Laboratory Phonology
Journal of the Association for Laboratory Phonology

Mooshammer, C., Krivokapić, J., & Belz, M. (2025). How final is final? The production and perception of utterance-medial and utterance-final boundaries. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 16(1), pp. 1–43. DOI: https://doi.org/10.16995/labphon.11679

Open Library of Humanities

# How final is final? The production and perception of utterance-medial and utterance-final boundaries

**Christine Mooshammer*,** Department of German Studies and Linguistics, Humboldt-Universität zu Berlin, Germany, christine.mooshammer@hu-berlin.de

**Jelena Krivokapić,** University of Michigan, USA, jelenak@umich.edu

**Malte Belz,** Department of German Studies and Linguistics, Humboldt-Universität zu Berlin, Germany, malte.belz@hu-berlin.de

*Corresponding author.

We examine the production and perception of two types of phrase-final prosodic boundaries, specifically, utterance-medial and utterance-final intonation phrase (IP) boundaries in German. These two types of boundaries are expected to differ in terms of general properties of the prosodic hierarchy, and properties of turn-taking and speech planning. In an articulatory magnetometer study and a perceptual rating study, we examine these boundaries in read speech, testing for temporal, spatial, and intonational properties of the boundaries. Only small and inconsistent differences were found in the temporal and spatial domains. The only robust difference is a lower f0 in the rhyme of the intonation contour for utterance-final IP boundaries compared to utterance-medial IP boundaries for five of the eight speakers. This is consistent with the results of the perception study, which indicate that listeners perceive subtle differences in the boundary-specific production, and that mean f0 during the rhyme and peak velocity were the information listeners used to determine utterance finality for the speakers producing the difference.

## 1. Introduction

The goal of this study is to examine the production and perception of intonation phrase boundaries in German that occur utterance-medially and utterance-finally, focusing on phrase-final properties. In contrast to phrase-medial boundaries (i.e., word boundaries), we use the term *utterance-medial* to refer to phrase-final boundaries when the speaker continues speaking after the boundary (Examples 1 and 2 below). The term *utterance-final* refers to phrase-final boundaries when the speaker ceases to speak after the boundary (Example 3).

(1) **phrase-medial:** Ich fuhr mit der *Bahn* am Donnerstag. Am Mittwoch wurde noch gestreikt.
'I took the *train* on Thursday. On Wednesday, there was still a strike.'

(2) **utterance-medial:** Ich fuhr mit der *Bahn*. Am Donnerstag musste ich in Frankfurt sein.
'I took the *train*. On Thursday, I had to be in Frankfurt.'

(3) **utterance-final:** Ich fuhr mit der *Bahn*.
'I took the *train*.'

Prosodic phrasing groups words into larger units which are used by both speakers and listeners for processing (Cole, 2015; Krivokapić, 2014; Shattuck-Hufnagel & Turk, 1996; Wagner & Watson, 2010). The number of prosodic phrases assumed varies by language, and is, even for the most examined language, English, not firmly agreed on, though most studies assume a minor and major category above the word level (see Beckman & Pierrehumbert, 1986; Shattuck-Hufnagel & Turk, 1996). For German, generally two prosodic categories above the level of the word are assumed as well, namely, the intermediate phrase (*ip*) and the intonation phrase (*IP*) (Grice et al., 2005).

Prosodic phrases are marked by boundaries and their phonetic correlates are fairly well established. At boundaries, acoustic segments and articulatory gestures are longer, and stronger prosodic boundaries often have pauses (see the overviews in Fletcher, 2010; Katsika, 2016; Paschen et al., 2022). Articulatory gestures become spatially larger phrase-initially, while phrase-final strengthening is not systematically observed (Byrd et al., 2006). Finally, gestures also become less overlapped at boundaries. The acoustic and articulatory effects are stronger at higher prosodic boundaries (for a potential exception in German, please see Kentner et al., 2023) and decrease with distance from the boundary (see the overviews in Byrd & Saltzman, 2003; Guitard-Ivent et al., 2021; Katsika, 2016; Paschen et al., 2022). In addition to temporal and spatial properties, boundaries are also marked by modulations of fundamental frequency (f0, e.g., for English, Silverman et al., 1992; for German, Peters, 2006), which at IP boundaries are referred to as *boundary tones*, and at ip boundaries as *phrase accents* (for German, see Grice et al., 2005).

The most salient information in the perception of prosodic boundaries is the presence of a pause, although final lengthening, f0 (e.g., pitch reset, boundary tones, height of pitch accent peaks), and creaky voice are also important correlates of boundary perception (see the overviews in Cole, 2015; Davidson, 2021; Kim, 2020; Petrone et al., 2017; Roy et al., 2017). The presence of a larger number of cues leads to stronger boundary perception (e.g., Brugos et al., 2018; Collier et al., 1993; De Pijper & Sanderman, 1994). Only two studies have examined kinematic correlates of boundary perception, and both of them examine American English (Krivokapić, 2007; Krivokapić & Byrd, 2012).

Krivokapić (2007) examines the perception of boundaries of different strengths (up to and including IP boundaries but not utterance-final boundaries) in CVCVC#VC strings and identifies for American English that listeners are most sensitive to the preboundary opening movement of the consonant immediately adjacent to the boundary, which contains most of the boundary related lengthening. Longer preboundary movement led to the perception of a stronger boundary. The next most informative movement was the opening movement of the postboundary consonant (which in the production showed shortening), and in this case, shorter movement led to the perception of a stronger boundary. Thus, the two movements with the most salient prosodic information were the movements that listeners were the most sensitive to (Krivokapić, 2007).

Krivokapić and Byrd (2012) examine the production and perception of IP boundaries (CV#CV strings) and find that kinematic measures spanning a boundary (the duration between the preboundary and postboundary vowel peak velocities and the duration between the preboundary and postboundary consonant peak velocities) are the best predictors of boundary strength perception, followed by the preboundary opening movement duration. Similarly to Petrone et al. (2017), Krivokapić and Byrd also find that listeners in the perception study do not systematically use information that was inconsistently used in the production study. Note that neither of the two kinematic studies take measures of f0 into account.

Not many studies exist for German prosodic boundary production, but the acoustic studies that exist also find final lengthening, pausing, and f0 changes as phonetic correlates of prosodic boundaries (e.g., Kentner & Féry, 2013; Kohler, 1983; Peters, 2003, 2006; Petrone et al., 2017). There are only two articulatory studies on German boundaries. Mücke and Hermes (2007), in a pilot study of two speakers of Viennese German, find significant lengthening of consonantal closing gestures closest to the boundary and limited evidence of displacement. In Belz et al. (2023), based on a subset of data that will be presented here, we examined lengthening at IP boundaries in comparison to word boundaries. We hypothesized that lax vowels are stretched to a far lesser degree by final lengthening than tense vowels, since in our previous studies, speech rate and word stress only affected the duration of tense vowels (Hoole & Mooshammer, 2002; Mooshammer & Geng, 2008). However, both vowel types lengthen to a similar degree in phrase-final utterance-medial position while the quantity contrast is still maintained. Furthermore, the

results show that for speakers of Northern German the effect of the boundary is such that at IP boundaries gestures lengthen and that the effect is strongest at the boundary and decreases with distance from it, while there is again only limited evidence of displacement (Belz et al., 2023).

The perception of prosodic boundaries in German is similarly understudied. In a production-perception study, Gollrad (2013) examines the effect of duration, f0, and pauses, and finds that pauses and boundary tones are used for IP boundary perception, while final lengthening may not be. For ip boundaries, Gollrad finds that listeners use final lengthening as the main indicator of a boundary. The relevance of the cues used in perception is closely linked to the presence of these cues in the production data, in the sense that the properties identified in the production task were the ones predominantly used in the perception task. Petrone et al. (2017) find pauses, f0, and final lengthening to be used in the perception of IP boundaries, with pauses being the most salient information in boundary perception, overriding final lengthening and f0 information. They further find that final lengthening is more relevant for the perception of the boundary than f0, and suggest that this might be due to f0 information being more variable in the production of boundaries. Peters (2006) identifies pauses as the most salient cue for utterance-medial boundaries, with final lengthening and f0 information used in boundary perception when there are no pauses. Peters also points out that for boundaries where lengthening is particularly prominent, listeners use it in perception even if pauses are present (see also Roy et al., 2017, who point out for English that while final lengthening is a strong cue for boundaries, this is only the case when the lengthening is quite strong).

Most studies examine boundaries within a sentence or boundaries within a larger corpus, but utterance-final IP boundaries, the topic of our study, have not been examined much, neither in production nor in perception. While Oller (1973), for English, finds final lengthening in utterances not followed by another sentence, he does not compare utterance boundaries to phrase boundaries that are not utterance-final. Berkovits (1993a, 1993b, 1994), in extensive studies of Hebrew, finds utterance-final lengthening but also does not examine utterance-medial phrase-final positions. Similarly, Kohler (1983) shows utterance-final lengthening in German, based on an acoustic study of one speaker, but does not compare this to utterance-medial phrase boundaries. An exception is Cambier-Langeveld (1997), who examines prosodic word, phonological phrase, IP, and utterance boundaries (where utterance boundaries are not followed by another utterance) in Dutch, and finds a general trend that prosodically higher boundaries lead to more lengthening. While there is no statistical comparison for the effect of different boundaries, for some of the target words examined there seems to be more lengthening utterance-finally compared to IP-finally.

For intonation, Berkovits (1984), examining phrase-medial and utterance-final boundaries, finds f0 lowering at utterance boundaries, for both English and Hebrew, but there is no comparison to utterance-medial phrase boundaries. Geluykens and Swerts (1994) examine the production and

perception of sentences in utterance-final versus utterance-medial position in Dutch and find that in utterance-final sentences, speakers produce different intonation contours than in utterance-medial sentences, and at utterance-final boundaries, they produce lower f0 than at utterance-medial boundaries. Listeners can distinguish the location of these sentences although it is not clear which prosodic properties of the utterances are decisive. Similarly, for German, Herman (2000) finds lower final f0 peaks in discourse-final position compared to identical sentences in discourse-medial position. This final lowering shows a steeper f0 decline than predicted by the declination in the final portion of an utterance (see, e.g., for English Arvaniti, 2007; Hirschberg & Pierrehumbert, 1986; Liberman & Pierrehumbert, 1984) and reflects some kind of discourse finality (Herman, 2000). However, Ladd (1988) finds that, in English, for three out of four speakers, utterance-final boundaries do not differ from utterance-medial boundaries in how low final f0 is, although one speaker does show utterance-final lowering. Thus, there is very little work that considers the properties of utterance-final compared to utterance-medial boundaries, especially for lengthening. The findings for f0, while not uniform, indicate f0 lowering turn-finally.

The perception of finality in utterance-final position for German is examined in Peters (2006). Listeners were asked to rate an utterance on a 7-point scale on whether a speaker intends to continue a syntactically and semantically complete sentence. F0, final rhyme duration, and voice quality of the stimulus sentences were manipulated in several steps. The results from 14 participants show that a falling f0 contour elicits significantly lower continuation ratings, but lengthening and phonation only affect the ratings significantly in combination with a falling f0 contour.

The current study specifically focuses on the production and perception of boundaries in utterance-final as compared to utterance-medial but phrase-final position. While, as outlined above, not much research has been done on this question, there is good reason to think that speakers produce and listeners perceive the differences between utterance-final and phrase-final boundaries. To begin with, there is evidence that listeners perceive and speakers produce boundaries in a gradient manner (e.g., for English, Korean, and Dutch Cho & Keating, 2001; Krivokapić & Byrd, 2012; Ladd, 1988; Swerts, 1997; Wagner, 2005; Wagner & Crivellaro, 2010). For example, Cho and Keating find that acoustic final lengthening across four boundaries in Korean (word, accentual phrase, intonation phrase, and utterance-medial boundaries) has a bi-modal distribution while initial articulatory contact across these boundaries shows a continuous increase, indicating a gradient distinction. Ladd finds for English that hierarchically higher IP boundaries show more declination reset and lengthening than hierarchically lower IPs, indicating IP boundaries can differ in strength. Similarly, Krivokapić and Byrd find for English that speakers can produce and listeners perceive IP boundaries of different strengths. Thus, it is reasonable to assume that speakers will produce and listeners perceive differences between utterance-medial and utterance-final IP boundaries even though they belong to the same prosodic category.

It is also possible that the utterance-final boundary differs systematically from other boundaries as it is the boundary where speakers completely stop speaking, and therefore there might be differences due to its communicative function and performance factors. Specifically, there is evidence that, in English an increase in lengthening is a phonetic correlate of turn-final boundaries compared to phrase-medial (word) boundaries (Local & Walker, 2012). Duncan (1972) also suggests, in an impressionistic evaluation of turn-taking behaviour in English, that there is final lengthening at turn ends. On the other hand, both Gravano and Hirschberg (2011) and Purse and Krivokapić (2023) find for English that segments/gestures at turn-final boundaries are shorter than at turn-medial boundaries. However, regardless of the direction of the effect, the utterance-final boundary is expected to differ from the utterance-medial IP. Regarding f0, speakers have been found to produce low f0 and a falling contour for end of turn utterances in declarative sentences (Geluykens & Swerts, 1994; Swerts & Geluykens, 1994), although the very end of the utterance does not necessarily have to differ from utterance-medial positions (Ladd, 1988). Research has also found that listeners are sensitive to phonetic correlates of turn-final prosodic boundaries (e.g., for Dutch Bögels & Torreira, 2015; Geluykens & Swerts, 1994). While we examine read speech in the present study, and conversation differs in many ways from read speech, we expect that these properties of turn ends might have their origins in the prosodic hierarchy. Thus, utterance-medial and utterance-final IPs might differ in production, with differences in lengthening and a more pronounced f0 at the end of the turn, and listeners might be sensitive to these differences.

Utterance-final boundaries might also show less lengthening than utterance-medial boundaries for another reason. At boundaries, speakers plan upcoming speech, and stronger boundaries (as indicated by longer pauses) are related to the speaker's need for more planning time for an upcoming utterance: longer pauses provide more time for speakers to plan the upcoming utterance (see the overview in Krivokapić, 2012). While the focus in these studies has been on pauses, and final lengthening has hardly been examined as it relates to planning, there is reason to assume that the effect of planning is also seen in final lengthening, since the boundary spans over more than just the pause (but see Ferreira, 1991; Krivokapić et al., 2022, for evidence, for English, against the hypothesis that planning takes place during final lengthening). Given that at utterance-final boundaries there is no further planning, speakers might produce less final lengthening at boundaries.

