Journal of Phonetics 97 (2023) 101225

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

The final lengthening of pre-boundary syllables turns into final shortening as boundary strength levels increase

Gerrit Kentner[a,b,*], Isabelle Franz[b,c], Christine A. Knoop[b], Winfried Menninghaus[b]

[a] *Goethe University Frankfurt, Germany*
[b] *Max Planck Institute for Empirical Aesthetics, Germany*
[c] *Hochschule für Gesundheit Bochum, Germany*

ARTICLE INFO

Article history:
Received 2 March 2022
Received in revised form 16 January 2023
Accepted 23 January 2023
Available online 16 February 2023

Keywords:
Pre-boundary lengthening
Prosodic phrasing
Pauses
Prosodic boundary
Declination

ABSTRACT

Phrase-final syllable duration and pauses are generally considered to be positively correlated: The stronger the boundary, the longer the duration of phrase-final syllables, and the more likely or longer a pause. Exploring a large sample of complex literary prose texts read aloud, we examined pause likelihood and duration, pre-boundary syllable duration, and the pitch excursion at prosodic boundaries. Comparing these features across six predicted levels of boundary strength (level 0: no break; 1: simple phrase break; 2: short comma phrase break; 3: long comma phrase break; 4: sentence boundary; 5: direct speech boundary), we find that they are not correlated in a simple monotonic fashion. Whereas pause duration monotonically increases with boundary strength, both pre-boundary syllable duration and the pitch excursion on the pre-boundary syllable are largest for level-2 breaks and decrease significantly through levels 3 to 5. Our analysis suggests that pre-boundary syllable duration is partly contingent on the tonal realization, which is subject to f0 declination as the utterance progresses. We also surmise that pre-boundary syllable duration reflects differences in planning complexity for the different prosodic and syntactic boundaries. Overall, this study shows that a simple monotonic correlation between pause duration and pre-boundary syllable duration is not valid.

1. Introduction

Speech streams are divided into larger constituents (e.g., utterances, discourse units) that contain smaller constituents (sentences within discourse units or within a speaker's turn, phrases within sentences, words within phrases, syllables within words). This hierarchical grouping of constituents underlying the sequence of speech sounds manifests itself in the rendition of prosodic boundaries, with stronger boundaries set between the higher-order (and therefore often larger) constituents, and weaker boundaries (or no boundaries) set between the smaller embedded constituents (e.g., Féry, 2017; Gee & Grosjean, 1983; Ladd, 2008). The distribution and varying strength of the prosodic boundaries in spoken text reflect both coherence and division between speech chunks.

We have developed and validated a coding manual (Franz et al., 2022) that, based on written texts, guides the annotation of several degrees of prosodic boundaries that may be realized in the oral renditions of these texts. Using this tool, we generated predictions regarding boundary strengths based on the syntactic and lexical structure of written sentences and the punctuation in these sentences, irrespective of the acoustic wave form of the spoken sentences (the latter was used to test the predictions in Franz et al., 2022). In the present study, we use those predictions along with the read-aloud prose texts on which they were evaluated. Studying select aspects of the acoustic features of the orally presented narratives, we address the following questions: What are the phonetic means that signal a prosodic boundary at the end of a given constituent (e.g., word, phrase, sentence)? Which component of the signal indicates the hierarchical level of the prosodic boundary, i.e., whether the unit is phrase-final, sentence-final, or utterance-final? Pursuing these questions with a special focus on the relation between pauses, the pre-boundary syllable duration, and the pitch/f0 range of the pre-boundary syllable at the different levels of prosodic boundaries allowed us to test and qualify the widely held hypothesis of final lengthening, that is, the hypothesis that phonemes and syllables

---

[*] Corresponding author at: Goethe University Frankfurt, Germany.
E-mail address: kentner@lingua.uni-frankfurt.de (G. Kentner).

https://doi.org/10.1016/j.wocn.2023.101225
0095-4470/© 2023 The Authors. Published by Elsevier Ltd.

directly preceding a prosodic boundary are lengthened compared to syllables that are not followed by a boundary, and that the degree of lengthening increases as the strength of the boundary increases (Byrd & Krivokapić, 2021; Byrd & Saltzman, 2003; Gussenhoven, 1992; Krivokapić, 2014, 2022; Turk, 2012, among others).

1.1. Phonetic expression of prosodic boundaries

Prosodic boundaries are marked by a variety of phonetic cues. The most obvious phonetic expression of a prosodic boundary is a pause between two chunks of speech. Strong boundaries are generally likely to be realized with a pause, and pauses at strong boundaries tend to be longer than pauses at weak boundaries (Cooper & Paccia-Cooper, 1980, Gee & Grosjean, 1983). However, not all prosodic boundaries result in a pause. In addition to pauses, prosodic boundaries are generally marked by a decrease in loudness, by boundary tones (pronounced excursions of the pitch contour preceding a boundary), and by a slow-down in the speech rate approaching the boundary (see, e.g., Cruttenden, 1997; Kim, 2020; Ladd, 2008, and references therein). Thus, phonemes and syllables directly preceding a prosodic boundary are lengthened compared to phrase-internal syllables. This latter phenomenon is known as “final lengthening” or “pre-boundary lengthening” (Gussenhoven & Rietveld, 1992; Krivokapić, 2007, 2014, 2022; Turk & Shattuck-Hufnagel, 2007; White, 2014; Schubö & Zerbian, 2020; White, 2002, among many others). Byrd and Saltzman (2003) conceptualize pre-boundary lengthening as “prosodically induced local slowing” of articulatory movements. Some evidence suggests that, in addition to pauses, pre-boundary lengthening might also be a universal speech cue that indicates upcoming prosodic boundaries (Fletcher, 2010; Seifart et al., 2021; Vaissière, 1983). Under this assumption, pre-boundary lengthening can be conceived as an articulatory necessity during the production of phrase boundaries. However, while pre-boundary lengthening is found across many different languages, it appears to be implemented in language-specific ways (Paschen et al., 2022; see Cho, 2016, for a review). Pre-boundary or final lengthening can also be thought of as the linguistic equivalent to stylistic dynamic changes such as *ritardando* or *ralletando*, which are regularly observed before constituent boundaries in music (Liberman, 2007; Todd, 1985).

Pre-boundary lengthening and pausing have been argued to elicit a single percept of a prosodic break, as listeners report hearing pauses even when there is no acoustic evidence for a pause in the signal (Martin, 1970). Since pre-boundary lengthening and pauses often cooccur, it has been claimed (or at least implied) that these prosodic boundary cues are positively correlated. Correspondingly, pre-pausal syllables have been found to be longer than syllables not followed by a pause (Rao, 2010, among others). Research suggests that pause likelihood and pause duration increase with prosodic boundary strength (Cooper & Paccia-Cooper, 1980; Gee & Grosjean, 1984; Peters, Kohler, & Wesener, 2005; Petrone et al., 2017; Mayer, Jasinskaja, & Kölsch, 2006; Yang, 2004, among many others), and so do pre-boundary syllable durations (Byrd, 2000; Peters et al., 2005; Petrone et al., 2017; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992; Cambier-Langeveld, 2000; Liberman, 2007). According to Byrd and Saltzman’s (2003) articulatory model, lengthening increases with boundary strength: “the stronger the boundary, the greater the degree of [. . .] local slowing that occurs for the boundary-adjacent articulatory gestures” (p. 160). In the same vein, Gussenhoven (1992) states that “an utterance-final syllable will get the durational benefit of being final in a large number of [prosodic] constituents” (p. 94).

However, the relationship between the two timing-related prosodic boundary cues is still relatively unclear. While there is consensual evidence for a monotonic increase in pause duration with increasing levels of prosodic boundary strength (e.g., Gee & Grosjean, 1983; Swerts, 1997; Yang, 2004), the data on pre-boundary lengthening are less straightforward. Wightman et al. (1992) report that, on their juncture scale that ranges from 0 (no break) to 6 (sentence boundary), pre-boundary phoneme and syllable rime durations systematically increase only up to index 4 (a sentence-internal phrase boundary) and that any further increase in segment duration at the boundary remains insignificant. Similarly, the data presented in Peters et al., 2005, p. 151, Fig. 1) that were gleaned from a corpus of quasi-spontaneous, unscripted German speech do not suggest that syllables at the end of a speaker’s utterance are systematically longer than utterance-internal phrase-final syllables; in a Spanish sample, Rao (2010) finds that, while pre-boundary syllables are generally lengthened before pauses, long pauses do not afford more pre-boundary syllable lengthening than short pauses.

Several other studies consider more than two levels/degrees of prosodic boundary strength as predictors of differences in pre-boundary lengthening (Byrd, 2000; Byrd & Saltzman, 1998; Ladd & Campbell, 1991 for English; Horne, Strangert, & Heldner, 1995 for Swedish; Cambier-Langeveld, 2000 for Dutch and English; Tabain, 2003; Tabain & Perrier, 2005; Michelas & D’Imperio, 2010, 2012 for French; Katsika, 2016 for Greek). These studies have investigated the phonetic expression of prosodic boundaries at various levels of the prosodic hierarchy in laboratory settings. The results provide some evidence that pre-boundary lengthening (and pause duration, in the case of Cambier-Langeveld, 2000) generally increases with prosodic boundary strength. In analyses of spontaneous speech data, Yang (2004) shows that pre-boundary lengthening is strongest for the final syllable before a pause. However, some studies also show that not all phonemes of the pre-boundary syllable or word are equally affected, with some phonemes apparently shortened rather than lengthened before stronger boundaries (e.g., Horne et al., 1995; Katsika, 2016). In addition, Betz and Wagner (2016) find that the last phoneme before a pause is affected most by pre-boundary lengthening.

There is also evidence suggesting that lengthening and pausing are not only correlated but in fact interact. Some researchers assume *cue trading*, where shorter pauses that provide relatively weak signals for a particular boundary are associated with more pronounced final syllable lengthening (Brugos, 2015, citing Fant & Kruckenberg, 1994). Conversely, relatively little lengthening would be expected when a longer pause suffices to signal boundary strength (Yang, 2007). In a similar vein, Ferreira (1993) has shown that pauses tend to be longer after words with inherently short vowels, which cannot easily be lengthened, than pauses after words with long, and hence more elastic, vowels in the same position. However, it is largely unknown whether such cue trading occurs across various levels of prosodic boundary strength. Kim (2020) also reports on speaker-specific differences regarding the way a given prosodic boundary is signaled, with some speakers using pre-boundary lengthening to a stronger extent than others.

1.2. The interplay of prosodic boundary cues: pre-boundary lengthening and pitch excursion

For listeners, pauses and pre-boundary lengthening together signal different degrees of cohesion or division between speech chunks, and they help in reconstructing the hierarchical organization of a linear string of syntactic and discourse constituents (Frazier et al., 2006). For speakers, on the other hand, pauses offer time to breathe. Speakers also use pauses to plan upcoming phrases and sentences (Ferreira, 1991; Krivokapić, 2014; Krivokapić et al. 2020; Krivokapić et al. 2022). Hesitations and filled pauses have been shown to be more likely the more complex the upcoming material is (or is expected to be; Arnold et al., 2007). Accordingly, both the length and complexity of preceding as well as upcoming material predict pause duration (Ferreira 1991; Kentner, 2007; Kentner & Féry, 2013; Krivokapić, 2007, Watson & Gibson, 2004; see Fletcher, 2010 for a review).

Apart from reflecting planning or production complexity, syllable lengthening might also be a corollary of intonational events. Prosodic boundaries are known to be accompanied by relatively pronounced pitch excursions, so-called *boundary tones* that are realized on the final syllables of prosodic or intonational phrases (Cruttenden, 1997; Ladd, 2008). Klatt (1975) suggests that increasing the final duration of segments enables a more effective manifestation of pitch excursions or boundary tones that signal prosodic boundaries of various strengths. Boundary tones can have rising or falling contours. Rising contours typically indicate that a given passage is not yet completed (e.g., Bolinger, 1989; Michalsky, 2017; Tyler, 2014). These contours are therefore called “continuation rises,” and they are typically produced at prosodic boundaries within a sentence. Falling contours, on the other hand, are typically found on syllables that end a declarative sentence or a paragraph. Since producing pitch excursions takes time, syllables bearing boundary tones are expected to be longer than syllables not bearing such tones. In addition, rising contours have been found to be significantly longer in duration than falling ones (Myers, 2003; Ohala & Ewan, 1973; Sundberg, 1979; Xu & Sun, 2002). Moreover, the pitch range that speakers can exploit for producing tonal events decreases as the sentence or utterance progresses (the so-called *declination* effect; see Cohen, Collier, & ‘t Hart, 1982; Collier & Gelfer, 1983; Ladd, 1984). Correspondingly, prosodic boundaries that are produced early on in a given sentence may be marked by greater pitch excursions (which take longer to execute) than sentence-final boundaries.

