Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Language redundancy effects on the prosodic word boundary strength in Standard German

Tianyi Zhao[1], *Tina Bögel*[1], *Alice Turk*[2], *Ricardo Napoleão de Souza*[2]

[1]University of Konstanz
[2]University of Edinburgh

tianyi.zhao@uni-konstanz.de, Tina.Boegel@uni-konstanz.de, a.turk@ed.ac.uk, R.N.deSouza@ed.ac.uk

Abstract

The Smooth Signal Redundancy Hypothesis proposes a complementary relationship between language redundancy and acoustic redundancy mediated via prosodic prominence and boundary structure. To test this hypothesis, the current study investigates the effects of lexical and repetition frequency on boundary-related segmental (+pause) duration patterns at prosodic word boundaries in Standard German. Results are consistent with predictions made by the Smooth Signal Redundancy Hypothesis, showing an inverse correlation between lexical frequency and duration that is mediated by prosodic structure: Word boundary-related target intervals for frequent words were shorter than the corresponding intervals for infrequent words, and effects on non-boundary-related intervals were not significantly different for frequent vs. infrequent words. These effects indicate a preference for producing stronger prosodic boundaries in case of low language redundancy. Repetition effects were attested for all intervals including the boundary-related intervals.

The effects appear to be stable across varying speech tempo by different speakers and targets, even when the factor of speech tempo is controlled for. This is consistent with the view that speech tempo, as a global factor that modulates the overall utterance, does not interfere with localized acoustic redundancy.

Index Terms: Smooth Signal Redundancy Hypothesis, German, language redundancy, prosodic boundary strength

1. Introduction

The aim of this paper is to test a prediction of the Smooth Signal Redundancy Hypothesis (SSRH, [1]) that speakers should plan stronger prosodic boundaries between words when the words have lower recognition likelihood due to linguistic and pragmatic factors, i.e., are less predictable during communication.

According to the SSRH ([2, 3, 4, 1]), the recognition likelihood (i.e., signal redundancy) of linguistic items in speech is spread evenly throughout the utterance to ensure successful communication between speakers and listeners. Signal redundancy refers to the combination of language redundancy (e.g., utterance length and predictability based on lexical, syntactic, semantic and pragmatic factors) and acoustic redundancy, i.e., recognition likelihood based on acoustic properties alone, such as duration, F0, acoustic correlates of loudness, and acoustic properties that signal phonemic distinctiveness such as formant structure and other spectral properties. Language redundancy and acoustic redundancy are assumed to be inversely correlated, in that items with lower language redundancy (e.g., lower lexical frequency) are correlated with higher acoustic redundancy (e.g., in form of longer duration) and vice versa. The relationship between language redundancy and acoustic redundancy is proposed to be mediated via prosodic structure, i.e., prominence [3, 4] and prosodic boundaries [1]. The purpose of our study is to explore this mediation between language and acoustic redundancy based on prosodic boundary strength. Relevant findings in the literature are consistent with this prediction. For example, [5] showed that phrase-final syllables in German were longer when they were less predictable from preceding context, where this duration difference was further modulated by boundary strength. A follow-up study ([6]) further confirmed the influence of contextual predictability on syllable and pause duration (with a trading relationship between the two for prominent items). In the current study, we focus on two frequency measures, i.e., lexical frequency and repetition, addressing the question of whether infrequent words and words mentioned for the first time are associated with stronger boundaries.

In general, there is a large body of literature assessing the influence of lexical frequency and repetition on duration. Most of these studies have reported effects of frequency (of words or morphemes) and repetition on the durations of whole words or morphemes ([7, 8, 9, 10, 11], but see [3, 12, 13, 14] who looked at smaller parts of words). For example, theories like the lexical-access-based model ([8]), which relates word duration to lexical activation, would assume that effects of lexical frequency and repetition are implemented via a production mechanism which coordinates lexical activation, strength or level of activation and articulation rate, i.e., it affects all parts of words to a comparable extent. In contrast, the SSRH primarily proposes effects of greatest magnitude on word edges and prominent syllables. There are also alternative explanations for varying pause durations due to language redundancy factors, e.g., pauses before longer, more complex phrases have been explained as providing time for planning ([15, 16]). Such explanations do not involve other acoustic cues of boundaries, for instance, final lengthening or f0 effects as SSRH predicts.

This paper will investigate the relationship between lexical frequency, repetition, and boundary strength (in the form of duration measures and pause) in Standard German with a particular focus on the analysis of which parts of the words are responsible for the effects. This work adds to a growing body of cross-linguistic word on language redundancy effects on speech acoustics.