We conducted an electromagnetic articulography study (Experiment 1) to evaluate the production of boundaries and a perception study (Experiment 2), in which participants rated the boundaries from Experiment 1. We examined the following questions:

Question 1: Do speakers of German distinguish between utterance-medial and utterance-final IP boundaries? Based on the above discussion, our prediction is that they will make a

distinction. It is, however, unclear what the direction of the changes will be. On the one hand, more extensive lengthening in utterance-final position could be expected due to the utterance being a hierarchically higher boundary and based on the findings from some turn-taking studies. On the other hand, less lengthening is expected if speakers also use final lengthening for planning purposes, since in utterance-final positions there is no further planning needed, and this also seems to be what some studies examining turn-taking find.

For f0, we expect utterance-medial boundaries to differ from utterance-final boundaries because of the previously observed final lowering.

Question 2: Do German listeners distinguish between utterance-medial and utterance-final IP boundaries for judging upcoming continuation? Our prediction is that, given the communicative relevance of utterance-final boundaries and the fact that listeners can distinguish boundaries of different strengths, including boundaries of the same category but of different strengths, listeners will distinguish between these two boundaries.

Question 3: Which information do listeners use in perception? To address this question, we examine the link between production and perception using a subset of the data. Since the most salient cue in boundary perception (pauses) cannot be used in this study, we expect that how strongly each of the other parameters affects perception will mostly depend on how salient these parameters are in the production of the speakers.

# 2. Production experiment

## 2.1. Method

### 2.1.1. Participants and recording procedure

Eight native speakers of German (4 male, 4 female), aged 23–28 years without noticeable speech impairments, participated in this study. They were informed about the methods and recording procedure, but they were not aware of the goal of the study. Data about their age, gender, language background, and possible health issues were collected, and they received a payment of 10 Euros per half hour. The participants were seated in a soundproof booth and instructed to read stimuli sentences from a computer screen. Each stimulus was presented on a screen in randomized order in blocks of five iterations. The onset of each stimulus was cued by a visual and an auditory signal. Participants could rest between blocks if they wished. The experimenter was visible to the speakers throughout the experiment through a window.

Acoustic data were recorded at 44.1 kHz using a shotgun microphone located in front of the speakers. Articulatory data were recorded by means of electromagnetic articulography (EMA), using the articulograph AG 501 (Carstens Electronics; for details on accuracy, see Savariaux et al., 2017). Movements of the tongue, jaw, and lips were recorded over time in

the three-dimensional space. Sensors were attached to the tongue tip (TT), the tongue mid (TM), the tongue back (TB), the lower incisors (JAW), and the upper and lower lips (UL and LL) with a medical adhesive. In addition, the tongue sensors were fixated with dental cement. Four reference sensors were placed just above the upper incisors, on the nasion and on the left and right mastoid part of the temporal bone. Additionally, to rotate the coordinate systems in a similar direction for all speakers, three sensors were attached to a protractor to record the bite plane just before the end of the experiment. The articulatory data were recorded at a sampling rate of 1250 Hz and then downsampled to 250 Hz for postprocessing in MATLAB v. R2013b. After low-pass filtering of the reference sensors at 20 Hz, the data were corrected for head movement and then rotated and translated to the recorded bite plane. Due to technical problems the reference sensor data of two participants could not be translated to the bite plane, so the data were rotated and translated to a fictional plane between the upper incisors and the nasion instead. The signals from all moving sensors were smoothed with a 50 Hz low-pass filter.

### 2.1.2. Material

The stimuli consisted of three monosyllabic minimal pairs (**Table 1**) of target words differing in vowel tenseness. The target words were embedded in carrier sentences for three boundary conditions: phrase-medial (1), utterance-medial (2), and utterance-final (3).

(1) **phrase-medial:** Ich fuhr mit der *Bahn* am Donnerstag. Am Mittwoch wurde noch gestreikt.
'I took the *train* on Thursday. On Wednesday, there was still a strike.'

(2) **utterance-medial:** Ich fuhr mit der *Bahn*. Am Donnerstag musste ich in Frankfurt sein.
'I took the *train*. On Thursday, I had to be in Frankfurt.'

(3) **utterance-final:** Ich fuhr mit der *Bahn*.
'I took the *train*.'

For phrase-medial and utterance-medial conditions the carrier sentences were identical in the material preceding the target word and for at least two words after the target word. The full list of stimuli is given in **Table 7**. In total, there were 18 stimulus and 9 filler sentences (to distract the participants from the prosodic patterns), which were repeated five times. Data for this study were collected together with data for another study. Note that while the English translation has a boundary at the comma, the German stimuli had only one utterance-medial prosodic boundary (the one at the full stop). A subset of the data, excluding the utterance-final position, were used for comparing final lengthening in tense and lax vowels. The results are published in Belz et al. (2023).

| | words | | |
|---|---|---|---|
| tense | Bahn [baːn] 'train' | Ruhm [ʁuːm] 'fame' | Stiel [ʃtiːl] 'stem' |
| lax | Bann [ban] 'ban' | Rum [ʁʊm] 'rum' | still [ʃtɪl] 'quiet' |

**Table 1:** Target words for experiments 1 and 2.

## 2.2. Measurements

### 2.2.1. Acoustic measurements

The recordings were transliterated and prealigned using WebMAUS (Kisler et al., 2012). The alignment was checked and corrected with Praat (Boersma, 2001). Every segment of each stimulus word was annotated on an interval tier. Another interval tier was added to annotate whether the participants realized the intended prosody per condition. All Praat TextGrids were converted into an EMU speech database (Winkelmann et al., 2017) and analysed with the package *emuR* (Winkelmann et al., 2016) in R (R Core Team, 2020). Two utterances with speech errors were excluded from all measurements. For the current analysis the word duration (in ms) WORDDUR was extracted based on hand-corrected annotations from WebMAUS.

### 2.2.2. F0 contours

Fundamental frequency tracks (f0) were added to the EMU speech database using the f0 tracker *praatToPitch2AsspDataObj* of the package *wrassp* (Bombien et al., 2020), with gender-specific f0 ranges (for females 40–400 Hz, for males 40–300 Hz). The relatively low threshold was chosen because close to phrase boundaries or at the end of an utterance glottalization is quite common. F0 contours were extracted for the rhymes of the test words. To minimize effects of the preceding and following partly voiceless contexts, only the middle parts from 20 to 90 percent of the time-normalized sequences were considered here. Twenty-eight of 686 trials were excluded by visual inspection due to obvious mistracking, and 20 trials had missing values due to voicelessness or glottalisation. Altogether, 7% of all contours were excluded. Following Sóskuthy (2021), we used log f0 of the remaining 638 contours. We residualized the log f0 per speaker, using the package *umx* (T. C. Bates et al., 2019). The parameter F0R was calculated as the averages of all residualized log f0 samples during the rhymes of the test words.

### 2.2.3. Articulatory data

Articulatory data were annotated using *mview* (Tiede, 2005), a MATLAB-based tool which allows for semiautomatic labelling of kinematic parameters of an articulatory gesture. In this paper we are mainly interested in the kinematic characteristics of the closing gesture for the final consonant, i.e., the movement of the consonantal articulator from the vowel towards the coda. The opening gesture following the coda consonant could not be analysed, even though it is

closer to the prosodic boundary and could therefore be more relevant to the research question. However, for the utterance-final condition and occasionally also for the phrase-final condition speakers frequently did not release the closure but kept the mouth closed. The tongue tip signal was analysed for the alveolar consonants /n/ and /l/. For the bilabial consonant /m/, the lip aperture (LA) was used. The LA signal was calculated as the Euclidean distance between the upper and lower lip signals. To quantify the effect of final lengthening on the shape of the velocity profile, the closing gestures of the last consonant in the target words were labelled and the tangential velocity extracted. The onset and offset of the closing gesture were labelled using a threshold criterion of 20% of the maximal tangential velocity. Eleven out of 233 lip gestures and two of 478 tongue tip gestures could not be analysed due to very small movements or problems detecting the onset of the movement.

The following parameters were calculated based on the MVIEW labelling procedure *findgest* (see **Figure 1** for a reference to the timestamps). Additionally, the acoustic duration of the target WordDur in ms was included in the subsequent analysis.

- closing duration ClosDur = gestural offset (3) – gestural onset (1) in ms
- displacement of the closing movement ClosDisp = three-dimensional Euclidean distance of tongue tip or LA positions from gestural onset (1) to the gestural offset (3) of the closing movement in mm
- velocity peak VelPeak = maximal velocity (2) during the closing movement in cm/s
- time to peak velocity T2VelPeak = time to maximal velocity (2) from gestural onset (1) in percent of ClosDur

The kinematic parameters above change significantly when comparing phrase-medial to utterance-medial prosodic boundaries (see Beckman et al., 1992; Belz et al., 2023, for a more detailed discussion on kinematic parameters). In general, apart from lengthening, stronger prosodic boundaries may also lead to larger displacements of the target segment (Tabain, 2003). We hypothesize that these parameters may also change when distinguishing utterance-medial from utterance-final boundaries, leading to longer gestures and longer time-to-peak velocity, and therefore include them in this study.

The parameters described so far are all so-called *magic moment measures*, which capture certain aspects of the shape of the velocity profiles. For finer details and a more holistic analysis we applied Generalized Additive Mixed Models (GAMMs) based on the velocity profiles (see 2.2.4). Therefore, the LA and tongue tip movements during the closing movement were extracted between onset and offset of the closing movement, the interval between (1) and (3) in **Figure 1**. For the alveolar consonants, the tangential velocity was calculated from the three-dimensional tongue tip signals; for the labial consonant, the LA signal was used. The velocity signals were

low-pass filtered at 20 Hz. As recommended in Wieling (2018), the velocity profiles were scaled for each speaker, i.e., the mean was subtracted and divided by the range per speaker. To investigate shape differences independently of temporal effects due to final lengthening, all 638 velocity profiles were time-normalized to 50 samples.

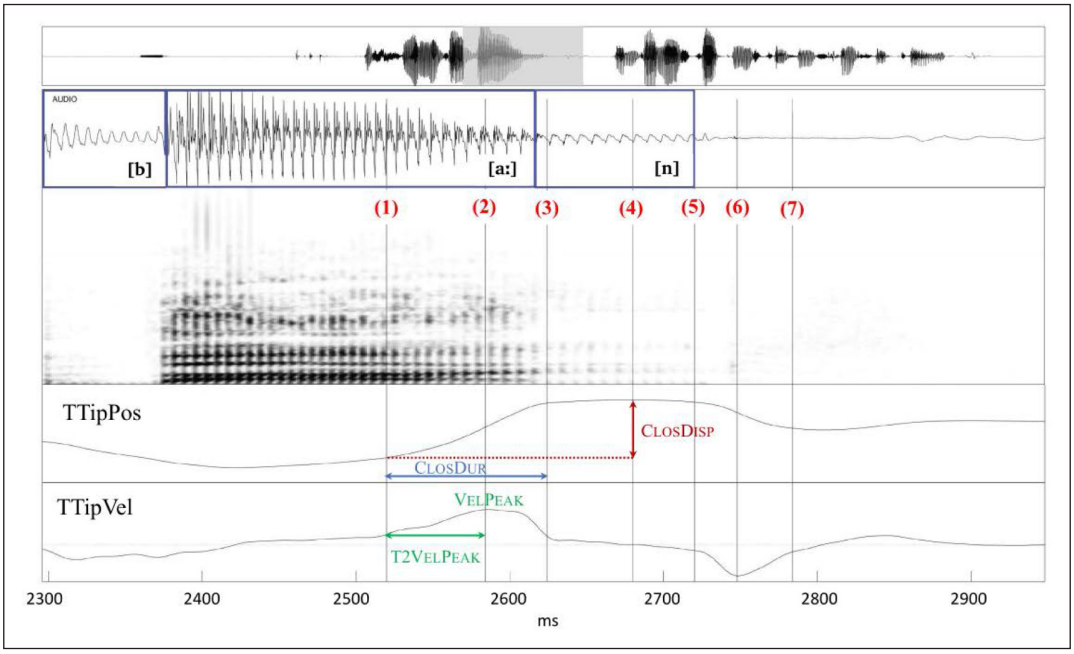


**Figure 1:** Labelling procedure for measuring kinematic parameters, exemplified for the articulatory movements for the preboundary consonant /n/ in the target word *Bahn*. The vertical tongue tip signal (TTipPos) and its velocity (TTipVel) were labelled for (1) gesture onset, (2) peak velocity of closing movement, (3) plateau onset, (4) point of maximum constriction, (5) plateau offset, (6) peak velocity of opening movement, and (7) gesture offset.

### 2.2.4. Statistics

For the production study we first calculated linear mixed effects models using the *lme4* package in R (see D. Bates et al., 2015; R Core Team, 2020). We tested whether the dependent variables WORDDUR, CLOSDUR, T2VELPEAK, VELPEAK, CLOSDISP, and F0R differ for the utterance-medial and utterance-final positions with participant as random effect. The phrase-medial position was only included in the figures for visualisation of the direction of effects but not tested statistically (see Belz et al., 2023, for a comparison between phrase-medial and utterance-medial position). Since the target word had a systematic and quite large effect on most of the kinematic parameters, the models were calculated separately for each target word. Outliers were excluded for items with residuals exceeding 2.5 times the standard deviation, for each target word-variable subset separately (see D. Bates et al., 2015). The effect size (Cohen's *d*) was calculated using the R

package *effsize* version 0.8.1 (Torchiano, 2020). To analyse the time variable velocity profiles of the tongue tip and lip closing movements and the f0 contour during the rhyme, we calculated Generalized Additive Mixed Models (GAMMs) following the recommendations of Wieling (2018). This non-linear regression method can identify systematic patterns in time-varying data while also modelling item- and participant-specific variation. We used R packages *msgv* for modelling (Wood, 2017) and *itsadug* for visualisation (van Rij et al., 2020). To find the appropriate models, we applied model comparisons as suggested in Wieling (2018) and Sóskuthy (2021). Further adjustments are explained in Sections 2.3.2 and 2.3.3.

