2025 ACTA UNIVERSITATIS CAROLINAE PAG. 139–156
PHILOLOGICA 3

VOWEL DURATION IN STRESSED AND UNSTRESSED SYLLABLES IN SPONTANEOUS ENGLISH

NELA BRADÍKOVÁ, RADEK SKARNITZL

ABSTRACT
Many phonetic "truths" are based on descriptions of controlled speech material, and verifying their validity in spontaneous productions is essential. The present study investigates vowel duration as an acoustic correlate of stress in spontaneous English, in communicatively motivated contexts. By analyzing British and American political debates, this study aims to verify previously reported tendencies – specifically, that stressed vowels are significantly longer than unstressed ones. Our analysis of 16 speakers, based on linear mixed effects models, confirms the significant effect of stress on vowel duration and also considers additional factors influencing segmental duration like vowel length, phrase-final position, vowel height, or the nature of the following segment. In addition to stress, multiple regression analysis identifies vowel length, phrase-final position and vowel height as the most influential vowel duration predictors. Despite the variability of spontaneous speech, vowel duration proves to be a reliable correlate of stress, supporting the findings from controlled-speech research.

Keywords: spontaneous speech; lexical stress; vowel duration; English

1. Introduction

Many findings in the speech sciences are based on descriptions of laboratory speech which is more or less controlled: speakers are asked to read sentences or even isolated words or pseudowords, with little linguistic creativity on their part. Such findings have been invaluable for developing theories of speech production, but they tend to be repeated, and it is only rarely that their generalizability is questioned. However, it is conceivable that "language rules" which have been formulated based on more or less tightly controlled speech materials may not hold in spontaneous speech. One of the goals of the present study is therefore to verify some of the claims about the sound patterns of English on speech material which may be regarded as truly spontaneous, naturalistic, and uttered with a clear communicative purpose.

Of course, spontaneous speech constitutes a challenge for researchers at several levels. The phonetic realization of segments may be extremely variable (Greenberg, 1999; Barry & Andreeva, 2001; Nakamura, Iwano, & Furui, 2008). Using Cauldwell's (2013) botanic metaphor, the sound shapes of individual words pronounced in the "jungle" of sponta-

https://doi.org/10.14712/24646830.2025.25

neous speech may very much differ from their canonical forms, which may be observed in the "greenhouse" or "garden". For instance, Johnson (2004) reported that the rate of syllable elision in three- to six-syllabic lexical words ranged between 26 and 59%, or that between 20 and 30% of segments deviated from the canonical form in lexical words longer than four phones (see also a summary of more studies in Tucker & Mukai, 2023). Such levels of reduction may lead to considerable difficulties in performing phonetic alignment at the segmental level (in other words, in identifying individual segments and their boundaries in the stream of speech). In turn, it may then be demanding to extract meaningful data from such material. Nevertheless, we are convinced that the validity of the findings reported in the literature must be put to the test in spontaneous speech. To do so, our present study addresses duration as a correlate of lexical stress in English.

1.1 Lexical stress in English

Correlates of lexical stress represent an area of speech science that has been researched for over 70 years (see van Heuven, 2019, for a summary). Duration is traditionally accepted as the primary acoustic correlate of lexical stress in English. In one of the first studies, Fry (1955) compared the duration and intensity of vowels in noun-verb pairs, such as object or contract in British English, with the target words embedded in sentences. His results showed that both dimensions are important for distinguishing between stressed and unstressed syllables. Lieberman (1960) relied on a similar speech material in American English and examined more acoustic correlates than Fry; he reports fundamental frequency (f_0) and peak amplitude to distinguish between stressed and unstressed syllables, with duration ranking third. The problem is, however, that some of the target words were embedded both in the nuclear and pre-nuclear position in the sentences (e.g., *Kinsey made a survey* and *Let's survey the field*), confounding prominence at the lexical and phrasal levels; the speakers in Lieberman's study were asked to read only the target word, as they would pronounce it in the sentence. In a study on Australian English, Adams and Munro (1978) used read sentences and report duration as "by far the most frequently used cue" (137). Crystal and House (1988) focused specifically on duration in their study of American English; they also relied on read sentences and confirmed duration's role in signalling word stress.

In more recent research, Bettagere (2010) investigated acoustic characteristics of lexical and emphatic stress in American English; he used a word list for the former and simple sentences in which speakers were prompted with a question to place emphatic stress correctly for the latter. Duration was again confirmed to be a more important cue for signalling both levels of stress than *f*0 or amplitude. Fuchs (2016) analyzed acoustic characteristics of lexical and phrasal stress (Fuchs uses the distinction stress and accent, respectively). He used read sentences of Standard British English speakers and found that duration was a correlate of phrase-level but not word-level stress. In what may probably be regarded, from the viewpoint of the analyzed speech material, as one of the most naturalistic studies, Eriksson and Heldner (2015) compared the acoustic characteristics of stressed and unstressed vowels in semi-spontaneous speech, phrase reading, and isolated word reading. Even in this study, however, the semi-spontaneous interview was recorded in a sound-treated studio with the experimenter, without any real-life communicative intent.

The objective of this study will therefore be to determine whether temporal differences between stressed and unstressed syllables – or, more precisely, stressed and unstressed vowels – can be observed in truly spontaneous speech, delivered with a clearly defined audience in the mind of the speakers, namely in a corpus of political debates in British and American English.

1.2 Factors in vowel duration

It is not surprising that the duration of vowels in a language is influenced by multiple factors, and lexical stress is only one of them. That is why this section first summarizes studies concerning various factors that affect vowel duration; many of these have been discussed for instance by Klatt (1976) or van Santen (1992).

