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

Differential effects of word frequency and utterance position on the duration of tense and lax vowels in German

Ivan Yuen, Omnia Ibrahim, Bistra Andreeva, Bernd Möbius

Department of Language Science and Technology, Saarland University
{ivyuen, omnia, andreeva, moebius}@lst.uni-saarland.de

Abstract

Acoustic duration is subject to modification from multiple sources, for example, utterance position [13] and predictability such as occurrence frequency at word and syllable levels [e.g., 2, 3, 4]. A study of German radio corpus data showed that these two sources interact to modify syllable duration. On the one hand, the predictability effect can percolate downstream to the segmental level, and this downstream effect is sensitive to phonological contrasts [9]. On the other, [6] showed that utterance-final lengthening is uniformly applied to tense and lax vowels in German. This then raises some questions as to whether the effects of the two sources of durational variation are uniformly applied or sensitive to phonological contrasts. The current study focused on the duration of tense and lax vowels in the stressed syllable of monosyllabic and disyllabic words in utterance-medial and utterance-final positions. Twenty German speakers participated in a question-answer elicitation task. A preliminary analysis of seven speakers showed effects of utterance position and word frequency, as well as interactions with vowel type, suggesting a non-uniform application of durational adjustments contingent on phonological vowel length. Interestingly, the frequency effect affects the duration of lax vowels, but utterance position affects the duration of tense vowels.

Index Terms: word frequency, predictability, vowel duration, production planning, utterance position

1. Introduction

Predictability is often reported to influence acoustic duration at both word and segmental levels [e.g., 2, 3, 4]. One type of predictability is related to occurrence frequency. A frequently-occurring unit is familiar and easy to access with short acoustic duration. Another source that is known to modify acoustic duration is prosodic structure. One of its types is utterance-final lengthening [e.g., 13]. A unit at the end of a prosodic phrase/clause will be lengthened. In [4], these different sources impacting on acoustic duration were analyzed in a study of English data from the Map Task corpus [1]. Acoustic duration was shown to be mostly accounted for by prosodic structure, not predictability (although various predictability measures in addition to occurrence frequency were examined in that study), leading to the postulated Smooth Signal Redundancy Hypothesis.

This hypothesis was further evaluated in a recent study [2] of German radio broadcast data from DIRNDL [8]. The study examined word-final syllable duration at the edge of an intonational phrase (IP), with two types of prosodic strength (weak vs. strong). Prosodic strength was defined in terms of pause duration. Trigram-based predictability (in terms of surprisal) of the word-final syllable was estimated from a trained language model. The findings showed that the effect of utterance-final lengthening is dependent on the interaction between prosodic strength and predictability. Less predictable syllables are longer at strong IP; more predictable syllables longer at weak IP. As the acoustic duration of a syllable changes, so will the duration of its sub-syllabic constituents. This downstream effect of predictability has been shown to interact with phonological contrasts in [9], e.g., consonant voicing in German, suggesting that predictability-induced acoustic duration might interact with phonemic identity. However, vowels and consonants were not systematically manipulated for voicing in that study. Therefore, there is a need to further test whether predictability will induce acoustic duration modification and whether the duration effect will percolate downstream to the segmental level during the phonetic encoding process of spoken language production.

A recent study examined the acoustic duration of tense vs. lax vowels in German in utterance-medial vs. utterance-final position [6], and observed that the acoustic duration of both tense and lax vowels undergoes utterance-final lengthening to a similar extent. This raises questions as to whether predictability and boundary might differentially influence the duration of a sub-syllabic unit, such as the vowel.

Drawing on these previous findings, the current study focused on occurrence frequency of the word/syllable as a predictability factor on the acoustic duration of tense and lax vowel durations in different utterance positions in German, with the goal of understanding how word/syllable predictability and prosodic positions might affect phonetic encoding. As a baseline, we expect that tense vowels will be longer than lax vowels, that the target vowels in utterance-final position will be longer than those in utterance-medial position, and that the target vowels in frequent words will be shorter than those in infrequent words. In light of [6], we predicted a uniform effect of utterance position on the acoustic duration of both vowel types, and assume a uniform effect of predictability. However, in light of [9], we expected the predictability effect to interact with the phonological vowel length contrasts. In light of [2], we expected the predictability effect to interact with the utterance position effect.

2. Method

2.1. Stimuli

The stimuli were varied in occurrence frequency: High vs. low for 2 types of vowels: Tense vs. lax in monosyllabic (see Table 1) and disyllabic words (see Table 2). Each target word occurred in two utterance positions, viz. medial vs. final. For instance, *Gift* (gloss: poison) occurred in medial position in *Das Gift is grün* (gloss: The poison is green) vs. final position in *Der Mann trinkt das Gift* (gloss: The man drinks the poison).