## 2.3. Results

### 2.3.1. Acoustic and kinematic parameters

In this section we investigate whether speakers show differences in production between the utterance-medial and the utterance-final position. **Figure 2** shows the distributions of the parameters, color-coded for the prosodic conditions. Phrase-medial position is distinct from the other positions for the word and closing gesture duration WordDur, ClosDur and for the residualized log f0 f0R. The phrase-medial position is included as for visualisation of the effect direction. The parameters VelPeak and ClosDisp show bimodal distributions (see **Figures 2** and **13**, Appendix). This can be attributed to larger and faster closing movements for the words *Bahn* and *Bann* with low vowels compared to the other target words with high vowels. As can be seen in **Figure 13** in the Appendix, the kinematic parameters peak velocity (VelPeak), displacement (ClosDisp), and duration (ClosDur) of the closing movement are highly correlated. The acoustically measured parameter WordDur is less strongly correlated with the other parameters because it is a more global measure, including all segments of the target word, whereas the others are localized to the final closing movement. For all parameter combinations there is a clear difference between phrase-medial (light blue lines in **Figure 13**) on the one hand, and utterance-medial (violet lines) and utterance-final (brown lines) on the other hand, but no differences between the latter two. Therefore, for the following statistical analysis, only the two levels utterance-medial and utterance-final were compared. Since some of the variables show a clear difference for the target words investigated here, the data had to be split accordingly. An alternative would be to include target word as random factor but this was rejected because of systematic effects of the target word on the parameters. For example, VelPeak, ClosDisp, and ClosDur vary with vowel height, WordDur with the segmental composition, and f0R is known to be higher for close vowels (intrinsic f0; see for German Hoole & Mooshammer, 2002). The results of the linear mixed effects models for each parameter (columns) and target word (rows) are presented in **Table 2**. The change of each parameter from utterance-medial to utterance-final position is given as $\beta$ for each target word. Except for the residualized log fundamental frequency

F0R, all parameters switch the sign for $\beta$ across words. WORDDUR is shorter in utterance-final than in utterance-medial position for the words with lax vowels *Bann*, *Rum* and *still* and the same or longer for all words with tense vowels. However, the differences are small and only significant for *Bahn* (see also **Figure 3**). This inconsistency across words is most obvious for the parameter CLOSDUR when comparing the words *Rum* and *Stiel*. Therefore, even though significant changes could be detected for some parameter-word combinations, these are spurious and inconsistent, with negligible or small values for Cohen's *d*. This is not the case for the residualized log fundamental frequency F0R (last panel in **Figure 3**): for all target words F0R is significantly lower in utterance-final position than in utterance-medial position with medium or larger effects as indicated by Cohen's *d* (see **Table 2**).

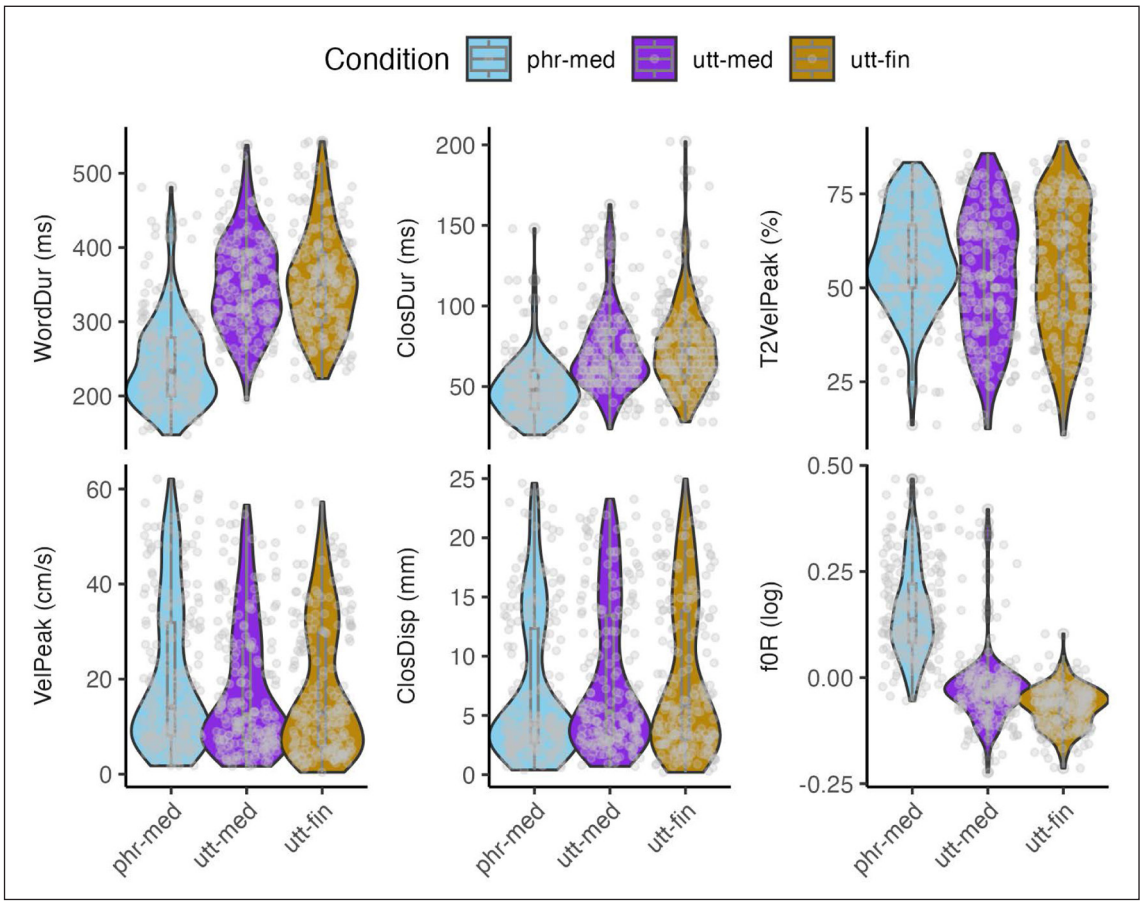


**Figure 2:** Violin plots with boxplots and data points of the acoustic and kinematic parameters word duration WORDDUR (ms), duration of the closing movement CLOSDUR (ms), time to peak velocity T2VELPEAK (%), peak velocity VELPEAK (cm/s), CLOSDISP (mm), and the residualized log f0 F0R.

| Word | | WordDur | ClosDur | T2VelPeak | VelPeak | ClosDisp | f0R |
|---|---|---|---|---|---|---|---|
| Bahn | *N* | 70 | 68 | 71 | 70 | 69 | 70 |
| | *β* | 15.18 | 0.34 | 1.50 | 0.23 | 0.28 | –0.03 |
| | *SE* | 5.62 | 2.97 | 2.22 | 1.11 | 0.34 | –0.01 |
| | sig | ** | | | | | ** |
| | Cohen's *d* | –0.39 | –0.13 | –0.17 | 0.07 | –0.01 | **0.58** |
| Bann | *N* | 71 | 71 | 71 | 71 | 71 | 71 |
| | *β* | –9.19 | –2.25 | –0.95 | 2.12 | 0.23 | –0.03 |
| | *SE* | 5.40 | 1.93 | 1.93 | 1.11 | 0.28 | –0.01 |
| | sig | | | | | | ** |
| | Cohen's *d* | 0.10 | 0.25 | 0.04 | –0.26 | –0.04 | **0.57** |
| Ruhm | *N* | 68 | 67 | 68 | 66 | 67 | 65 |
| | *β* | 13.00 | 3.29 | 1.47 | –1.12 | –0.50 | –0.03 |
| | *SE* | 6.56 | 4.84 | 3.56 | 0.39 | 0.19 | 0.01 |
| | sig | | | | ** | * | ** |
| | Cohen's *d* | –0.33 | –0.08 | –0.1 | **0.68** | **0.52** | **0.64** |
| Rum | *N* | 74 | 72 | 72 | 73 | 72 | 73 |
| | *β* | –1.12 | –11.77 | 0.06 | –0.48 | –0.46 | –0.05 |
| | *SE* | 5.52 | 3.69 | 2.38 | 0.45 | 0.16 | 0.01 |
| | sig | | ** | | | ** | *** |
| | Cohen's *d* | 0.01 | 0.39 | –0.01 | 0.11 | 0.33 | **0.99** |
| Stiel | *N* | 63 | 62 | 64 | 63 | 63 | 64 |
| | *β* | 10.78 | 8.06 | 4.37 | 0.25 | 0.61 | –0.02 |
| | *SE* | 7.38 | 4.12 | 3.39 | 0.79 | 0.32 | 0.01 |
| | sig | | | | | | ** |
| | Cohen's *d* | –0.34 | –0.44 | –0.33 | –0.05 | –0.40 | **0.54** |
| still | *N* | 66 | 67 | 68 | 66 | 67 | 66 |
| | *β* | –1.10 | 5.42 | 2.48 | 0.48 | 0.37 | –0.03 |
| | *SE* | 6.35 | 3.8 | 3.45 | 0.50 | 0.28 | 0.01 |
| | sig | | | | | | ** |
| | Cohen's *d* | –0.01 | –0.21 | –0.15 | –0.11 | –0.19 | **0.68** |

**Table 2:** Results of linear mixed effects models for the parameters WordDur, ClosDur, T2VelPeak, VelPeak, ClosDisp, and f0R, split by target word (rows). N = number of items, *β* = estimate of change from utterance-medial to utterance-final, SE = estimate of standard error, Cohen's *d* in bold for medium and large effects ($d > |0.5|$). For each subset, the outliers were excluded by the criterion described in section 2.2.4, therefore the number of items differs for each cell.
***$p < .001$; **$p < .01$; *$p < .05$.

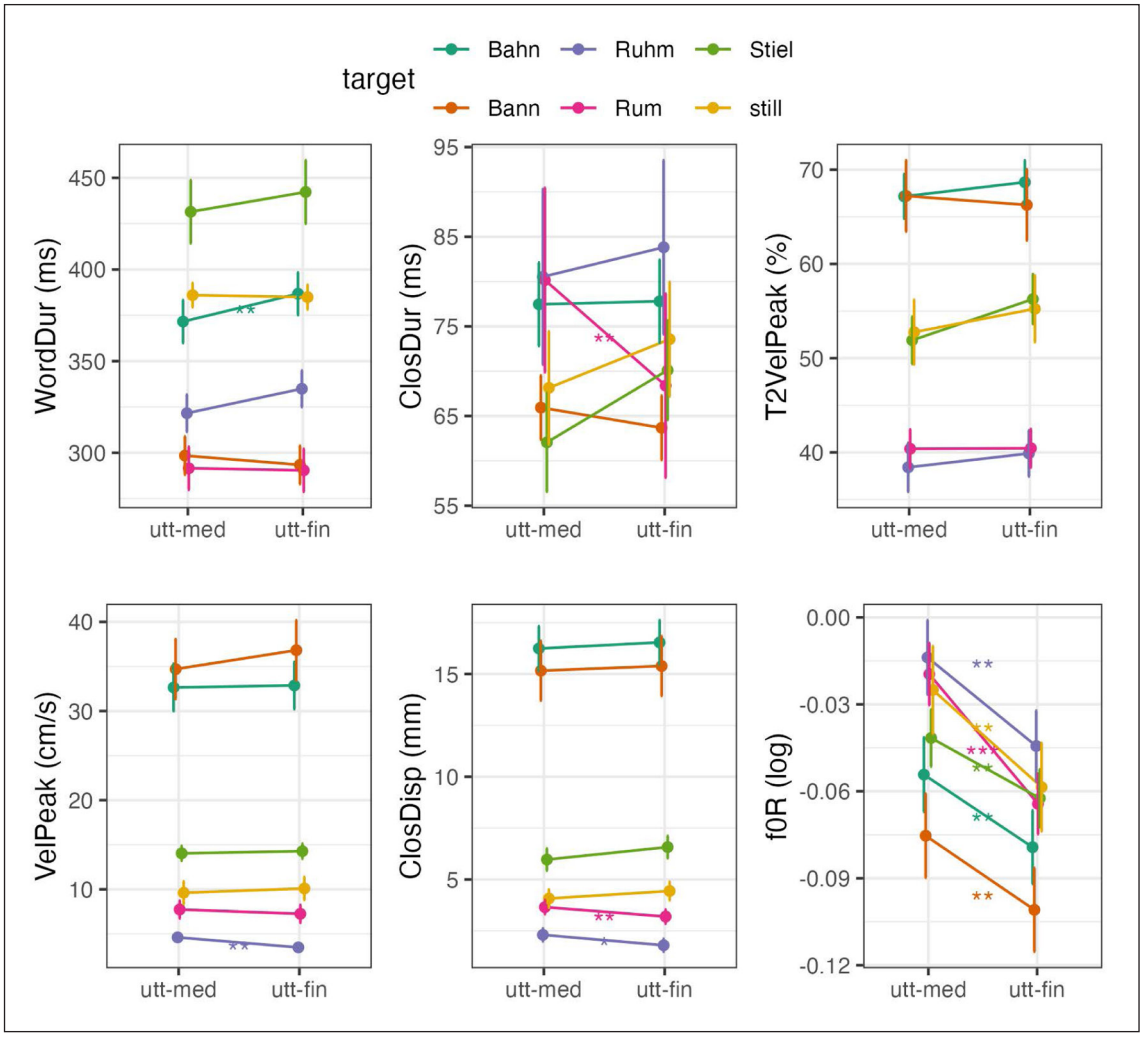


**Figure 3:** Effect plots: Results from linear mixed models, color-coded for each target word, for the parameters WORDDUR, CLOSDUR, T2VELPEAK, VELPEAK, CLOSDISP, and F0R. P-values $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

To summarize so far, utterance-medial position can be distinguished consistently from the utterance-final position by a higher fundamental frequency. The other parameters do not show any consistent differences across the investigated words. One possible reason for this negative outcome could be that the measured kinematic parameters CLOSDUR, T2PVELPEAK, CLOSDISP, and VELPEAK are not sensitive to finer details of the velocity profiles. For example, T2VELPEAK does capture certain aspects such as the symmetry of the velocity profile, but there might be other shape differences relevant for the distinction between the two final positions. Therefore, in the following exploratory section the shape of the velocity profiles will be investigated in more detail by using Generalized Additive Mixed Models.