Remaining at the suprasegmental level, the duration of segments in general is affected by the position of the word within a prosodic phrase. The most widely examined of these effects is that of phrase-final deceleration (also referred to as phrase-final lengthening), with many studies confirming the finding that syllables at the ends of prosodic phrases tend to be longer in duration (e.g., Lehiste, 1972; Wightman et al., 1992; Byrd & Saltzman, 2003, among others). Specifically, Wightman et al. showed that the lengthening affects the rhyme of the phrase-final syllable and that its degree depends on the depth of the prosodic break; Crystal and House (1988) observed the effect of pauses, but also of final consonants.

Apart from prosodic influences, several factors affecting vowel duration have been documented at the level of individual segments and their interactions. It is logical that vowel length is a crucial factor: phonologically long vowels (including diphthongs) will on average be longer than phonologically short vowels. However, finer distinctions need to be mentioned as well. First, we have to account for what has been called intrinsic vowel duration: open vowels like [a ɑ] are known to have inherently longer duration than close vowels like [I u] (e.g., Peterson & Lehiste, 1960; House, 1961; Solé & Ohala, 2010). Second, vowel duration in English is significantly affected by the phonological voicing of the subsequent consonant. Since English maintains the phonological contrast between voiced and voiceless obstruents in the word-final position (e.g., *meat* and *mead*, *face* and *phase*) and phonetic voicing cannot serve as a reliable cue, the duration of the preceding vowel has become phonologized (Kohler, 1984; Kluender, Diehl, & Wright, 1988). Luce and Charles-Luce (1985) asserted that vowel duration is the most reliable correlate of phonological voicing of word-final stops. While the shorter duration of vowels before phonologically voiceless consonants than before voiced ones seems to be quasi-universal, vowel duration is exploited to a considerably larger extent in English as the major cue to final consonant voicing (Chen, 1970); this phenomenon is typically referred to as pre-fortis shortening.

Returning to our research questions, despite the extensive research on vowel duration and stress in English, there remains a gap in understanding these phenomena in truly spontaneous speech. This study aims to bridge this gap by examining the reported results concerning vowel duration in stressed and unstressed syllables, while also considering multiple segmental and prosodic factors.

2. Method

2.1 Material

As has already been mentioned above, our objective in this study was to analyze spontaneous speech delivered with a real communicative purpose to a clearly defined audience. These requirements are fulfilled by recordings of political debates. At the same time, we aimed to examine standard British and American speech, which would allow us to make at least tentative generalizations about the two major standard variants: Southern British English and the General American accent.

The recordings were obtained from publicly accessible archives of the BBC programme Westminster Hour (www.bbc.co.uk/programmes/b006s624) and C-SPAN network (www.c-span.org) for the British and American recordings, respectively. The material consisted of recordings of 16 speakers in total: eight British English speakers (four females, four males) and eight American English speakers (four females, four males). Within each group, neither the British nor the American speakers displayed any significant regional features in their speech.

For each speaker, the speech material analyzed in this study spanned approximately 200 words of spontaneous speech, which amounts to roughly 60–100 seconds per recording. In total, the material provided 4,927 tokens of stressed and unstressed vowels. The speech material is summarized in Table 1, which shows the number of words and vowels depending on the number of syllables.

Table 1 Summary of the analyzed speech material.

	number of syllables	word tokens	vowel instances
monosyllabic words	1	2507	
polysyllabic words	2	632	1264
	3	260	780
	4	74	296
	5	16	80

2.2 Analysis

The recordings were first transcribed and automatically segmented using the P2FA forced-alignment tool (Yuan & Liberman, 2008). Overlapping speech was excluded from subsequent analyses. All speech sound boundaries were manually adjusted based on phonetically motivated criteria (Machač & Skarnitzl, 2009). In the next stage, all syllables had to be labelled as either stressed or unstressed. Naturally, we were interested in actual realizations of stress, and not in canonical or dictionary forms; in some treatments, one would say that the task was to identify accented syllables (but see a different use of the term accent in Section 1.1). To do this, we used careful auditory analysis, relying on an alternation of broader- and narrower-context listening. Since the material consisted of spontaneous speech, it was natural that some ambiguous cases appeared, and these were resolved in a joint analysis session by both authors. Lastly, we identified prosodic boundaries – in ToBI (Beckman & Ayers Elam, 1997), these would correspond to both major (BI4) and minor (BI3) prosodic phrases – to distinguish between syllables in the phrase-final and non-final stress groups. All these analyses were conducted in Praat (Boersma & Weenink, 2024), which was also used to extract vowel durations and other relevant information using a dedicated script.

To evaluate the statistical significance of stress and other segmental and prosodic factors on vowel duration, we built a linear mixed-effects (LME) model using R (R Core Team, 2024) and the *lme4* package (Bates et al., 2015). Initially, we constructed a model with absolute vowel duration as the dependent variable which, however, when checked for homoscedasticity, suggested that the data were heteroscedastic. Of course, that is not surprising, since duration values are known to be positively skewed. Therefore, we subsequently constructed a model using log-transformed duration values as the dependent variable. As fixed effects, we included STRESS (with the levels being stressed or unstressed), the quality of the following segment (coded as FOLLOWING, with the levels open syllable, voiced coda, voiceless coda), VOWEL LENGTH (short, long, diphthong), VOWEL HEIGHT (close, mid, open), phrase-FINAL position (phrase-final or -internal), WORD LENGTH, and the VARIETY of English (British, American). Finally, as random effects, we included by-SPEAKER and by-WORD intercepts, as well as by-SPEAKER random slopes for the effect of STRESS. The complete structure of the random effects was thus (1 + Stress | Speaker) + (1 | Word); this accounts for the possibility that every speaker treats the difference between vowel durations in stressed and unstressed syllables differently. The statistical significance of the fixed effects was ascertained using likelihood ratio tests, comparing the fit of the complete model to that of a model without the given predictor. When appropriate, post-hoc pairwise comparisons were conducted using the *emmeans* package (Lenth, 2024): estimated marginal means (emmeans) were computed from the LME model, with Tukey's method applied to adjust for multiple comparisons.

