Effects of meter, genre and experience on pausing, lengthening and prosodic phrasing in German poetry reading

Petra Wagner, Simon Betz

Phonetics Work Group, Faculty of Linguistics and Literary Studies & CITEC, Bielefeld University, Germany

petra.wagner@uni-bielefeld.de

Abstract

The adequate and pleasant delivery of poetic speech remains a challenge for humans and machines alike. The present corpus study analyzes factors and strategies that characterize the stylistic expression of poetry and prose by professional actors as well as laypersons with musical training, focusing on their pausing, lengthening and intonation at verse boundaries. Our results show a clear influence on speakers' experience in modulating their speech with respect to prosodic timing: professional actors systematically insert more and more diverse prosodic boundaries and pauses than laypersons, and make strategic use of lengthening at verse endings in poetic speech. Our results further point out the relevance of pausing and lengthening as a time-buying strategy that enhances speech fluency, and we make tentative suggestions for modeling (poetic) speech in expressive speech synthesis.

Index Terms: poetry, meter, expressive speech, speech style, speaker experience

1. Introduction

Reciting poetry adequately is difficult for humans and machines alike, and what constitutes an adequate, successful or pleasant poetic style is hitherto not well understood. The present study therefore aims at enriching our knowledge on the genre specific delivery of poetry vs. prose, with an emphasis on pausing and lengthening behaviour. We particularly target speakers' background expertise in poetry reading to better understand what constitutes a more or less pleasant, adequate, or successful, delivery of poetry reading. Currently, little is known about the influence of the level of expertise on poetic speech delivery, with studies typically focusing either on layperson productions [1, 2], or individual professional speakers [3, 4].

Most related research within phonetics has concentrated on the rhythmic realizations of different poetic meters such as iambic vs. trochaic rhythms across genres, with the main findings being that poetic speech delivery tends to be slower than comparable prose [5, 6, 1]. Also, poetic speech shows an - albeit small - tendency to emphasize rhythmic structure by a meter-specific adaptation of timing patterns [6, 7]. A long-standing belief has been that iambic and trochaic prosodic feet can be differentiated acoustically according to the psychoacoustic rules of the iambic-trochaic law, with trochees being expressed by (foot initial) high intensities, and iambs by (foot final) lengthenings. This view has been recently called into question by [8], who claim that both iambic and trochaic foot are expressed by the general prosodic rules of prominence and grouping, not by meter-specific acoustic-phonetic correlates.

More recently, genre-specific tendencies in the rhythmic-prosodic delivery of poetry vs. prose were further confirmed in eye-tracking studies, which showed that poetry reading is less prone to visual "word skipping" as compared to prose reading [1]. Interestingly, the genre-specificity of poetry reading is also successfully used in improving school children's reading fluency [9].

In addition to rhythmic structure, some attention has been paid to the realization of pauses in poetry and prose. A recent study on the distribution of pauses in Czech prose and poetry readings found surprisingly similar strategies for pausing across the two genres [2], even if readers did not have access to the verse structure of the text in the prose reading condition, and pausing is a strong predictor of perceived pleasantness across varieties of Portuguese poetry declamations [10, 11]. A qualitative study on the delivery of a single professional poetry reader in Brazilian Portuguese also revealed a strong link between pause realization and textual cues to pauses or phrasing [4]. However, the author also noted a strongly expressive use of pauses. Beyond revealing the linguistic structure, the professional speaker used pauses to signal anxiety and thoughtfulness, differentiated by form, duration modulation and the absence or presence of breathing noise.

The strong link between text and pausing has also been addressed by existing computational approaches to metricalization and pause prediction of poetic speech [12, 13]. Differentiated models of poetic pausing particularly need to take into account cases where text and pausing behaviours are not well aligned. In European poetry, *enjambement* is a common stylistic device to elicit pauses visually by introducing line breaks that indicate verse endings in places where a pause is syntactically and semantically inadequate. [3] explored, how this "pausing dilemma" was realized by a professional speaker of German. This speaker realized enjambement in a way that employed (phrase-final) lengthening, but avoided a silent pause or intonational cues of phrase finality. This strategy can be called a phonetically hybrid realization, with markers of phrase-finality and continuity present at the same time [14]. It is yet unclear whether this strategy of enjambement resolution is an individual speaker's choice, or can be generalized further. In their automatic analysis of enjambement realization, [15] found culture-specific preferences of whether enjambements co-occured with a pause, or not, while different literary traditions may also favour their own recitation styles [3].