2. Method

2.1. Materials

Our experiment compared durational patterns on target words with high vs. low frequency elicited in structurally similar carrier phrases. All targets were read three times by the participants during the experimental session. To create the target

 10.21437/SpeechProsody.2024-219

words, ten geographic attributive words were chosen as word stems and were derived into nouns that refer to people from a specific region (all with plural feminine inflections, e.g., *Pariserinnen* ('women from Paris') and *Japanerinnen* ('women from Japan'). The target words were designed to be long (five syllables, with main lexical stress on the second syllable) to isolate boundary-related intervals and stressed-syllable-intervals from each other and from other parts of the word, in order to better determine the location of the relevant effects.

The frequency measures of these words were retrieved from the online corpus *Referenz und Zeitungskorpora* in DWDS (the German Digital Lexical System of the Academy [17], cf. [18]); Only geographic attributive words with more than 3000 hits (raw values) were used. The list of targets was additionally checked in a pilot experiment to ensure that the target words were familiar to the participants.

For each word, a paired nonsense word was generated by alternating the onset of the stressed syllable of the German word in a consistent manner (e.g., voiced plosives were changed to different voiced plosives). The other segments in the words remained the same. All nonsense words were phonotactically permissible in German, e.g., *Pariserinnen-Pabiserinnen*, *Japanerinnen-Jatanerinnen*). In total, ten target pairs (=20 targets) were created. The nonsense words can be assumed to have extremely low lexical frequency measures (=*infrequent*) compared to the real German words (=*frequent*), as participants should have never encountered these nonsense words in their daily life. The target sentences were constructed with a numeral, the target and a verb phrase (see Table 1). The sentence consisted of four prosodic words.

Table 1: *The target pair 'Berlinerinnen–Berbinerinnen' in their carrier sentences. The numeral remained the same within each sentence pair. Verbs in the carrier sentences were altered to reduce similarity between target sentences.*

Frequency	Target
frequent	Drei Berlinerinnen fahren zusammen 'Three women from Berlin travel together'
infrequent	Drei Berbinerinnen kochen zusammen 'Three women from Berbin cook together'

2.2. Participants

Thirty German native speakers (6 male and 24 female; mean age = 24.2, *SD* = 4.3) participated in the experiment. They were recruited at the University of Konstanz and received a small payment afterwards. None had any reported language impairments.

2.3. Procedure

The study included a background questionnaire and a recording session. The target sentences were individually presented on a screen and participants were instructed to read them out aloud. At the beginning of the experiment, participants were required to listen to a target word list (produced by a female German native speaker who is one of the authors) in order to support correct pronunciation of the nonsense words, especially in terms of the stress pattern. No visual form of the word list was given.

All 20 target sentences from the materials were randomized into one block along with two fillers per target to prevent the creation of a contrastive focus. Each block was repeated three times by participants during the experiment. Participants were unaware of the purpose of the study. All instructions were given prior to the recording and no further instruction was provided during the actual experiment. Each session took around 20 minutes and was recorded with a condenser microphone in a sound-proof booth in the Phonlab at the University of Konstanz (digitized with a sampling rate of 44.1 KHz, 16-Bit, stereo).

2.4. Data processing and analysis

There were ten target pairs (=20 target sentences) repeated three times per speaker. For the investigation of lexical and repetition frequency effects, the current analysis was only based on the first and the third repetition (sentences with wrong stress patterns or any other incorrect productions were excluded, resulting in 992 sentences in total). All target sentences were first automatically segmented via MAUS [19, 20], then manually checked and corrected with Praat [21] according to the syllable structure of the target and segmentation reliability following standard segmentation criteria [22]). Identifiable pauses[1] preceding and following the target words were annotated, as they might signal the existence of stronger prosodic boundaries. The phrasal prominence located mostly on the target word, with some exceptions on the numerals or verbs.

Three groups of measures were extracted, as illustrated in Tier 3 – Tier 5 in Figure 1[2];

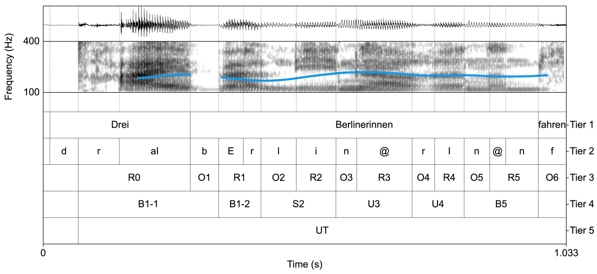

Figure 1: *An example of the annotation scheme for the sentence 'Three women from Berlin travel together' highlighting the relevant intervals annotated for the analysis (Tier 3 – Tier 5; Tier 1 shows the words, Tier 2 the annotation generated by MAUS).*

Tier 3: the respective onsets and rhymes of the targets.