A linear mixed effects model and likelihood ratio tests inform us about the significance of the individual factors – that is, about whether the factors' contribution to the overall model is significant. However, we were also interested in the relative contribution of the individual predictors, in how they compare in determining the final duration of the vowels. For that reason, we conducted a series of stepwise multiple linear regression (MLR) analyses using the *lm* function in R. Log-transformed vowel duration was used as the dependent variable, and the same factors as with LME served as predictors. We employed backward elimination and stepwise selection in both directions, using the Akaike Information Criterion (AIC) as the guiding metric to evaluate the contribution of each predictor to the model.

All plots were generated using the *ggplot2* package (Wickham, 2016). Note that the results of LME modelling are based on the log-transformed values of vowel duration. However, for the sake of more transparent interpretability of the results, the boxplots used to illustrate the effect of individual variables will depict absolute vowel durations.

3. Results and discussion

3.1 Statistical modelling of effects on vowel duration

The main objective of this study was to ascertain the difference in the duration of vowels depending on whether they appear in stressed or unstressed syllables in spontaneous speech. Our findings are consistent with previous research (see Section 1), with STRESS found to have a highly significant effect on vowel duration ($\chi^2(1) = 47.2$, $p < 0.0001$); detailed results of LME modelling are provided in the Appendix. Stressed vowels are indeed longer in duration than unstressed vowels, across both varieties of English, as can be seen in Figure 1. Note that the figure shows absolute duration values, whereas the statistical model is based on the log-transformed values of duration (see Section 2.2).

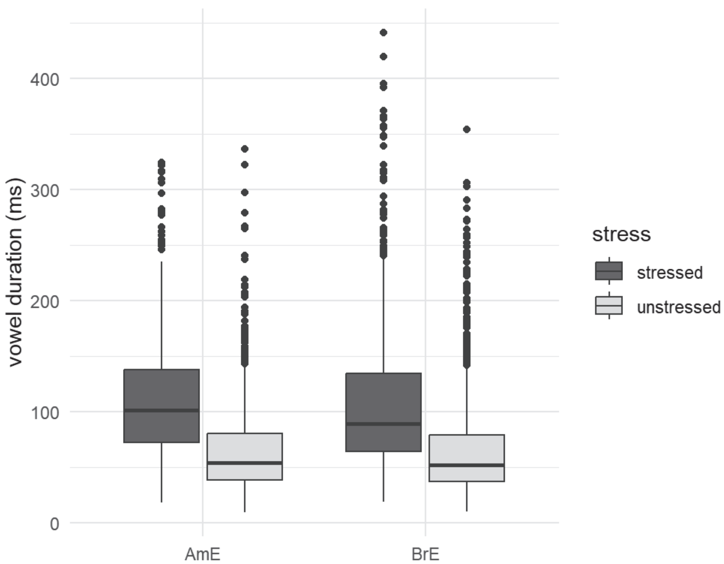

Figure 1 Absolute duration of stressed and unstressed vowels in the two varieties.

Vowel durations in stressed and unstressed syllables were not a priori expected to differ between British and American English. Although, based on inspecting Figure 1, there seems to be a small difference in the duration of stressed vowels, with those in American English slightly longer than in British English, VARIETY did not have a significant effect on vowel duration in our model ($\chi^2(1) = 2.7$, $p = 0.1$). Speaker identity was treated as a random factor in the analysis, but as seen in Figure 2, there were no prominent differences between the productions of individual speakers: the absolute duration of stressed vowels was consistently longer than that of unstressed vowels across speakers, and to a rather similar degree. It is interesting to observe the range of outlier values seen in Figure 2; this further showcases the variability and complexity of spontaneous speech. Given that no effect of VARIETY was observed in our model, this factor will not be included in subsequent visualizations.

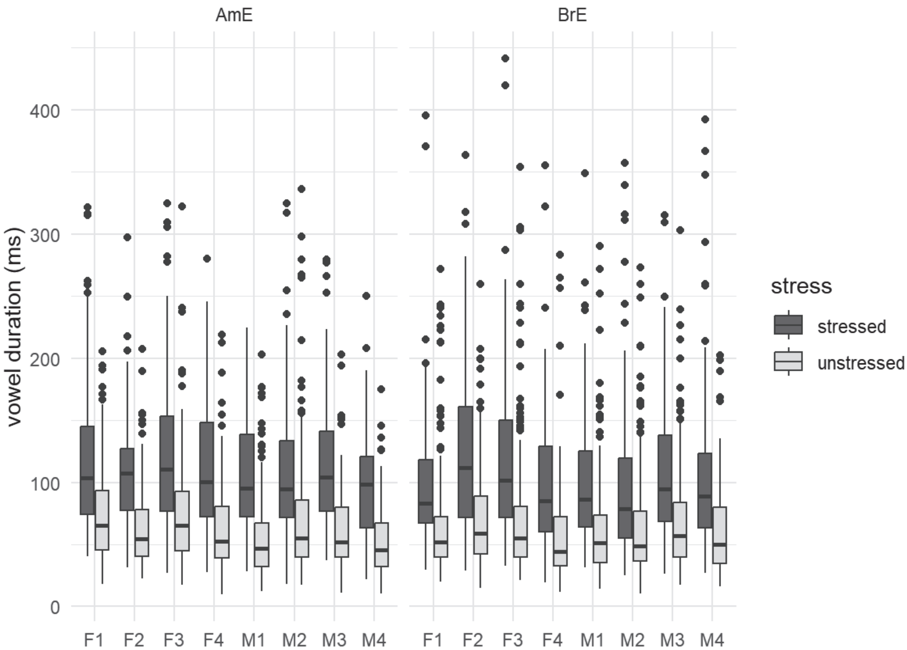

Figure 2 Absolute duration of stressed and unstressed vowels in individual speakers of the two varieties.

