Polysyllabic shortening and word-final lengthening in English

Andreas Windmann[1], *Juraj Šimko*[2], *Petra Wagner*[1]

[1]Faculty for Linguistics and Literary Studies, Bielefeld University, Germany
[2]Institute of Behavioral Sciences, University of Helsinki, Finland
[1]andreas.windmann@uni-bielefeld.de,petra.wagner@uni-bielefeld.de [2]juraj.simko@helsinki.fi

Abstract

We investigate Polysyllabic shortening effects in three prosodic domains, the word, the inter-stress interval (ISI) and the narrow rhythm unit (NRU), in a large corpus of English broadcast speech. Results confirm and extend earlier findings, indicating that these effects are interpretable as artifacts of word-final lengthening. We do, however, find effects compatible with the assumption of eurhythmic principles in speech production.
Index Terms: speech timing, speech rhythm

1. Introduction

Polysyllabic shortening (PS) denotes the alleged property of syllable or vowel duration to be inversely related to the number of syllables in some larger prosodic unit. For example, [2] found that /iː/ is shorter in *speedy* than in *speed*, and shorter still in *speedier*. Similar effects have been observed in other languages as well [4, 5, 6]). These effects appear to suggest a tendency on part of speakers to keep durations of larger prosodic units constant. In particular, PS in the interval between the onsets of two consecutive stressed syllables (Inter-Stress Interval; ISI) is predicted by the isochrony hypothesis for "stress-timed" languages such as English, Dutch or Swedish, which were claimed by this hypothesis to place stressed syllables at temporally regular intervals [7, 8]. Subsequent research (e.g. [9]) has falsified the strong form of this hypothesis, showing for various languages that ISI duration is a linear function of the number of component syllables. Yet, as pointed out by [10], this does not preclude the existence of PS effects in the ISI.

[11] have suggested an alternative interpretation. They point out that previous experimental evidence for PS came almost exclusively from pitch-accented words, and find that in English, PS at the word level is indeed near-absent in unaccented contexts. They take this to suggest that the effect is an epiphenomenon of accentual lengthening: since pitch accent lengthens all syllables in accented words in English, the accentual lengthening, according to this interpretation, has to be "shared out" among the individual syllables in longer words. On this view, PS would be a mere corollary of word prominence, and does not require an explanation based on tendencies towards periodicity. One caveat is that [11] do observe a very subtle but statistically reliable effect in the direction of PS also in unaccented words with initial stress.

Some corpus studies have found PS at the ISI level in English, compatible with weak versions of the isochrony hypothesis [12, 13, 14, 15]. Results by [16], however, suggest that this correlation only holds as long as word boundary locations are not taken into account. [16] reports that the duration of stressed vowels in accented words is not a function of the number of syllables in the word or in the ISI, but of the number of syllables in the interval between the stressed syllable onset and the right boundary of the word.

Results by [11] and [16] are compatible with two interpretations. One is that they are indicative of genuine PS effect which does not operate at the word or ISI level, but on the interval between the onset of a stressed syllable and the following word boundary. This interval has been termed *word rhyme* [11] or *Narrow Rhythm Unit* (NRU) [17, 18], and [17] explicitly proposed it as the domain of a temporal equalization process in English. According to this theory, speakers of English attempt at regularizing NRU duration, which would predict a PS effect in this unit. Syllables not contained in an NRU, i.e., unstressed syllables occurring before the main stress of the word they are part of and thus within the so-called *anacrusis*, are produced as rapidly as possible according to [17]'s model.

An alternative interpretation, suggested by [11], is that the observed pattern is the result of a progressive word-final lengthening effect: vowels are longest when they are directly adjacent to a word boundary, and become shorter with added intervening syllables. For stressed vowels, both hypotheses are indistinguishable – the number of syllables between the stressed syllable onset and the right word boundary is, by definition, the same as the syllable count in the NRU. For unstressed vowels, however, it is possible to pit the number of syllables in the NRU against the number of syllables to the right word boundary: the NRU compression hypothesis predicts the vowel in the word-final syllable to be shorter in "Minister" (trisyllabic NRU) than in "Mister" (bisyllabic NRU), whereas the progressive word-final lengthening hypothesis predicts no such difference.

