# VARIABILITY IN SPACE, COMPENSATION IN TIME: EFFECTS OF AGING ON TONGUE BODY KINEMATICS IN PROSODIC FOCUS MARKING

Doris Mücke[1], Tabea Thies[1,2], Philipp Buech[3], Noemi Furlani[1], Janine Schreen[1,2], Anne Hermes[3]

[1]IfL-Phonetics, University of Cologne; [2]Department of Neurology, University Hospital Cologne; [3]Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne Nouvelle, Paris
doris.muecke@uni-koeln.de

## ABSTRACT

The study investigated how the movements of the tongue vary across different ages for prosodic focus marking. We collected articulatory recordings from 44 German speakers, aged 19 to 79 years, and analyzed the phonetic characteristics of vowel production, such as position, velocity, and duration of the vocalic gesture. Across all ages we found more extreme vocal tract configurations to signal prominence, but the variability of the target positions in the vowels was considerably higher with advancing age. The spatial modifications were accompanied by longer vocalic movements of the tongue body under prominence induced by longer deceleration phases for all speakers. While younger speakers showed a fine-grained temporal marking for all degrees of prominence under investigation (background, broad and contrastive focus), older speakers produced durational differences only to differentiate between unaccented and accented conditions (background vs. broad/contrastive focus). Interestingly, the highlighting strategies were less fine-grained but considerably strengthened with advancing age.

**Keywords:** prosodic prominence, aging, speech kinematics, hyper-articulation, vowel production

## 1. INTRODUCTION

In West-Germanic languages, the main prominence of the sentence is usually placed on the focussed parts [1], while other parts within the sentence are considered out-of-focus, i.e., in background position. It has been shown that focus structure determines the position of the nuclear pitch accent and that pitch is a dominant parameter to distinguish between unaccented (background) and accented (broad/narrow/contrastive focus) parts of the utterance on the level of speech production and perception [1, 2]. Other constant phonetic parameters of prosodic modulation involve more subtle, but systematic changes in articulation. Prosodic modulations of vowels and consonants lead to longer, larger and faster articulatory movements to signal prominence in terms of hyper-articulation [3, 4, 5].

To adapt to communicative demands, speakers modulate multiple phonetic cues of the rather flexible prosodic system to differentiate between different degrees of prosodic prominence. A recent study [6] compared different phonetic variables for focus marking. The authors report that the relative importance of the phonetic variables differ for comparisons across accentuation (background vs. broad focus) and within accentuation (broad vs. narrow vs. contrastive focus). Spatio-temporal characteristics of tongue body kinematics play a moderate role to encode the opposition between accented and unaccented syllables (e.g., background vs. broad/contrastive focus), while modifications across different focus types with nuclear accents on the same syllable are more subtle, e.g., broad vs. contrastive focus. The authors argue that tongue body positions are already peripheral under accentuation even in broad focus so there might not be enough space for further adjustments [6].

Prosodic prominence requires a high amount of physical control. Previous studies on prominence marking involving localized hyper-articulation strategies are restricted to younger or middle-aged adults. Thus, little is known about how prosodic strategies evolve with advancing age. For gross motor control, such as limb movement, aging involves the loss of flexibility and muscular strength resulting in smaller and slowed down movements [7]. Further, it can affect the execution of a movement leading to asymmetrical movement patterns in terms of prolonged deceleration phases and greater motor variability decreasing the degree of accuracy of goal-directed movements in older adults [8]. Studies on speech motor control report on a slowed-down speech system when measuring articulation rates in the acoustic domain [9, 10], while spatial modifications in terms of a more centralized vowel space are rather unclear [11]. Interestingly, there is some evidence from an acoustic speech production study, that variability in older speakers may increase, at least in the temporal domain [12]. Kinematic studies of speech production found slower vocalic tongue body movements in older speakers accompanied by asymmetrical velocity profiles in terms of longer deceleration phases [13]. However, tongue body velocities are either reported to decrease, or to be unaffected by aging [13, 14]. For prosodic focus marking, there is one small articulatory study investigating effects of prominence on aging including 4 older and 4 younger speakers [14, 15]. The results indicate that older speakers might prolong vocalic tongue body movements and the respective deceleration phase to signal prominence in a stronger way than younger speakers.

A clear picture of articulatory focus marking under advancing age is still missing. Hence, the present study investigates focus-related modulations of the tongue in 44 German speakers of different ages by using electromagnetic articulography.