Along with stress, proximity to a prosodic boundary is one of the suprasegmental factors which is known to influence vowel duration (see Section 1.2). In our analysis of spontaneous speech, phrase-FINAL position proved to have a significant effect on (log-transformed) vowel duration ($\chi^2(1) = 661.4$, $p < 0.0001$). The effect can also be observed in the absolute duration values shown in Figure 3. As the graph shows, vowels are longer in the phrase-final position, and the difference is even more evident in stressed vowels; this is supported by the significance of the interaction between phrase-FINAL position and STRESS ($\chi^2(1) = 65.9$, $p < 0.0001$). It is important to note that we did not distinguish between prosodic break types (BI3 and BI4) or between phrase-final vowels followed by a pause and those occurring within a longer stretch of speech. It can be assumed that the presence of a pause could further increase the difference between the duration of phrase-final and phrase-internal vowels (both stressed and unstressed). This assumption would however be worth confirming through further research.

The next factor considered in our analysis was VOWEL LENGTH. Originally, we operated with a two-level factor (short and long vowels, where the latter included diphthongs). However, as shown in Figure 4, diphthongs turned out to be considerably longer than long monophthongs in our data. For that reason, a three-way distinction was used in the model, which proved that VOWEL LENGTH has a significant effect on vowel duration ($\chi^2(2) = 573.0$, p < 0.0001). Figure 4 also shows that the effect of phonological vowel length is consistent across stressed and unstressed vowels, and the interaction between STRESS and VOWEL LENGTH is not significant ($\chi^2(2) = 2.0$, $p > 0.3$).

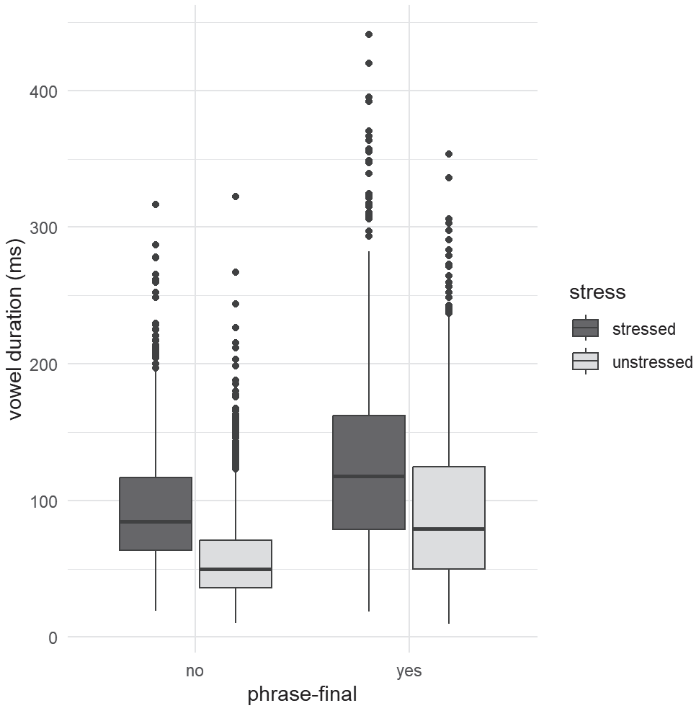

Figure 3 Absolute duration of stressed and unstressed vowels in phrase-final and phrase-internal positions.

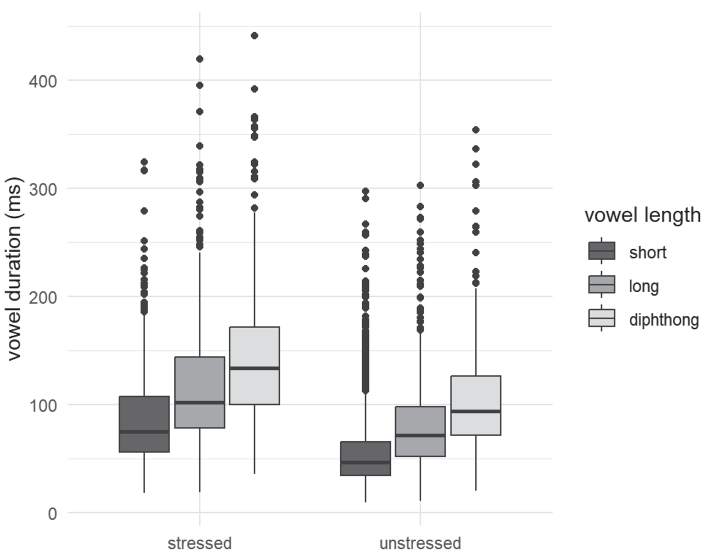

Figure 4 Absolute duration of stressed and unstressed vowels, depending on vowel length.

Another effect that we considered was intrinsic vowel duration, coded as VOWEL HEIGHT in this analysis, which was found to be significant ($\chi^2(2) = 184.4, p < 0.0001$). The results shown in Figure 5 suggest that the relationship between vowel duration and vowel height differs within stressed and unstressed syllables. This is also confirmed by the significant interaction between STRESS and VOWEL HEIGHT ($\chi^2(2) = 40.4, p < 0.0001$). To be specific, vowel duration varies systematically across the three height categories in stressed syllables in accordance with tendencies reported in the literature (see Section 1.2). However, within unstressed syllables, mid vowels are shorter in duration than high vowels, although a post-hoc pairwise comparison, calculated using the *emmeans* function with Tukey's adjustment, shows that the difference falls short of statistical significance ($p > 0.2$). This tendency can presumably be explained by the fact that the unstressed mid vowels include a lot of schwas, which are likely to be most reduced not only spectrally but also in the temporal domain, and thus would be shorter in duration than high vowels.