Interestingly, the findings by [3] on actively avoiding silent pauses in poetry reading show similarities with research on conversational pausing and disfluency research. In particular, it was found that word internal lengthenings, produced with an ongoing phonation and little pitch movement, are a frequent and unobtrusive device to delay speech production, and can even be successfully employed in synthetic speech [16, 17]. That way, speakers actively delay the production of content with-

out physically interrupting the generation of a speech signal, and aid perceptual fluency by an unobtrusive hesitation. Hesitations are often thought to be indicative of ongoing speech planning processes [18], but they can also act as an attention guiding device [17], express different epistemic states [19], act as focusing device [20], or indicate the continuation of a temporally delayed message as a floor holding device. Extreme forms of lengthenings have also been reported from other areas of expressive speech production such as football commentaries [21]. Summing up, lengthenings share many functions and forms with those attributed to different types of pauses in poetry, and have the added benefit of being perceived as less disturbing than filled pauses produced by humans or machines alike [22, 17].

Our research objectives for the current study are two-fold: First, we want to better understand whether and how professional actors differ from laypersons in their realization of prosodic boundaries and pauses. Second, we want to understand better whether and how these differences show in genre-specific reading, namely in Germany poetry as opposed to prose. To take into account potential meter specific aspects of timing, we furthermore differentiate iambic and trochaic meters. We will discuss the implications of our findings for expressive speech synthesis and overall fluency in speech delivery.

2. Methods and Materials

2.1. Corpus

We used corpus recordings of 12 native speakers of German. Six (3f, 3m) of these are professional actors with extensive training in poetry and prose recitation (henceforth: actors), the other six (3f, 3m) are active singers in a gospel choir, with extensive training in rhythmic performance, but not public reading (henceforth: laypersons). All speakers read aloud a set of 10 German poems from the 19th and 20th century, from various authors, which are prototypical representatives of 4 different meters (2 dactyls, 3 iambs, 3 trochees, 2 poems with a varying meter, i.e., "lieder")). Each speaker first read the set of poems without the original type setting, as a regular prose text (henceforth: prose), and in a second recording in the original typeset containing verses and stanzas (henceforth: poetry). Prior to the poetry recordings, all speakers were informed to uphold the intended meter of each poem, but without specifying how this should be done. During poetry recordings, a percussionist controlled that intended meters were audibly realized. The current study is only concerned with the recordings of bisyllabic meters[1], trochees (n=13,610 syllables) and iambs (n=23,310 syllables), distributed across the two genres (poetry: n=18,465 syllables, prose: n=18,455 syllables) and levels of experience (actors: n=18,833 syllables, laypersons: n=18,087 syllables).

2.2. Annotations and feature extraction

All recorded materials are manually annotated for the following features:

- syllable boundaries and 'silent' pauses
- (poetic) foot boundaries

---

[1]The analyzed iambic poems were: Der Blinde (Georg Heym), First monologue by Iphigenie from "Iphigenie auf Tauris" (Johann Wolfgang von Goethe), Vom Schnee (Durs Grünbein); trochaic poems were: Der Träumer (Christian Morgenstern), Bei den Wassern Babels saßen (Heinrich Heine), Legende von der Entstehung der Buches Lao Te King auf dem Weg des Laotse in die Emigration (Bertolt Brecht)

- verse boundaries
- prosodic phrase boundaries
- poetic meter
- canonical and phonetically realized syllable
- pitch accent type (GToBI convention, [23])
- phrase boundary type (GToBI convention, [23])

For reasons of comparison, verse boundaries and foot boundaries are compared at identical places across both genres, even though prose strictly speaking does neither have poetic feet nor verses. However, given the strong text-verse alignment found in previous studies, we consider such a comparison reasonable. Speakers' level of experience (actors, laypersons) as well as the read genre (prose, poetry) was extracted for the analysis as well. Several further acoustic-phonetic and categorical features were extracted out of the annotated recordings. For the present analysis, we particularly used

- foot duration (log normalized)
- syllable and pause duration (log normalized)
- contour of prosodic boundary (rising, falling), extracted out of the GToBI annotations,
- strength of prosodic boundary (ip, IP), with "H-" and "L-" mapped onto intermediate "ip" boundaries, and all others onto strong "IP" boundaries.