Tier 4: durations of the six broader intervals where some of the onsets and rhymes were merged (e.g., *R0* and *O1* in Tier 3 were merged into one boundary-related interval *B1-1*).

Tier 5: interval *UT*[3] (for the calculation of speech tempo and modelling of the durational changes at the utterance level).

Intervals were merged on the basis of two motivations: 1) The focus of the current study was to disentangle effects at the boundary-related intervals (*B1-1*, *B1-2* and *B5*) from other syllables (S2, U3 and U4; see Figure 1) throughout the targets,

[1] The pauses were identified based on the silence intervals automatically annotated by MAUS (see [23] for reference).

[2] Note that the focus of the annotation in this study was to determine the landmarks associated with consonant and vocalic intervals, i.e., the onset and offset of consonant constrictions, not the onsets and rhymes themselves.

[3] The *UT* interval here is not fully displayed due to limited space.

which was why we measured boundary-related intervals separately from other intervals. The boundary-related intervals consisted of the rhyme interval for the pre-boundary syllable, any pause, and the onset constriction interval for the post-boundary word. This is arguably also the locus of the greatest magnitude of boundary-related lengthening in Germanic languages (see [24, 25]). 2) In case of segmental difficulties, some onsets and offsets of segments could not be reliably distinguished, e.g., the onset [dʁ] and rhyme [eɪ] in the numeral *drei* ('three'), or the onset [n] and the rhyme [ən] in the word-final syllable *B5*.

Given that speakers are highly variant in speech tempo, which is likely to have an impact on their duration measures, a reference for the speech tempo across speakers was determined: A longer interval for the whole utterance ranging from the rhyme of the first syllable to the penultimate syllable was measured to estimate the speech tempo of individual speakers (as partially presented in Figure 1). Speaker tempo was calculated as follows: The interval was first divided by the syllable count within this interval, resulting in one value for every single target. The values of all targets produced by the same speaker were summed up and the mean of the sum was taken as the reference for the speaker tempo.

To analyze durational differences between frequent and infrequent targets, linear mixed effects regression models (lmer) were used with lexical frequency and repetition frequency as fixed effects, speaker tempo and item as random effects (random slopes were added for speaker tempo when necessary) with the Satterthwaite approximation implemented in the R-library lmerTest [26, 27]. The duration of the utterance was additionally modelled to detect overall durational changes. Insignificant interactions were removed and the best-fit models were determined by model comparison via the *ANOVA* test. A chi-square test was applied to detect the association between pause insertions and frequency measures. A paired Wilcoxon rank sum test was conducted to compare speech tempo across repetition.

3. Results

3.1. Speech tempo

As Table 2 shows, participants varied greatly in speech tempo: In the first repetition, the fastest speaker was around 130 ms faster than the slowest speaker (approximate to the average syllable duration, *SD* of all speakers' tempo = 22 ms). The tempo discrepancy was reduced in the third repetition (the fastest speaker was 114 ms faster than the slowest speaker, *SD* = 21 ms). The significant difference of speaker tempo across repetition was confirmed by a paired Wilcoxon rank sum test (V = 123046, $p < 0.0001$).

Table 2: *Distribution of the varying speaker tempo (in ms) across repetition (see the calculation of speech tempo in Section 2.4).*

	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	*SD*
Repetition 1	155	177	190	193	201	285	22
Repetition 3	137	160	174	173	186	251	21

In order to control for the potential duration differences caused by individual speaker tempo in different repetitions, the calculated tempo values were further used in the duration analysis to account for speaker variation.

3.1.1. *Pause at the word boundaries*

As illustrated in Figure 2, participants were more likely to pause at the word boundaries for infrequent than frequent words, and number of insertions considerably decreased by repetition. For the pauses at both prosodic word boundaries (i.e., left and right edges of the target word), chi-square tests showed a significant association between lexical frequency and pause insertions in the first repetition ($\chi^2 = 7.8$, df = 1, $p < 0.01$) as well as in the third repetition ($\chi^2 = 3.92$, df = 1, $p < 0.05$). In addition, the placement of pauses can be significantly associated with times of repetition ($\chi^2 = 40.28$, df = 1, $p < 0.0001$).