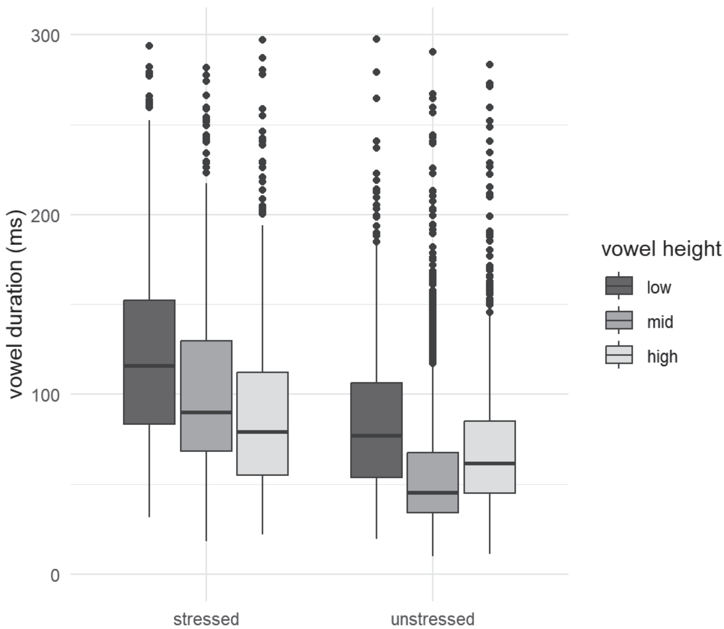

Figure 5 Absolute duration of stressed and unstressed vowels, depending on vowel height.

The next effect we were interested in examining was that of WORD LENGTH, expressed as the number of syllables in a word. Since English is a language traditionally referred to as stress-based, which involves the temporal compression of unstressed syllables within a stress group, increasing word length should, other things being equal, be reflected in shorter vowel durations. Our analysis supports this: the effect of WORD LENGTH turned out to be significant ($\chi^2(1) = 75.8, p < 0.0001$). More detailed results, presented in Figure 6, reveal a clear difference between the absolute durations of vowels in monosyllabic and polysyllabic words. More specifically, it is particularly stressed vowels in monosyllabic words which are longer in duration than those in polysyllabic words. However, post-hoc pairwise comparisons show that both stressed and unstressed vowels in disyllabic words are significantly longer than those in four-syllabic words ($p < 0.05$); the comparison of di- and three-syllabic words was only marginally significant ($p = 0.08$).

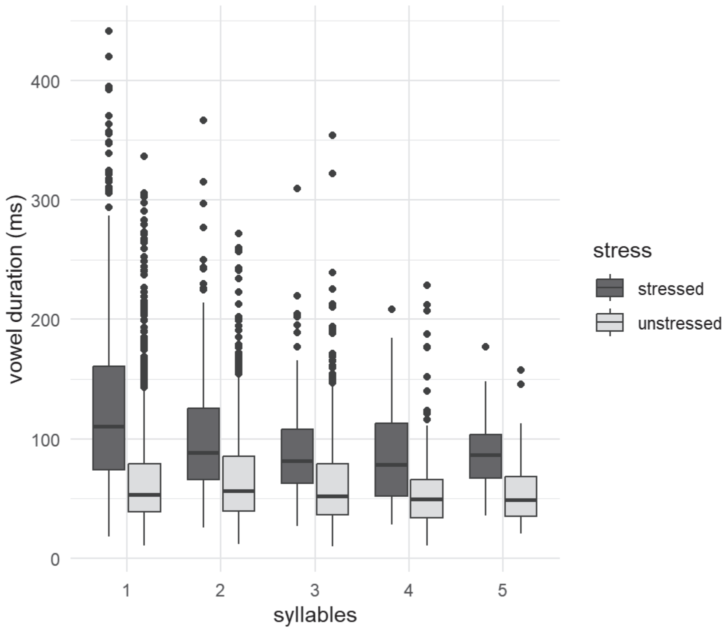

Figure 6 Absolute duration of stressed and unstressed vowels, depending on the number of syllables in a word.

The last factor, whose effect on vowel duration we aimed to examine, was the nature of the FOLLOWING segment. As described in Section 1.2, English uses vowel duration to cue the difference between fortis and lenis obstruents in the coda. It is therefore not surprising that our spontaneous speech data confirm the significance of this factor ($\chi^2(2) = 52.7$, $p < 0.0001$). Note that the voiced group comprises both lenis obstruents and sonorants in the coda (i.e., *maid* as well as *main*). A more detailed analysis of Figure 7 suggests that the significance of the factor FOLLOWING is mostly due to the longer duration of vowels in open syllables (e.g., *May*), rather than by pre-fortis shortening; however, post-hoc pairwise comparisons (conducted on the log-transformed values of the LME model) confirm the statistical significance between pre-voiced and pre-voiceless (fortis) vowels in both stressed and unstressed syllables as well ($p < 0.001$).

It is interesting to probe the duration of pre-fortis and pre-voiced vowels further because, despite the significance of the pairwise comparisons reported above, we expected to see a greater difference in a contrast which cues phonological distinctions with a high functional load. In Figure 8, we only show the absolute duration data for monosyllabic words, and it is obvious that the duration difference is considerably more pronounced in the stressed syllables. In other words, the effect of pre-fortis shortening is most salient in stressed monosyllabic words (*mate* ['mĕɪt] and *maid* ['meɪd̥]), while it may