## 2. METHODOLOGY

*Speakers and Recordings*: Speech data of 44 German speakers (19 females, 25 males) aged between 19 and 79 years were recorded with an Electromagnetic Articulograph (AG 501). Due to the world pandemic situation[1], there are no speakers between the ages of 31 and 44. However, based on the literature on effects of aging on speech motor control, it can be assumed that this should be a rather uncritical time span for our research question. For all speakers, sensors were placed on the lower and upper lip, tongue tip and tongue body. The tongue sensors were placed approximately $1cm$ and $4cm$ behind the tongue tip. Reference sensors on the cartilage behind the ears were used for head correction and rotation on the midsagittal plane.

*Speech material*: Ten target words were elicited in an interactive animated game scenario to keep sentence prosody as natural as possible. The target words were disyllabic girl names (’$C_1V_1.C_2V_2$) with stress on the first syllable. All target words were embedded in a carrier sentence, such as “Der Opa hat der Mali/Mila/Mula/Lani/Lina/Luna gewunken” (English: “*The grandpa waved to Mali/Mila/Mula/Lani/Lina/Luna*”) and appeared in three different focus conditions, i.e., background (target word is unaccented), broad and contrastive focus (target word bears the nuclear pitch accent). $V_1$ contains one of the five peripheral vowels: /i:, e:, a:, o:, u:/. In total 1320 target words went into the analysis (10 target words $\times$3 focus conditions $\times$44 speakers).

*Kinematic tongue body measures of $V_1$ production*: For temporal measures we calculated (i) movement duration (in $ms$), i.e., the gestural activation interval for $V_1$ from start to maximum target, (ii) maximum velocity (in $mm/s$), (iii) duration of the acceleration phase (in $ms$), i.e., the interval from start to peak velocity and (iv) duration of the deceleration phase (in $ms$), i.e., the interval from peak velocity to target ($mm/s$). In a mass-spring model, the acceleration phase can be related to the stiffness of the movement [4]. For positional measures, the (v) vertical and (vi) horizontal tongue body positions according to the window method were obtained for $V_1$ [16]. Therefore, all position values of the articulatory tongue body trajectory during the first half of the acoustic vowel were averaged to detect the articulatory position for $V_1$.

*Statistics*: The statistical analysis was done in R v4.2.2 [17]. Generalized additive mixed models (GAMMs) were performed using the packages `mgcv` v1.8-40 [18] and `itsadug` v2.4.1 [19]. The GAMMs were applied to the duration, velocity, acceleration and deceleration of the vocalic tongue body movement. The model structure was the same for all measurements: focus condition (background, broad, contrastive) was entered as a fixed parameter and as a smooth term by age. Random intercepts for the preceding consonant and vowel were added. The formula for the model structure was the same for all measurements: `parameter ~ focus condition + s(age,by=focus condition) + s(vowel,bs='re') + s(consonant,bs='re')`.

## 3. RESULTS

### 3.1. Spatial modifications

For the spatial analysis, speakers have been divided into five groups to provide a better overview of the descriptive data: 19-30 years (n=17), 45-59 years (n=11), 60-69 years (n=8), and 70-79 years (n=8). Fig. 1 provides vowel charts for the kinematic tongue body positions, separately for focus conditions (background, broad focus, contrastive focus), vowel type (/i:, e:, a:, o:, u:/) and the five age groups.

The figure demonstrates that all age groups show more extreme vocal tract configurations from background to contrastive focus to enhance vowel place features in terms of hyper-articulation (Fig. 1; from left to right). This is especially the case for the low vowel /a/, which is directly related to the opening of the oral cavity to expand sonority. The main difference between the age groups lies in the variability of the target positions. While in the younger group, inter-speaker variability is low with a clear distinction between neighboring vowels, the vowel dispersion considerably increases with advancing age resulting in greater overlapping vowel positions, e.g., for /i - e/ or /u - o/.