2.3. Statistical analyses

In order to assess the influence of genre, speakers' experience and meter on their realization of pausing and phrasing, we built (generalized) linear mixed regression models to predict (1) foot durations, (2) syllable durations, (3) pause durations, (4) pause distributions, (5) boundary strength as well as (6) boundary contours. We used *verse finality* (final, nonfinal), *genre* (poetry, prose), and *experience* (actors, laypersons) as fixed factors, and *poem* and *speaker* as random factors (random intercepts due to convergence issues). All statistical analyses were carried out in R [24], with the *lme, lmerTest* packages [25, 26] for model building. For model reduction and optimization, we used the *buildmer* package [27], starting from maximal model with all factors and interactions, reducing it until a model is reached that converges but does not compromise model accuracy (log-likelihood comparison). Post-hoc contrasts were calculated using *emmeans* [28] (Tukey), plots were generated with [29, 30]. Due to space restrictions, a selection of effects will be discussed and reported below. The full models and anonymized data sets (excluding the recordings) and analysis scripts can be made available upon request.

3. Results

3.1. Pause distributions

Actors produce more pauses than laypersons ($est = 0.46842, se = 0.03, t = 14.14, p < 2e - 16$). The model also reveals a 3-way interaction of *verse finality*, *meter* and *experience* ($est = -0.78, t = -2.31, p = 0.021$), but no effect of *genre* (cf. Fig. 1). This interaction reveals a similar tendency both for laypersons and actors alike: speakers have a low probability to insert pauses at verse boundaries in iambic contexts, but lean more towards pausing at verse boundaries in trochaic contexts ($p < 0.0001$) for all contrasts). The predicted probability for avoiding pauses at verse boundaries in iambic contexts is approximately $\geq 70\%$.
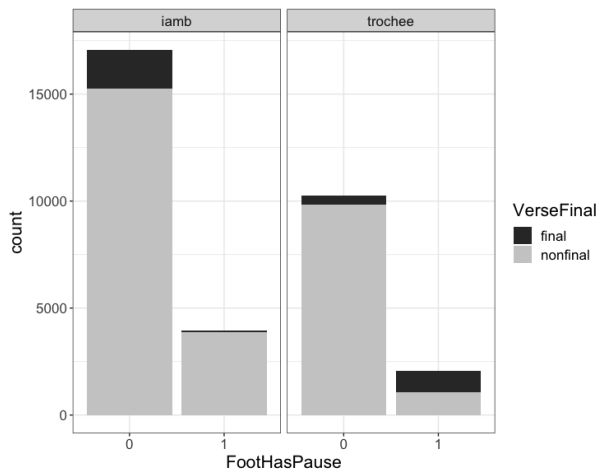
Figure 1: *Number of occurrences of pauses across different meters and verse boundaries. Pauses are avoided at verse boundaries in iambic contexts, but tend to occur in verse final trochaic contexts.*

Figure 2: *Distributions of pause durations (log) of actors and laypersons across genres. Actors produce shorter pauses in poetry reading, while laypersons produce (slightly) longer pauses in poetry reading.*

3.2. Pause durations

The model reveals a main effect of *verse finality* on pause duration: pauses occurring at verse boundaries are longer ($est = 0.72, se = 0.10, t = 7.32, p = 3.10e - 13$). There is an interaction of *experience* and *genre* ($est = -0.24, se = 0.04, t = -5.43, p = 6.17e - 08$), which shows significant contrasts (cf. Fig. 2): actors produce longer pauses in prose ($p < 0.0001$), while laypersons produce longer pauses in poetry ($p < 0.004$).

3.3. Syllable durations

The model reveals a main effect of *verse finality* and *genre* on syllable duration: In poetry ($est = 1.52, se = 7.99e - 03, t = 57.45, p < 2e - 16$) and before verse boundaries ($est = 1.52, se = 7.67, t = 19.81, p < 2e - 16$), syllables are lengthened, independent of experience.

3.4. Foot durations

The model reveals a 3-way interaction ($est = -1.57e - 01, se = 2.64e - 02, t - 5.95, p = 2.89e - 09$) of *genre*, *experience* and *verse finality*: while verse final feet are lengthened by both actors and laypersons in prose ($p < 0.0001$ for both contrasts), only actors show a systematic verse final lengthening effect in poetry reading ($p < 0.0001$, cf. Fig. 3). That is, laypersons show a particular poetry reading style that restricts verse final lengthening to the pre-boundary syllable.