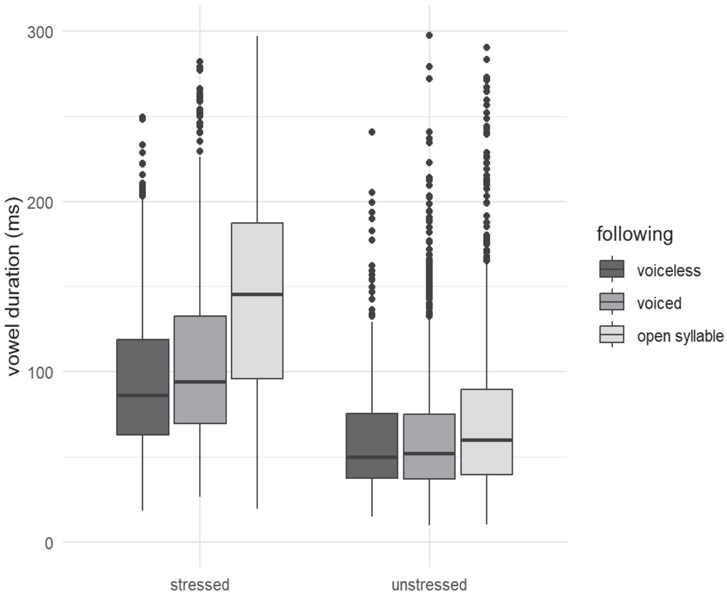

Figure 7 Absolute duration of stressed and unstressed vowels, depending on the nature of the following segment.

be less salient in longer word pairs (for example, *fickle* ['fĭkḷ] and *figure* ['fɪgə] or *sightline* ['săɪtlaɪn] and *sideline* ['saɪdlaɪn]).

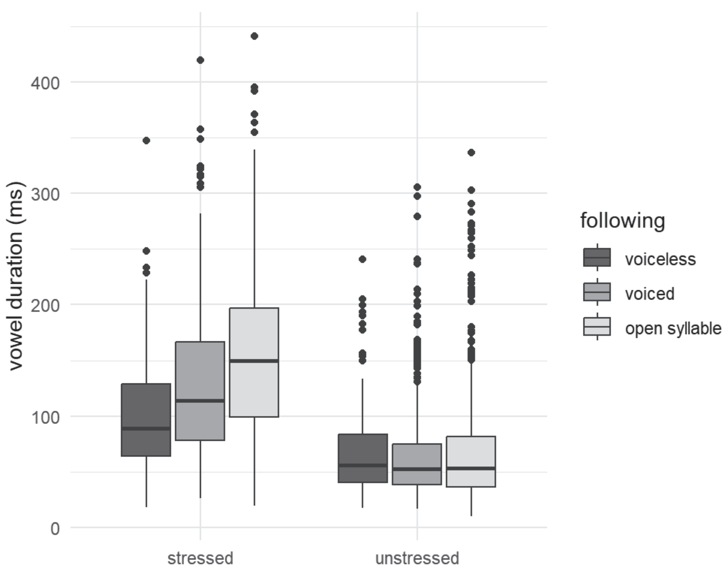

Figure 8 Absolute duration of stressed and unstressed vowels in monosyllabic words, depending on the nature of the following segment.

Having discussed the significance of each of the factors individually using LME modelling, in the next section we will assess their relative importance for vowel duration using multiple regression analysis (MRA). The drawback of MRA is, however, that random factors like the speakers' identity are not considered.

3.2 Relative contribution of predictors to vowel duration

As mentioned in Section 2.2, the log-transformed vowel duration functioned as the dependent variable in the MR analysis. We used the Akaike Information Criterion (AIC) to assess the relative importance of the factors affecting vowel duration. The AIC compares different models (i.e., models with different sets of predictors) to each other, whereby the model with the lowest AIC is regarded to be the best trade-off between accuracy (goodness of fit) and complexity (the number of predictors). An accompanying metric, frequently reported in MLR analysis, is R^2_{adj} (the adjusted coefficient of determination), which is used to evaluate the explanatory power of a regression model; more specifically, it corresponds to the percentage of variance explained by the model, adjusted for the number of predictors used in the model.

The overall model, with all predictors included, was significant: $F(10, 4909) = 395.4$, $p < 0.0001$, $R^2_{adj} = 0.445$. In other words, the complete model explained 44.5% of the variance in the vowel duration data. Through a bidirectional stepwise analysis, the metric ranks the predictors according to the degree of explained variance in the MLR model from highest to lowest. The AIC values of predictors and the corresponding R^2_{adj} values, listed in the order of explained variance, are presented in Table 2 and visualized in Figure 9.

Table 2 Results of bidirectional stepwise MLR, with AIC and R^2_{adj} values in decreasing order of explained variance (see text).

predictor	AIC	R^2_{adj}
intercept	−5145.9	
vowel length	−6430.4	0.230
phrase-final	−7178.6	0.339
stress	−7724.5	0.408
vowel height	−7906.9	0.430
syllables	−8005.5	0.441
variety	−8019.0	0.443
quality of the following segment	−8032.5	0.445

Starting with the null model with only the intercept value, adding vowel length results in a considerable decrease in AIC, indicating that vowel length alone accounts for a significant amount of explained variance of vowel duration, approximately 23%. The addition of the phrase-final condition further lowers the AIC and increases the degree of explained variance to 33.9%. Incorporating the affiliation of the vowel to a stressed vs. unstressed syllable adds another nearly 7% of explained variance, and subsequently including vowel height brings the explained variance to 43%.

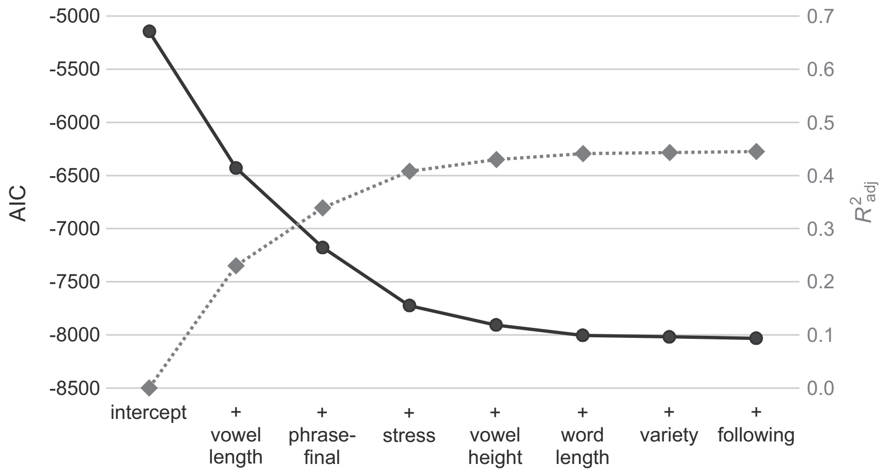

Figure 9 Changes in the Akaike Information Criterion (AIC, in black circles, axis on the left) and the adjusted coefficient of determination (R^2_{adj}, in grey diamonds, axis on the right) in the multiple regression model (see text).