Based on these assumptions, one might expect that the syllables preceding sentence-internal prosodic boundaries—which are typically associated with rising contours—exhibit longer durations than those that close off a sentence or paragraph. Additionally, on this account, tonal differentiations of boundary strengths might be expected to be more reliably signaled in early positions of the sentence than in later positions. However, it is not yet known what effects the declination-related or declination-induced restriction of tonal excursions may have on the duration of the respective syllables. It has been posited that syllable lengthening at prosodic boundaries is a corollary of the production of boundary tones (Klatt 1975). Lindblom (1968) suggests that pre-boundary lengthening is a compensatory signal for the decrease in loudness and pitch towards the end of a sentence. If that is the case, we would expect longer pre-boundary syllable durations in sentence-final positions (which are more strongly affected by declination) than in sentence-initial or sentence-medial positions. On the other hand, as suggested above, pre-boundary syllable durations and tonal excursions might be interdependent and therefore undulate in tandem over the course of the utterance.

In light of these assumptions, we compared the phonetic realization of prosodic boundaries that have various degrees of boundary strength and specifically studied the relationship between pre-boundary lengthening and pause durations. In addition, we examined the tonal realization of pre-boundary syllables. We analyzed a corpus of four prose texts, each read aloud by eight professional speakers (Franz et al., 2022).

2. Methods

The corpus as well as parts of the present analysis are based on, and therefore partially coincide with, those we report in Franz et al. (2022). We indicate where we diverge from that model. Our re-use of the same speech sample that was earlier harnessed for validating the boundary predictions is mainly motivated by convenience: The fairly time-consuming coding process was already completed for these texts, and the data we had gathered for the purpose of an independent study were readily available. In the present study, we use this sample to examine the specific relationship between two phonetic boundary cues, pausing and pre-boundary syllable duration, as well as the effects of boundary strength on the f0 excursions for the pre-boundary syllables. We did not elaborate on these issues in the earlier validation study. The prose texts with annotations, the acoustic data extracted from the speech recordings, and the analysis script are available at the Open Science Framework repository (https://osf.io/6hwtf/).

The speech samples under examination, that is, the prose texts that are each read aloud by eight speakers, are the opening pages of four famous German novels or novellas from the 19th and early 20th century (*Effi Briest* by Theodor Fontane, *Die Wahlverwandschaften* by Johann Wolfgang von Goethe, *Der Prozess* by Franz Kafka, and *Die Marquise von O.* by Heinrich von Kleist).

As instances of literary prose, the linguistic forms of the texts differ significantly from ordinary and spontaneous language use, especially with respect to features such as lexis and syntactic complexity. While all four texts constitute artful prose, they vary considerably in terms of formal complexity and style. As a coarse indicator of this variation, Table 1 indicates the variations with respect to the range and the average number of syllables per sentence for the different texts.

Table 1
Descriptive statistics for the four prose texts.

Author	Fontane	Goethe	Kafka	Kleist	Overall
Number of sentences examined	99	101	98	76	374
Number of words examined	1503	1544	1816	1727	6590
Number of syllables examined	2599	2672	3013	3080	11,364
Number of syllables per sentence					
Minimum	1	3	3	4	1
Maximum	134	100	170	119	170
Mean (*SD*)	26.3 (29)	26.5 (19)	30.7 (29)	40.6 (22)	30.4 (26)
Median	17	20	21.5	37.5	22.5
Mode	11	8	10	46	8

2.1. Recordings

We reanalyzed recordings of excerpts from the four novels by Fontane, Goethe, Kafka, and Kleist; each of the excerpts were read by eight professional speakers (four male, four female). At the time of the recordings, all speakers were enrolled in a university program for rhetoric and professional stage reading. Each speaker was seated in a sound-attenuated recording booth with a Neumann U87 Ai Studio microphone placed ∼ 30 cm from their mouth and read the four text excerpts aloud. All of the speakers were familiar with the texts and had diligently prepared the reading before the recordings took place. The excerpts contained roughly 1500–1800 words each, with the lengths of the excerpts ranging from 1503 (Fontane) to 1816 words (Kafka). In total, we analyzed roughly-six hours of speech in which the eight readers produced more than 90 000 syllables.

2.2. Data preparation

When slips of the tongue or disfluencies occurred during the recordings, the speaker was asked to re-read the affected sentence or paragraph, starting at least one sentence before the faulty passage. The defective stretches of speech were later erased and replaced by the corrected renditions. The durations of the original pauses that preceded the to-be-corrected passages were preserved. We then used Web-MAUS (Kisler et al., 2012) to automatically segment the cleaned recordings into words and syllables. A number of Web-MAUS routines were employed for this purpose, with the original written text and the sound recordings as input. First, we used grapheme-to-phoneme conversion (G2P) to create a word-by-word phonemic transcript. We then used phoneme-to-syllable assignment (PHO2SYL) to determine phonemic syllables that, together with the individual phonemes, we mapped onto the sound and force-aligned. A small number of manual corrections of the MAUS segmentation were necessitated when the G2P routine produced implausible transcripts; these involved several rare words (e.g., *verleumdet*, “slandered,” or place names like *Hohencremmen*) and words with obsolete orthographies (e.g., *ergetzt*, “regaled”; current standard orthography: *ergötzt*). The WebMaus procedure’s assignment of individual phonemes to syllables was left unchanged as long as the number of syllables was correct. This means that ambisyllabic consonants after short lax vowels (e.g., the [m] in [tsɪmɐ] *Zimmer*, “room”) were in most cases associated with the first syllable, leaving the following syllable without an onset ([tsɪm.ɐ]). The durations of all annotated syllables as well as all pauses were extracted using praat, version 6.1.21 (https://www.fon.hum.uva.nl/praat/).

We used an awk script to align each spoken syllable with the citation-form syllable that had been annotated for predicted phrasing and prominence in Franz et al.’s (2022) coding manual. The predicted values for citation-form syllables without a spoken equivalent were ignored. The syllables to which this mainly applied were unstressed, such as the first syllable of *Geschenk* (“present,” citation form [gə.ʃɛŋk], realized as monosyllabic [kʃɛŋk]) or the last syllable of *haben* (“have,” citation form [ha:.bən], realized as monosyllabic [ha:m]). For the present study, we also discarded the final syllables and following pauses (if applicable) of questions, as interrogativity has been shown to affect syllable duration above and beyond the effect of pre-boundary lengthening (Michalsky, 2017). The four prose texts contained 35 questions, and as these were read by eight speakers each, 35 × 8 = 280 of the 2992 sentence-final word transitions (∼9%) were discarded.

2.3. Determination of pauses

Silent intervals between words that were detected by the automatic speech segmentation system (>50 ms) were treated as pauses. We are not aware of any agreed-upon threshold for the duration of phonetic silence that counts as an actual pause. While Trouvain (2003) uses a threshold of 100 ms and Zellner (1994) suggests a value of around 200 ms for clearly perceivable pauses, de Pijper and Sandermann (1994, p. 2043, Table III) show that even relatively short silent intervals (<100 ms) may be perceived as pauses. We therefore decided to include all silent intervals that passed the threshold of 50 ms as pauses in our analysis. We are aware that some of the silent intervals that are shorter than 100 ms may have resulted from the occlusion of articulators during the production of plosives and hence may not constitute genuine pauses. Our analysis does not consider the distributions of plosive sounds at the beginning of words; however, we have no reason to believe that these systematically covary with the different levels of boundary strength, so they are unlikely to confound our results.

2.4. Coding of boundaries and pre-boundary syllables

To predict the location and strength of prosodic boundaries, we employed Franz et al.’s (2022) coding system for the annotation of expectable prosodic features (prominence and phrasing) in written prose texts. This system distinguishes five degrees of boundary strength based on orthographic (punctuation) and (morpho-)syntactic features of the text. Apart from simple word transitions that are not assumed to afford prosodic breaks, Franz et al.'s system distinguishes four degrees of boundary strength, with the strongest boundary being a sentence boundary (level 4, determined by sentence-final punctuation marks). Commas in general demarcate weaker breaks. Specifically, commas preceded by three or more words were coded as level 3 (long comma phrase) breaks, whereas short comma phrases (with only one or two preceding words) were coded as level 2 breaks. Level 1 boundaries are those word transitions that do not come with punctuation marks but are still predicted to afford prosodic boundaries. To predict level 1 boundaries, we used the topological model for German sentence structure (Wöllstein, 2014). The topological model distinguishes core parts of the sentence (the so-called sentence bracket that embraces the middle field) from peripheral ones (pre-field and post-field). In canonical declarative sentences, the inflected verb marks the beginning of the core part, which is generally preceded by a syntactic constituent that represents the pre-field. If the pre-field contains two or more words of which at least one is a content word (noun, verb, adjective, or adverb), a level 1 boundary was set before the tensed verb. Similarly, constituents that follow the main clause and may be introduced by conjunctions, prepositions, or comparative conjunctions afford a level 1 break if these constituents are longer than four syllables and contain a content word. Level 1 breaks were also set at constituent boundaries within the above-mentioned topological fields (pre-field, middle field or post-field) when these exceeded a maximum of either 12 syllables or five words (see Franz et al., 2022, for details).

For the present study, we additionally coded whether specific sentence boundaries mark the boundary of direct speech, i.e., a transition from the narrator's voice to the direct speech of a figure and vice versa, or a transition from the direct speech of one figure in the text to the direct speech of another figure. These sentence boundaries were coded as level 5 breaks. We expected these breaks to be stronger than the level 4 breaks because they not only mark the end of a single sentence but also of an entire speech act—which may, moreover, consist of more than one sentence.

All other word transitions were assumed not to involves prosodic break and were therefore considered neutral (level 0). Notably, level 0 in the present analysis includes both the cliticizations or "linkages" and the neutral word transitions that Franz et al. (2022) distinguished as level -1 and level 0, respectively, as this additional distinction was not relevant for the purposes of the present study's research question.[1]

Use of Franz et al.'s (2022) coding system allowed us to study the length of pre-boundary syllables while controlling for syllable prominence. Here we ignored several minor distinctions that are assumed in Franz et al.'s annotation of syllable prominence. Specifically, Franz et al. distinguish nine predicted degrees of syllable prominence, which are mainly computed by adding values for lexical stress (primary stress > secondary stress > unstressed with full vowel > reduced/schwa syllable) to values for the part of speech (nouns > adjectives > verbs > auxiliary verbs and particles > other function words), along with values for certain syntactic configurations that predict phrasal and sentence accents (for details, please refer to the coding manual). However, some of these predicted differences turned out to lack consistent phonetic correlates (e.g., the difference between the two highest degrees of prominence). Therefore, and for reasons of simplicity, we adopted a categorization for the present study that merely distinguishes (a) unstressed syllables, (b) syllables predicted to have lexical stress but no phrasal accent, and (c) stressed syllables predicted to bear phrasal or nuclear accents, according to the coding manual.

2.5. Validation

Franz et al. (2022) evaluated the annotated features in the written texts against the actual read-aloud renditions. To this end, the duration, f0 excursion, and loudness of each syllable, along with the duration of each pause, were extracted from the read-aloud versions of the prose texts, and the phonetic measures were compared to the annotations of the written texts. This evaluation validated the originally assumed degrees of boundary strength in the written texts as corresponding to systematically increasing degrees of pause duration in the read-aloud versions. Franz et al.'s phonetic analysis also shows that pre-boundary syllable duration does not increase in tandem with pause duration across all degrees of boundary strength. Using a slightly revised version of the annotation protocol, the present study follows up and elaborates on this finding.

3. Results

The results in Sections 3.1 and 3.2 partially overlap with findings reported in Franz et al. (2022). Here, we add the results for direct speech boundaries (level 5, which are not considered in Franz et al., 2022). Additionally, we examine in detail the correlation between pauses and pre-boundary syllable duration in various conditions in Section 3.3. In Section 3.4, we report on the f0 range for the pre-boundary syllable. In Section 3.5, we examine the effect of declination by comparing the pre-boundary syllable duration and the f0 range for syllables in early parts of the sentence with those of syllables in later parts.