### Shape of the velocity profiles

As described in Section 2.2.3, the input data for the GAMM analysis consist of the scaled and time-normalized velocity profiles for the final closing movement. The raw and time-normalized velocity profiles are shown in the Appendix, **Figure 14**, and the averaged profiles of the 8 speakers can be seen in **Figure 4** for each target word. Most profiles exhibit a single peak and the curves for the phrase-medial position are usually higher than for the other two conditions. For model comparison, we followed the suggestions by Wieling (2018). Since the residuals show strong deviations from the normal distributions at both tails, the data were modelled by adding the model specification `family = "scat"` in R, i.e., the scale-t family for heavily tailed data. Furthermore, correction for the autocorrelation in the time series data was calculated and included in the final model. The following model for the two final conditions was selected by model comparison:

```
bam(vel.norm ~ Condition +
   s(Time, by = Condition, k = 24, bs = "tp") +
   s(Time, Speaker, bs = "fs", m = 1, k=21, xt="cr") +
   s(Time, target, bs = "fs", m = 1, k=21, xt="cr"),
   rho = 0.9400473, AR.start = start.event,
   discrete = T, family = "scat")
```

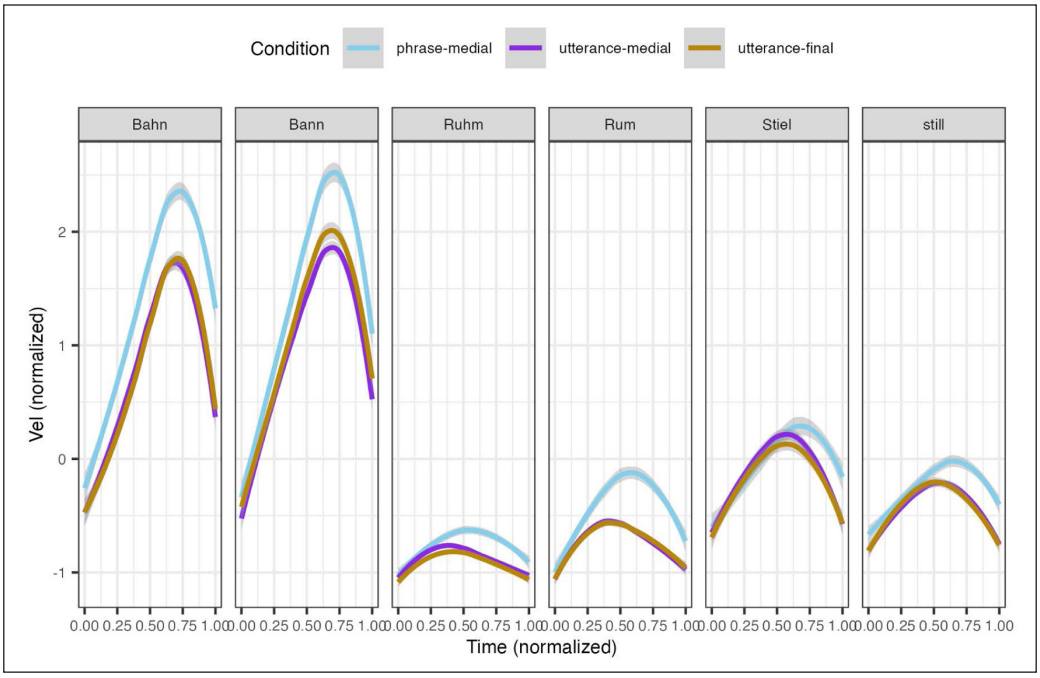


**Figure 4:** Velocity profiles of the closing movement per target word, color-coded for condition.

where `vel.norm` is the time series of the velocity profile; `Condition`, the position; `Participant`, the participant, and `target`, the target word. The variable `Time` in the terms entered here means that velocity profiles varied over time, target word, and participant. The model did not improve if `Condition` was included in these terms. As shown in **Figure 5a** the velocity profiles are not affected by prosodic position, i.e., the modelled profiles for the utterance-medial (violet) and the utterance-final position (brown) overlap completely. The difference between the two curves is not significant (see **Figure 5b**). **Table 3** also indicates that the velocity profiles do not differ significantly in parametric and in smooth terms.

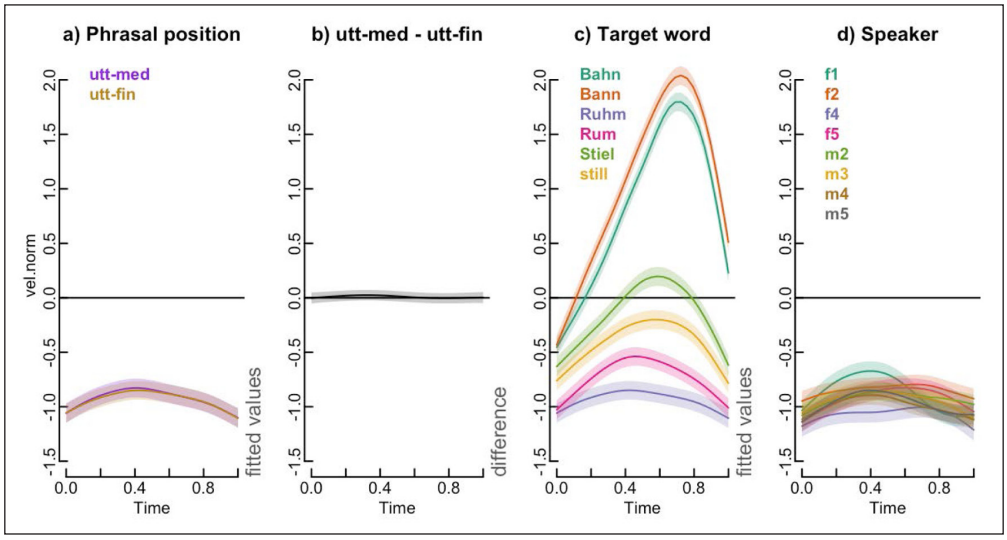


**Figure 5:** Modelled velocity profiles of the closing movement for a) the two final conditions, b) the difference between the two final positions, c) the target words, and d) the speakers. For better visibility the random effects are not included, therefore the widths of the confidence bands are smaller. The lines are extracted from the fitted GAMM.

| **A. parametric coefficients** | **Estimate** | ***SE*** | ***t*** | ***p*** |
|---|---|---|---|---|
| (Intercept) | –0.3151 | 0.1549 | –2.0339 | * |
| Conditionutt-fin | –0.0090 | 0.0218 | –0.4141 | |
| **B. smooth terms** | **edf** | **Ref.df** | ***F*** | ***p*** |
| s(Time):Conditionutt-fin | 3.6185 | 4.9271 | 1.2075 | |
| s(Time,subject.id) | 53.6115 | 167.0000 | 1.9058 | *** |
| s(Time,target) | 90.8564 | 125.0000 | 95.9777 | *** |

**Table 3:** Approximate significance of parametric and smooth terms for the GAM model of velocity profiles.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

In **Figure 5c**, the effect of the target word over time is shown. As was also observed above in Section 2.3.1, the target words *Bahn* and *Bann* with low vowels show larger displacements and peak velocities than words with close vowels. Because the differences for the target words are very prominent, the data were split by target word and the individual models were calculated per target word. The residuals for the models for *Bann* and *Ruhm* were heavily tailed, so these models were calculated with the model specification `family = "scat"`:

```
bam(vel.norm ~ Condition +
   s(Time, by=Condition, k=20, bs="tp")  +
   s(Time, Speaker, bs="fs",m=1, xt="cr", k=20),
   rho = rhoval, AR.start = start.event, discrete = T)
```

The results per target word are visualized in **Figure 6**. The red line at the bottom indicates periods of significant differences. To correct for multiple comparisons, confidence intervals are set at 2.58 times the standard error. Only for the target word *Bann* a short period shows a significantly larger value for the utterance-medial position. For the other target words the velocity profiles do not differ between the two positions.

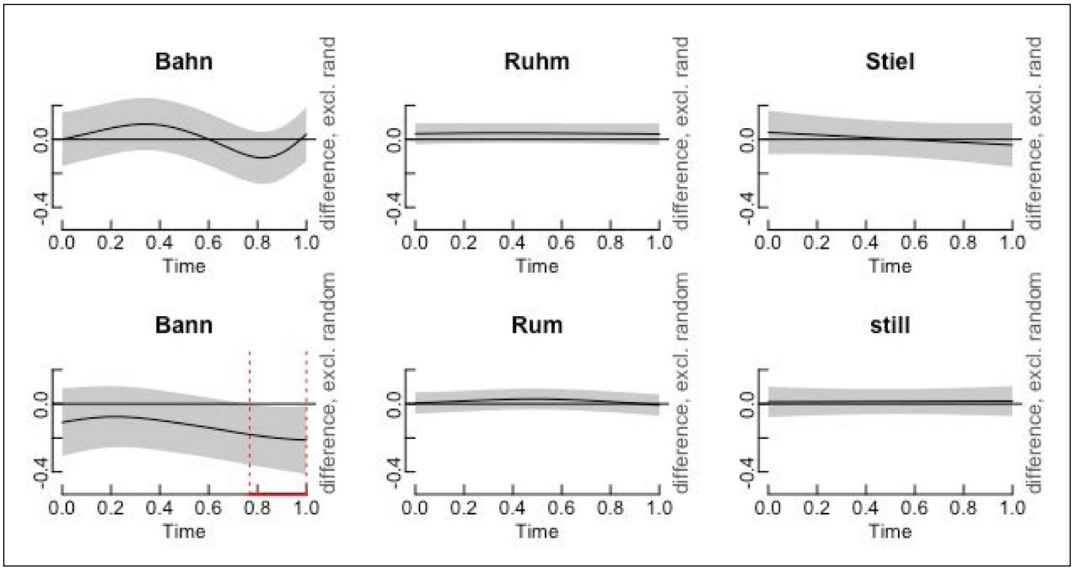


**Figure 6:** Difference trajectories with confidence bands of 2.58 times the standard error for utterance-medial vs. utterance-final position. The red line indicates the areas with significant differences between the two conditions. The lines are extracted from the fitted GAM.

The results indicate that the velocity profiles are larger with larger velocity peaks for the phrase-medial position (as shown in **Figures 14** and **4**) but that there is only a single significant difference between the utterance-medial and utterance-final positions for a short period of time.

There also seems to be no difference in shape, such as peakedness or skewness, for the different conditions.

### 2.3.3. f0 contours

**Figure 7** shows the contours of the residualized log f0 contours per speaker, averaged for each prosodic position.[1] The rhymes in phrase-medial position (shown as light blue lines) are generally realized with a higher f0 that is often flat or slightly rising, except for participant f4, who shows a falling contour. In the utterance-medial and utterance-final positions (violet and brown lines), the f0 is mostly falling, with the exception of participant f4, who shows a falling-rising pattern. For most speakers the contour shapes for utterance-medial and utterance-final are similar and run in parallel, with the exception of m3. This participant produced the first block with a different pattern for intended utterance-medial test words with rising intonation in the rhyme and for some phrase-medial test words with a following pause, which were identified as IP boundaries by the annotators. These items still sounded natural. Because of this variation, the confidence interval in **Figure 7** is rather wide for this speaker (see also **Figure 15** in the Appendix, speaker m3, violet lines). For 5 out of the 8 speakers there is a pronounced difference in height with lower contours in utterance-final position (brown lines) than in utterance-medial position (violet lines). Speakers f5, m4, and m5 differentiate very little or not at all between the two conditions.

To test whether the observed differences are statistically significant, several GAM models were compared, following the suggestions by Wieling (2018) with corrections for auto-correlation and heavily tailed distributions. As suggested by Sóskuthy (2021), the smooth terms were fitted with cubic regression splines that slightly reduce type I errors for pitch data with large *k* values. By model comparison, the following model was selected:

```
bam(T1_resid ~ Condition +
   s(Time, k = 20, by = Condition, bs = "tp") +
   s(Time, Speaker, by = Condition, bs = "fs", m = 1, k=21, xt="cr") +
   s(Time, target, by = Condition, bs = "fs", m = 1, k=21, xt="cr"),
   rho = 0.9585071, AR.start = start.event,
   discrete = T, family = "scat")
```

where `T1_resid` is the time series of the residualized log f0 contour; `Condition`, the ordered factor position; `Speaker`, the participant, and `target`, the target word.

[1] **Figure 15** in the appendix shows the time-normalized raw f0 contours per speaker and prosodic condition.

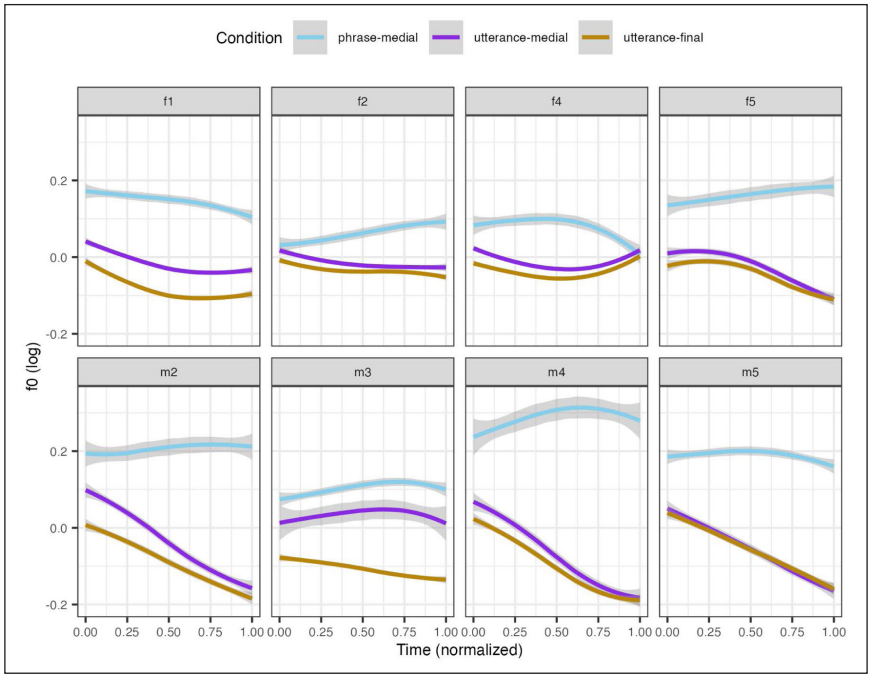


**Figure 7:** Averaged residualized log fundamental frequency contours per speaker during the rhyme, color-coded for condition.

**Figure 8a** shows the modelled f0 contours for the rhymes in the two final positions, and **Figure 8b**, the difference between them. As shown in the top part of **Table 4** and in **Figure 8b**, the utterance-medial position is significantly larger than the utterance-final position over the complete rhyme (see red line at the bottom). F0 contours differ for the target words with lower peaks for the words with low vowels (see **Figure 8c**) due to vowel-intrinsic f0 differences (see Hoole & Mooshammer, 2002, for German).