10.21437/SpeechProsody.2024-90

To ensure that the high vs. low frequency stimuli were as similar as possible in terms of syllable structure, consonantal voicing, and vowel identity, we were constrained to arrive at a set of stimuli as systematic and balanced as we could for the monosyllables and disyllables.

Table 1: *Monosyllabic word list.*

	High Frequency Word/Syl	Low Frequency Word/Syl
Lax	*Kind*	*Gift*
	Mensch	*Mett*
Tense	*Tag*	*Tal*
	Zahl	*Zahn*

In monosyllabic words, word and syllable frequency covary; whereas in disyllabic words, word frequency varies while keeping its constituent syllable frequency constant. By including the two types of words, we wanted to disentangle word from syllable frequency effects. We were interested in the acoustic vowel duration in the stressed syllable of the embedding words. If word frequency triggers durational adjustment, we will *not* expect word frequency to interact with the number of syllables. However, the presence of an interaction will imply that the source of the durational adjustment might be attributable to the syllable frequency.

Table 2: *Disyllabic word list.*

	High Frequency Word	Low Frequency Word
Lax	*Nüsse*	*Mücke*
	Mütze	*Pfütze*
	Katze	*Tatze*
Tense	*Fehler*	*Feder*
	Meter	*Feta*
	Vater	*Faser*

2.2. Participants

20 participants took part (9M, 11F, mean age = 23 years), with remuneration for their participation. A subset of the data from 7 speakers was analyzed.

2.3. Procedures

The production experiment consisted of two parts: a picture-naming task followed by a verbal sentence response generation task (i.e. test phase). Each task began with practice items to ensure that participants knew what to do. The experiment was created on Labvanced for presentation of our stimuli and experimental prompts, as well as for randomization. In the picture-naming task, participants were instructed to name a pictured object, with the goal of ensuring that the intended target word will be used for and associated with the corresponding object. If participants used a word/label different from the intended target word, they were corrected and instructed to repeat the target word. Testing only began when they could successfully use the intended target word for the corresponding picture.

In the test phase, participants were instructed to generate a verbal sentence response to a question as soon as possible, using the intended target word. In each trial, the picture of an object was presented first and the participants were instructed to click on a button to hear an auditory prompt question which elicited their verbal sentence response. Two types of prompt questions were used. For instance, *Welches Gift is grün?* (Which poison is green) elicited target words in utterance-medial position; *Was trinkt der Mann?* (What does the man drink) target words in utterance-final position. Although the latter also induced accentuation, it is not likely to confound our question(s) about whether predictability-induced duration interacts with tense or lax vowels, because all stimuli will be subject to accentuation and utterance-final lengthening. The stimuli were repeated once. Their responses were recorded using Audacity at a sampling rate of 44.1KHz.

2.4. Annotation

Annotation was carried out first by a phonetically trained undergraduate research assistant at two levels, viz. vowel and word, in Praat [7], with reference to waveform and spectrogram. The first author then cross-checked the annotation. At the vowel level, the beginning of target vowels was identified in terms of clear sharp F1 and F2 onset, and the end of target vowels in terms of clear F2 offset. The annotated vowel boundaries took into consideration additional acoustic landmarks from the immediately adjacent consonant in the onset and coda positions, for example, frication noise for onset fricatives, the closure, release/aspiration for plosives, nasal formants for nasals, the closure/release phases for affricates, or the sharp turning point in F2 transition/lateral release for laterals. Other additional acoustic cues were used to identify laterals, such as a change in the amplitude of a waveform arising from the secondary resonance chamber.

Word level annotation is guided by specific acoustic landmarks associated with different consonant types in syllable onset or coda position. The onset of burst release was used to identify the beginning of words with an onset plosive consonant. The onset of frication was used to identify those with an onset fricative consonant. The onset of voicing and nasal formants were used to identify those with an onset nasal consonant. The offset of a disyllabic word was identified as the end of F2, because it always ends with an unstressed vowel. Since a monosyllabic word ends with a coda consonant, its word boundary was identified according to consonant types. For coda plosive consonants, the end of release/aspiration was used. For coda fricatives, the end of frication was used. For coda nasals, the end of nasal formants was used. For laterals, the end of F2 was used. Figure 1 exemplifies the annotated boundaries of the stressed vowel in the word *Kind* (gloss: child).