3.1. Likelihood of pauses and pause duration at word transitions

The likelihood and duration of pauses increase with each degree of boundary strength, as shown in Table 2 and Fig. 1. Table 2 shows that pauses were registered at only 3.9% of the neutral (level 0) word transitions, while 98.3% of the direct speech (level 5) boundaries afforded a pause, with pause likelihood increasing monotonically along the phrasing scale. Fig. 1 shows the pause durations (values rescaled to a log-transformed y-axis) broken down by the six degrees of the phrasing scale, using the median duration as the average marker.

We used two mixed effects models (Bates et al., 2015) in the statistical computing environment R (R Core Team, 2020) to examine what the numbers in Table 2 and visual inspection of the plot suggest, namely, that the likelihood and durations of pauses increase monotonically along the boundary strength scale, and the distributions of the pause durations are signifi-

---

[1] Of all neutral word transitions, roughly 53% are of the linkage or cliticization type. These word transitions are assumed not only to be not conducive to prosodic breaks, but also to reject prosodic breaks. Typical examples are the transitions between a determiner and a noun (e.g., *das Haus*, "the house"), or between a verb and a pronominal object (e.g., *holt es*, "bring it").

Table 2
Number of transitions and frequency of pauses at each degree of boundary strength.

Level	0 Neutral	1 Simple phrase break	2 Short comma phrase	3 Long comma phrase	4 Sentence boundary	5 Direct speech boundary
N of transitions	38,505	4472	1216	5472	2064	712
N of pauses	1516	404	471	2484	1523	700
Percentage of pauses	3.9%	9.0%	38.7%	45.4%	73.8%	98.3%

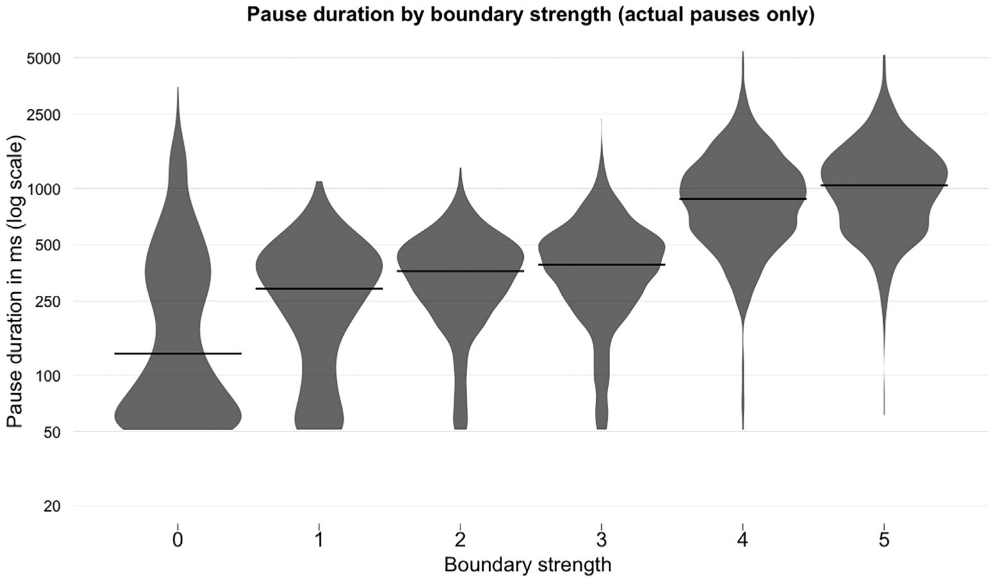

Fig. 1. Violin plot with pause durations in ms (log scale), broken down by the boundary strength index. The average line represents the median duration. The plot was created with the yarrr-package in R, version 0.1.5 (https://CRAN.R-project.org/package=yarrr). This plot shows only the data for actual pauses, that is, word transitions at which pauses were detected (word transitions without pauses are ignored).

cantly different from one level to the next. The first model is a generalized linear mixed model that takes the presence (coded as 1) vs absence (0) of a pause at each word-to-word transition in the prose texts as the dependent variable. The second model is a linear model that uses only the durations of actual pauses as the dependent variable (i.e., boundaries without pauses are excluded). As pause duration distributions are known to be skewed (e.g., Campione & Véronis, 2002), we used log-transformations to better fit the model assumptions. In both models, the contrasts between the boundary strength indices (levels 0–5), coded as successive differences contrasts, were entered as the main effect of interest. We also included several covariates that may affect pause durations: the length of the preceding and following words (both measured as the number of phonemes in the citation form of the word), and the duration (in ms, log-transformed) of the word-final syllable at the word transition. The text and author (*n* = 4) and the speaker (*n* = 8) were included as grouping terms (so-called random intercepts). The two models' coefficients of interest (all fixed effects) are presented in Table 3.

Not only do the two models confirm the significant stepwise monotonic increase in pause likelihood and duration along the boundary strength scale, but a comparison of the two models also reveals an interesting effect of the pre-boundary syllable duration: While there is a relatively strong positive effect of the pre-boundary syllable duration on the pause likelihood, the effect of the pre-boundary syllable duration on the pause duration is smaller and, crucially, negative. This finding motivated further investigation of the pre-boundary syllable duration in relation to the various levels of boundary strength (in Section 3.2) and to pauses (in Section 3.3).

Table 3
Coefficients of interest for the two models.

Generalized linear mixed model for the likelihood of a pause:

Coeff.	Estimate	*SE*	*z*-value	*p*-value
Boundary 1–0	0.47995	0.06243	7.688	<0.001
Boundary 2–1	1.92526	0.08486	22.689	<0.001
Boundary 3–2	0.29769	0.07041	4.228	<0.001
Boundary 4–3	1.38002	0.06334	21.789	<0.001
Boundary 5–4	2.75979	0.29895	9.232	<0.001
Preceding word length	0.10813	0.00610	17.724	<0.001
Pre-boundary syll. duration	1.13853	0.03951	28.815	<0.001
Following word length	−0.16792	0.00810	−20.72	<0.001

Linear mixed model for the pause duration (transitions with actual pauses only):

Coeff.	**Estimate**	***SE***	***t*-value**
Boundary 1–0	0.279221	0.040935	6.821*
Boundary 2–1	0.335604	0.04905	6.842*
Boundary 3–2	0.124075	0.03609	3.438*
Boundary 4–3	0.830988	0.02408	34.507*
Boundary 5–4	0.159484	0.03382	4.716*
Preceding word length	0.021113	0.00293	7.198*
Pre-boundary syll. duration	−0.044063	0.01842	−2.392*
Following word length	−0.058392	0.00413	−14.127*

Note. * t-values > |2| indicate that the alpha-level, i.e., the risk of falsely rejecting the null hypothesis, is smaller than 0.05.

3.2. Pre-boundary syllable durations

As the left panel of Fig. 2 shows, compared to word-final syllables at neutral word transitions (boundary strength index 0, i.e., those with no prosodic boundary predicted), syllables preceding prosodic boundaries (boundary strength indices 1–5) have a longer duration. However, the increase in the pre-boundary syllable duration does not correlate monotonically with the predicted boundary strength. Instead, while the word-final syllable durations increase from level 0 to level 2 (small comma phrase), they decrease from level 2 to level 5. This nonmonotonic pattern holds for predicted breaks both without (Fig. 2, left panel) and with a following pause (Fig. 2, right panel): Syllables preceding word transitions without a pause are generally shorter than pre-pausal syllables. When a pause follows, syllables at neutral word transitions (boundary strength index 0) appear, on average, to have durations that are as long as or even longer than the durations of syllables at the strongest prosodic boundaries (sentence and direct speech boundaries, indices 4 and 5, respectively). It is important to note that there are relatively few data points with pauses at boundary level 0 and hardly any data points ($n = 12$) without a pause at boundary level 5 (see Table 2 for details).

We employed a linear mixed model with word-final syllable durations as the (log-transformed) dependent variable. The contrasts between the boundary indices (levels 0–5), coded as successive differences contrasts, were the main effects of interest. The word length of the current and the following word (both measured by the number of phonemes in the citation form of the word) and the pause duration (log transformed when > 0) were entered as covariates. Again, the text and author and the speaker were used as grouping variables (random intercepts). This model confirms significant differences between all successive boundary strength indices (Table 4). Crucially, while the coefficients for the contrasts between levels 1 and 0 and between levels 2 and 1 are positive, the remaining contrasts yield negative coefficients. Apart from the boundary strength scale, the pause duration affects the pre-boundary syllable duration such that pre-boundary syllables preceding a pause have a longer duration than pre-boundary syllables that do not precede a pause.

Table 4
Linear mixed model for word-final syllable durations.

Coeff	Estimate	*SE*	*t*-value
Boundary 1–0	0.1158243	0.0074657	15.514*
Boundary 2–1	0.1477854	0.0151771	9.737*
Boundary 3–2	−0.0692896	0.0146834	−4.719*
Boundary 4–3	−0.1505582	0.0123752	−12.166*
Boundary 5–4	−0.050882	0.020556	−2.475*
Word length	−0.001344	0.0008084	−1.662
Pause duration (log)	0.037218	0.0013146	28.312*
Following word length	−0.0031857	0.0007804	−4.082*

Note. * t-values > |2| indicate that the alpha-level, i.e., the risk of falsely rejecting the null hypothesis, is smaller than 0.05.

3.3. Correlation of pause duration and pre-boundary syllable duration

The scatterplots in Fig. 5 show the correlations between the pause durations (x-axis) and the durations of word-final syllables (y-axis). The left panel shows all word-final syllable durations, including word transitions without pauses (indicated by pause duration = 0 on the x-axis); the right panel shows data for only the word transitions with actual pauses. Apart from the linear correlation (dashed line in Fig. 3), we fitted a generalized additive model (GAM) to the data, using the default parameters of ggplot2 (Wickham, 2016). GAMs are designed to find a smooth response function instead of simple linear model coefficients (intercept and slope). To do this, GAMs use piecewise cubic splines to model nonlinear relationships between the two variables. When all data points are consid-

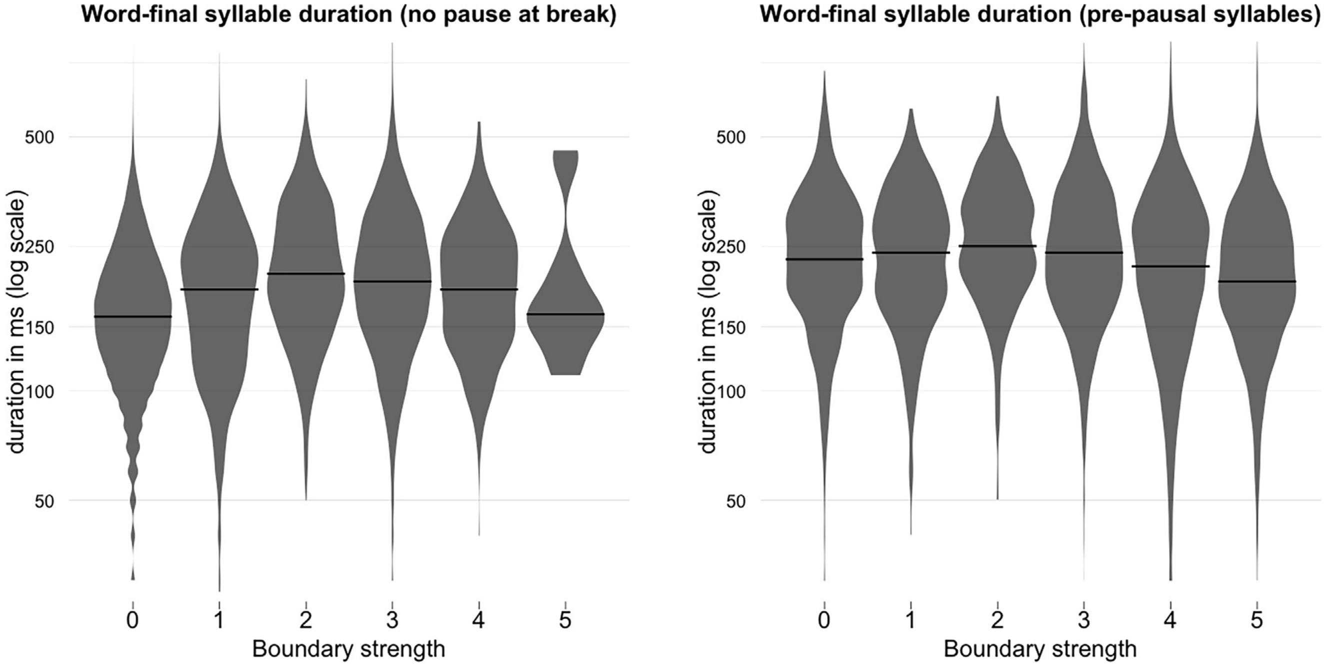

Fig. 2. Violin plot for word-final syllable durations in ms (log scale), broken down by boundary strength index. The left panel shows a violin plot for the subset of data points after which no pause was registered. The right panel shows a violin plot for the subset of data points with a following pause. Note that for boundary strength value 5, there are only 12 data points without a following pause (left panel). The remainder of the violin plots represent more than 400 data points (see Table 1).