Results of a corpus analysis by [19] appear to favor the word-final lengthening hypothesis: PS in the NRU, as observed in earlier studies on the same data [18, 20] does no longer hold if NRU-initial, -final and -medial phones are analyzed separately. However, this study did not control for prominence, and, additionally, conflated consonants and vowels. Since the final phone of an NRU will presumably often be a coda consonant, it is not clear how to interpret the result of the study, given that [11] found only nuclei, not coda consonants to shorten with syllable count in the NRU (or distance to the right word boundary). In this paper, we will report a reanalysis of the same data, in order to assess possible influences of NRU length on vowel duration whilst controlling for prominence and positional factors. We will also investigate possible effects of syllable count in the word and in the ISI, in order to provide a replication of the studies by [11] and [16] on somewhat more naturalistic data. The remainder of this paper is structured as follows: in Section 2, we discuss the data and methods used in our study. Results of the analysis are presented in Section 3 and discussed in Section 4. Section 5 presents concluding remarks.

2. Data and Method

The data in our study come from the Aix-MARSEC corpus [21]. It comprises approximately $5^1/_2$ hours of automatically segmented and prosodically transcribed broadcast speech, produced by 17 male and 36 female speakers of British English. Analyses were carried out on vowel durations, using the existing segmentation of the data. A number of measures were taken in order to avoid confounding of results. Vowels from utterance-initial and final words were excluded from the analysis, so as to avoid potential effects of initial and final lengthening. We also discarded data from a number of words in the corpus for which stress was marked on more than one syllable, as it is not clear how to define units such as the ISI in such cases. Analyses were carried out on vowel durations z-normalized by phoneme label, in order to factor out inherent vowel duration differences.

We controlled for two prosodic variables: first, a variable termed PROMINENCE was defined using the existing prosodic transcription of the corpus, comprising three prominence levels, stressed accented (1), stressed unaccented (2) and unstressed (3). These will be referred to as *S +Acc*, *S -Acc* and *U*, respectively. Second, we identified word-final vowels and defined a control variable WITHIN-WORD POSITION with the levels *final* and *non-final*. Vowels from monosyllabic words were counted as final. The experimental variables of interest, finally, were the number of syllables in the ISI, word, and NRU. We did not control for any other variables, such as the phonological environment of a vowel, syllable type, word class, or between-speaker variation. We assume that such potential confounding variables not accounted for should be randomly distributed with respect to our experimental variables, or that they should cancel each other out to some extent. As we shall see, our control methods are rigorous enough to yield consistent results, which are in good agreement with findings from more controlled studies.

Vowel durations were analyzed using quantile regression, as implemented in the R package *quantreg* [22]. Quantile regression allows for computing median estimates, which arguably yields a more accurate representation of vowel durations than techniques that provide mean estimates, as vowel duration distributions typically exhibit a considerable positive skew. We applied a stepwise analysis procedure: first, we fitted a model with the factors PROMINENCE (*S +Acc/S -Acc/U*) and WITHIN-WORD POSITION (*final/non-final*) to the data. This model will be referred to as *basicmodel*. We then created a dummy variable, referred to as CONTROL, which comprised all combinations of factor levels of PROMINENCE and WITHIN-WORD POSITION. In a second analysis step, we constructed three separate regression models, one for each of the three constituent types, word, ISI, and NRU. In each of these models, slopes for vowel duration by syllable count in the respective constituent type were nested within the levels of CONTROL, using R's "/" operator [23]. Thus, PS effects were tested separately within the subsets of the corpus defined by the combinations of PROMINENCE and WITHIN-WORD POSITION, so that confounding by these factors was eliminated.

The three models used for testing the individual constituent types will be referred to as *wordmodel*, *isimodel*, and *nrumodel*. Data from cells as defined in these models that contained less than 100 observations were discarded. After all exclusions, approximately 40000 vowels remained to be analyzed in *wordmodel* and *isimodel*. For *nrumodel*, there were additional exclusions, as detailed below. We applied Bonferroni correction, hence $\alpha = 0.002$ (9 comparisons in *basicmodel* + 16 nested slopes in the other three models = 25 individual comparisons).