3.5. Boundary strengths

There are main effects of *experience*, *meter* and *verse finality* on boundary strengths, but none of *genre*: verse endings trigger stronger boundaries ($est = -2.63, se = 0.87, z = -3.03, p = 0.002$), trochees trigger stronger prosodic boundaries ($est = 0.91, se = 0.27, z = 3.36, p = 0.001$), and actors are more likely to produce less strong prosodic boundaries ($est = 2.25, se = 0.29, z = 5.84, p = 5.28e - 09$). A look at the overall distribution of boundaries reveals that actors overall employ both final and non-final prosodic boundaries, while laypersons tend to use strong prosodic boundaries almost exclusively. Both speaker groups are consistent in relying on strong

boundaries for verse endings, and less strong boundaries elsewhere (cf. Fig. 4).

3.6. Boundary contours

Due to convergence issues, interactions could not be tested, but a visual inspection indicated a possible interaction of *genre* and *experience* (cf. Fig. 5), with actors producing more falling contours in prose, and laypersons overall preferring rising contours, while almost exclusively pausing in poetry. The model reveals a main effect of *verse finality* and *genre*, but no effects for *meter* or *experience*: In prose, speakers tend to produce more falling contours ($est = 1.25, se = 0.10, z = 12.20, p < 2e - 16$), and at verse boundaries, rising contours are more likely ($est = -2.55, se = 1.05, z = -2.43, p = 0.02$).

4. Discussion and Conclusions

Overall, actors produce more pauses and more prosodic boundaries. Unlike laypersons, they sytematically use intermediate as well as strong prosodic boundaries, while laypersons tend to rely on verse final strong prosodic boundaries. Also, actors systematically distribute their preboundary lengthening throughout the entire foot in poetry reading, while laypersons restrict their preboundary lengthening to the preboundary syllables. Taken together, actors have access to a diverse bundle of time management measures, which they strategically use to optimize the timing of their delivery. Comparable phenomena have been described in spontaneous speech production, where speakers are found to "buy time", when speech delivery needs to be delayed (cf. Introduction). We argue that similar time management strategies provide actors with a rich set of opportunities to realize meaningful, style-related modulations which cost time both in planning and realization. One of these "costly", style-related modulations reserved for poetry reading may be the shorter pauses that actors produced in poetry reading. A possible reason for these shorter pauses may be the aim of obtaining a higher degree of perceptual fluency in poetry recitation, which is likely to be disrupted by overly long silent pauses. This interpretation is strengthened further by the finding that actors strategically
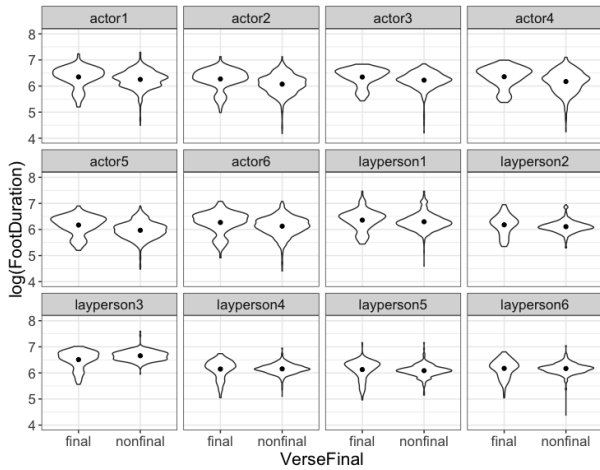
Figure 3: *Distributions of foot durations (log) across all speakers in poetry readings, comparing verse final and non final position. Laypersons do not show systematic final lengthening of feet at verse boundaries.*

Figure 4: *Distributions of boundary strengths at verse endings (final) and elsewhere (non final), for 6 actors and 6 laypersons. Laypersons produce prosodic boundaries almost exclusively at verse endings, and rarely produce intermediate boundaries.*

produced significantly more rising contours in poetry, probably aiding the impression of textual continuation, while laypersons do not differentiate between genres, but favour rising boundary contours overall. Again, this points towards a richer repertoire of actors when shaping the prosodic *gestalt* during their performances. As we did not investigate the text-prosody relationships, our study cannot contribute to shedding further light on the stylistic phenomenon of *enjambement*. Still, our results add further evidence that professional speakers make strategic use of verse final foot lengthening together with pause modulations in poetry reading, which is in line with the findings by [3] on *enjambement* resolution. Maybe somewhat surprisingly, meter had very little impact on the prosodic realizations across genres and level of experience, with two notable exceptions: as opposed to iambs, trochees attract stronger boundaries as well as pauses, while iambs tend to "repel" pauses. Currently, these findings are difficult to interpret straightforwardly, but possibly, the "built in" final lengthening in iambs may be interpreted as a "time buying" lengthening, which makes adding another pause unnecessary. This factor may be an important aspect in the "driving" rhythm often attributed to iambic structures.