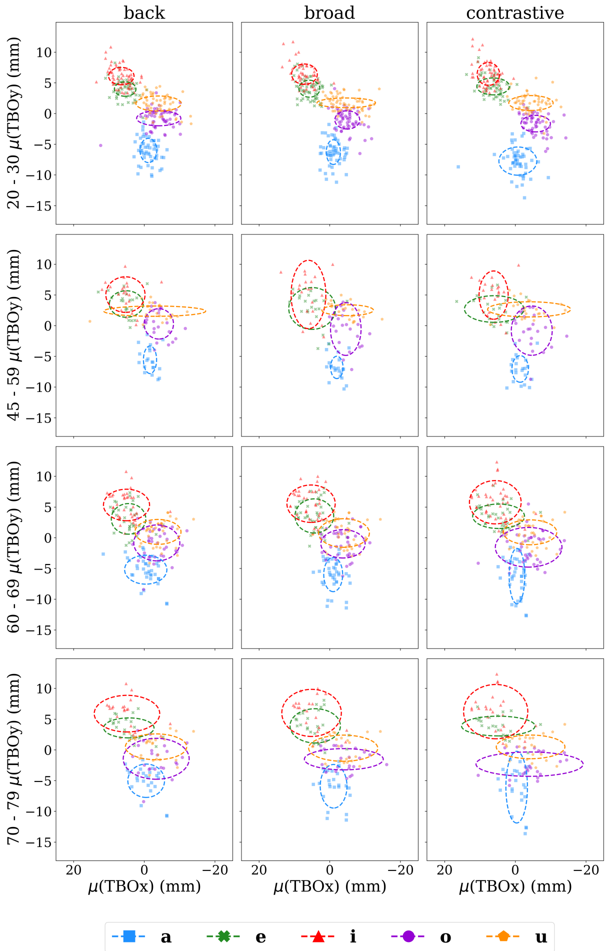


**Figure 1:** Articulatory vowel space based on vertical and horizontal tongue body positions during the target vowels. Centers of the ellipses are based on the means, width and height are based on the variances. Each row represents a different age group.

### 3.2. Temporal modifications

Fig. 2 visualizes the estimates of focus structure on the vocalic tongue body durations. The horizontal axis shows the age of the speakers. The upper plot shows the shaded confidence bands (95% confidence interval) for the movement durations separately in the background (red), broad focus (green) and contrastive focus (blue). The following plots (b-d) show differences between the focus conditions. Note, that the differences between 'background-broad' and 'background-contrastive' reflect the comparison across accentuation, while 'broad-contrastive' implies a comparison within accentuation. The red-shaded vertical areas in the difference plots indicate that the confidence intervals of the compared conditions are significantly different from zero. Regarding the fixed effect, the GAMM for duration revealed a significant effect of focus condition ($p < 0.001$). No data was available for speakers between 31-44 years.

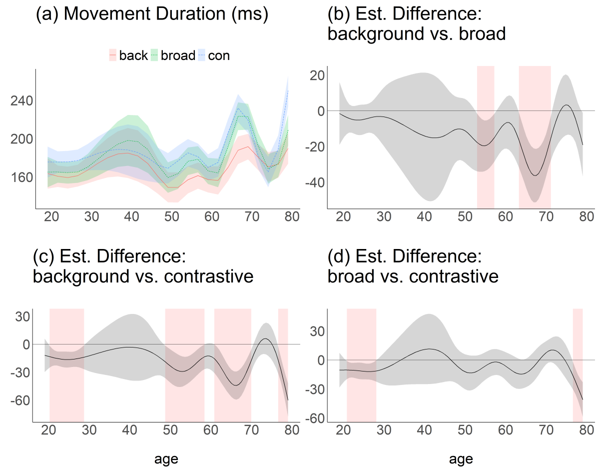


**Figure 2:** **(a)–(d)** Non-linear smooths and difference plots for durations of the vocalic tongue body movement.

The younger speakers below 30 years increase the vocalic tongue body movement on average by $16ms$ across ('background-contrastive') and $11ms$ within accentuation ('broad-contrastive'). The middle-aged and older speakers, however, differentiate only across accentuation at an age older than 48 years. Interestingly, the articulatory strengthening of accented syllables in terms of kinematic vowel prolongations are much stronger for speakers beyond 60 years (about $28ms$ from background to contrastive). When taking the background condition as the baseline in the upper plots (2(a), red-shaded bands), we can see that overall durations of vocalic movements increase with age (about $24ms$), but the prominence marking appears to have an 'extra boost' on the prominent syllable (about $33ms$).

The maximum velocity of the tongue body movement during vowel production is another parameter we investigated. However, the GAMM (see Fig. 3(a)) revealed no effect of focus condition ($p > 0.05$). Moreover, no differences between the focus conditions across age were observed on the maximum speed. There is a trend of decreased maximum speed in speakers older than 60 years when compared to younger and middle-aged speakers (about $13mm/s$ slower for older speakers), though the pattern is not clear.