The remaining three predictors only provide minor changes to AIC and therefore contribute less new information to the model. If we were aiming for a parsimonious but still effective model of vowel duration in our spontaneous English data, the following four predictors should be included:

- vowel length: the distinction between short monophthongs, long monophthongs, and diphthongs affects vowel duration to the largest extent
- position within the phrase: vowels in the last stress group of a prosodic phrase are significantly longer than phrase-internal vowels
- stress: vowels in stressed syllables are longer than those in unstressed syllables
- vowel height: the correlate of intrinsic vowel duration, with open (low) vowels longer than close (high) vowels

4. General discussion

The present study examined vowel duration as the primary acoustic correlate of lexical stress in English. Its primary aim was to determine whether the long-held relationship between stress and vowel duration, which was observed on more or less controlled speech materials, would be verified in truly spontaneous speech, namely in political debates broadcast in the United Kingdom and the United States. In the most general sense, the study thus examined whether stressed vowels are longer in duration than unstressed vowels even in spontaneous English. In addition, we wanted to see how other factors, known to affect the duration of vowels, modulated their relationship with lexical stress.

Our results confirm that vowel duration may be considered an important cue for the distinction between stressed and unstressed syllables; this relationship holds both globally (Figure 1) and at the level of individual speakers (Figure 2). At the same time, no difference was observed between our British and American speakers, although Figure 1 suggests a tendency for stressed syllables to be shorter in British English; this may confirm the impression that the British politicians were, overall, speaking slightly faster than the American ones.

As for the modulating factors on the segmental level, our analysis included the effect of vowel length (considering the temporal differences of short and long monophthongs and diphthongs), intrinsic vowel duration (conceptualized as vowel height), and the shortening of vowels before fortis consonants (which is used in English to cue the voicing contrast in the syllabic coda). All three segmental factors were found to affect vowel duration, in line with the findings reported in the literature (see Section 1.2). However, the relationship between these factors and stress contrast differs. The duration difference between stressed *and* unstressed vowels holds for vowel length (Figure 4), and the importance of this factor is confirmed by its first ranking in the stepwise multiple regression analysis (cf. Table 2 and Figure 9). With vowel height, the exceptionally short nature of *schwa* vowels blurs the relationship somewhat but, overall, the relationship also holds (Figure 5). The effect of coda fortis consonants was the least obvious: it is manifested to the greatest extent in the stressed syllables of monosyllabic words (Figures 7 and 8).

Regarding prosodic factors that modulate the effect of lexical stress on vowel duration, we focused on phrase-final deceleration and word length. Our results suggest that vowels in phrase-final stress groups were longer than those in phrase-internal stress groups, and this difference affects both stressed and unstressed vowels (Figure 3). In comparison, vowel duration operates slightly differently in monosyllabic words, with particularly stressed vowels being considerably longer in monosyllabic words (see Figure 6).

While this study examined the relations between the presence of a prosodic boundary and vowel duration, it is important to note that we did not differentiate various depths of prosodic boundary (ToBI 3 and 4) or between phrase-final stress groups that were followed by a pause and those that were immediately followed by another prosodic phrase. For a complete understanding of phrase-final deceleration, it would be worth adding another level of the phrase-FINAL factor, which would correspond to vowels in prepausal stress groups, and determining whether its effect on vowel duration in spontaneous speech would be greater.

As emphasized throughout this study, the purpose of our research was to analyze real spontaneous speech. Since spontaneous speech is a highly complex phenomenon and many differences have been observed between canonical forms and spontaneous realizations, we had expected our main findings to be less clear-cut. However, our results demonstrate that duration is indeed a crucial cue to the stressed vs. unstressed distinction in English spontaneous speech – naturally, along with other correlates like peripheral vs. reduced (centralized) vowel quality.

Unsurprisingly, the analysis of spontaneous speech proved to be methodologically laborious; this concerns both phonetic segmentation (where segments one expects may not be realized at all) and the identification of stressed (accented) syllables. It is obvious that a meticulous auditory and acoustic analysis (see Section 2) was essential to carry out these steps and make our analysis valid. Although the extent of the speech material which could be examined for this study was relatively restricted, we believe that it was sufficient for the results to be reliable and that the presented study provides interesting insight into stress and vowel duration in spontaneous English.

REFERENCES

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica*, *35*, 125–126. https://doi.org/10.1159/000259926

Barry, W., & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, *31*(1), 51–66. https://doi.org/10.1017/S0025100301001050

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

Beckman, M.E., & Ayers Elam, G. (1997). *Guidelines for ToBI labelling*, version 3. The Ohio State University Research Foundation.