3. Results

We will begin by discussing the results of *basicmodel*. Planned comparisons showed non-word-final *S +Acc* vowels to be significantly longer ($t = 3.93; p < 0.0001$) and *S -Acc* vowels to be significantly *shorter* ($t = -7.84; p < 0.0001$) than *U* vowels in the same position. In word-final position, *S +Acc* vowels are also longer than *U* vowels ($t = 17.84; p < 0.0001$); the difference between word-final *S -Acc* and word-final *U* vowels is not significant at the $\alpha = 0.002$ level ($t = 2.40; p = 0.016$). *U* vowels are longer in word-final than non-word-final position ($t = 14.67; p < 0.0001$), as are *S +Acc* ($t = 19.20; p < 0.0001$) and *S -Acc* ($t = 15.10; p < 0.0001$) vowels. There is also evidence for an interaction: the durational difference between *S +Acc* and *U* vowels is greater ($t = 8.66; p < 0.0001$) in word-final than in non-word-final position. The durational difference between *S -Acc* and *U* vowels tends to be smaller in non-word-final than in word-final position ($t = -2.58; p = 0.009$). Thus, there are reliable effects of prominence and within-word position on vowel duration and an interaction between both. Results are graphed in Figure 1.

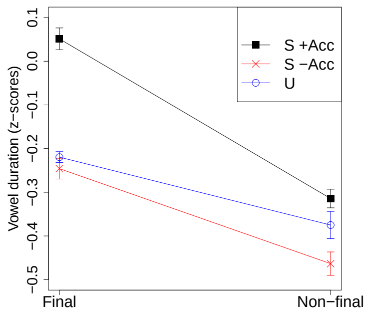

Figure 1: *Z-normalized vowel duration (medians and 95% confidence intervals) by prominence and within-word position in the Aix-MARSEC corpus.*

The surprising finding of greater *U* than *S -Acc* vowel durations is most likely an artifact of the z normalization: as is well-known, the distribution of English vowel phonemes in stressed and unstressed syllables is near-complementary; most vowel phonemes in the corpus appear almost exclusively either in stressed or in unstressed syllables (with the exception of short high vowels, for which stress indeed makes little difference). Since the z-score normalization sets the mean durations of all vowel phonemes to zero, duration differences between stressed and unstressed vowels are largely eliminated. The category of *S -Acc* vowels in particular comprises mainly those observations from the lower end of the stressed vowel duration distribution, so that *S -Acc* vowels appear to be even shorter than *U* vowels. Lexically stressed and unstressed vowel durations are thus not directly comparable using our data and method.

Figure 2 graphs z-normalized vowel durations (medians and 95% confidence intervals) by syllable count in the ISI (the interval between two consecutive stressed syllable onsets), for all data (left panel), and separately for word-final (middle panel) and non word-final vowels (right panel). The different colors denote the three levels of prominence. The individual trajectories in the middle and right panel correspond to the nested slopes in *isimodel*. This way of presenting the data highlights the benefits of our nested analysis: as long as the data are pooled across within-word positions, there seem to be clear effects of syllable count in the ISI, especially in S +Acc vowels. Once within-word position is controlled, a different picture emerges: for word-final vowels, there is some evidence compatible with PS at the ISI level in *U* and *S-Acc* vowels, which is corroborated by *isimodel* yielding significant negative slopes for syllable count in the ISI in word-final *U* ($t = -5.04; p < 0.0001$) and, tentatively, in word-final *S -Acc* vowels ($t = -3.15; p < 0.05$; note that these effects may be underestimated by the slopes of our model, which assume linear effects of duration by syllable count). None of the remaining nested slopes are significant.