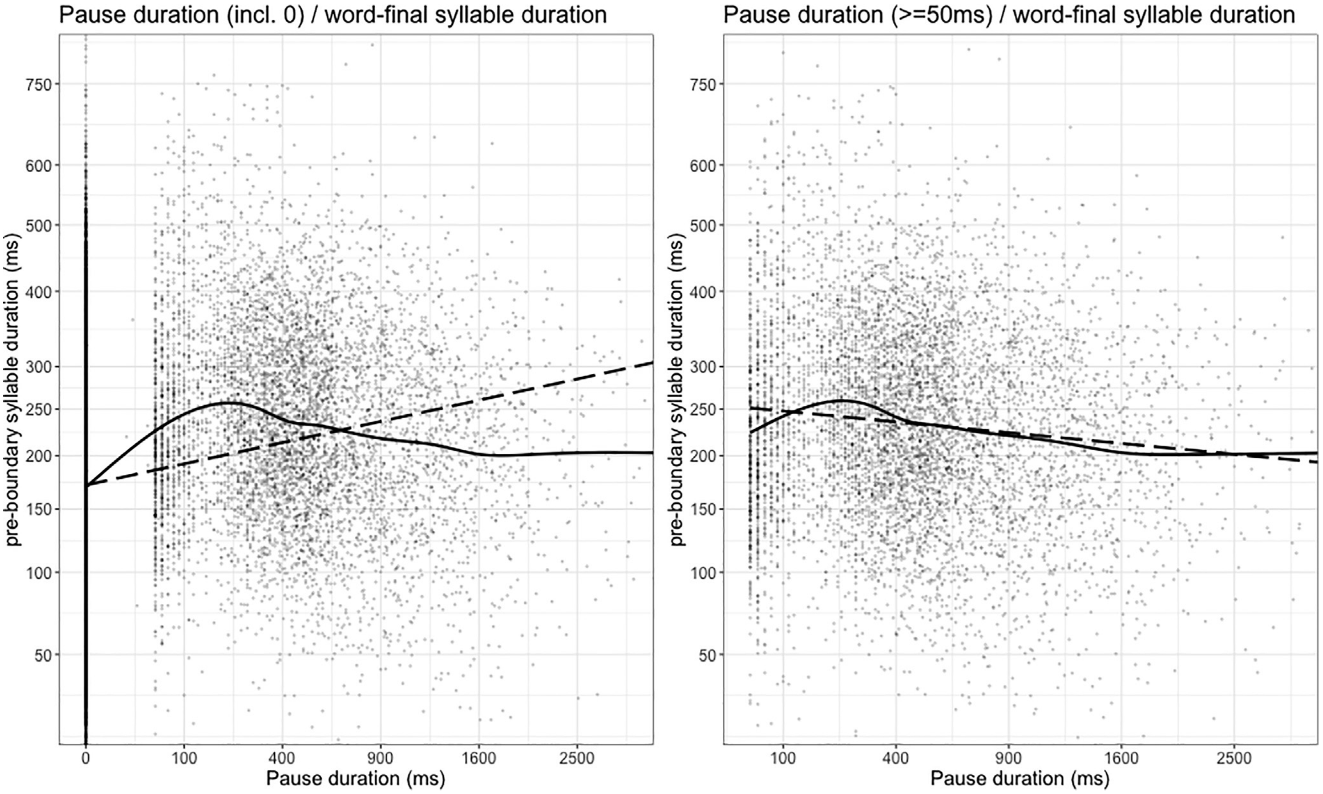

Fig. 3. Scatterplots showing the correlation between the pause duration (x-axis) and pre-boundary syllable duration (y-axis). The left panel shows all relevant data points (including syllables not followed by a pause). The right panel shows the subset of data points that are followed by actual pauses. The dashed line depicts a linear correlation, while the black curve shows a GAM smoother.

ered (including word transitions without pauses), a comparison with the GAM model (black curve) reveals that a simple linear correlation (dashed line) between the two variables underestimates word-final syllable lengthening before short pauses but overestimates the lengthening before long ones. The GAM smoother suggests that the strongest pre-boundary syllable lengthening effect is found at pauses ranging between 150 and 300 ms. Longer pauses imply distinctly weaker lengthening effects. In fact, as the right panel shows, when only pre-pausal syllables are considered, the GAM (black curve) mostly fits a negative linear correlation (dashed line), at least for pauses > 300 ms; a positive correlation of the pause duration and pre-pausal syllable lengthening can be observed for pauses up to 300 ms. The strongest lengthening effect is observed when word-final syllables without following pauses (minimum value of the GAM smoother at pause duration = 0, left panel of Fig. 3) are compared to syllables preceding short pauses (peak of the GAM smoother).

This general nonmonotonic pattern appears to be stable across a number of conditions. First, all eight speakers follow this general pattern (Fig. 4, left panel, with separate GAM smoothers fit for each reader/speaker). Second, all four texts have similar correlations of the pause duration and the pre-boundary syllable duration (Fig. 4, right panel, with separate GAM smoothers fit for each text). We note that for very long pauses (>1500 ms), the GAM smoothers for the different texts diverge ostensibly. However, these pauses constitute only a small percentage of the data points (1% of all word transitions, about 5% of all word transitions with a pause).

Third, the nonmonotonic relationship can also be observed across different levels of the word-final syllable's prominence. It appears for the generally shorter unstressed syllables (Fig. 5, left panel), for unaccented stressed syllables (Fig. 5, middle panel), and for syllables bearing a phrasal accent (Fig. 5, right panel).

Finally, the GAM smoothers produce comparable nonmonotonic, inverted u-shaped patterns for word-final syllables that differ in terms of the number of spoken phonemes that make up that syllable (Fig. 6, left panel), and also for word-final syllables of words that differ in terms of the number of syllables (Fig. 6, right panel). The plot in the left panel of Fig. 6 suggests that the number of phonemes affects the overall duration of the syllable but has hardly any effect on the correlation with pause duration. The word length (measured by number of syllables) also does not have a strong effect on the shape of the correlation. However, for monosyllabic words, the syllable duration is clearly longer. This is likely due to the fact that the final and only syllable of monosyllabic words usually bears stress, whereas word-final syllables of polysyllabic words tend to be unstressed. Moreover, the peak of the correlation with pause duration appears to be protracted by more than 100 ms for monosyllabic words compared to polysyllabic ones.

3.4. Pitch range of pre-boundary syllables

To explore the relationship between the predicted boundary strength, pre-boundary syllable duration, and potential tonal events in the vicinity of the boundaries, we measured the pitch range for each word-final syllable. To do this, we computed the

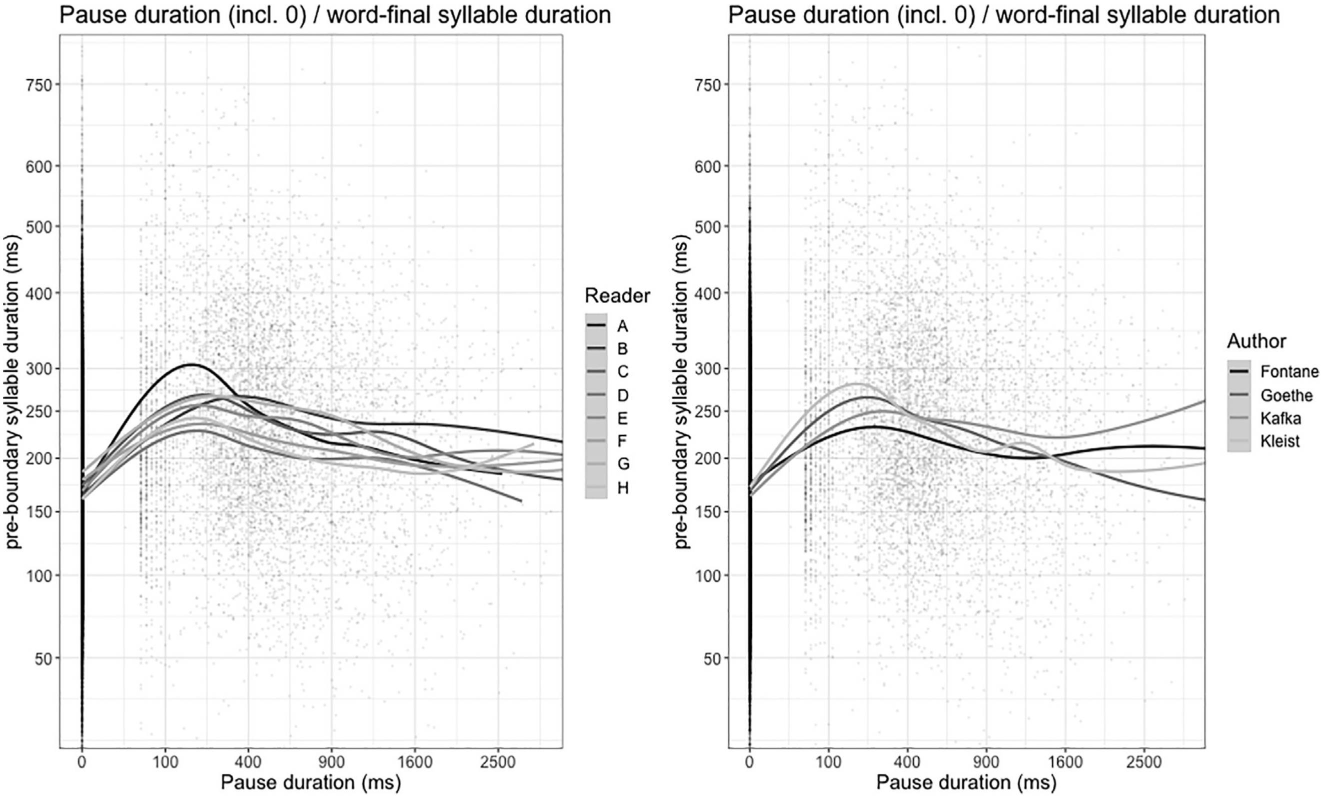

Fig. 4. Scatterplots showing the correlation between the pause duration (x-axis) and the pre-boundary syllable duration (y-axis). The left panel shows separate superimposed GAM smoothers for each of the eight different speakers, while the right panel shows separate GAM smoothers for the four texts.