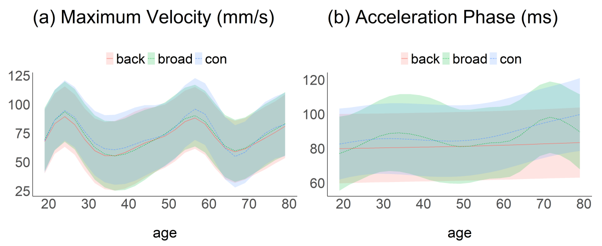


**Figure 3:** Non-linear smooths for (a) maximum velocity and (b) acceleration phase of the tongue body movement.

Fig. 3(b) displays the effect of focus conditions on the acceleration phase of the vocalic tongue body movement. Interestingly, most speakers do not modulate the acceleration phase under prominence. The GAMM for testing differences in acceleration phases revealed a significant effect of focus condition ($p < 0.001$). Significant differences were found between 'background-broad' in the age of 65-77 years (about $12ms$ longer in broad focus) and between 'background-contrastive' from 59-79 years (about $12ms$ longer in contrastive).

The deceleration phase appears to be a strong prominence parameter across all speakers (see Fig. 4). Especially younger speakers systematically prolong the deceleration phase across and within accentuation (e.g., 11ms for maximum diverging focus structure), while most speakers older than 52 years show modulations only between accented and unaccented parts of the utterance (about $8ms$ longer from background to broad and $14ms$ longer from background to contrastive). There is a clear trend for longer deceleration phases related to advanced age, as can be taken from the red-shaded confidence intervals in background conditions: When being compared to speakers younger than 30 years, we find $11ms$ longer deceleration phases for speakers older than 60 years and $18ms$ longer for speakers beyond 70 years. The GAMM revealed only a significant effect of the focus condition ($p < 0.001$) on the deceleration phase.

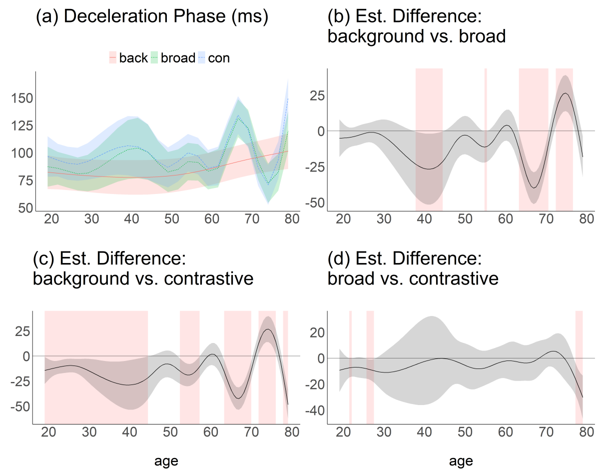


**Figure 4:** **(a)–(d)** Non-linear smooths and difference plots for deceleration phase of the vocalic tongue body movement.

## 4. DISCUSSION & CONCLUSION

Prosodic focus marking implies modifications of the supralaryngeal system. In line with the literature, more extreme tongue configurations are produced to distinguish between accented and unaccented syllables as well as within accentuation. It cannot be confirmed that vowel spaces are more reduced in older speakers, but inter-speaker variability and therefore accuracy of the spatial targets in the vertical and horizontal dimensions appear to be rather high. Even though all speakers produce more peripheral vowels under prominence, the variability in the spatial tongue positions considerably increases with advanced age. Older speakers show more overlap between neighboring vowels than younger speakers, and this process already starts for middle-aged speakers.

The spatial modifications are accompanied by longer vocalic movement durations of the tongue body. Vocalic prolongations are especially induced by longer deceleration phases of the gestural activation interval, while stiffness-related acceleration phases remain stable. The temporal modulations are more fine-grained for younger speakers, i.e., below 30 years, as they differentiate between all degrees of prominence. In contrast, the strength of prosodic modulations are enhanced especially in the temporal domain for speakers older than 60 years. This aging-induced process already starts in middle-aged speakers. We interpret the temporal adjustments as compensation strategies for the increasing variability in the spatial dimension within older speakers. Durational adjustments of vowels in accented target syllables are known to be effective since they lead to an increase in sonority on the perceptual level [20]. We conclude that the results on prosodic focus marking strategies with increasing age reveal a plasticity of the prosodic system that can be adapted through life [21].