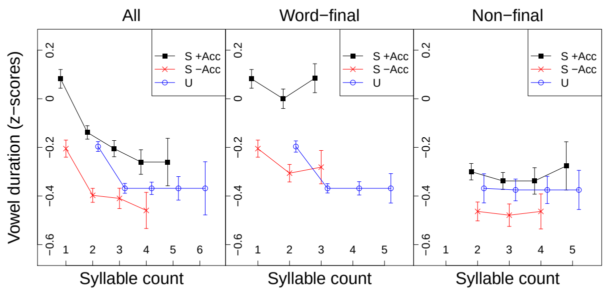

Figure 2: *Z-normalized vowel duration by prominence level, within-word position and syllable count in the ISI.*

Figure 3 graphs z-normalized vowel durations from the MARSEC corpus as a function of the number of syllables in the *word*, in the same fashion as Figure 2 above. The pattern of results is the same as in the ISI analysis: as long as within-word position is not controlled, there seems to be a shortening effect of the number of syllables in the word on vowel duration, particularly for *S +Acc* vowels. Once within-word position is controlled, the effect of word length on vowel durations turns out to be essentially random. None of the nested slopes for syllable count in *wordmodel* turned out significant at $\alpha = 0.002$. This, by the way, shows that our policy of counting vowels from monosyllabic words as final is justified, as they pattern durationally with final vowels from longer words. There is a quite distinct durational pattern in word-final *U* vowels. We will provide a possible interpretation of this pattern below.

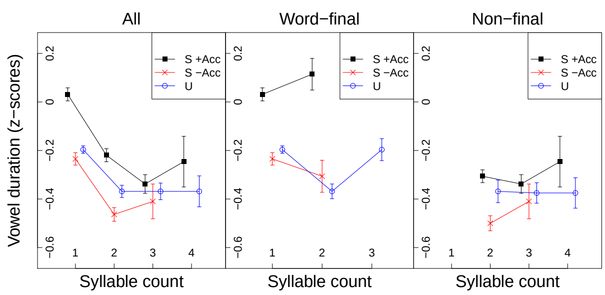

Figure 3: *Z-normalized vowel duration by prominence level, within-word position and syllable count in the word.*

Figure 4 graphs z-normalized vowel durations from the MARSEC corpus as a function of the number of syllables in the *NRU* (the interval between the onset of a stressed syllable and the following word boundary), in the same fashion as in Figures 2 and 3 above. For the NRU analysis of unstressed vowel duration, we excluded observations from syllables in *anacruses*, i.e., unstressed syllables occurring before the stressed syllable within a word (or within words that do not contain a stressed syllable at all), as these are not part of the NRU according to [17]'s model. Since syllable count in the NRU is, by definition, one for word-final stressed vowels, these were also excluded from *nrumodel*, but are shown in Figure 4. These exclusions lead to approximately 19000 observations being included in *nrumodel*.

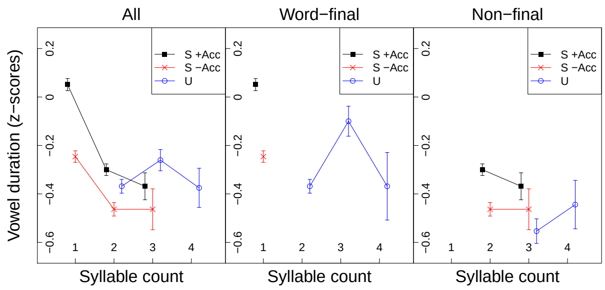

Figure 4: *Z-normalized vowel duration by prominence level, within-word position and syllable count in the NRU.*

Inspection of Figure 4 tentatively suggests a progressive word-final lengthening effect in *S +Acc* vowels, in accordance with results by [16] (recall that for stressed vowels, syllable count in the NRU is isomorphic to the number of syllables between the vowel and the right word boundary). The difference between *S +Acc* vowels from bi- and trisyllabic NRU, and, hence, between penultimate and antepenultimate *S +Acc* vowels, however, is not significant at $\alpha = 0.002$ ($t = -1.93; p = 0.054$). Crucially, there are no effects compatible with PS at the NRU level in unstressed vowels. The nested slope for word-final *U* vowels suggests a *lengthening* effect ($t = 10.56; p < 0.0001$), but graphical presentation of results in the middle panel of Figure 4 indicates a more complex pattern, similar to the result observed in the word-level analysis.