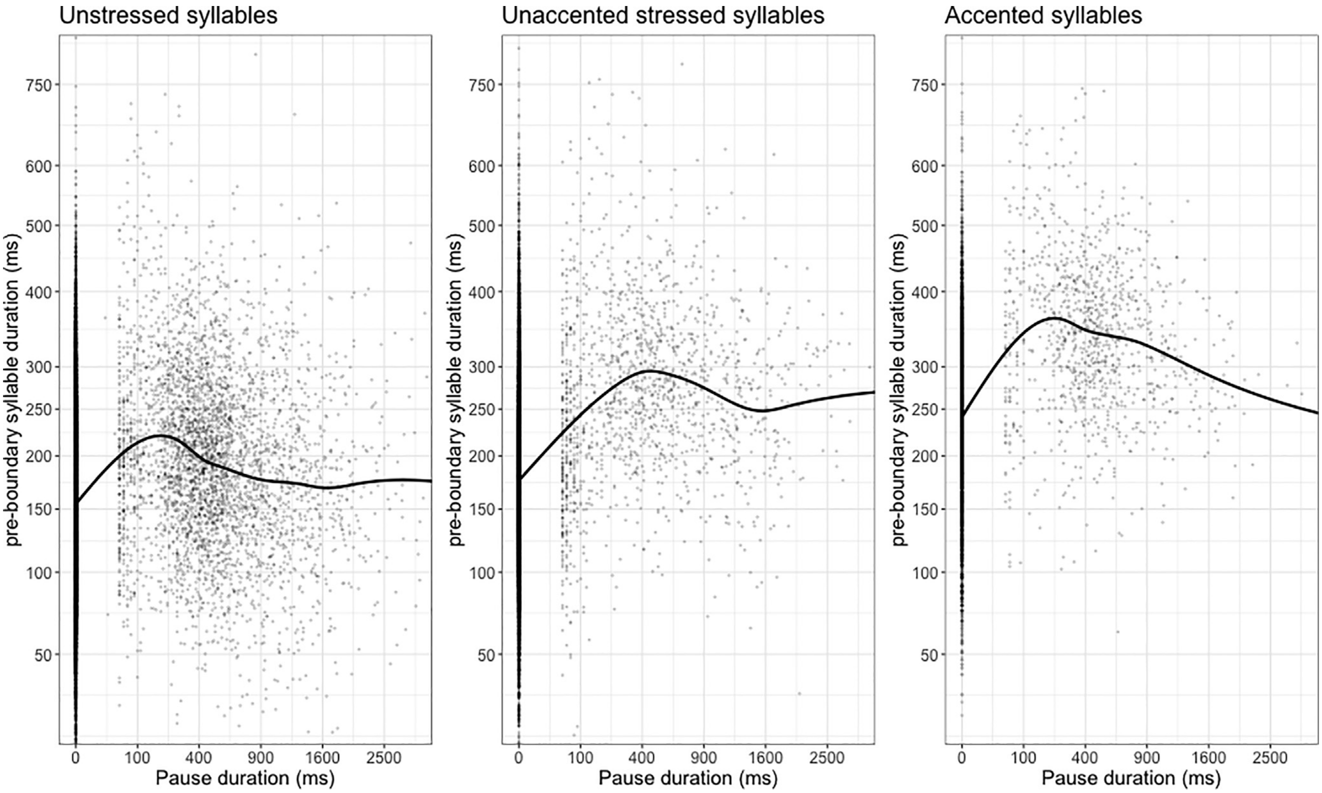

Fig. 5. Scatterplots with superimposed GAM smoothers (black curves) showing the correlation between the pause duration (x-axis) and pre-boundary syllable duration (y-axis), broken down by the prominence of the pre-boundary syllable (left panel: unstressed syllables; middle panel: stressed unaccented syllables; right panel: accented syllables).

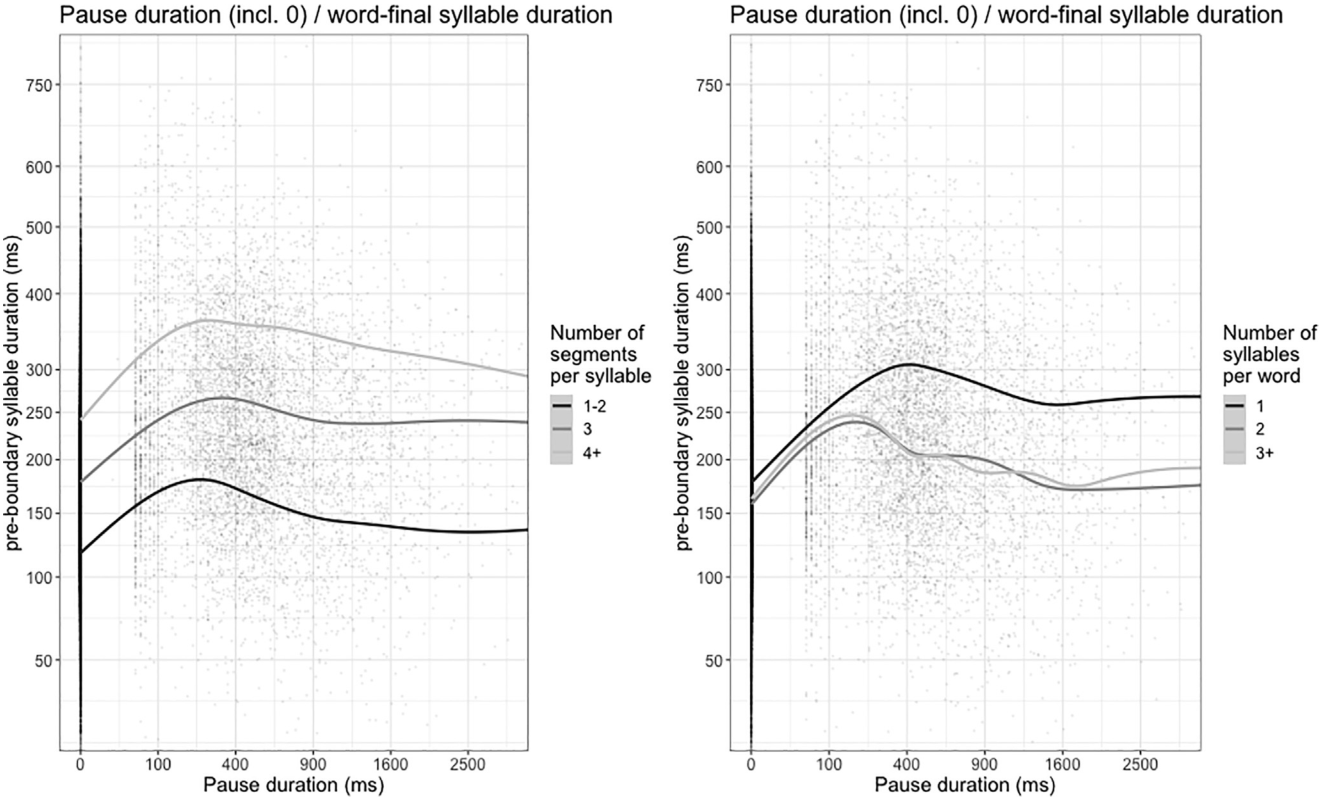

Fig. 6. Scatterplots showing the correlation between the pause duration (x-axis) and pre-boundary syllable duration (y-axis), with separate GAM smoothers for different numbers of segments in the word-final syllable (left panel), and separate GAM smoothers for different numbers of syllables within the pre-boundary word (right panel).

difference in semitones between maximum f0 and minimum f0 on the syllables; the difference was set to 0 when no f0 was detected by the praat algorithm. This measure, although rather coarse, allowed us to gauge the extent (but not the direction) of pitch excursions on all word-final syllables. Fig. 7 shows the pitch range broken down by the five predicted boundary strength values for syllables not followed by a pause (left panel) and for pre-pausal syllables (right panel). Like the case

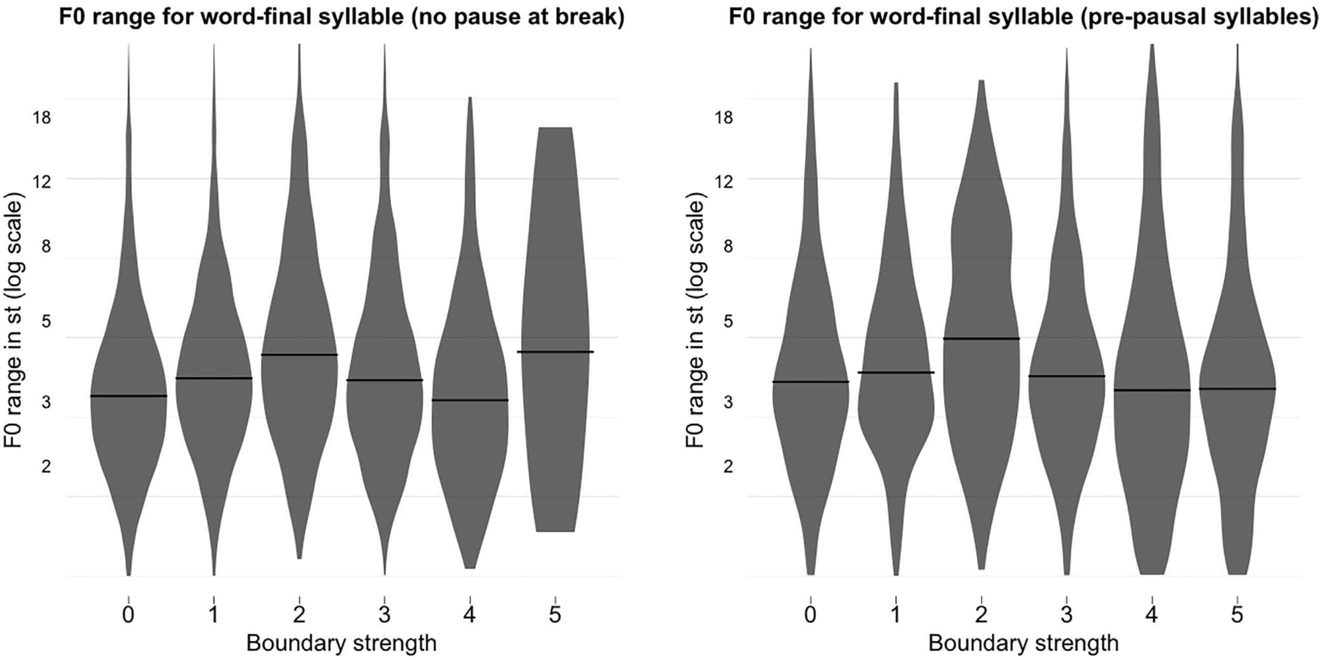

Fig. 7. Violin plots for f0 ranges in semitones (log scale) broken down by the boundary strength index. Average lines represent the median. The left panel shows a violin plot for the subset of data points after which no pause was registered. The right panel shows a violin plot for the subset of data points with a following pause. Note that, for boundary strength level 5, there are only 12 data points without a following pause (left panel). The remainder of the violins represent more than 400 data points (see Table 2).

of pre-boundary syllable duration (see Fig. 2), the results show generally greater pitch excursions for pre-pausal syllables (right panel) compared to syllables not followed by a pause (left panel). In both cases, there is a stepwise increase in the pitch range values from boundary strength level 0 up to level 2, and a decrease from level 2 to level 5 boundaries. The only ostensible exception to this pattern concerns the highest boundary level 5 in the cases where no pause follows the pre-boundary syllable (rightmost violin plot in the left panel of Fig. 7). However, as noted above, there are only 12 data points for this particular case, whereas each of the other violins represent at least 400 data points (see Table 2). A linear model with the f0 range (in semitones, log transformed) as the dependent variable and otherwise the same independent variables (boundary strength level coded as a contrast of successive differences; lengths of the pre-boundary and post-boundary words, both measured by the number of phonemes in the citation form of the word; pause duration (log); and the text and author and the speaker as grouping terms) largely confirms what visual inspection of the violin plots suggests: We find significant differences for four out of five successive contrasts (the difference between level 4 and level 5 breaks is not significant), with the f0 range increasing from level 0 up to level 2 breaks and then decreasing beyond that. The model parameters of interest are presented in Table 5.

We also examine the correlation between the pause duration and f0 range of the pre-boundary syllables. The left panel of Fig. 8 shows the f0 range for all word-final syllables, including word transitions without pauses (pause duration = 0 on the x-axis), while the right panel shows data for only the word transitions with actual pauses. Apart from the linear correlation (the dashed line in Fig. 8), we fit a GAM to the data (black curve). When all data points are considered (including word transitions without pauses, left panel), the linear correlation suggests an overall positive correlation between the extent of the pitch excursion and the pause duration. However, a comparison with the GAM smoother (black curve) reveals that the simple linear correlation (dashed line) between the two variables underestimates the pitch excursion before short pauses (<400 ms) but overestimates the f0 range before long pauses (900–1200 ms). The data corresponding to very long pauses (>1500 ms) are scarce (about 1% of all word transitions) and hence difficult to interpret. When only pre-pausal syllables are considered (right panel), the GAM (black curve) more closely fits a negative-going linear correlation (dashed line), at least for pauses > 300 ms; a positive correlation of the pause duration and pre-pausal pitch excursion is still observable for pauses up to 300 ms.

In sum, the data profiles are similar when plotting pre-boundary syllable durations and f0 ranges against boundary strength values (see Figs. 2 and 7), and likewise when plotting pre-boundary syllable durations and pitch excursions against pause durations (see Figs. 3 and 8). The longest pre-boundary syllable durations and the highest f0 ranges are observed for level 2 breaks and for pauses of around 300 ms.

Table 5
Linear mixed model for f0 range (semitones) of pre-boundary syllables.