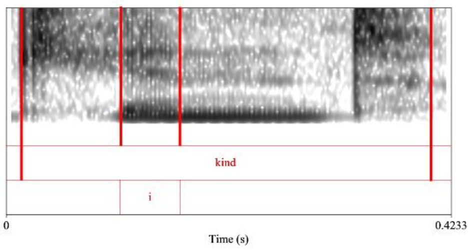

Figure 1: *A sample annotation of a target vowel in a monosyllabic word 'Kind' in utterance-final position*

3. Results

A total of 533 items were analyzed, excluding incorrect items (n = 21), false start (n = 4) and items ending with rising intonation (n = 2). Linear mixed effects modelling was carried out in R [12] using the lme4 [5] and lmerTest [10] packages. Two omnibus analyses were run, using raw vowel duration (ms) and vowel ratio as the respective dependent variable. Vowel ratio (as defined by vowel duration/word duration) was included to minimize the possibility of confounding individual word duration. Predictors included Utterance position (Medial vs. Final), Vowel type (Tense vs. Lax), Word frequency (High vs. Low), Number of syllables (Mono vs. Di), and their interactions. Various models were constructed and compared using Chi-square test to determine the maximal random structure. The results of the vowel ratio from the omnibus analysis revealed three significant main effects and 2-way interactions, with one significant 3-way interaction among Vowel type, Number of syllables and Word frequency (F = 4.5, df = 1, p = .045 *). The model structure was ~ Utterance position * Vowel type * Number of syllables * Word frequency + (Utterance position + Number of syllables | speakers) + (Utterance position | items). Note that the 3-way interaction using raw vowel duration (ms) as the dependent variable did not reach statistical significance.

To better understand the significant 3-way interaction observed in vowel ratio, we separately analyzed monosyllabic and disyllabic words. In each analysis, we examined both raw vowel duration (ms) and vowel ratio patterns.

3.1. Monosyllabic words

Word frequency significantly interacted with Vowel type to affect vowel ratio (Table 3), but such an interaction was absent from the raw vowel duration measurement (Table 4). As seen in Figure 2, the Word frequency-by-Vowel type interaction in monosyllabic words arises because only the lax vowels exhibit the effect of Word frequency, not the tense vowels. The same pattern was observed in utterance-final and utterance-medial positions.

Table 3: Vowel ratio in *monosyllabic words. The model is ~ Utt. position * Vowel type * Word frequency + (Vowel type | speaker) + (1 | item).*

Factors	F	df	p
Utt. position (UP)	9.3	1	*.003***
Vowel type (V)	122.4	1	*<.0001****
Word frequency (WF)	6.2	1	*.05**
UP * V	3.6	1	.06
UP * WF	2.2	1	.14
V * WF	8.3	1	*.03**
UP * V * WF	.2	1	.66

Table 4: Vowel duration (ms) in *monosyllabic words. The model is ~ Utt. position * Vowel type * Word frequency + (Utt. position + Vowel type | speaker) + (Utt. position | item).*

Factors	F	df	p
Utt. position (UP)	22.3	1	*.007***
Vowel type (V)	90.4	1	*.0001****
Word frequency (WF)	2.6	1	.24
UP * V	16.3	1	*.04**
UP * WF	.01	1	.93
V * WF	2.4	1	.25
UP * V * WF	2.1	1	.25

However, as seen in Figure 3, raw vowel duration is subject to the interaction of Utterance position and Vowel type, with the effect of Utterance position more pronounced for the tense vowels. There was no significant effect of Word frequency on raw vowel duration.

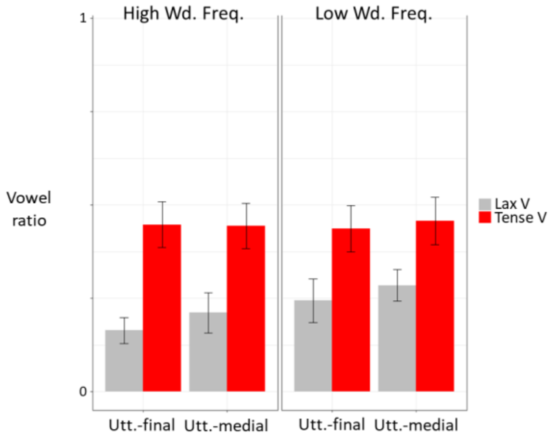

Figure 2: *Vowel ratio of lax vs. tense vowels in monosyllabic words with high vs. low word frequency of occurrence in utterance-final vs. utterance medial positions with +/- 1 SD*