We find it particularly striking that many of the abovementioned strategies that characterize actors' reading style were present across *all* professional speakers, and were lacking across *most* laypersons. Even though all our speakers impressionistically differed strongly in their individual performances, actors shared the ability of higher diversity and strategic lengthening and pause distribution. We therefore plan to investigate the perceptual consequences of these strategies, the expectations being that actors' performances are perceived as more pleasant to listen to, due to a higher degree of diversity and a higher degree of differentiated expressiveness. We argue that our results also make some interesting suggestions for how to best exploit the diverse pausing and lengthening strategies in speech synthesis: Pausing and local tempo adjustments by lengthening remain to be an often overlooked aspect in TTS, and should be modeled explicitly, as they may buy valuable time that can be used strategically both by humans as well as technical systems [17]. Since our results show that identical texts may lead to genre specific realizations with respect to pausing,
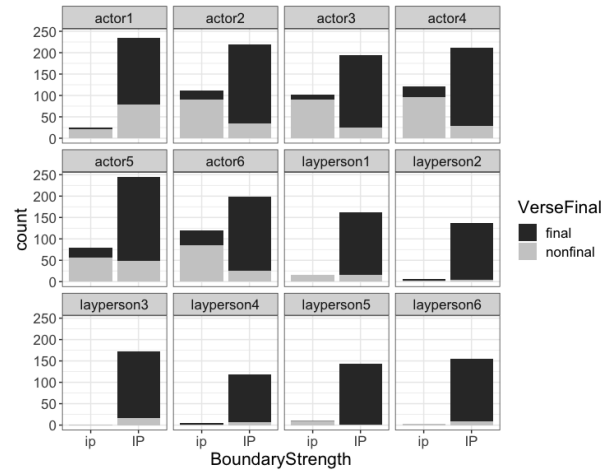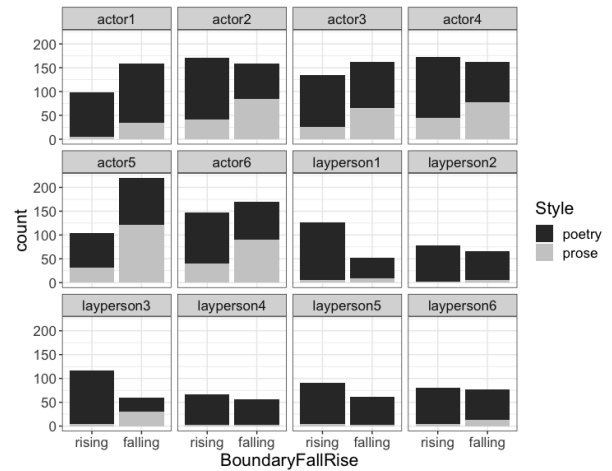

Figure 5: *Distributions of boundary contours across genres, for 6 actors and 6 laypersons.*

tempo and the distribution of prosodic phrases and lengthenings, it may not be sufficient to rely on an automatic connection between text and appropriate reading style generation during training. Rather, a variety of genres or reading styles may have to be taken into account. Furthermore, it may be advisable to investigate different levels of professionalism in reading the training materials. While overly expressive productions often attributed to professional actors may introduce undesired effects, it may be useful to have access to their fine grained strategic adjustments in style. In any case, it would be interesting to compare actors' performances to those of other professional readers, e.g., newsreaders, in providing suitable training materials.

5. Acknowledgements

6. References

[1] S. Blohm, S. Versace, S. Methner, V. Wagner, M. Schlesewsky, and W. Menninghaus, "Reading poetry and prose: Eye movements and acoustic evidence," *Discourse Processes*, vol. 59, no. 3, pp. 159–183, 2022.

[2] P. Šturm and J. Volín, "Occurrence and duration of pauses in relation to speech tempo and structural organization in two speech genres," *Languages*, vol. 8, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2226-471X/8/1/23

[3] N. Schauffler, F. Schubö, T. Bernhart, G. Eschenbach, J. Koch, S. Richter, G. Viehhauser, T. Vu, L. Wesemann, and J. Kuhn, "Prosodic realisation of enjambment in recitations of German poetry," *Proc. Speech Prosody, Lissabon (Portugal)*, 2022.