As can be seen in **Figures 7** and **8d**, speakers vary substantially in the realisation of this difference. Therefore, GAM models were calculated per speaker. The data by speakers f2, m3, m4 and m5 show heavily tailed distributions of the residuals, which is why, for f2, m4 and m5, the data were modelled by adding the model specification `family = "scat"`. This was not possible for speaker m3, probably due to the bimodal distribution of the extracted contours. Corrections for autocorrelation were determined per speaker.

```
bam(T1_resid ~ Condition +
   s(Time, k = 20, by = Condition, bs = "tp") +
   s(Time, target, bs = "fs", m = 1, k=21, xt="cr"),
   rho = rhoval, AR.start = start.event,
   discrete = T)
```

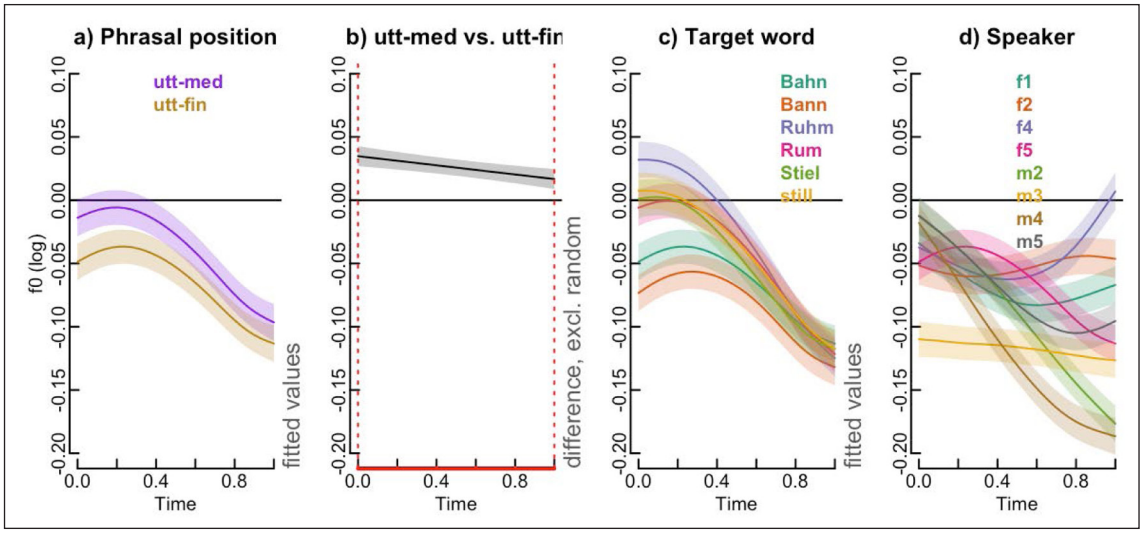


**Figure 8:** Modelled f0 contours during rhymes with confidence bands for **a)** utterance-medial and utterance-final position, **b)** the difference between utterance-medial and utterance-final position, **c)** different target words, and **d)** speaker. The red line in b) indicates the period with significant differences between the two lines. The lines are extracted from the fitted GAMM.

| **A. parametric coefficients** | **Estimate** | ***SE*** | ***t*** | ***p*** |
|---|---|---|---|---|
| (Intercept) | –0.0573 | 0.0139 | –4.1388 | *** |
| Condition.L | –0.0182 | 0.0023 | –7.8489 | *** |
| **B. smooth terms** | **edf** | **Ref.df** | ***F*** | ***p*** |
| s(Time):Conditionutt-fin | 1.0005 | 1.0010 | 14.9527 | *** |
| s(Time, speaker) | 53.9396 | 167.0000 | 1501.6795 | *** |
| s(Time,target) | 24.0605 | 125.0000 | 3.2342 | *** |

**Table 4:** Approximate significance of fixed and smooth terms for the f0 GAM for utterance-medial and utterance-final position.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

**Figure 9** shows the modelled differences of the residualized log f0 contours per speaker. The red line at the bottom of each plot indicates periods of significant differences. To correct for multiple comparisons, confidence intervals are set at 2.58 times the standard error. For speakers f1, f4, m2, and m3, the f0 contours differ significantly over the complete rhyme. For speaker f5, the final 20% of the rhyme is not significant, and for m4 neither is the final 75%. Speaker m5 uses similar f0 contours for the two positions.

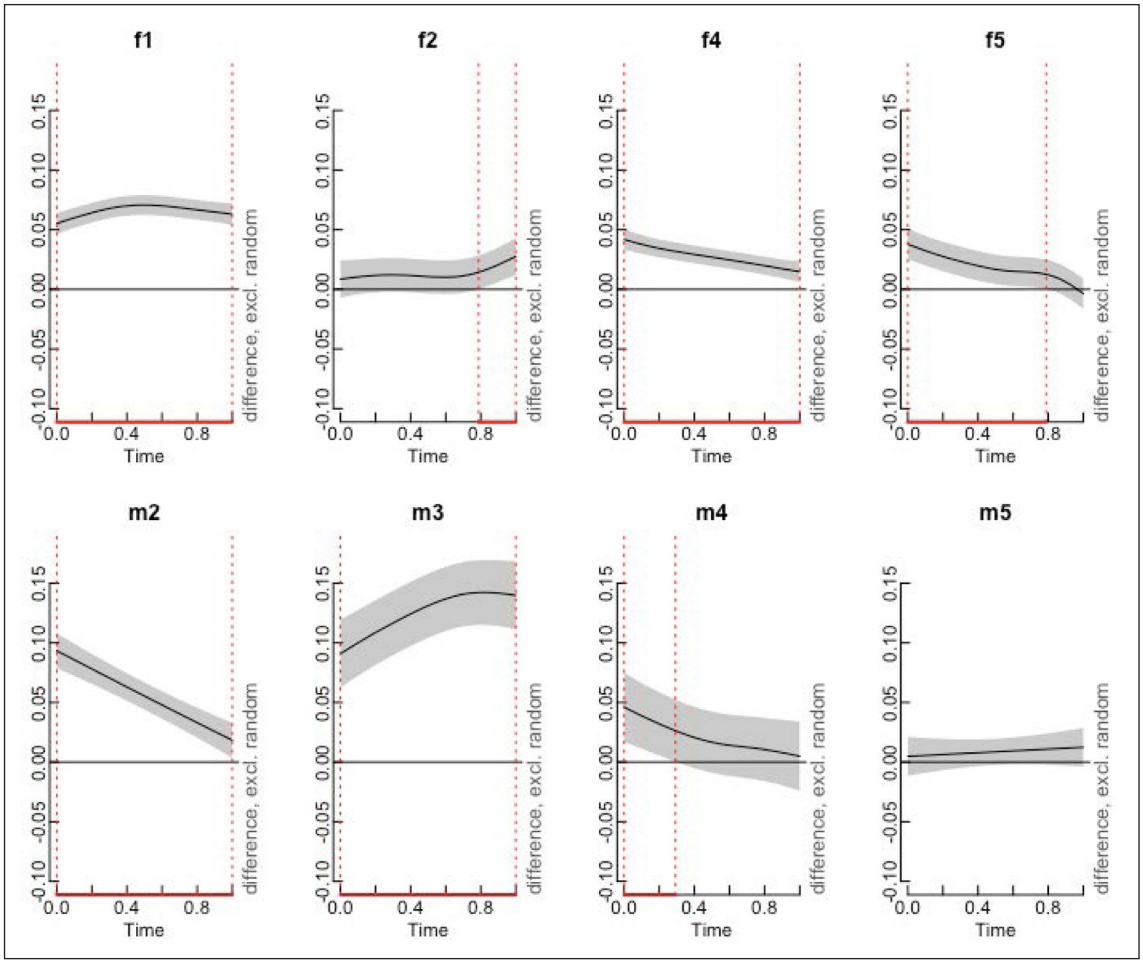


**Figure 9:** Difference f0 contours during rhymes for each speaker with confidence bands of 2.58 times the standard error for utterance-medial vs. utterance-final position. The red line indicates the areas with significant differences between the two conditions. The lines are extracted from the fitted GAMM models per speaker.

## 2.4. Summary production

As shown visually, the kinematic and acoustic parameters differ for the phrase-medial position compared to utterance-medial and utterance-final, i.e., the closing gesture is shorter and faster, the word duration is shorter and the f0 is higher than in phrase-medial position. However, the kinematic parameters as well as the velocity profiles of the final closing gesture did not differ systematically for the utterance-medial and utterance-final positions. There are some inconsistent and small differences for WORDDUR, VELPEAK, and CLOSDISP. However, these are not consistent across target words and also change direction. Much more consistent is the effect on the F0R within the rhyme. In the utterance-final condition the mean f0 is significantly lower than in the utterance-medial position. Furthermore, the f0 contours differ significantly across

the whole rhyme with larger values for the utterance-medial position. This is also confirmed for most individual speakers with varying degrees of time stretches. Therefore, f0 differentiates the two final positions most consistently. We now turn to the question of whether listeners use this parameter as a cue for continuation.

## 3. Perception experiment

The goal of the perception experiment is to investigate whether listeners perceive the observed differences in production and use them to judge whether the speaker will continue to speak after the end of a phrase. Therefore we carried out a rating experiment. We included the phrase-medial condition to show that the participants generally understood the task.

### 3.1. Methods

#### 3.1.1. Stimuli and task

For each of the 8 speakers, one trial of the five iterations they produced during the production experiment was determined for every position (utterance-medial/utterance-final) for the six target words by listening to the stimuli and by visual inspection of the f0 contours. The criteria for selection were no hesitation within the utterance and an intonation contour closest to the speaker's average. All selected utterances were cut at the end of the target word and postprocessed by ramping down the intensity to avoid sudden jumps due to cutting the audio stream.

The experiment was designed as an online experiment with the configurator Percy (see Draxler, 2017). First, the listeners were asked for their gender, age, and native language. Then the experiment started. After playing each utterance, the following question appeared on the display: "Do you think the speaker will continue to speak?" ('Glauben Sie, dass der Sprecher noch weitersprechen wird?') together with a Likert scale from 1 (no) to 7 (yes). The first two trials contained the target word *Beet* and were played to the participants to allow them to adjust the loudness and to familiarize them with the task. Utterances including *Beet* were also used as fillers to distract them from the minimal pairs.

A total of 168 stimuli (3 prosodic conditions $\times$7 target words $\times$8 speakers) were presented, of which 144 were experimental stimuli. The experiment was designed to last 30 minutes, and it took the listeners a median of 16.5 minutes to complete.

#### 3.1.2. Participants

The rating tests were carried out in two different locations: the phonetics laboratory and via internet. For the part taking place in the phonetics lab, the experiment was promoted via https://lingex-zas.de (an online recruitment platform for linguistic experiments). Participants (from now on referred to as *listeners*) were compensated with 5 Euros. Twenty-seven listeners were tested

in a sound-attenuated booth in the phonetics laboratory to ensure that they focused on the task and finished the experiment. These results could then be compared to the results in the online experiment, ensuring that all listeners really did pay attention to the task. A link for the online experiment was sent to several email lists. Altogether 72 listeners volunteered, 45 online and 27 in the laboratory. Within the online experiment 17 listeners did not finish the experiment and were therefore excluded. Furthermore, one person who participated online did not pay attention to the stimuli and was also excluded. A total of 33 female and 21 male listeners were included in the following analysis, 27 in the lab and 27 online. For speaker m3, listeners rated one utterance-medial stimulus from the first block very high for continuation, resembling a phrase-medial pattern, which is why this stimulus was treated as an outlier and excluded from the analysis.

### 3.1.3. Statistics

The phrase-medial condition was compared to the utterance-medial and utterance-final condition by a visual inspection of the ratings, clearly indicating huge differences in rating. To test whether the ratings for the two conditions under investigation (utterance-medial and utterance-final) differed significantly, a cumulative link mixed model (CLMM) was constructed using the *ordinal* package version 2022.11.16 (Christensen, 2022). The model contained *condition* (phrase-final, utterance-final) and *controlled* (internet, lab) as fixed effects with sum-coded contrasts and their interaction. Random intercepts for participants, stimuli, and target speaker were added, together with random slopes per condition for listeners. The effects are visualised by extracting the predicted probabilities from the model using *ggeffect()* from the *ggeffects* package (Lüdecke, 2018).

The full model translates to:

```
clmm(ratings ~   condition * controlled +
                      (1 + condition | listeners) +
                      (1 | stimulus) + (1 | target.speaker,
                      data = p.sub))
```

## 3.2. Results

As expected, the phrase-medial condition was rated higher for continuation than the utterance-medial and utterance-final condition. **Figure 10a** shows the mean ratings per speaker. As can be seen, listeners perceive a difference of utterance-medial and utterance-final stimuli except for speakers *f2*, *m4*, and *m5*. Therefore, there was a clear perceptual difference between phrase-medial and the other two conditions, which indicates that the listeners understood the task and were able to rate the stimuli for continuation. The mean ratings of prosodic condition per test

location can be seen in **Figure 10b**. While phrase-medial stimuli show high proportions for continuation, the picture is less clear for utterance-medial and utterance-final stimuli. Both are rated low for continuation, with lower means for utterance-final stimuli at the lower end of the range (no continuation). For all three conditions, the ratings are higher in the lab compared to the online ratings.

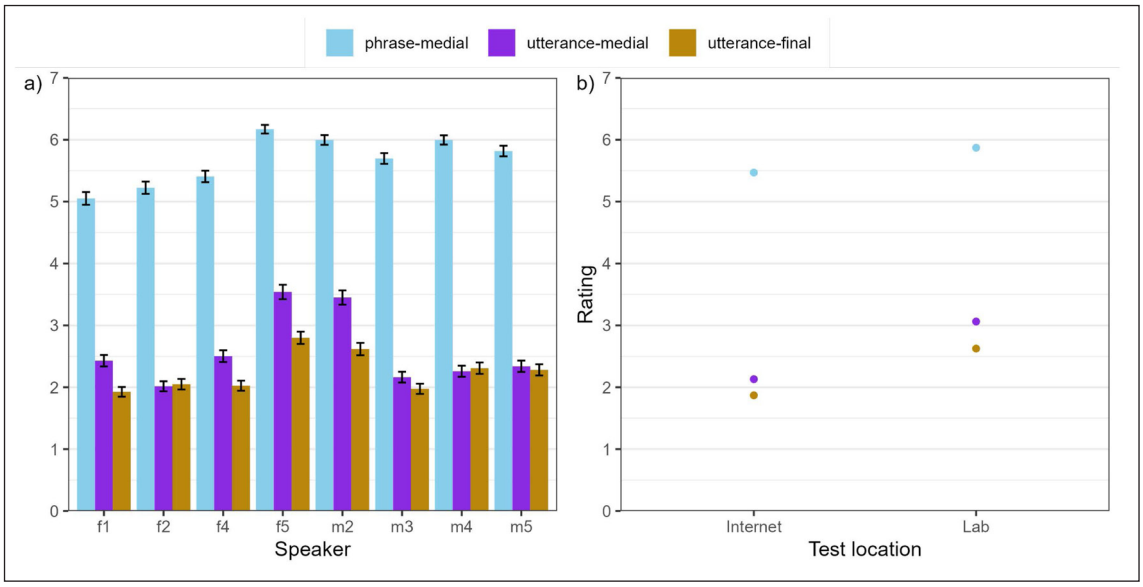


**Figure 10:** Results of the rating experiment: a) mean ratings of test location, b) mean ratings per speaker with standard errors.