Coeff.	Estimate	*SE*	*t*-value
Boundary 1–0	0.10950	0.00939	11.667*
Boundary 2–1	0.16156	0.01906	8.479*
Boundary 3–2	−0.16611	0.01844	−9.006*
Boundary 4–3	−0.10821	0.01624	−6.665*
Boundary 5–4	−0.006	0.02771	−0.216
Preceding word length	−0.00538	0.00103	−5.239*
Pause duration (log)	0.00926	0.00168	5.525*
Following word length	−0.00529	0.00098	−5.379*

Note: * t-values > |2| indicate that the alpha-level, i.e., the risk of falsely rejecting the null hypothesis, is smaller than 0.05.

3.5. Effects of declination on pre-boundary syllable f0 and duration

F0 contours are subject to declination effects: Declination means that speakers generally use a higher f0 range for syllables early in a given clause or sentence; in contrast, syllables towards the end of longer stretches of speech typically show a lower f0 ceiling and, consequently, relatively little f0 excursion. This might in turn affect syllable duration: As f0 excursions take time, syllables with greater f0 excursions are expected to have longer durations than syllables with little f0 movement. On the other hand, it has been suggested that speakers compensate for declination effects by lengthening syllables at the end of utterances (Lindblom, 1968).

To explore the effects of declination on the pre-boundary syllable f0 range and duration, we compare the various boundary strength levels in early regions of the sentences with those in later regions. To illustrate this point, we arbitrarily determined that “early” regions are those that are shorter than the overall mode of sentence length in our corpus, namely, shorter than eight syllables (see Table 1). We compared the boundaries in the early regions with those in the later regions, which involve all other syllable positions (i.e., from the eighth syllable onwards) within a given sentence. The results are shown in Figs. 9 and 10. These plots reveal a marked difference in terms of both the pre-boundary syllable duration and the f0 range between early and late regions in the profiles for the three higher levels of boundary strength: For early regions (left panel of Fig. 9), the pre-boundary syllable duration increases monotonically from boundary strength level 0 to boundary strength level 5. In contrast, for later regions (right panel of Fig. 9), we observe the familiar inverted u-shaped pattern, with progressively shorter pre-boundary syllable durations from level 2 to level 5 boundaries. This u-shaped pattern can also also observed for the f0 range of pre-boundary syllables in the late regions (right panel of Fig. 10). For the early regions (left panel of Fig. 10), the f0 ranges increase from boundary level 0 up to level 2, and they appear to plateau across boundary strength levels 2 to 5.

These observations suggest that the sentence region has a substantial effect on the way the various boundary levels are phonetically expressed on the pre-boundary syllable, in terms of both duration and f0 range. To substantiate this finding, we supplement the two linear mixed models reported in Sections 3.2 and 3.4 (Tables 4 and 5) with a fixed effect for the sentence position (the number of the syllable after which the boundary was set, counting from the start of the sentence) and a coefficient for the interaction between sentence position and boundary strength level. The word length of the current and following words and the pause duration (log transformed when > 0) were entered as covariates. Again, the text and author and the

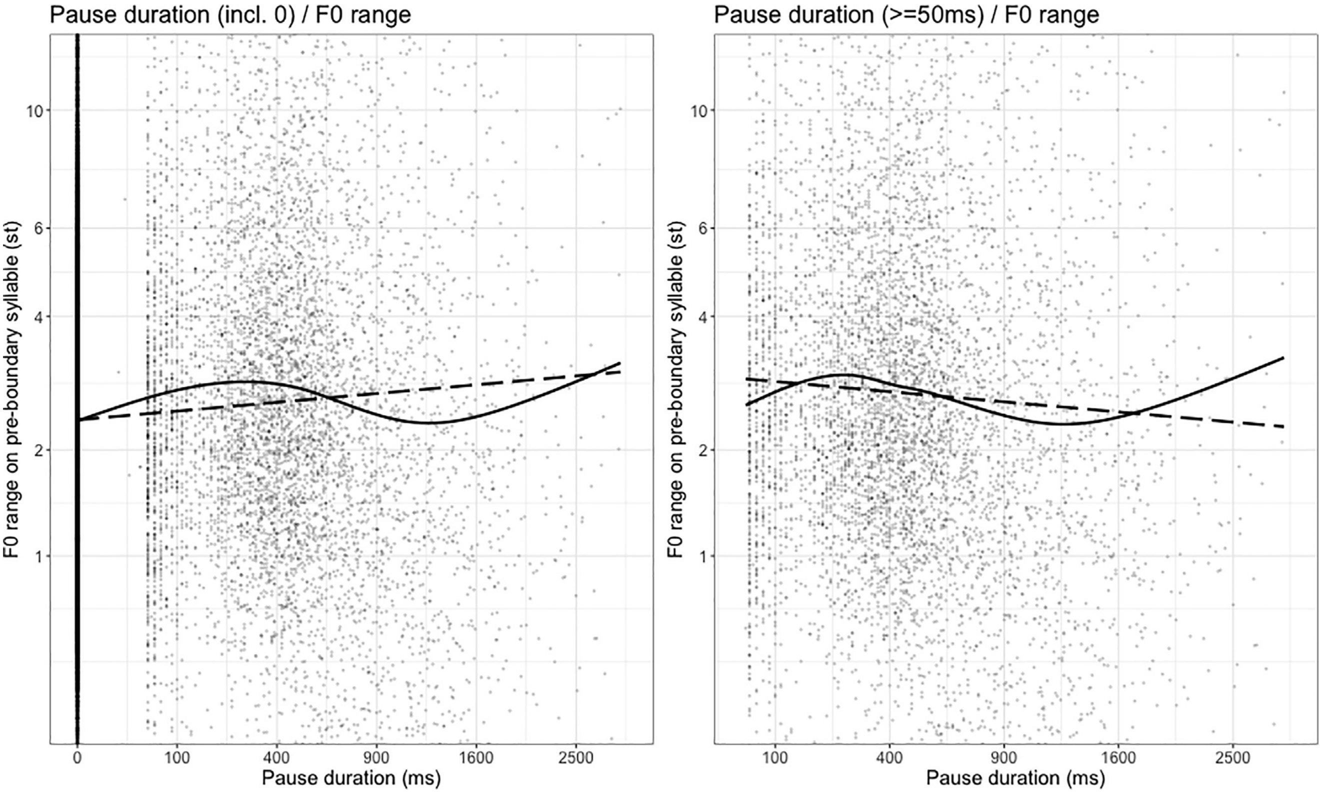

Fig. 8. Scatterplots showing the correlation between the pause duration (x-axis) and f0 excursion of the pre-boundary syllable (y-axis). The left panel shows all relevant data points (including syllables not followed by a pause). The right panel shows the subset of data points with actual pauses. The dashed line depicts a linear correlation, while the black curve shows a GAM smoother.

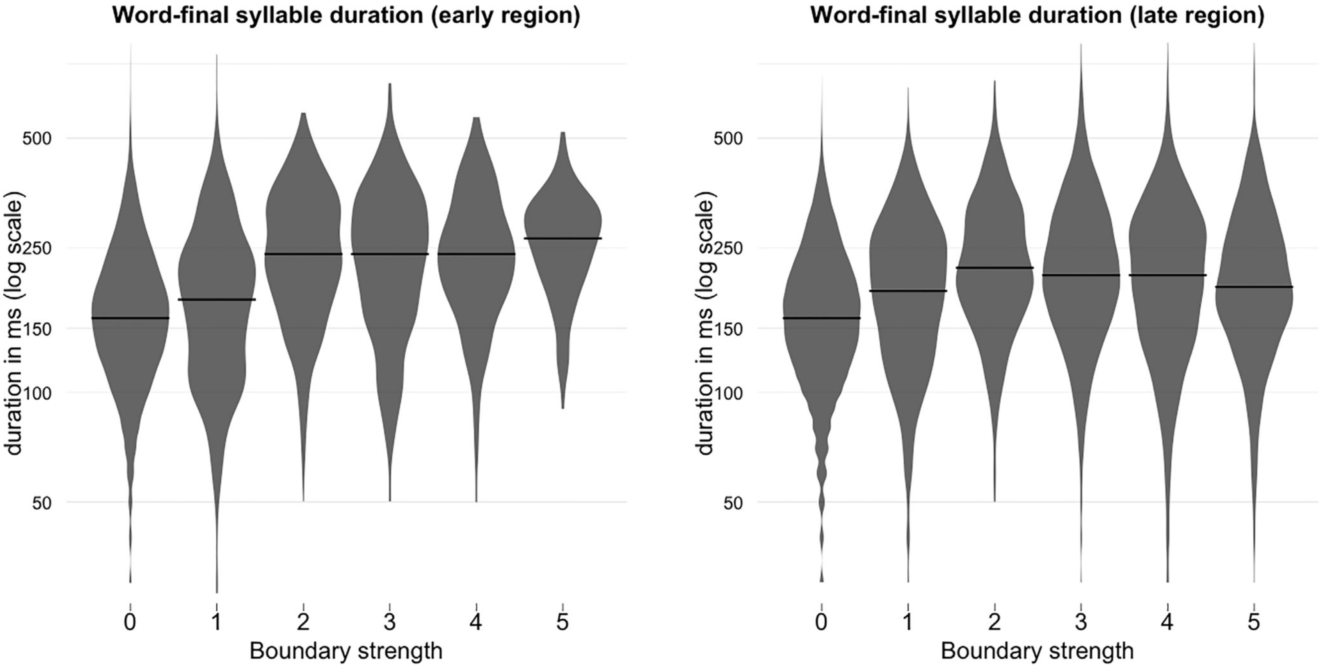

Fig. 9. Violin plots for the pre-boundary syllable duration (log scale) broken down by the boundary strength index. Average lines represent the median. The left panel shows a violin plot for boundaries set within early regions of the sentences (within the first seven syllables of each sentence). The right panel shows the values for the remaining data points.

speaker were used as grouping variables (random intercepts). The coefficients of interest for these augmented models are presented in Tables 6 and 7. The two models reveal a significant effect of the sentence position and a significant interaction between the boundary strength level and sentence position on the syllable duration and f0 range. The models thus confirm

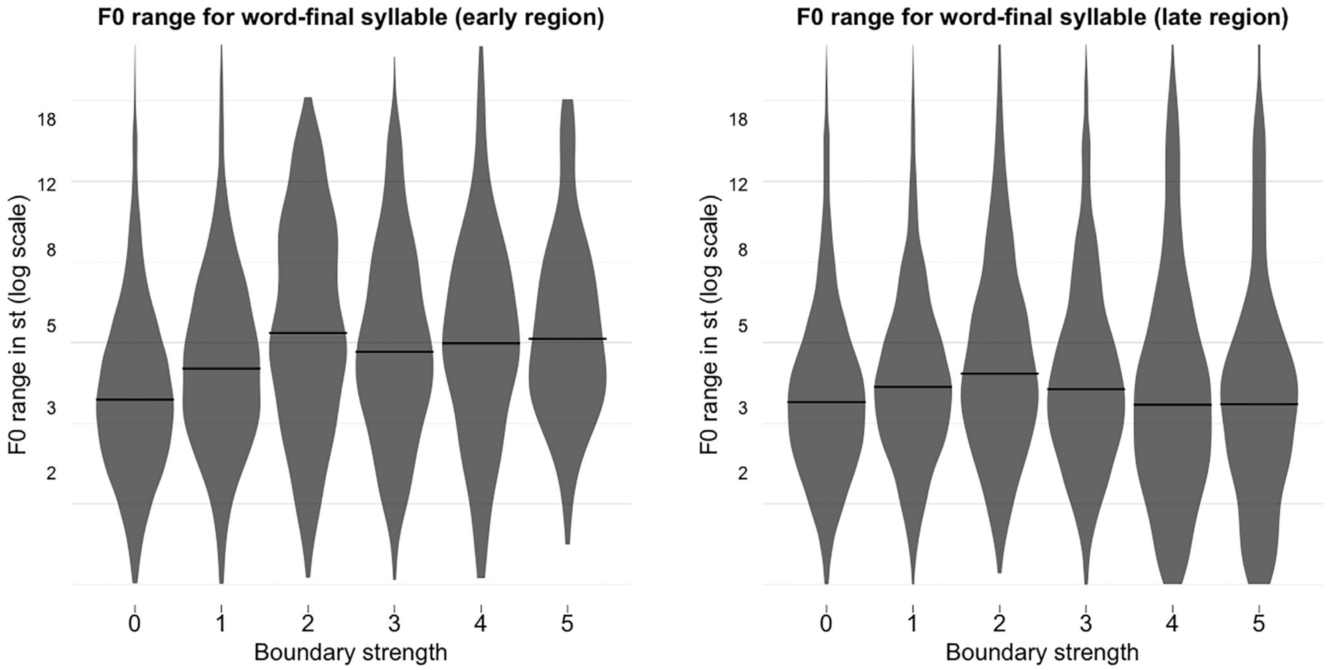

Fig. 10. Violin plots for the f0 excursion of the pre-boundary syllable (in semitones, log scale) broken down by the boundary strength index. Average lines represent the median. The left panel shows a violin plot for boundaries set within the early regions of the sentences (within the first seven syllables of each sentence). The right panel shows the values for the remaining data points.