## 5. ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation (DFG) in the collaborative research center SFB1252 Prominence in Language (Project-ID 281511265) as well as by the French ANR (ANR-10-LABX-0083, LABEX-EFL).


## 6. REFERENCES


[1] D. R. Ladd, *The Structure of Intonational Meaning: Evidence from English*. Indiana University Press, 1980.

[2] S. Baumann and B. Winter, "What makes a word prominent? Predicting untrained German listeners' perceptual judgments," *Journal of Phonetics*, vol. 70, pp. 20–38, Sep. 2018.

[3] B. Lindblom, "Explaining Phonetic Variation: A Sketch of the H&H Theory," in *Speech Production and Speech Modelling*, ser. NATO ASI Series, W. J. Hardcastle and A. Marchal, Eds. Dordrecht: Springer Netherlands, 1990, pp. 403–439.

[4] T. Cho, "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English," *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3867–3878, Jun. 2005, publisher: Acoustical Society of America.

[5] J. Harrington, J. Fletcher, and M. E. Beckman, "Manner and place conflicts in the articulation of accent in Australian English," in *Papers in Laboratory Phonology: Volume 5*, ser. Papers in Laboratory Phonology, M. B. Broe and J. B. Pierrehumbert, Eds. Cambridge: Cambridge University Press, 2000, vol. 5, pp. 40–55.

[6] S. Roessig, B. Winter, and D. Mücke, "Tracing the Phonetic Space of Prosodic Focus Marking," *Frontiers in Artificial Intelligence*, vol. 5, 2022.

[7] C. J. Ketcham and G. E. Stelmach, *Movement Control in the Older Adult*. National Academies Press (US), 2004, publication Title: Technology for Adaptive Aging.

[8] E. A. Christou, "Aging and Variability of Voluntary Contractions," *Exerc Sport Sci Rev*, vol. 39, no. 2, pp. 77–84, Apr. 2011.

[9] J. D. Amerman and M. M. Parnell, "Speech timing strategies in elderly adults," *Journal of Phonetics*, vol. 20, no. 1, pp. 65–76, Jan. 1992.

[10] P. H. Ptacek, E. K. Sander, W. H. Maloney, and C. C. R. Jackson, "Phonatory and Related Changes with Advanced Age," *Journal of Speech and Hearing Research*, vol. 9, no. 3, pp. 353–360, Sep. 1966, publisher: American Speech-Language-Hearing Association.

[11] D. D'Alessandro and C. Fougeron, "Changes in Anticipatory VtoV Coarticulation in French during Adulthood," *Languages*, vol. 6, no. 4, p. 181, Dec. 2021, number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[12] P. Tremblay, I. Deschamps, P. Bédard, M.-H. Tessier, M. Carrier, and M. Thibeault, "Aging of speech production, from articulatory accuracy to motor timing," *Psychol Aging*, vol. 33, no. 7, pp. 1022–1034, Nov. 2018.

[13] A. Hermes, J. Mertens, and D. Mücke, "Age-related Effects on Sensorimotor Control of Speech Production," in *Interspeech 2018*. ISCA, Sep. 2018, pp. 1526–1530.

[14] D. Mücke, T. Thies, J. Mertens, and A. Hermes, "Age-related effects of prosodic prominence in vowel articulation," in *12th International Seminar on Speech Production*, New Haven, United States, Dec. 2020, pp. 126–129.

[15] T. Thies, A. Hermes, and D. Mücke, "Compensation in Time and Space: Prominence Marking in Aging and Disease," *Languages*, vol. 7, no. 1, p. 21, Mar. 2022, number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[16] S. Roessig, *Categoriality and continuity in prosodic prominence*. Language Science Press, Jan. 2021, publication Title: Language Science Press.

[17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[18] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society (B)*, vol. 73, no. 1, pp. 3–36, 2011.

[19] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting time series and autocorrelated data using gamms," 2022, r package version 2.4.1.

[20] M. Beckman, J. Edwards, and J. Fletcher, "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Gesture, Segment, Prosody*, ser. Papers in Laboratory Phonology, D. R. Ladd and G. J. Docherty, Eds. Cambridge: Cambridge University Press, 1992, pp. 68–89.

[21] E. De Leeuw and C. Celata, "Plasticity of native phonetic and phonological domains in the context of bilingualism," May 2019, accepted: 2019-10-16T14:01:23Z Publisher: Elsevier.