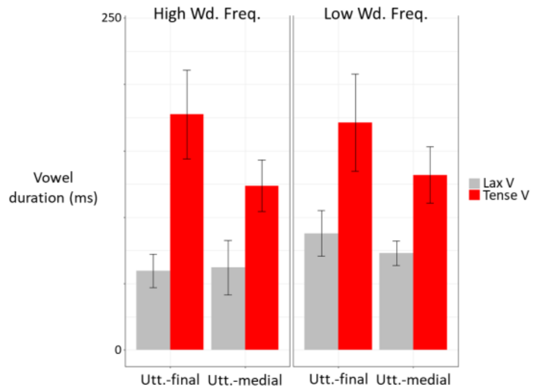

Figure 3: Raw *duration (ms) of lax vs. tense vowels in monosyllabic words with high vs. low word frequency of occurrence in utterance-final vs. utterance medial positions with +/- 1 SD*

3.2. Disyllabic words

Neither the main effect of Word frequency nor any of its interaction with other factors reached statistical significance in terms of vowel ratio or raw vowel duration (ms) (see Tables 5 and 6 respectively).

Table 5: Vowel ratio in *disyllabic words. The model is ~ Utt. position * Vowel type * Word frequency + (Utt. position + Vowel type | speaker) + (Utt. position + Word frequency | item).*

Factors	F	df	p
Utt. position (UP)	7.7	1	*.02**
Vowel type (V)	21	1	*.002***
Word frequency (WF)	.13	1	.73
UP * V	21.3	1	*.0006****
UP * WF	.007	1	.93
V * WF	.3	1	.59
UP * V * WF	1.5	1	.25

Table 6: Vowel duration (ms) in *disyllabic words. The model is ~ Utt. position * Vowel type * Word frequency + (Vowel type | speaker) + (Vowel type | item).*

Factors	F	df	p
Utt. position (UP)	66.8	1	<.0001***
Vowel type (V)	20.2	1	*.001***
Word frequency (WF)	.01	1	.93
UP * V	80.4	1	*<.0001***
UP * WF	.9	1	.33
V * WF	.6	1	.45
UP * V * WF	.04	1	.84

Although the effect of Word frequency did not significantly influence the measures of raw vowel duration and vowel ratio in disyllabic words, both measures are subject to the interaction of Utterance position and Vowel type. While utterance positions affect the vowel ratio measure for lax vowels (Figure 4), utterance positions affect the raw duration of tense vowels (Figure 5).

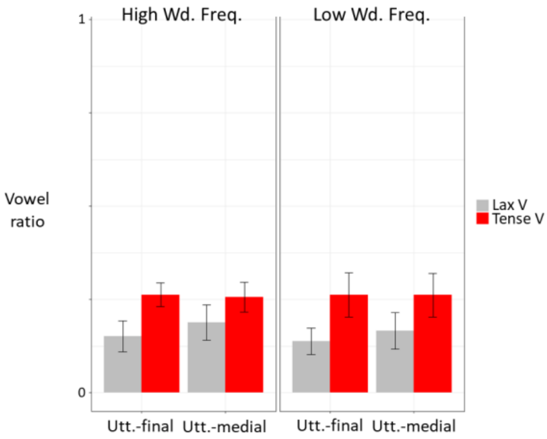

Figure 4: Vowel ratio *of lax vs. tense vowels in disyllabic words with high vs. low word frequency of occurrence in utterance-final vs. utterance medial positions with +/- 1 SD*

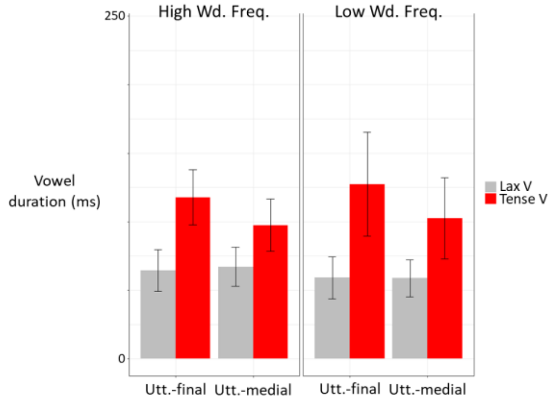

Figure 5: Raw *duration (ms) of lax vs. tense vowels in disyllabic words with high vs. low word frequency of occurrence in utterance-final vs. utterance medial positions with +/- 1 SD*

The separate by-Number-of-syllables analyses then indicate that the significant omnibus 3-way interaction containing Word frequency (using vowel ratio as the dependent measure) is primarily driven by the monosyllabic words.