[4] I. C. Viola and S. Madureira, "The roles of pause in speech expression," in *Proceedings of the IV Conference on Speech Prosody. Brasil*, 2008, pp. 721–4.

[5] A. Kruckenberg and G. Fant, "Iambic versus trochaic patterns in poetry reading," *Nordic Prosody VI*, pp. 123–135, 1993.

[6] J. Bröggelwirth, "Ein rhythmisch-prosodisches modell lyrischen sprechstils," Ph.D. dissertation, Universitäts-und Landesbibliothek Bonn, 2007.

[7] P. Wagner, "Meter specific timing and prominence in german poetry and prose," *Understanding prosody: The role of context, function and communication*, pp. 219–237, 2012.

[8] E. Moghiseh, M. Sonderegger, and M. Wagner, "The iambic-trochaic law without iambs or trochees: Parsing speech for grouping and prominence," *The Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 1108–1129, 2023.

[9] T. Rasinski, W. H. Rupley, and W. D. Nichols, "Synergistic phonics and fluency instruction: The magic of rhyming poetry!" *New England Reading Association Journal*, vol. 44, no. 1, p. 9, 2008.

[10] P. A. Barbosa, "Pleasantness and wellbeing in poem declamation in European and Brazilian Portuguese depends mostly on pausing and voice quality," *Frontiers in Communication*, vol. 7, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomm.2022.855177

[11] ——, "The dance of pauses in poetry declamation," *Languages*, vol. 8, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2226-471X/8/1/76

[12] T. Baumann, H. Hussein, and B. Meyer-Sickendiek, "Style detection for free verse poetry from text and speech," in *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 1929–1940. [Online]. Available: https://aclanthology.org/C18-1164

[13] H. Hussein, B. Meyer-Sickendiek, and T. Baumann, "Free cerse and beyond: How to classify post-modern spoken poetry," 05 2020, pp. 690–694.

[14] R. Tsur and C. Gafni, "Enjambment–irony, wit, emotion. a case study suggesting wider principles," *Studia Metrica et Poetica*, vol. 5, no. 2, pp. 7–28, 2018.

[15] H. Hussein, B. Meyer-Sickendiek, and T. Baumann, "Automatic detection of enjambment in German readout poetry," *Proceedings of the Speech Prosody, Poznan, Poland*, 2018.

[16] S. Betz, J. Voße, S. Zarrieß, and P. Wagner, "Increasing recall of lengthening detection via semi-automatic classification," in *Proceedings of Interspeech 2017*. Stockholm, Sweden: ISCA, 2017, pp. 1084–1088.

[17] S. Betz, B. Carlmeyer, P. Wagner, and B. Wrede, "Interactive hesitation synthesis: modelling and evaluation," *Multimodal Technologies and Interaction*, vol. 2, no. 1, p. 9, 2018.

[18] R. J. Lickley, "Fluency and disfluency," *The handbook of speech production*, pp. 445–474, 2015.

[19] A. Kirkland, H. Lameris, E. Székely, and J. Gustafson, "Where's the uh, hesitation? the interplay between filled pause location, speech rate and fundamental frequency in perception of confidence," in *Proceedings of Interspeech*, 2022, pp. 18–22.

[20] W. L. Chafe, "Some reasons for hesitating," in *Temporal variables in speech*. De Gruyter Mouton, 2011, pp. 169–180.

[21] J. Trouvain, "Between excitement and triumph-live football commentaries in radio vs. tv." in *ICPhS*, 2011, pp. 2022–2025.

[22] O. Niebuhr and K. Fischer, "Do not hesitate!-unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance." in *Proceedings of Interspeech*, 2019, pp. 544–548.

[23] M. Grice and S. Baumann, "Deutsche Intonation und GToBl," *Linguistische Berichte*, pp. 267–298, 2002.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[25] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[26] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[27] C. C. Voeten, *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*, 2023, r package version 2.8. [Online]. Available: https://CRAN.R-project.org/package=buildmer

[28] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2022, r package version 1.7.3. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[29] D. Lüdecke, "ggeffects: Tidy data frames of marginal effects from regression models." *Journal of Open Source Software*, vol. 3, no. 26, p. 772, 2018.

[30] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org