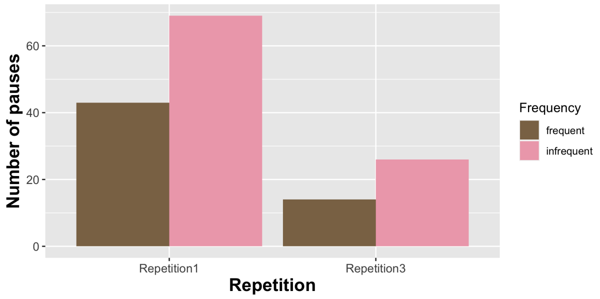

Figure 2: *Number of pauses inserted at both word boundaries among frequent and infrequent targets across repetition.*

3.1.2. *Lexical frequency effects, repetition effects and interaction*

Linear mixed effects regression models were applied to the annotated segments and intervals in Figure 1 (Tier 3 – Tier 5) respectively. To begin with, target sentences as a whole significantly increased in duration when they contained infrequent targets (ß = 0.05, *SE* = 0.014, *t* = 3.309, $p < 0.01$), and decreased if repeated (ß = -0.087, *SE* = 0.025, *t* = -3.41, $p < 0.01$).

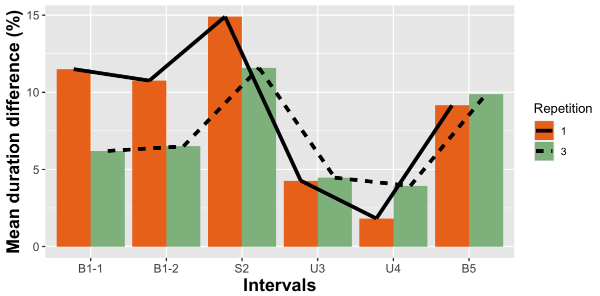

Figure 3: *Mean duration differences (%) of the intervals across frequency condition and repetition respectively, including the target word and both boundaries. Repetition is color-coded and tendency shown by lines.*

Among the three boundary-related intervals, both *B1-1* and *B5* showed significant effects of lexical frequency (*B1-1*: ß = 0.15, *SE* = 0.063, t = 2.42, $p < 0.05$; *B5*: ß = 0.02, *SE* = 0.005, t = 4.05, $p < 0.001$), revealing a significant difference in duration at the boundaries between frequent and infrequent targets (see Figure 3). Repetition effects were further attested for *B1-1* (ß = -0.16, *SE* = 0.047, *t* = -3.36, $p < 0.01$, as hinted by the comparatively large durational differences in Figure 4) than *B5* (ß = -0.03, *SE* = 0.014, t = -1.83, $p = 0.07$, approaching significance). For *B1-2*, lexical frequency was insignificant ($p > 0.1$), deviating from the visualization in Figure 3, but repetition was significant (ß = -0.048, *SE* = 0.021 $t = -2.26$, $p < 0.05$).

Additionally, there was a significant interaction between lexical frequency and repetition for *B1-1* ($\chi^2 = 5.25$, df = 1, $p < 0.05$). To understand the nature of this interaction, data were split by lexical frequency and the repetition effect was tested in these subsets. Results revealed a consistent duration reduction for both frequency conditions in the third repetition, corresponding to the numerical pattern of *B1-1* in Figure 4.

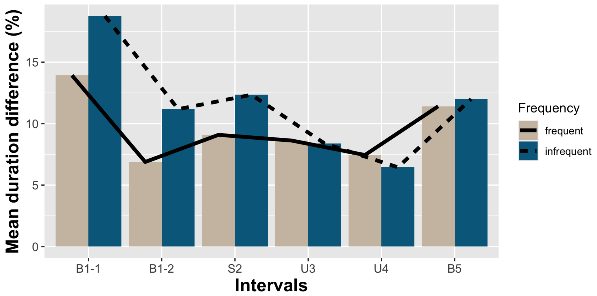

Figure 4: *Similar to Figure 3 but with the emphasis on mean duration differences between repetitions.*