Table 6
Coefficients of interest for the augmented model for the pre-boundary syllable duration.

Coeff.	Estimate	*SE*	*t*-value
Boundary 1–0	1.27E-01	7.66E-03	16.6*
Boundary 2–1	1.55E-01	1.52E-02	10.17*
Boundary 3–2	−5.22E-02	1.50E-02	−3.49*
Boundary 4–3	−1.31E-01	1.28E-02	−10.27*
Boundary 5–4	−5.17E-02	2.06E-02	−2.51*
Sentence position	4.72E-04	9.50E-05	4.97*
Preceding word length	−1.55E-03	8.12E-04	−1.91
Pause duration (log)	3.75E-02	1.32E-03	28.47*
Following word length	−3.40E-03	7.82E-04	−4.34*
Boundary strength : Sentence position	−4.46E-04	6.65E-05	−6.7*

Note. * t-values > \|2\| indicate that the alpha-level, i.e., the risk of falsely rejecting the null hypothesis, is smaller than 0.05.

Table 7
Coefficients of interest for the augmented model for the pre-boundary syllable f0 excursion.

Coeff.	Estimate	*SE*	*t*-value
Boundary 1–0	1.28E-01	9.65E-03	13.31*
Boundary 2–1	1.64E-01	1.91E-02	8.59*
Boundary 3–2	−1.31E-01	1.88E-02	−6.97*
Boundary 4–3	−7.63E-02	1.67E-02	−4.58*
Boundary 5–4	−1.35E-02	2.77E-02	−0.49
Sentence position	−5.91E-05	1.20E-04	−0.49
Preceding word length	−4.92E-03	1.03E-03	−4.78*
Pause duration (log)	1.01E-02	1.68E-03	6.03*
Following word length	−5.11E-03	9.85E-04	−5.19*
Boundary strength : Sentence position	−7.01E-04	8.69E-05	−8.06*

Note. * t-values > \|2\| indicate that the alpha-level, i.e., the risk of falsely rejecting the null hypothesis, is smaller than 0.05.

that the boundary strength level and sentence position interactively affect the phonetic expression of the pre-boundary syllable, in terms of both its duration and its f0 excursion. Specifically, in the models for both duration and f0 excursion, the negative coefficients of the interaction between the boundary strength level and the sentence position reflect what the figures for the "late" regions (right panels in Figs. 9 and 10) suggest, namely, that the phonetic expression of boundary strength decreases as the sentence progresses.

4. Discussion

The results of this study complement and partially qualify previous findings on the phonetic realization of prosodic boundaries across different levels of boundary strength.

First, the analysis confirms that both the likelihood of a pause and the pause duration monotonically increase along the six-step scale of predicted levels of prosodic boundary strength, with significant differences between all levels. Pauses at word transitions that do not correspond to predicted prosodic boundaries (level 0) are relatively rare, and, if present, they are shorter (<150 ms on average) than pauses at predicted boundaries. Conversely, at the strongest prosodic boundary, where the reader enacts a shift from the narrator's voice to the direct speech of a literary figure (or vice versa, or between the voices of different literary figures), pauses are very frequent (with 98% likelihood almost compulsory), and these pauses are the longest (∼1 sec on average).

Second, in contrast to pause likelihood and durations, word-final syllable durations show a nonmonotonic relationship to predicted boundary strength. Whereas word-final syllables at neutral word transitions (boundary strength level 0) are on average shorter than word-final syllables preceding (predicted) prosodic boundaries, the word-final syllable duration increases only up to boundary level 2 (a small comma phrase), on average, and then decreases from level 2 to level 5 boundaries (direct speech boundaries). This decrease in the word-final syllable duration from level 2 phrases to word-final syllables at level 5 boundaries attenuates pre-boundary syllable durations. However, this phenomenon depends on the sentence position: the decrease in pre-boundary syllable durations for the higher-level boundaries does not hold for "early" boundaries, i.e., those that are produced within the range of the first seven syllables in the sentence. That is, the sentence position and boundary strength interactively affect pre-boundary syllable durations.

Third, when considering actual pause durations instead of predicted boundary strength, pre-pausal syllables are longer than word-final syllables that are not followed by a pause. This holds across all six levels of predicted prosodic boundary strength (although the comparison is complicated by the uneven distribution of pauses across the different boundary strengths). In general, pre-pausal syllable durations positively correlate only with pause durations of up to around 300 ms. For pauses longer than 300 ms, we observe a marked negative correlation between the pause duration and the final syllable duration. This general nonmonotonic pattern appears to hold irrespective of the speaker and text and the prominence, word length, and phonetic content of the word-final syllable.

Similar result patterns were obtained for the tonal expression of the boundaries: The average f0 excursion of the pre-boundary syllable increases from neutral word transitions (level 0) to small comma phrase breaks (level 2) and then decreases from level 2 to level 5 boundaries. Again, the decrease in the f0 excursion for the higher-level boundaries does not hold when the corresponding boundaries occur in early regions of the sentence. In general, the f0 range of pre-boundary syllables is greater when they precede a pause compared to pre-boundary syllables that are not followed by a pause.

Taken together, our results point to a non-contradictory combination of opposite effects. On the one hand, pre-pausal syllable durations are generally longer and bear greater f0 excursions than word-final syllables that are not followed by pauses. This additive relationship probably serves to render the conclusion of a phrase or a sentence more salient. At the same time, considering the pre-boundary syllable duration and the duration of the following pause as interacting variables, the pre-pausal syllable duration actually decreases as the pause duration increases with the predicted boundary strength. As we have shown (see Table 2 and Fig. 1), pause likelihood and duration correlate with the boundary strength scale, which in turn corresponds to the degree of finality of the respective stretches of speech (0: word-final < 1, 2, 3: phrase-final < 4: sentence-final < 5: discourse-final). Since the pre-boundary syllable duration does not correlate with this scale in a monotonic fashion, our findings partially revise the hypothesis of *final lengthening* for pre-boundary syllable duration. Specifically, for the stronger prosodic boundaries, we instead diagnose a final shortening.

The nonmonotonic relation between the pre-boundary syllable duration and the boundary strength and/or pause duration is unexpected in light of the extant literature on pre-boundary lengthening. Specifically, our results do not square with the widely held view that pre-boundary syllable duration increases progressively with boundary strength (Byrd & Saltzman, 2003; Gussenhoven, 1992, among others).

In contrast, the similar nonmonotonic pattern for the f0 excursions of word-final syllables at the various boundary strength levels can be related to known facts about intonation: The stronger boundaries are clause-final or sentence-final boundaries. These boundaries are therefore likely to be set in regions that are typically affected by declination, that is, where speakers have a more limited f0 range at their disposal compared to sentence-initial or sentence-medial regions, which are not (or at least less) affected by declination. f0 excursions are therefore predicted to be greater at prosodic breaks in sentence-medial positions (e.g., at boundary level 2) than in clause-final or sentence-final positions (roughly corresponding to boundary levels 3–5). All other things being equal, producing greater f0 excursions affords or necessitates more time than producing smaller excursions. In addition, we reason that sentence-final f0 excursions tend to have falling contours (at least in the case of the declarative sentences that we examine here), whereas non-final or sentence-medial boundaries readily afford f0 rises (so-called continuation rises). The literature on tone and intonation maintains that producing f0 rises takes more time than producing falling contours. Again, this suggests that, in sentence-medial positions, pre-boundary syllables—which tend to bear rising contours—require more time than pre-boundary syllables—which tend to have falling contours—at the end of a sentence.

For all of these reasons, the correlation we observed between pre-boundary syllable duration and the corresponding f0 range supports the view that the unexpected nonmonotonic pattern in the temporal domain is, at least partially, a consequence of the likewise nonmonotonic but quite expectable pattern in the tonal domain. However, our analysis cannot exclude a relationship in the opposite direction: It is also possible that more extended f0 excursions are not the reason for pre-boundary lengthening but the consequence of it. Under this view, pre-boundary lengthening affords the time necessary for f0 excursions to be executed; shortening, in turn, may lead to the truncation of f0 contours that would otherwise be fully expressed.

Whatever the relationship between pauses and pre-boundary syllable duration, these measures might also be governed by a psycholinguistic factor, namely, speech planning: Speakers use prosodic breaks to mark the end of long or complex constituents and to plan upcoming phrases and sentences. While pause duration is conditioned by the complexity of both the preceding and the following material, pre-boundary syllable lengthening necessarily happens during the execution of the current speech chunk, and therefore likely reflects the planning effort related to it. If pre-boundary lengthening reflects current planning effort or production complexity, we can assume that the effect of pre-boundary lengthening will be stronger at prosodic boundaries *within* larger speech units (say, at boundary strength level 1 or 2) that still require planning effort than at prosodic boundaries that close off these major chunks and for which planning has therefore been completed.

We believe it is plausible that the decreasing word-final syllable durations at the larger boundaries reflect diminishing production complexity for the current stretch of speech, whereas the pause durations reflect the combined complexity of the preceding and the following material. This assumption is, at first glance, not fully compatible with extant research on pre-boundary syllable duration—which, in contrast to our results, does find some evidence for a monotonically positive correlation between pre-boundary lengthening and prosodic boundary strength (Cambier-Langeveld, 2000, for Dutch; Michelas & D'Imperio, 2012; Michelas & D'Imperio, 2010, for French; Liberman, 2007, reporting on the 3 + 4 phrasing of American English telephone numbers). Other findings suggest a plateauing of the final-lengthening effect at higher-level prosodic boundaries (e.g., Peters et al., 2005; Wightman et al., 1992). Crucially, however, we are not aware of previous research that shows a systematic decrease in pre-boundary syllable duration and hence a *final shortening*.

A closer look at the studies by Michelas & D'Imperio (2012); Cambier-Langeveld (2000); Michelas & D'Imperio (2010), and Liberman (2007) reveals a difference that might be important for the interpretation of the present results: In these studies, the critical prosodic breaks were all set relatively early after the onset of the sentences (after the second word in Michelas & D'Imperio, 2012; after the third in Liberman, 2007; and after the fifth word in Cambier-Langeveld, 2000). The effects of declination might therefore be relatively small. The sentences were read aloud, and therefore, quite plausibly, mostly pre-planned. Moreover, as is typical for controlled experiments, the sentences were not only relatively short but structurally predictable, especially since some experiments rely on multiple repetitions of the same sentences. We suggest that in such experimental settings, differences in production complexity at the different boundary strength levels may barely emerge, as effects of planning are likely to be minor.

In contrast, although also read aloud, the sentences in our study are part of an ongoing narrative and vary greatly in length. Prosodic boundaries at the various levels are distributed widely over the sentences. Under these conditions, differential effects of incremental planning for the different boundary strength levels may materialize. However, when focusing on the early regions of the sentences in the present data set (see the left panels of Figs. 9 and 10), neither the pre-boundary syllable duration nor the f0 excursion significantly decrease for the higher levels of boundary strength. This particular finding is in line with the aforementioned experiments that produced progressive pre-boundary lengthening along the boundary strength scale or at least plateauing at the higher levels. Arguably, within the first couple of words that are read aloud, differential effects of incremental planning for the different boundary strength levels are mostly mute. The question of whether this also holds for spontaneous speech is beyond the scope of this study.