To compare the two final levels, a cumulative linked mixed model (CLMM) was calculated, including only the utterance-medial and utterance-final conditions. Both condition and test location showed a significant effect in the model, but the interaction was not significant. The model indicates that utterance-final stimuli are rated significantly lower for continuation than utterance-medial stimuli (**Table 5**).

**Figure 11a** illustrates the percentage of ratings across condition and test location in accordance with the CLM model. It can be observed that ratings 1 and 2 (i.e., indicating minimal expected continuation) constitute over 50% of the ratings in both test locations and conditions, and over 60% of the ratings of utterance-final stimuli. **Figure 11b** shows the predicted probabilities for the ratings across condition and test location. Once more, listeners who participated online exhibited stronger preferences for rating extreme values, whereas listeners in the laboratory demonstrated a tendency towards greater caution. Nevertheless, the results obtained from both test locations indicate a higher probability of expecting no further continuation after utterance-final than utterance-medial stimuli.

| | CLMM |
|---|---|
| Condition Utterance-final | –0.53 (0.10) *** |
| Location Lab | –1.54 (0.39) *** |
| Cond.Utterance-final:Loc.Lab | 0.05 (0.12) |
| 1\|2 | –0.22 (0.28) |
| 2\|3 | 1.00 (0.28) *** |
| 3\|4 | 1.72 (0.28) *** |
| 4\|5 | 2.50 (0.28) *** |
| 5\|6 | 3.27 (0.29) *** |
| 6\|7 | 4.22 (0.29) *** |
| AIC | 14257.10 |
| Num. obs. | 5130 |
| Groups (stimulus) | 95 |
| Groups (userid) | 54 |
| Groups (target.speaker) | 8 |

**Table 5:** Predictions for listeners' ratings on a scale of no continuation (1) to continuation (7) for utterance-medial (reference level) and utterance-final conditions and location (controlled: Internet vs. lab), with standard error in parentheses.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

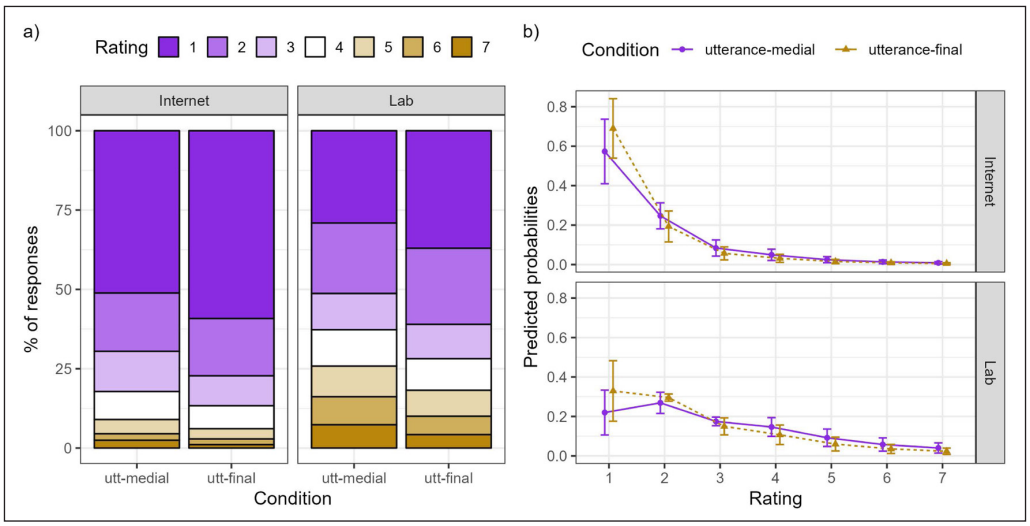


**Figure 11: a)** Percentage of ratings (1 = no continuation, 7 = continuation) per condition (utterance-medial vs. utterance-final), **b)** predicted probabilities of ratings per condition and test location.

Finally, we investigate effects specific to the target speakers producing the stimuli, calculating Kruskal-Wallis rank sum tests for the ratings of the utterance-medial and utterance-final condition. We found significant effects for five of eight speakers: *f1* ($\chi^2 = 17.6$, $p < .001$), *f4* ($\chi^2 = 14.1$, $p < .001$), *f5* ($\chi^2 = 19.5$, $p < .001$), *m2* ($\chi^2 = 26.6$, $p < .001$), *m3* ($\chi^2 = 33.7$, $p < .001$). This means that either the other three speakers have produced fewer cues for the listeners to rely on or that the cues were too subtle. In this case, the Likert scale might have been too coarse to measure the cues.

### 3.3. Summary

To summarize, we find evidence for a perceptual difference between phrase-medial, utterance-medial, and utterance-final stimuli by comparing their continuation ratings. Phrase-medial stimuli are rated significantly higher for continuation than utterance-medial and -final stimuli. Further, utterance-final stimuli are rated significantly lower for continuation than utterance-medial stimuli, suggesting that listeners can perceive the phonetic cues indicating the boundary position. A more fine-grained analysis reveals that listeners perceive differences in this boundary position for five of the eight speakers.

## 4. Perception-production link

In the following section we investigate whether acoustic and articulatory parameters of the stimuli used in the perception experiment can predict the rating results of the perception study. The aim is to identify relevant production parameters that listeners use for distinguishing between the utterance-medial and utterance-final positions.

### 4.1. Material and method

First, we selected the production data linked to the 144 trials used for the perception experiment (see Section 3.1.1). Sixteen of the stimulus trials were excluded from the previous articulatory and f0 analyses as outliers (see Sections 2.2.2 and 2.2.3). Furthermore, one stimulus in utterance-medial position showed exceptionally large ratings for continuation with a mean score of 6.4 and was therefore excluded as an outlier. The following analysis is based on the remaining 84 trials, 40 in utterance-medial position and 44 in utterance-final position. For these, the rating score was averaged over the 54 listeners.

To investigate which parameter best predicts the rating scores, stepwise regression models was calculated by the function `stepwiseAIC` from the R package MASS (see Venables & Ripley, 2002) which chooses the best model based on an AIC criterion by forward and backward selection. The dependent variable were the averaged scores and the measured parameters VelPeak, ClosDisp, ClosDur, T2VelPeak, f0R and WordDur were used as predictors. To deal with collinearity

between the predictors we followed Winter (2019) and calculated the variance inflation factor using *vif* from the package *car* (Fox & Weisberg, 2019). Values exceeding 4 indicate extensive collinearity and, as suggested by Winter (2019), one of these predictors should be excluded.

## 4.2. Results

A regression model for the two final conditions was calculated. To reduce effects of collinearity the variance inflation factors were calculated following the recommendations in Winter (2019). Since the values of 11.35 for ClosDisp and 10.78 for VelPeak exceed the threshold of 4 and both parameters are strongly affected by vowel height, we excluded ClosDisp from further analysis.

The results of this model are shown in **Table 6**. Three parameters are selected as relevant for predicting the continuation ratings. The articulatory parameter ClosDur does not significantly affect the continuation ratings. The second parameter VelPeak is significantly larger, that is, the movement is faster for larger (i.e., less final) continuation ratings (see **Figure 12a**). Similarly, f0R is higher for higher continuation ratings. The variable VelPeak does not differ consistently for the two final positions in our full production dataset (see **Table 2**). To test whether this also holds for the reduced dataset of 84 stimuli used for the perception experiment, we calculated the following linear mixed effects model

```
lmer(VelPeak ~ Condition * target + (1|speaker))
```

| Model | Estimate | Standard error | *p* |
|---|---|---|---|
| (Intercept) | 2.6348 | 0.1682 | *** |
| ClosDur | –0.0030 | 0.0018 | |
| VelPeak | 0.0153 | 0.0040 | *** |
| f0R | 4.7234 | 1.2744 | *** |
| $R^2$ | 0.2684 | | |
| Adj. $R^2$ | 0.2410 | | |
| Num. obs. | 84 | | |

**Table 6:** Results of stepwise regression models for utterance-medial and utterance-final positions with the estimate of the intercept for averaged continuation ratings, the slope, the standard error and the significance of the slope. A positive slope indicates that larger phonetic parameter values predict larger continuation ratings.
$^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

and found a significant effect for the target but not for condition. This is confirmed by post hoc tests, using the R package `emmeans` (Lenth, 2024) which show that target words do not differ significantly for condition (see **Table 8** in the Appendix). Furthermore, Cohen's *d* for VelPeak is

negligible with a value of 0.073 (compared to a medium effect of 0.68 for F0R). As can be seen in **Figure 12**, left panel, the violet and brown data points also show no consistent separation for the reduced dataset used in the perception experiment. However, listeners seem to be sensitive to this articulatory parameter and tend to rate utterances with larger peak velocities as less final.

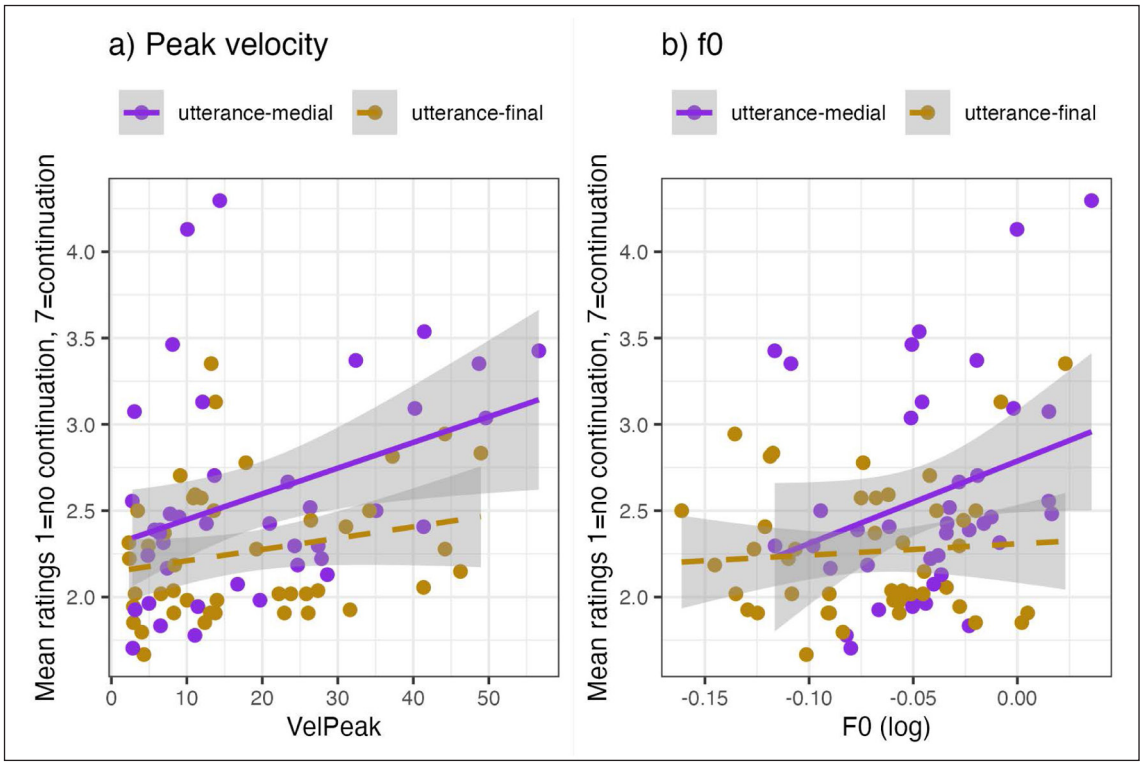


**Figure 12:** Scatterplots with regression lines and confidence intervals for rating scores and a) peak velocity and b) z-scored f0 for utterance-medial and utterance-final position.

## 4.3. Summary

We find evidence of a significant relationship between production data and continuation ratings. The explained variance for the regression model for utterance-medial and utterance-final position is 24%. Even keeping in mind that the regression models are based on a very small data set (84 trials), the result suggests that the rating scores can be partly predicted from the measured parameters, i.e., the impression of whether a speaker might continue speaking is signaled by temporal kinematic and tonal parameters. The selected parameters that best predict the continuation ratings of utterance-medial and utterance-final positions are the peak velocity of the prepausal closing gesture and the averaged log f0 of the rhyme. The effect of peak velocity on the continuation ratings seems to be less consistent than for the averaged f0 since the former is not systematically distinguished for the two final positions in the production data.

## 5. Discussion

We set out to answer the following questions concerning the production and perception of finality as well as the link between them:

- Question 1: Do speakers of German distinguish between utterance-medial and utterance-final IP boundaries?
- Question 2: Do German listeners distinguish between utterance-medial and utterance-final IP boundaries for judging upcoming continuation?
- Question 3: Which information do listeners use in perception?

Before we discuss these results, we briefly summarize the findings on phrase-medial (word) boundaries, partly based on Belz et al. (2023), where we found, for a subset of the current data, shorter closing gesture durations and shorter acoustic segment durations for phrase-medial compared to utterance-medial IPs for German. In the study at hand, larger excursions of the velocity profile and higher f0 values with level or rising contours were shown visually for the phrase-medial position compared to the final positions. The other kinematic parameters varied strongly with the target word. Listeners seem to use this information and rate phrase-medial stimuli higher for continuation than the final positions.

Turning now to the question of finality, for production (*research question 1*), our hypothesis was that utterance-medial and utterance-final IP boundaries will differ. For lengthening, we suggested two possible directions of effect: one is that in utterance-final position there is no upcoming utterance to plan, and therefore speakers do not need time to plan, and thus there might be less final lengthening than at utterance-medial IP boundaries. Alternatively, there might be more final lengthening and more pronounced f0 lowering at utterance-final IP boundaries, because the utterance-final boundary indicates a hierarchically higher prosodic category, and as such should show stronger phonetic correlates of boundaries, as has been found in many studies for utterance-medial prosodic boundaries (e.g., Cho and Keating, 2001; Fougeron and Keating, 1997; Krivokapić and Byrd, 2012; Ladd, 1988; Wagner, 2005). Surprisingly, the only systematic difference we found in production was for fundamental frequency, with lower f0 produced at utterance-final than at utterance-medial IP boundaries. The kinematic parameters and velocity profiles investigated here failed to show consistent effects. It is possible that this is because our speakers read the stimuli rather than produced them spontaneously, i.e., the utterance was provided to them, and this might reduce the effect of planning. While speakers of course have to plan an upcoming utterance in both spoken and in read speech, the amount of planning is reduced in read speech (see the discussion in, e.g., Ferreira, 1991; Krivokapić et al., 2022), and thus the potential effect might not appear as strongly in this case.