The rest of intervals throughout the targets, including the rhyme of the first syllable in the target, the stressed syllable, the third and penultimate syllables did not show any significant effects of lexical frequency (all $p > 0.1$, see the visualized differences in Figure 3 as a reference). However, they all showed significant repetition effects (*S2*: ß = -0.053, *SE* = 0.022, t = -2.39, $p < 0.05$; *U3*: ß = -0.053, *SE* = 0.022, t = -2.45, $p < 0.05$; *U4*: ß = -0.017, *SE* = 0.008, t = -2.18, $p < 0.05$). Notably, the stressed syllable *S2* had a significant interaction between lexical frequency and repetition, but the duration of *S2* was significantly shorter in the third repetition regardless of the frequency condition (frequent: ß = -0.057, *SE* = 0.021, t = -2.68, $p < 0.01$; infrequent: ß = -0.08, *SE* = 0.021, t = -3.88, $p < 0.001$), comparable to the results of the first boundary-related interval *B1-1* (Figure 4). The significance or insignificance of effects maintained when speaker tempo was excluded as a random factor.

4. Discussion

The results of the lexical frequency are in line with the proposals by SSRH: First of all, the significant effects of lexical frequency at the prosodic word boundaries confirmed the complementary relationship between lexical frequency (as one of the key redundancy measures) and boundary strength. Targets with a high frequency yielded shorter duration and fewer pauses, indicating weaker boundary strength, whereas both durational measures and pause insertions at the boundaries increased significantly for targets with low frequency, signalling stronger boundary strength. More importantly, the lexical frequency effects were exclusively significant in the boundary-relevant intervals of the target words. The lack of significant effects in the other intervals throughout the target words provides further support for SSRH, in that the effects were not spread evenly throughout the whole target, but rather concentrated at the word boundaries.

By contrast, repetition effects were found for all the tested intervals (although the effect for *B5* was marginal), indicating that repetition tends to reduce the overall duration of a word or an utterance rather than being restricted to the prosodic boundaries (see the lexical retrieval account in [8, 9]). However, it is worth mentioning that the numerical representation in Figure 4 showed greater effects in the boundary-related intervals (*B1-1* and *B5*), giving some support to the view that first mentions, like less frequent words, may have stronger boundaries. It is possible that two types of effects co-exist on these words, a) a repetition-related practice effect (cf. [28]), which shortened all intervals within the word, and b) a language-redundancy-related boundary strength effect, which preferentially shortened the boundary-related intervals. There were significant interactions between lexical frequency and repetition in the first boundary-related interval as well as the stressed syllable. However, upon a closer look, durations for both were consistently reduced by repetition regardless being frequent or infrequent. These results (along with the corresponding magnitudes of effects in Figure 4) should be further assessed. In addition, it would be meaningful to examine the phrasal prominence pattern for the target sentences to see if differences in phrasal prominence pattern could explain some of the (insignificant) variability in duration on the stressed syllable for frequent vs. infrequent targets.

Speech tempo, although greatly varying between speakers, did not seem to be interfering with effects of lexical frequency on the duration of boundary-related intervals, as the significance of lexical frequency effects was maintained even if speaker tempo as a factor was controlled for. In other words, speech tempo modulates duration globally for the whole utterance and would not interfere with the local, inverse correlation between language and acoustic redundancy, which is in accordance with proposals regarding speech tempo in [1, 29].

Another factor to take into account is the lab speech setting. Following the predictions made by the SSRH, speakers manipulate prosodic cues to moderate the redundancy variation across linguistic items throughout an utterance to achieve a smooth signal, i.e., an efficient and robust communication. Given the current experimental procedure (i.e., a reading task), it might be the case that the redundancy effects were weaker than they would have been in a non-reading task, as speakers may not be aiming for real communication during recording.

5. Conclusion

Summing up, lexical frequency affects the boundary-related interval duration, i.e., prosodic boundary strength in German, but not the other intervals across the target word. The results of lexical frequency confirmed the inverse correlation between language redundancy and acoustic redundancy mediated by prosodic structure as proposed by the Smooth Signal Redundancy Hypothesis. Furthermore, repeated items were generally produced with shorter duration and fewer pauses were inserted at the prosodic word boundaries for infrequent items and items upon the first mention. In addition, repetition effects were not restricted to the boundaries, but reduced the overall duration of the whole utterance. These effects appear to be stable across varying speech tempo of different speakers and targets.

6. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with project BO 3388/4-1 and by the Arts and Humanities Research Council (AHRC) with project AH/W010801/1. We would like to thank the PhonLab, part of the core facilities LingLab at the University of Konstanz, for the use of their recording facilities.

7. References

[1] A. Turk, "Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis," *Laboratory Phonology*, pp. 227–262, 2010.

[2] M. Aylett, "Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech," Ph.D. dissertation, University of Edinburgh, 2000.