Bettagere, R. (2010). Differences in acoustic characteristics of stress patterns in American English. *Perceptual and Motor Skills*, *110*(2), 339–347. https://doi.org/10.2466/pms.110.2.339-347

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer (Version 6.4)*. Retrieved from www.praat.org

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

Cauldwell, R. (2013). *Phonology for listening: Teaching the stream of speech*. Speech in Action.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*, 129–159. https://doi.org/10.1159/000259312

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America*, *83*, 1574–1585. https://doi.org/10.1121/1.395912

Eriksson, A., & Heldner, M. (2015). The acoustics of word stress in English as a function of stress level and speaking style. *Proceedings of Interspeech 2015*, 41–45. https://doi.org/10.21437/Interspeech.2015-9

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*(4), 765–768. https://doi.org/10.1121/1.1908022

Fuchs, R. (2016). The acoustic correlates of stress and accent in English content and function words. *Proceedings of Speech Prosody 2016*, 435–439. https://doi.org/10.21437/SpeechProsody.2016-89

Greenberg, S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176. https://doi.org/10.1016/S0167-6393(99)00050-3

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, *33*(9), 1174–1178. https://doi.org/10.1121/1.1908941

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.), *Proceedings of the first session of the 10th international symposium on spontaneous speech: Data and analysis* (pp. 29–54). The National International Institute for Japanese Language.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59*, 1208–1221. https://doi.org/10.1121/1.380986

Kluender, R. K., Diehl, R. L., & Wright, B. A. (1988). Vowel length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, *16*, 153–169. https://doi.org/10.1016/S0095-4470(19)30480-2

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, *41*, 150–174. https://doi.org/10.1159/000261721

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*(6B), 2018–2024. https://doi.org/10.1121/1.1913062

Lenth, R. (2024). *emmeans: Estimated Marginal Means, aka least-squares means*. R package version 1.10.3. Retrieved from https://CRAN.R-project.org/package=emmeans

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 451–454. https://doi.org/10.1121/1.1908095

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, *78*(6), 1949–1957. https://doi.org/10.1121/1.392651

Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Epocha.

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language*, *22*, 171–184. https://doi.org/10.1016/j.csl.2007.07.003

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693–703. https://doi.org/10.1121/1.1908183

R Core Team (2024). *R: A language and environment for statistical computing* (version 4.4.2). R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Solé, M. J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (eds.), *Laboratory phonology 10* (pp. 607–655). De Gruyter Mouton. https://doi.org/10.1515/9783110224917.5.607

Tucker, B. V., & Mukai, Y. (2023). *Spontaneous speech*. Cambridge University Press.

van Heuven, V. J. (2019). Acoustic correlates and perceptual cues of word and sentence stress: towards a cross-linguistic perspective. In R. Goedemans, J. Heinz, & H. van der Hulst (eds.), *The study of word stress and accent: theories, methods and data* (pp. 15–59). Cambridge University Press. https://doi.org/10.1017/9781316683101.002

van Santen, J. P. H. (1992). Contextual effects on vowel duration. *Speech Communication*, *11*, 513–546. https://doi.org/10.1016/0167-6393(92)90027-5

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

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

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*, 5687–5690.

APPENDIX

Results of the linear mixed-effects (LME) model

```
Linear mixed model fit by REML ['lmerMod']
Formula: logdur ~ stress + variety + following + vowel_length + vowel_heigh
t + final + word_length + (1 + stress | speaker) + (1 | word)
   Data: data

REML criterion at convergence: 5592.1

Scaled residuals:
```

Min	1Q	Median	3Q	Max
−4.2128	−0.6209	0.0008	0.6171	4.0259

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
word	(Intercept)	0.028624	0.16919	
speaker	(Intercept)	0.007411	0.08609	
	stressunstressed	0.003076	0.05546	0.10
Residual		0.159165	0.39895	

Number of obs: 4920, groups: word, 1271; speaker, 16

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	5.011059	0.047249	106.055
stressunstressed	−0.304312	0.020069	−15.163
varietyBrE	−0.083741	0.047242	-1.773
followingvoiced	−0.140363	0.023037	−6.093
followingvoiceless	−0.180285	0.024891	−7.243
vowellengthlong	−0.220653	0.026139	−8.441
vowellengthshort	−0.513550	0.022568	−22.756
vowelheightlow	0.212663	0.023517	9.043
vowelheightmid	−0.065959	0.017435	−3.783
finalyes	0.399319	0.014973	26.670
wordlength	−0.075482	0.008593	−8.784

RESUMÉ

Řada fonetických „pravidel“ je založena na analýzách kontrolovaného materiálu nebo z dat získaných cíleně pro experimentální účely. Pro úplné porozumění těmto přijímaným „pravdám“ je však nezbytné ověřit jejich platnost i ve spontánní řeči. Tato studie se zabývá trváním samohlásek jako akustickým korelátem slovního přízvuku v angličtině ve spontánní řeči produkované v komunikačně motivovaných kontextech. Za cíl si klade ověřit dřívější poznatky o souvztažnosti přízvuku a trvání vokálů, a to v nahrávkách osmi amerických a osmi britských mluvčích účastnících se politických debat. Konkrétně studie ověřuje hypotézu, že přízvučné vokály mají signifikantně delší trvání než vokály nepřízvučné. Pomocí lineárních smíšených modelů naše analýza potvrzuje statisticky významný vliv slovního přízvuku na trvání vokálů. Výzkum zároveň zohledňuje další faktory ovlivňující segmentální trvání, jako jsou

fonologická délka daných vokálů (tedy zda se jedná o krátký či dlouhý monoftong nebo diftong), pozice v rámci prozodické fráze, vokalická výška jakožto korelát inherentního trvání nebo charakter následujícího segmentu. Z krokové vícenásobné regresní analýzy vyplývá, že kromě přízvuku jsou nejvýznamnějšími prediktory trvání vokálů fonologická délka vokálu, finální pozice v prozodické frázi a vokalická výška; tyto čtyři faktory vysvětlují 43 % variability v trvání našich samohlásek. Navzdory vysoké variabilitě charakteristické pro spontánní řeč se trvání vokálů ukazuje jako spolehlivý korelát přízvuku, což potvrzuje dřívější závěry získané z výzkumu kontrolované řeči.

Nela Bradíková
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
bradikovanela@gmail.com

Radek Skarnitzl
Department of Czech Language and Literature
Faculty of Education
University of Hradec Králové
radek.skarnitzl@uhk.cz