4. Discussion

As expected, tense vowels are longer than lax vowels in both absolute and relative terms. Both vowel types exhibit stronger lengthening in utterance-final than utterance-medial position (in both absolute and relative terms). Admittedly, this effect might have been exaggerated by accentuation in the former. In partial agreement with our prediction, the effect of Word frequency only reached statistical significance for monosyllabic words, when vowel ratio was analyzed. No effect of Word frequency was observed for disyllabic words. This difference suggests that the source of this effect might be due to the covarying syllable frequency in the monosyllabic words vs. the controlled syllable frequency in the disyllabic words.

Counter to the expectations motivated by [6], a significant interaction was found between Utterance position and Vowel type for disyllabic words (in both absolute and relative terms). That is, *utterance position affects the raw duration of the tense vowels, not the lax vowels. Yet utterance position affects the ratio of the lax rather than the tense vowels in a word.* If the lax vowels are subject to some durational constraint such that the effect of utterance-final position on the word does not percolate down to the lax vowels, this could explain why the ratio measure affects the lax vowels. As for monosyllabic words, the interaction was present only when raw vowel duration was analyzed. Note though that there is a trend for the interaction towards statistical significance when vowel ratio served as the dependent variable. These observations do not support the interpretation that the effect of Utterance position is applied uniformly to tense and lax vowels in the current study. The different observations between the 2 studies could be due to different measures used with different research questions. [6] examined proportional lengthening in different utterance positions, whereas the current study used vowel/word ratio. Instead, our data favour the interpretation that the effect of utterance positions seems to be constrained by the phonological vowel length contrasts, in line with [e.g., 9, 11].

Moreover, we observed a statistically significant interaction between Word frequency and Vowel type on the measured vowel ratio in monosyllabic words, consistent with findings in [9] on German voicing contrasts. The interaction probably arises because *word frequency affects the duration of lax vowels more than tense vowels*. One limitation is that the analysis is preliminary, with more data required to verify the reported observations.

5. Conclusions

Our findings showed that the word frequency effect might arise from covarying syllable frequency. Both predictability and utterance position affect absolute vowel duration and relative vowel ratio (i.e., vowel/word duration), and these factors interact with the phonological vowel length contrasts (tense vs. lax) in German. Intriguingly, word frequency affects the duration of lax vowels more than tense vowels; but utterance positions affect the duration of tense vowels more than lax vowels. These data seem to suggest that word/syllable level and utterance level processes might be associated with different types of vowel length contrasts during phonetic encoding in spoken language production planning.

6. Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 232722074-SFB1102.

7. References

[1] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The Hcrc Map Task Corpus. *Language and Speech*, *34*(4), 351-366.

[2] B. Andreeva, B. Möbius, B. and J. Whang. 2020. “Effects of surprisal and boundary strength on phrase-final lengthening,” in *Proceedings 10th International Conference on Speech Prosody*, Tokyo, Japan, 2020, pp. 146-150.

[3] I. Arnon, and U. Cohen-Priva, U. “More than words: The effect of multi-word frequency and constituency on phonetic duration,” *Language and Speech*, 56, 349-371. 2013.

[4] M. Aylett and A. Turk. “The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech,” *Language and Speech*, 47, 31-56. 2004.

[5] D. Bates, M. Mächler, B. Bolker and S. Walker. “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, 67, pp. 1-48. 2015.

[6] M. Belz, J. Krivokapić and C. Mooshammer. “Interaction between phrasal structure and vowel tenseness in German: An acoustic and articulatory study,” *Language and Speech*, 66, 3-34. 2023.

[7] P. Boersma and D. Weenink. Praat: doing phonetics by computer. Version 6.1.53. 2021.

[8] K. Eckart, A. Riester and K. Schweitzer, K. “A Discourse Information Radio News Database for Linguistic Analysis,” in C. Chiarcos, S. Nordhoff & S. Hellmann (Eds.) *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata* (pp.65-75). 2012.

[9] O. Ibrahim, I. Yuen, B. Andreeva and B. Möbius. “The effect of predictability on German stop voicing is phonological selective,” in *Proceedings Speech Prosody*, Lisbon, Portugal, 2022. pp. 669-673.

[10] A. Kuznetsova and P. B. Brockhoff and R. H. B. Chrsitensen. “ lmerTest Package: Tests in Linear Mixed Effects Models,” *Journal of Statistical Software*, 82, 12, pp. 1-26. 2017.

[11] S. Nakai, A. Turk, K. Suomi, S. Granlund, R. Ylitalo and S. Kunnari. “Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish,” *Journal of Phonetics*, 40, 796-807. 2012.

[12] R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria). 2022.

[13] C. Wightman, S. Shattuck-Hufnagel, M. Ostendorf and P. Price. “Segmental durations in the vicinity of prosodic phrase boundaries,” *Journal of the Acoustical Society of America*, 91 (3), 1707-1717. 1992.