We also discussed the possibility that the utterance-final boundary might show different phonetic properties from the utterance-medial IP boundary as it is a turn-ending boundary. Only a few studies have examined this question, and they find different results. Local and Walker (2012) do find lengthening in a spontaneous speech study, while Gravano and Hirschberg (2011) and Purse and Krivokapić (2023) do not. We do not find any lengthening (or shortening) effect, and this might also be because we examined read speech, and therefore there was little reason for speakers to produce these utterances as part of a dialogue.

On the other hand, it is unclear why we did not see a lengthening effect distinguishing the prosodic hierarchy (i.e., lengthening of the utterance-final boundary compared to the utterance-medial one). A possible reason for the lack of any consistent effect might be that, in this study, the closing movement at the phrase edge was analysed. The following opening movement might be more sensitive to the subtle differences in boundary strength, given that this is where the effects of the boundary are known to be strongest in production (Belz et al., 2023; Byrd & Saltzman, 2003; Byrd et al., 2006). However, since in utterance-final position the closure was often unreleased, the opening movement could not be analysed. Thus, we are possibly missing important kinematic properties of boundaries and their perception. But we think this is not the reason for the lack of effect, as we do observe final lengthening on the closing duration when comparing utterance-medial and utterance-final boundaries to the phrase-medial boundary (while we did not conduct a statistical analysis, **Figure 13** indicates a difference, and Belz et al. (2023) analyse phrase-medial versus phrase-final utterance-medial boundaries and find lengthening). Thus, we would expect the boundary effect to be present on the closing movement for all boundaries, yet we do not observe a difference between utterance-medial and utterance-final boundaries. We therefore suggest that the lack of difference between utterance-medial and utterance-final boundaries is an accurate reflection of the properties of the boundaries, rather than an artefact of the lack of opening movement. One possible explanation for these findings (i.e., that there is no difference in the temporal properties of the two boundaries) is that both show a lengthening effect related to their structural position, but the utterance-medial IP also has a lengthening effect related to planning; this is only a small effect (due to relatively little planning in reading), and as a result, the two boundaries end up not differing. Under this interpretation, we see an effect of both planning and structure (see Purse & Krivokapić, 2023, for a similar argument). This interpretation is also consistent with the finding that f0 shows evidence that the utterance-final IP is hierarchically higher than the utterance-medial IP.

The most consistent phonetic correlate of finality in our study was f0, which was lower at utterance-final boundaries than at utterance-medial boundaries for the majority of our speakers. Thus, confirming our hypothesis in this respect, we found evidence that the hierarchically higher prosodic phrase is distinguished from a hierarchically lower one. Although the difference was found for the entire duration of the rhyme, it was not consistent across speakers, as three of

eight speakers did not show a pronounced difference (see **Figure 7**). The result deviates from the finding of Ladd (1988), who found no difference in f0 between strong and weak phrasal boundaries in English in read speech. However, the results are in line with f0 differences in studies of read speech in English for discourse-medial vs. discourse-final boundaries (Herman, 2000), in German for strong versus weak phrasal boundaries (Petrone et al., 2017), and in Dutch for turn-final and topic-final vs. non-turn-final and non-topic-final boundaries (Geluykens & Swerts, 1994). The similarities between these different prosodic boundaries can perhaps be seen as evidence that one of the correlates of signalling prosodic boundaries is hyperarticulation due to prosodic strengthening. While this effect has not been found systematically in previous studies (see the overview in Byrd et al., 2006), a number of studies do find it, e.g., diphthongs are hyperarticulated in utterance-final positions in English, Japanese, and Chinese (Zhang, 2022), and evidence that vowels are strengthened phrase-finally have been found in French (Tabain, 2003) and English (Fougeron & Keating, 1997). Furthermore, tonal f0 range is expanded in utterance-final position in Yoloxóchitl Mixtec (DiCanio et al., 2021). Therefore, a hyperarticulation effect could be reflected in lower f0 in utterance-final (i.e., stronger) boundary positions, possibly to enhance the effect of finality. However, we found no evidence for hyperarticulation within the kinematic parameters, such as ClosDisp, ClosDur, VelPeak. Another possible explanation for f0 lowering could be physiologically induced, as speakers may relax their respiratory support at the end of an utterance, i.e., if no further utterances are following, changing from speech to quiet breathing mode. This might affect the f0 declination leading to steeper declination slope and a lower f0 minimum at the end. However, we did not find any research that would substantiate this hypothesis. We will investigate the interplay between f0 and breathing in the near future.

For perception (research question 2), our prediction was that listeners will distinguish between utterance-medial and utterance-final boundaries, based on studies showing that listeners can distinguish between boundaries of different strengths (e.g., Gollrad, 2013; Krivokapić and Byrd, 2012; Petrone et al., 2017; Wagner and Crivellaro, 2010; Wightman et al., 1992), and that they can identify utterance-final boundaries (Geluykens & Swerts, 1994; Peters, 2006). And indeed, listeners judged stimuli produced in utterance-final position as conveying significantly less continuation than stimuli in utterance-medial position, confirming previous results on discourse-finality by, e.g., Herman (2000) and Peters (2006). However, the caveat remains that this result only held for five of eight speakers in the experiment. A further caveat is the holistic presentation of the stimuli, which allows listeners to draw on a range of cues, including syntactic and contextual cues as well as other acoustic cues (e.g., voice quality), to determine when the current speaker is about to conclude their utterance. In this sense, listeners formed judgments based on their overall impression of the stimuli. In light of the foregoing, it is advised that the results not be overinterpreted as reflecting a precise set of perceptual cues for utterance-medial versus utterance-final distinctions. However, the experiment has demonstrated that

listeners are able to distinguish between utterance-medial and utterance-final stimuli if the speakers in question have used acoustic cues to discern these stimuli (see also the discussion on the production-perception link below). It would be beneficial for future studies to adopt more focused designs that isolate individual prosodic elements or investigate listeners' attention to particular cues in a more controlled manner.

Regarding the link between production and perception (research question 3), we expected salient production parameters to have a more pronounced effect on perception than inconspicuous parameters. Confirming that prediction, a positive correlation was found between f0 (the parameter identified in the production study) and the continuation ratings, i.e., the higher the f0, the higher the rating for continuation. Five out of the eight speakers were distinguished by the listeners. These speakers also produced the final rhyme with a higher f0 in the utterance-medial position than in the utterance-final position. The other three speakers showed no significant difference in production (m5) or only for a very short period (f2 and m4), which seems to be too short for listeners to use as a cue. In terms of interindividual variation, our results suggest a strong link between production and perception, as listeners only discriminate between the two final positions for speakers that produce a difference in f0. As to why this interindividual variation exists, our experiment has too few participants to answer this question.

We further found a positive correlation between the kinematic parameter peak velocity and the continuation ratings, i.e., the faster the tongue movement, the higher the rating for continuation (indicating that a speaker might continue speaking). However, peak velocity is not used systematically by the speakers: it only differed significantly between the two final conditions for the target word *Ruhm* (see **Table 2**), and there was only a negligible effect in the reduced dataset for the perception stimuli. Therefore, we assume that listeners use this cue less consistently than mean f0.

To conclude, our study is one of the few studies to examine the production and perception of utterance-final boundaries. Our findings confirm previous studies on utterance-final lowering, both for production and perception. A lower f0 in the final rhyme of the utterance can signal discourse-finality and can be seen as evidence for hyperarticulation. The results for the temporal and spatial properties show no difference between utterance-medial and utterance-final boundaries. We suggest that this is the result of a structural and planning effect on lengthening combining at utterance-medial boundaries, compared to only a structural effect at utterance-final boundaries.

# Appendix

## Stimuli

| Target | Condition | Stimulus |
|---|---|---|
| Bahn | phrase-medial | *Ich fuhr mit der **Bahn** am Donnerstag. Am Mittwoch wurde noch gestreikt.* |
| | | 'I took the **train** on Thursday. On Wednesday, there was still a strike.' |
| | utterance-medial | *Ich fuhr mit der **Bahn**. Am Donnerstag musste ich in Frankfurt sein.* |
| | | 'I took the **train**. On Thursday, I had to be in Frankfurt.' |
| | utterance-final | *Ich fuhr mit der **Bahn**.* |
| | | 'I took the **train**.' |
| Bann | phrase-medial | *Der König verhängte einen **Bann** am Donnerstag. Am Mittwoch war er vorbei.* |
| | | 'The king declared a **ban** on Thursday. It was lifted on Wednesday.' |
| | utterance-medial | *Der König verhängte einen **Bann**. Am Donnerstag fing er an.* |
| | | 'The king declared a **ban**. It was lifted on Thursday.' |
| | utterance-final | *Der König verhängte einen **Bann**.* |
| | | 'The King declared a **ban**.' |
| Stiel | phrase-medial | *Sie kaufte Blumen mit dickem **Stiel** am Abend. Da sind sie oft im Angebot.* |
| | | 'She bought flowers with thick **stem**s in the evening. At that time, they are often on sale.' |
| | utterance-medial | *Sie kaufte Blumen mit dickem **Stiel**. Am Abend waren sie im Angebot.* |
| | | 'She bought flowers with thick **stem**s. In the evening, they were on sale.' |
| | utterance-final | *Sie kaufte Blumen mit dickem **Stiel**.* |
| | | 'She bought flowers with thick **stem**s.' |
| still | phrase-medial | *Das Kind war ganz **still** am Abend. Es war sehr müde.* |
| | | 'The child was very **quiet** in the evening. It was very tired.' |
| | utterance-medial | *Das Kind war ganz **still**. Am Abend war es sehr müde.* |
| | | 'The child was very **quiet**. In the evening, it was very tired.' |
| | utterance-final | *Das Kind war ganz **still**.* |
| | | 'The child was very **quiet**.' |

(Contd.)

| Target | Condition | Stimulus |
|---|---|---|
| Ruhm | phrase-medial | *Die TV Show brachte ihr viel **Ruhm** auf der Party. Sie wurde dort auch gezeigt.* |
| | | 'The TV show brought her a lot of **fame** at the party. It was also shown there.' |
| | utterance-medial | *Die TV Show brachte ihr viel **Ruhm**. Auf der Party wurde sie auch gezeigt.* |
| | | 'The TV show brought her a lot of **fame**. At the party it was also shown there.' |
| | utterance-final | *Die TV Show brachte ihr viel **Ruhm**.* |
| | | 'The TV show brought her a lot of **fame**.' |
| Rum | phrase-medial | *Sie tranken sehr viel **Rum** auf der Party. Es gab auch Whiskey.* |
| | | 'They drank a lot of **rum** at the party. There was also whiskey.' |
| | utterance-medial | *Sie tranken sehr viel **Rum**. Auf der Party gab es auch Whiskey.* |
| | | 'They drank a lot of **rum**. At the party there was also whiskey.' |
| | utterance-final | *Sie tranken sehr viel **Rum**.* |
| | | 'They drank a lot of **rum**.' |

**Table 7:** Stimulus sentences used in this experiment.

## *Additional tables*

| Target word | *β* | *SE* | df | *t* | *p* |
|---|---|---|---|---|---|
| Bahn | 0.9415 | 2.9353 | 65.15 | 0.321 | .7494 |
| Bann | 0.4909 | 2.8256 | 65.00 | 0.174 | .8626 |
| Ruhm | 0.5537 | 3.6810 | 65.32 | 0.150 | .8809 |
| Rum | –1.0254 | 3.1823 | 65.48 | –0.322 | .7483 |
| Stiel | 4.8380 | 2.9343 | 65.12 | 1.649 | .1040 |
| still | –1.1192 | 2.9353 | 65.15 | –0.381 | .7042 |

**Table 8:** Results of from post hoc tests, comparing the effect of prosodic position on VELPEAK per target word, based on 84 stimuli used for the perception experiment.
Degrees-of-freedom method: kenward-roger.

## Additional figures

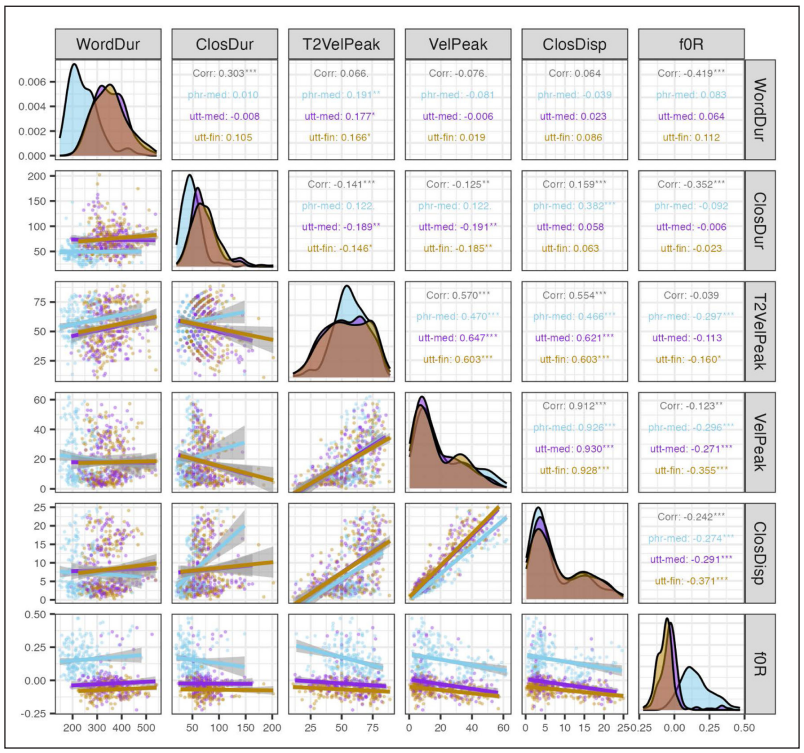


**Figure 13:** Scatterplots for the acoustic and kinematic parameters word duration WORDDUR (ms), duration of the closing movement CLOSDUR (ms), time to peak velocity T2VELPEAK (%), peak velocity VELPEAK, CLOSDISP (cm/s), and the residualized log f0 F0R, color-coded for condition: light blue phrase-medial, violet utterance-medial, brown utterance-final.

