarticulation. Under prominence, the lips are opened to a greater extent, expanding the sonority of the vowel, and tongue body targets are partly hyperarticulated. However, there are also differences in the encoding of prosodic prominence when comparing loud and habitual speech. Both speakers tend to use more sonority expansion in loud speech and more localised hyperarticulation in habitual speech.

The data on loud speech underline the flexibility of prominence marking in supralaryngeal articulation, in that particularly those dimensions are modulated that are less restricted phonologically or physiologically. For instance, the lips encode prosodic structure more freely than the tongue body and to a greater extent than in habitual speech. Moreover, in cases where the strategies of sonority expansion and localised hyperarticulation conflict in the vertical tongue body movement, the horizontal movement may enhance the vowel's place feature. Both findings indicate the importance of expanding sonority in loud speech. Further corroborating the notion of complementarity, the results show that not only are more prominent entities strengthened, less prominent entities are additionally weakened in their production, finally making the prominent entity stand out more.

To conclude, kinematic modifications for the encoding of prominence are found in habitual and loud speech. The two levels of local and global vocal effort variation go hand in hand in order to increase intelligibility and successfully convey a message. The articulatory data from two exemplary speakers provide evidence for flexible and complementing strategies with the goal to encode prominence in loud speech. The relative importance of the highlighting strategies shifts: In these speakers, sonority expansion appears to be stronger in loud speech, while localised hyperarticulation seems to be stronger in habitual speech. This is likely to be triggered by the complex interplay of physiological constraints of vocal tract configurations and the demands of the communication process. The data of the two speakers is preliminary. It is possible that speakers differ in the way they prioritise the articulatory cues to express prominence in loud speech, which will be assessed in the analysis of the entire corpus of 20 speakers.

Author note
This work was funded by the Deutsche Forschungsgemeinschaft as part of the CRC 1252 "Prominence in Language", project A04 "Dynamic modelling of prosodic prominence".

References

Beckman, Mary, Jan Edwards & Janet Fletcher. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In Gerard J. Docherty & Robert D. Ladd (eds.), *Gesture, segment, prosody*, 68–89. Cambridge: Cambridge University Press.

Boersma, Paul & David Weenink. 2021. Praat: doing phonetics by computer.

Cho, Taehong. 2005. Prosodic strengthening and featural enhancement: Evidence from acousticand articulatory realizations of /A,i/ in English. *The Journal of the Acoustical Society of America* 117(6). 3867–3878.

Geumann, Anja. 2001. Invariance and variability in articulation and acoustics of natural perturbed speech. In Philip Hoole (ed.), *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM),* vol. 38, 265–393. Munich: Institut für Phonetik und Sprachliche Kommunikation, Ludwig-Maximilans-Universität München.

de Jong, Kenneth. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America* 97(1). 491–504.

Lindblom, Björn E. 1990. Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer Academic Publishers.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc & Michael Wagner. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Proceedings of INTERSPEECH*, *20-24 August, Stockholm, Sweden*, 498–502.

R Core Team. 2020. R: A language and environment for statistical computing.

Winkelmann, Raphael, Klaus Jaensch, Steve Cassidy & Jonathan Harrington. 2021. emuR: Main Package of the EMU Speech Database Management System. R package version 2.3.0.