[3] M. Aylett and A. Turk, "The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech," *Language and Speech*, vol. 47, no. 1, pp. 31–56, 2004.

[4] ——, "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei," *Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3048–3058, 2006.

[5] B. Andreeva, B. Möbius, and J. Whang, "Effects of surprisal and boundary strength on phrase-final lengthening," in *Proceedings of the 10th International Conference on Speech Prosody 2020*, 2020, pp. 146–150.

[6] I. Yuen, O. Ibrahim, B. Andreeva, and B. Möbius, "Non-uniform cue-trading: Differential effects of surprisal on pause usage and pause duration in German," in *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague: Guarant International, 2023, pp. 619–623.

[7] M. Pluymaekers, M. Ernestus, and R. H. Baayen, "Articulatory planning is continuous and sensitive to informational redundancy," *Phonetica*, vol. 62, pp. 146–159, 2005.

[8] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language*, vol. 60, no. 1, pp. 92–111, 2009.

[9] T. Q. Lam and D. G. Watson, "Repetition is easy: Why repeated referents have reduced prominence," *Memory and Cognition*, vol. 38, no. 8, pp. 1137–1146, 2010.

[10] J. M. Kahn and J. E. Arnold, "A processing-centered look at the contribution of of givenness to durational reduction," *Journal of Memory and Language*, vol. 67, pp. 311–325, 2012.

[11] T. Bögel and A. Turk, "Frequency effects and prosodic boundary strength," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds., 2019, pp. 1014–1018.

[12] P. Mousikou and K. Rastle, "Lexical frequency effects on articulation: A comparison of picture naming and reading aloud," *Frontiers in psychology*, vol. 6, p. 1571, 2015.

[13] Z. Malisz, E. Brandt, B. Möbius, Y. M. Oh, and B. Andreeva, "Dimensions of segmental variability: Interaction of prosody and surprisal in six languages," *Frontiers in Communication*, vol. 3, 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomm.2018.00025

[14] O. Ibrahim, I. Yuen, M. van Os, B. Andreeva, and B. Möbius, "The combined effects of contextual predictability and noise on the acoustic realisation of German syllables," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 911–920, 2022.

[15] J. Krivokapić, "Prosodic planning in speech production," *Speech planning and dynamics*, pp. 157–190, 2012.

[16] J. Krivokapić, W. Styler, and B. Parrell, "Pause postures: The relationship between articulation and cognitive processes during pauses," *Journal of Phonetics*, vol. 79, p. 100953, 2020.

[17] Berlin-Brandenburgischen Akademie der Wissenschaften, "Digitales Wörterbuch der deutschen Sprache," https://www.dwds.de/d/wb-dwdswb, 2023, accessed: 12.04.2023.

[18] W. Klein and A. Geyken, "Das Digitale Wörterbuch der Deutschen Sprache (DWDS)," *Lexicographica*, vol. 26, pp. 79–96, 2010.

[19] F. Schiel, "Automatic Phonetic Transcription of Non-Prompted Speech," in *Proc. of the ICPhS*, San Francisco, August 1999, pp. 607–610.

[20] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[21] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program, Version 6.4.01]," 2023, available at http://www.praat.org/.

[22] A. Turk, S. Nakai, and M. Sugahara, "Acoustic segment durations in prosodic research: a practical guide," in *Methods in Empirical Prosody Research*, S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schliesser, Eds. Berlin, New York: De Gruyter, 2006, pp. 1–28.

[23] BAS WebServices, "General Help (FAQs)," accessed: 03.04.2024. [Online]. Available: https://clarin.phonetik.uni-muenchen.de/BASWebServices/help/

[24] T. Cambier-Langeveld, "The domain of final lengthening in the production of Dutch," *Linguistics in the Netherlands*, vol. 14, no. 1, pp. 13–24, 1997.

[25] L. White, "Communicative function and prosodic form in speech timing," *Speech communication*, vol. 63, pp. 38–54, 2014.

[26] R. Baayen, D. Davidson, and D. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, no. 4, pp. 390–412, Nov 2008, special Issue: Emerging Data Analysis.

[27] A. Kuznetsova, P. Brockhoff, and R. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, Dec 2017.

[28] E. Keuleers, K. Diependaele, and M. Brysbaert, "Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono-and disyllabic words and nonwords," *Frontiers in psychology*, vol. 1, p. 174, 2010.

[29] A. Turk and S. Shattuck-Hufnagel, *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. Oxford University Press, 02 2020.