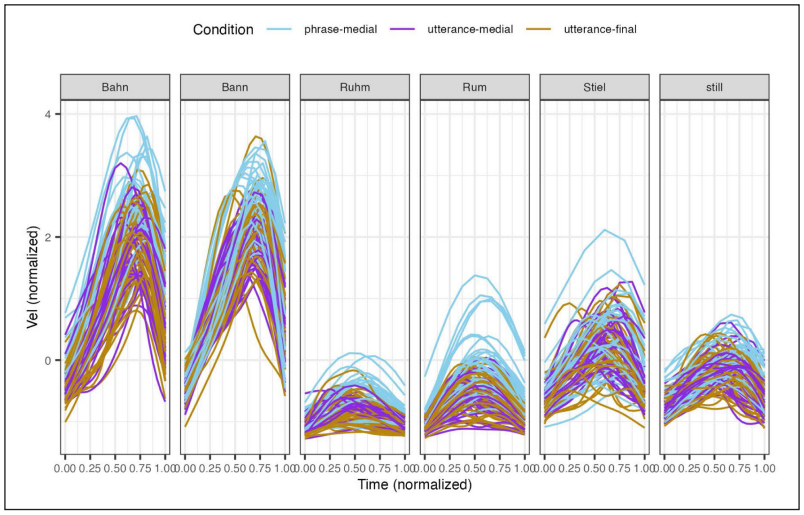


**Figure 14:** Time-normalized velocity profiles of the final closing gesture per word, colour-coded for condition.

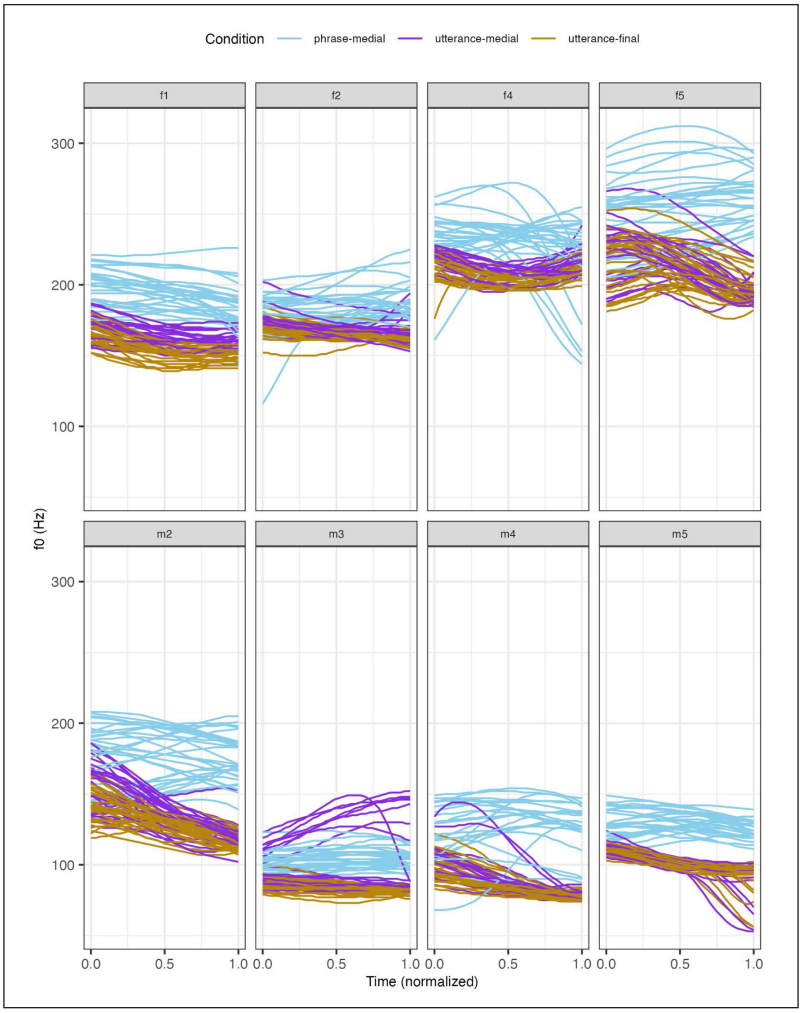


**Figure 15:** Time-normalized fundamental frequency contours during the rhyme per speaker, colour-coded for condition.

## Data availability statement

The aggregated data and R markdown scripts are available at https://osf.io/sa7vu/.

## Ethics and consent

The EMA recordings were approved by the ethics committee of the German Linguistic Society (Deutsche Gesellschaft für Sprachwissenschaft) in September 2014. Informed consent to participate in the study was obtained from all participants.

## Acknowledgements

We thank Melanie Weirich for assisting with the EMA recordings and stimuli, our undergraduate research assistents Anja Riemenschneider, Alina Zöllner, and Patricia Weber for their help

with data processing and labeling, Christoph Draxler for assisting us with the Percy platform, and Oksana Rasskassova for discussing previous versions of this work. We also thank Daniela Palleschi for help with statistical modelling and the interpretation of cumulative linked mixed models. Finally, we thank the editor and the two anonymous reviewers for their very helpful comments and suggestions.

## Competing interests

The authors have no competing interests to declare.

## References

Arvaniti, A. (2007). On the presence of final lowering in British and American English. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes. vol. 2: Experimental studies in word and sentence prosody* (pp. 317–347). Mouton de Gruyter. https://doi.org/10.1515/9783110207576.2.317

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, T. C., Neale, M. C., & Maes, H. H. (2019). Umx: A library for structural equation and twin modelling in r. *Twin Research and Human Genetics*, *22*, 27–41. https://doi.org/10.1017/thg.2019.2

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In M. E. Beckman & J. Kingston (Eds.), *Papers in Laboratory Phonology II* (pp. 68–86). Cambridge University Press. https://doi.org/10.1017/CBO9780511519918.004

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255. https://doi.org/10.1017/S095267570000066X

Belz, M., Rasskazova, O., Krivokapić, J., & Mooshammer, C. (2023). Interaction between phrasal structure and vowel tenseness in German: An acoustic and articulatory study. *Language and Speech*, *6*(1), 3–34. https://doi.org/10.1177/00238309211064857

Berkovits, R. (1984). Duration and fundamental frequency in sentence-final intonation. *Journal of Phonetics*, *12*(3), 255–265. https://doi.org/10.1016/S0095-4470(19)30882-4

Berkovits, R. (1993a). Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech*, *36*(1), 89–98. https://doi.org/10.1177/002383099303600105

Berkovits, R. (1993b). Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics*, *21*(4), 479–489. https://doi.org/10.1016/S0095-4470(19)30231-1

Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech*, *37*(3), 237–250. https://doi.org/10.1177/002383099403700302

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9), 341–345.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. https://doi.org/10.1016/j.wocn.2015.04.004

Bombien, L., Winkelmann, R., & Scheffers, M. (2020). *Wrassp: An r wrapper to the assp library* [R package version 0.1.9].

Brugos, A., Breen, M., Veilleux, N., Barnes, J., & Shattuck-Hufnagel, S. (2018). Cue-based annotation and analysis of prosodic boundary events. *Speech Prosody 2018*. https://doi.org/10.21437/SpeechProsody.2018-50

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*, *120*(3), 1589–1599. https://doi.org/10.1121/1.2217135

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands*, *14*(1), 13–24. https://doi.org/10.1075/avt.14.04cam

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155–190. https://doi.org/10.1006/jpho.2001.0131

Christensen, R. H. B. (2022). Ordinal—regression models for ordinal data [R package version 2022.11-16. https://CRAN.R-project.org/package = ordinal].

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Collier, R., de Pijper, J. R., & Sanderman, A. (1993). Perceived prosodic boundaries and their phonetic correlates. *HUMAN LANGUAGE TECHNOLOGY: Proceedings of a Workshop held at Plainsboro, New Jersey, March 21–24, 1993*. https://doi.org/10.3115/1075671.1075750

Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews: Cognitive Science*, *12*(3), e1547. https://doi.org/10.1002/wcs.1547

De Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, *96*(4), 2037–2047. https://doi.org/10.1121/1.410145

DiCanio, C., Benn, J., & Castillo García, R. (2021). Disentangling the effects of position and utterance-level declination on the production of complex tones in Yoloxóchitl Mixtec. *Language and Speech*, *64*(3), 515–557. https://doi.org/10.1177/0023830920939132

Draxler, C. (2017). PercyConfigurator-perception experiments as a service. *Proceedings of the INTERSPEECH*, 823–824.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283. https://doi.org/10.1037/h0033031

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233. https://doi.org/10.1016/0749-596X(91)90004-4

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 524–602). Blackwell. https://doi.org/10.1002/9781444317251.ch15

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740. https://doi.org/10.1121/1.418332

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Geluykens, R., & Swerts, M. (1994). Prosodic cues to discourse boundaries in experimental dialouges. *Speech communication*, *15*(1–2), 69–77. https://doi.org/10.1016/0167-6393(94)90042-6

Gollrad, A. (2013). *Prosodic cue weighting in sentence comprehension: Processing German case ambiguous structures* [Doctoral dissertation, Universität Potsdam].

Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, *25*(3), 601–634. https://doi.org/10.1016/j.csl.2010.10.003

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic Typology* (pp. 55–83). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0003

Guitard-Ivent, F., Turco, G., & Fougeron, C. (2021). Domain-initial effects on C-to-V and V-to-V coarticulation in French: A corpus-based study. *Journal of Phonetics*, *87*, 101057. https://doi.org/10.1016/j.wocn.2021.101057

Herman, R. (2000). Phonetic markers of global discourse structures in English. *Journal of Phonetics*, *28*(4), 466–493. https://doi.org/10.1006/jpho.2000.0127

Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 136–144. https://doi.org/10.3115/981131.981152

Hoole, P., & Mooshammer, C. (2002). Articulatory analysis of the German vowel system. In P. Auer, P. Gilles, & H. Spiekermann (Eds.), *Silbenschnitt und Tonakzente* (pp. 129–152). M. Niemeyer. https://doi.org/10.1515/9783110916447.129

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics*, *55*, 149–181. https://doi.org/10.1016/j.wocn.2015.12.003

Kentner, G., & Féry, C. (2013). A new approach to prosodic grouping. *Linguistic Review*, *30*(2), 277–311. https://doi.org/10.1515/tlr-2013-0009

Kentner, G., Franz, I., Knoop, C. A., & Menninghaus, W. (2023). The final lengthening of pre-boundary syllables turns into final shortening as boundary strength levels increase. *Journal of Phonetics*, *97*, 101225. https://doi.org/10.1016/j.wocn.2023.101225

Kim, J. (2020). *Individual differences in the production and perception of prosodic boundaries in American English* [Doctoral dissertation, University of Michigan]. https://doi.org/10.1121/1.5147830

Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via web services: The use case WebMAUS. *Proceedings Digital Humanities 2012*, 30–34. Hamburg, Germany.

Kohler, K. J. (1983). Prosodic boundary signals in German. *Phonetica*, *40*, 89–134. https://doi.org/10.1159/000261685

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*(2), 162–179. https://doi.org/10.1016/j.wocn.2006.04.001

Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 157–190, Vol. 1). Mouton de Gruyter.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical transactions of the Royal Society. Series B, Biological sciences*, *369*(1658). https://doi.org/10.1098/rstb.2013.0397

Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, *40*(3), 430–442. https://doi.org/10.1016/j.wocn.2012.02.011

Krivokapić, J., Styler, W., & Byrd, D. (2022). The role of speech planning in the articulation of pauses. *The Journal of the Acoustical Society of America*, *151*(1), 402–413. https://doi.org/10.1121/10.0009279

Ladd, D. R. (1988). Declination "reset"and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America*, *84*(2), 530–544. https://doi.org/10.1121/1.396830

Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.10.4]. https://CRAN.R-project.org/package = emmeans

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). MIT Press.

Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association*, *42*(3), 255–280. https://doi.org/10.1017/S0025100312000187

Lüdecke, D. (2018). Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Mooshammer, C., & Geng, C. (2008). Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association*, *38*(2), 117–136. https://doi.org/10.1017/S0025100308003435

Mücke, D., & Hermes, A. (2007). Phrase boundaries and peak alignment: An acoustic and articulatory study. *Proceedings of the 16th International Congress of Phonetic Sciences*, 997–1000.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America*, *54*(5), 1235–1247. https://doi.org/10.1121/1.1914393

Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, *94*, 101179. https://doi.org/10.1016/j.wocn.2022.101179

Peters, B. (2003). Multiple cues for phonetic phrase boundaries in German spontaneous speech. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1795–1798.

Peters, B. (2006). *Form und Funktion prosodischer Grenzen im Gespräch* [Doctoral dissertation, Christian-Albrechts Universität Kiel].

Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, *61*, 71–92. https://doi.org/10.1016/j.wocn.2017.01.002

Purse, R., & Krivokapić, J. (2023). The kinematic properties of prosodic boundaries in conversational turn-taking. *Proceedings of the 23rd International Congress of Phonetic Sciences*.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1). https://doi.org/10.5334/labphon.108

Savariaux, C., Badin, P., Samson, A., & Gerber, S. (2017). A comparative study of the precision of Carstens and Northern Digital Instruments Electromagnetic Articulographs. *Journal of Speech, Language, and Hearing Research*, *60*(2), 322–340. https://doi.org/10.1044/2016_JSLHR-S-15-0223

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A Prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. https://doi.org/10.1007/BF01708572

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). Tobi: A standard for labeling English prosody. *Second International Conference on Spoken Language Processing*. https://doi.org/10.21437/ICSLP.1992-260

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, *101*(1), 514–521. https://doi.org/10.1121/1.418114

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and speech*, *37*(1), 21–43. https://doi.org/10.1177/002383099403700102

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *The Journal of the Acoustical Society of America*, *113*(5), 2834–2849. https://doi.org/10.1121/1.1564013

Tiede, M. (2005). *MVIEW: Software for visualization and analysis of concurrently recorded movement data*. Haskins Laboratory.

Torchiano, M. (2020). *Effsize: Efficient effect size computation* [R package version 0.8.1]. https://doi.org/10.5281/zenodo.1480624

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). Itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs [R package version 2.4].

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (4th) [ISBN 0-387-95457-0]. Springer. http://www.stats.ox.ac.uk/pub/MASS4/

Wagner, M. (2005). *Prosody and recursion* [Doctoral dissertation, Massachusetts Institute of Technology].

Wagner, M., & Crivellaro, S. (2010). Relative prosodic boundary strength and prior bias in disambiguation. *Proceedings of the 5th Speech Prosody*. https://doi.org/10.21437/SpeechProsody.2010-250

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945. https://doi.org/10.1080/01690961003589492

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wightman, C. W., Shattuck–Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717. https://doi.org/10.1121/1.402450

Winkelmann, R., Harrington, J., & Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*. https://doi.org/10.1016/j.csl.2017.01.002

Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2016). emuR: Main Package of the EMU Speech Database Management System. Retrieved June 12, 2019, from https://rdrr.io/cran/emuR/

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge. https://doi.org/10.4324/9781315165547

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9781315370279

Zhang, M. (2022). *Prosodic influences on the acoustics of vowel sequences* [Doctoral dissertation]. https://www.proquest.com/dissertations-theses/prosodic-influences-on-acoustics-vowel-sequences/docview/2719126842/se-2