In any event, we would like to emphasize that, irrespective of all the differences between a read-aloud, artful prose text and quotidian spontaneous speech, the material used in our study represents a style of narrative discourse that is a valid and established kind of oral text production. The present investigation therefore complements previous studies on pre-boundary lengthening that use decontextualized experimental sentences, thereby enriching our understanding of speech production more generally. We acknowledge that the rather specific nature of the speech sample, i.e., literary prose texts read aloud by professional speakers enrolled at an academy for stage reading, may raise concerns about the generalizability of our results. We cannot fully exclude the possibility that the effects reported here are peculiar to the particular speakers we examined. However, we deem it unlikely that speakers who are specifically trained in developing an individual artistic style would be so consistent in their realization of pre-boundary syllables and pauses (see the left panel of Fig. 4) or that the observed interactive effect between sentence position and boundary strength on pre-boundary syllable duration and f0 excursion would be the result of professional training. We assume that such an effect would hardly be amenable to conscious attention and training and is therefore more likely to reflect a more general characteristic of spoken speech. Evaluating this assumption is, however, a matter for future research.

5. Conclusion

This study on the realization of prosodic boundaries in prose texts read aloud shows that the durations of pauses at prosodic boundaries and the durations of pre-boundary syllables are not correlated in a simple monotonic fashion: Whereas pause durations increase along the scale of predicted prosodic boundary strength, the pre-boundary syllable durations and pitch excursion of the pre-boundary syllable follow an inverted u-shaped pattern. The longest durations at breaks correspond to the end of small comma phrases, and syllable durations actually decrease as the degree of finality of the prosodic boundaries increases (long comma phrase > sentence > direct speech boundary). Our analysis suggests that the decrease in pre-boundary syllable duration for the higher levels of prosodic boundary strength is related to the effects of f0 declination: the lower f0 excursions in sentence-final positions compared to sentence-medial positions likely take less time to execute, and pre-boundary syllable duration consequently decreases.

The results reported here are also compatible with the idea that pre-boundary syllable durations reflect current planning complexity, because this planning complexity decreases as the degree of finality of the prosodic boundary increases. We propose that the above-mentioned divergent findings regarding the so-called "final lengthening" can be ascribed to the fact that the prosodic boundaries examined in these previous laboratory studies were produced relatively early (after the first couple of words) within self-contained single sentences. For this reason, those findings cannot readily be extended to the many other contexts in which boundaries of different strengths are normally produced. We close by emphasizing the importance of examining more complex spoken prose when studying the phonetic realization of the various degrees of prosodic boundary strength.

CRediT authorship contribution statement

Gerrit Kentner: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Isabelle Franz:** Data curation, Conceptualization, Writing – review & editing. **Christine A. Knoop:** Data curation, Writing – review & editing. **Winfried Menninghaus:** Funding acquisition, Writing – review & editing.

Acknowledgements

We gratefully acknowledge the advice and support of Muralikrishnan, Dominik Thiele, and Antonia Walsch regarding the collection and processing of the audio data. Julia Merrill and Kerstin Kipp initiated contact with the professional speakers from the Hochschule für Musik und Darstellende Kunst, Stuttgart. Marianne Pouplier, Jelena Krivokapić, and two anonymous reviewers provided critical and encouraging feedback on the initial manuscript. Not least, we would like to thank the professional speakers, Irene Baumann, Dominik Eisele, Elias Hartung, Paula Scheschonka, Ramon Schmid, Jonathan Springer, Lina Syren, and Dorothee Wolfsberger, for providing us with audiobook-quality recordings.

References

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say *thee uh* you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 914. https://doi.org/10.1037/0278-7393.33.5.914.

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.

Betz, S., & Wagner, P. (2016). Disfluent lengthening in spontaneous speech. In: Jokisch, O. (Ed.). *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2016* (pp. 135-144). Dresden: TUDpress.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.

Brugos, A. M. (2015). *The interaction of pitch and timing in the perception of prosodic grouping*. [Doctoral dissertation, Boston University]. Boston University Digital Repository. https://open.bu.edu/handle/2144/14012.

Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics, 7*(1), 31–53. https://doi.org/10.1146/annurev-linguistics-030920-050033.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics, 26*(2), 173–199. https://doi.org/10.1006/jpho.1998.0071.

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

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica, 57*(1), 3–16. https://doi.org/10.1159/000028456.

Cambier-Langeveld, G. M. (2000). *Temporal marking of accents and boundaries* [Doctoral dissertation, University of Amsterdam]. University of Amsterdam Digital Academic Repository. https://hdl.handle.net/11245/1.172937.

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *Proceedings of Speech Prosody 2002*. Aix-en-Provence, France. https://www.isca-speech.org/archive_open/sp2002/sp02_199.pdf.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*(3), 120–141. https://doi.org/10.1111/lnc3.12178.

Cohen, A., Collier, R., & 't Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica, 39*(4–5), 254–273. https://doi.org/10.1159/000261666.

Collier, R., & Gelfer, C. E. (1983). Physiological explanations of F0 declination. In M. P. R. Van den Broecke & A. Cohen (Eds.) (pp. 354–360). Dordrecht: Foris.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

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.

Fant, G., & Kruckenberg, A. (1994). Notes on stress and word accent. *Speech, Music, and Hearing: Quarterly Progress and Status Report* (Vol. 2, pp. 125–144). Department of Speech, Music and Hearing, Royal Institute of Technology.

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.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review, 100*(2), 233–253. https://doi.org/10.1037/0033-295x.100.2.233.

Féry, C. (2017). *Intonation and prosodic structure*. Cambridge University Press.

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* (2nd ed., pp. 521–602). Wiley. https://doi.org/10.1002/9781444317251.ch15.

Franz, I., Knoop, C. A., Kentner, G., Rothbart, S., Kegel, V., Vasilieva, J., ... Menninghaus, W. (2022). Prosodic phrasing and syllable prominence in spoken prose: A validated coding manual. *OSF Preprints*. https://doi.org/10.31219/osf.io/h4sd5.

Frazier, L., Carlson, K., & Clifton, C. Jr, (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*(4), 411–458. https://doi.org/10.1016/0010-0285(83)90014-2.

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8*(1), 59–85. https://doi.org/10.1207/s15516709cog0801_3.

Gussenhoven, C., & Rietveld, A. C. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics, 20*(3), 283–303. https://doi.org/10.1016/s0095-4470(19)30636-9.

Gussenhoven, C. (1992). In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, & J. R. Rennison (Eds.), *Phonologica 1988: Proceedings of the 6th International Phonology Meeting* (pp. 89–99). Cambridge University Press.

Horne, M., Strangert, E., & Heldner, M. (1995)K. Elenius & P. Branderud (Eds.). *Prosodic boundary strength in Swedish: Final lengthening and silent interval duration, 95*, 170–173.

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. *The Linguistic Review, 30*(2), 277–311. https://doi.org/10.1515/tlr-2013-0009.

Kentner, G. (2007). Length, ordering preference and intonational phrasing: Evidence from pauses. *Proceedings of Interspeech 2007*. Antwerp. https://doi.org/10.21437/Interspeech.2007-693.

Kim, J. (2020). *Individual differences in the production and perception of prosodic boundaries in American English* [Doctoral dissertation, University of Michigan]. https://hdl.handle.net/2027.42/162927.

Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via web services: The use case WebMAUS [Paper presentation]. *Digital Humanities Conference 2012*, Hamburg, Germany.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3*(3), 129–140. https://doi.org/10.1016/S0095-4470(19)31360-9.

Krivokapić, J., Styler, W., & Parrell, B. (2020). Pause postures: The relationship between articulation and cognitive processes during pauses. *Journal of Phonetics, 79*, 100953. https://doi.org/10.1016/j.wocn.2019.100953.

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.

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. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1658), 20130397. https://doi.org/10.1098/rstb.2013.0397.

Krivokapić, J. (2022). Prosody in articulatory phonology. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic theory and practice* (pp. 213–236). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/10413.003.0008.

Ladd, D. R., & Campbell, W. N. (1991). *Theories of prosodic structure: Evidence from syllable duration, II*, 290–293. https://www.coli.uni-saarland.de/groups/BM/phonetics/icphs/ICPhS1991/.

Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology, 1*, 53–74 https://www.jstor.org/stable/4615382.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Liberman, M. (2007). Ritardando al fine. *Language Log* (November 2007). http://itre.cis.upenn.edu/~myl/languagelog/archives/005133.html.

Lindblom, B. (1968). Temporal organization of syllable production. *Quarterly Progress and Status Report, 9*(2–3), 1–5. Stockholm: KTH Department of Speech, Music, and Hearing.

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior, 9*, 75–78. https://doi.org/10.1016/S0022-5371(70)80010-X.

Mayer, J., Jasinskaja, E., & Kölsch, U. (2006). Pitch range and pause duration as markers of discourse hierarchy: Perception experiments. *Proceedings of Interspeech 2006*. Pittsburg. https://www.isca-speech.org/archive/pdfs/interspeech_2006/mayer06_interspeech.pdf

Michalsky, J. (2017). *Frageintonation im Deutschen: Zur intonatorischen Markierung von Interrogativität und Fragehaltigkeit*. Berlin: de Gruyter.

Michelas, A., & D'Imperio, M. (2010). Durational cues and prosodic phrasing in French: Evidence for the intermediate phrase. *Proceedings of Speech Prosody 2010*. Chicago. http://sprosig.org/sp2010/papers/100881.pdf.

Michelas, A., & D'Imperio, M. (2012). When syntax meets prosody: Tonal and duration variability in French accentual phrases. *Journal of Phonetics, 40*(6), 816–829. https://doi.org/10.1016/j.wocn.2012.08.004.

Myers, S. (2003). F_0 timing in Kinyarwanda. *Phonetica, 60*(2), 71–97. https://doi.org/10.1159/000071448.

Ohala, J. J., & Ewan, W. G. (1973). Speed of pitch change. *The Journal of the Acoustical Society of America, 53*(1), 345.

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., Kohler, K. J., & Wesener, T. (2005). In K. J. Kohler, F. Kleber, & Peters (Eds.). *Prosodic structures in German spontaneous speech* (35, pp. 143–184). Kiel: IPDS: AIPUK.

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.

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

Rao, R. (2010). *Final lengthening and pause duration in three dialects of Spanish. Selected proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology* (pp. 69–82). Somerville, MA: Cascadilla Proceedings Project.

Schubö, F., & Zerbian, S. (2020). Phonetic content and phonological structure affect pre-boundary lengthening in German. *Proceedings of Speech Prosody 2020.* Tokyo. https://doi.org/10.21437/SpeechProsody.2020-23.

Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., Himmelmann, N. P., & Bickel, B. (2021). The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. *Linguistics Vanguard, 7*(1). https://doi.org/10.1515/lingvan-2019-0063.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics, 7*(2), 71–79. https://doi.org/10.1016/S0095-4470(19)31040-X.

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.

Tabain, M., & Perrier, P. (2005). Articulation and acoustics of /i/ in preboundary position in French. *Journal of Phonetics, 33*(1), 77–100. https://doi.org/10.1016/j.wocn.2004.04.003.

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.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*(1), 33–57. https://doi.org/10.2307/40285321.

Trouvain, J. (2003). *Tempo variation in speech production. Implications for speech synthesis*. Institute of Phonetics, Saarland University.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001.

Turk, A. (2012). The temporal implementation of prosodic structure. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 242–253). Oxford University Press.

Tyler, J. (2014). Rising pitch, continuation, and the hierarchical structure of discourse. *University of Pennsylvania Working Papers in Linguistics, 20* (1), 36 https://repository.upenn.edu/pwpl/vol20/iss1/36/.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53–66). Heidelberg: Springer.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19* (6), 713–755. https://doi.org/10.1080/01690960444000070.

White, L. (2002). *English speech timing: A domain and locus approach* [Doctoral dissertation, University of Edinburgh]. Edinburgh Research Archive. https://era.ed.ac.uk/handle/1842/23256.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication, 63*, 38–54. https://doi.org/10.1016/j.specom.2014.04.003.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Heidelberg: Springer.

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.

Wöllstein, A. (2014). Topologisches Satzmodell. In J. Hagemann & S. Staffeldt (Eds.), *Syntaxtheorien: Analysen im Vergleich* (pp. 143–164). Tübingen: Stauffenburg.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America, 111*(3), 1399–1413. https://doi.org/10.1121/1.1445789.

Yang, L. C. (2004). Duration and pauses as cues to discourse boundaries in speech. *Proceedings of Speech Prosody 2004.* Nara, Japan. http://sprosig.org/sp2004/PDF/Yang.pdf

Yang, L. C. (2007). Duration, pauses and the temporal structure of Mandarin conversational speech http://www.icphs2007.de/conference/Papers/1687/1687.pdf.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). Hoboken, NJ: John Wiley.