Figure 5, finally, clarifies why PS effects are observed in uncontrolled data: shown are the percentage of word-final vowels as a function of constituent length in the MARSEC corpus: for example, 100% of the stressed vowels in monosyllabic ISI come from word-final syllables, which is not surprising, given that a monosyllabic ISI is defined as a primary stressed syllable followed by another primary stressed syllable, so that there is necessarily a word boundary intervening. In bisyllabic ISI, this proportion is only about 60% for stressed syllables, and it decreases further with increasing ISI length. The resulting trajectories bear a striking resemblance to the durational results obtained without controlling for within-word position, particularly for stressed vowels. As for unstressed vowels, results are not obviously related to the proportion of word-final observations, and we will argue below that there is another factor that needs to be taken into account.

4. Discussion

To summarize, results of the corpus analysis do not support PS effects on vowel duration at the level of any of the constituents investigated. Such effects seem to be pervasive if within-word

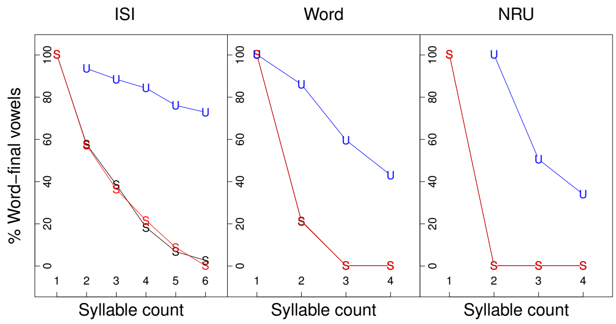

Figure 5: *Percentage of vowels that are word-final by prominence and syllable count in ISI, word, and NRU (black: S +Acc; red: S -Acc; blue: U).*

position is not controlled. Once position within the word is accounted for, apparent shortening effects are no longer observed. It may be argued that our $\alpha = 0.002$ criterion is overly conservative, but in most cases, the effects of constituent length are not even in the *direction* of PS. Importantly, we have shown that this is also the case at the NRU level, which was not entirely clear from previous investigations.

The data do provide evidence for two localized lengthening effects, accentual and word-final lengthening, lending support to [24]'s domain-and-locus approach towards speech timing. As argued above, the effect of lexical stress cannot be assessed in the MARSEC corpus, due to the complementary distribution of vowel phonemes in stressed and unstressed syllables. One caveat is that our analysis does not definitively establish the word as the trigger of the final lengthening effect – it may be the case that the lengthening of word-final vowels is really instantiated by some intermediate phrasal constituent that has simply not been marked in the corpus annotation. This may be a distinct possibility, given that some authors (e.g. [25]) are skeptical about word-final lengthening in the absence of higher-level boundaries. Yet, the general conclusion remains that apparent PS effects are an artifact of such localized lengthening phenomena. We do not subscribe to [19]'s interpretation of "lengthening of the initial and final phoneme of each Narrow Rhythm Unit": the established categories of prominence and word or phrasal boundaries suffice to describe prosodic timing in English, and there is no need to posit units such as the NRU.

The analysis revealed two durational patterns not obviously accounted for by prominence or final lengthening: first, word-final *S -Acc* vowels are lengthened in monosyllabic ISI, and word-final U vowels are lengthened in bisyllabic ISI compared to ISI with greater syllable count. Similar patterns have been observed in earlier experimental studies (e.g. [2, 26]). While these findings are in the direction of PS, this is not the preferable explanation – no comparable tendency is observed in non-word-final vowels, and in either case, the difference only resides in the comparison between vowels from ISI with minimum versus larger syllable count. A unified explanation may be suggested based on the fact that the difference in either case is whether the critical syllable is followed by a stressed or an unstressed syllable across the word boundary: for final stressed syllables, mono- and bisyllabic ISI correspond to $S\#S$ and $S\#U$ sequences, respectively ('#' denoting the word boundary). For unstressed final syllables, bi- and trisyllabic ISI correspond to $SU\#S$ vs. $SU\#US$ sequences. Following [24], this effect may be glossed as "stress-adjacent lengthening", perhaps indicative of a kind of low-level boundary marking.

Second, the by-NRU analysis revealed an alternating duration pattern in U vowels, with vowels from trisyllabic NRU being longer than vowels from bi- and tetrasyllabic NRU in word-final position, and, conversely, vowels from tetrasyllabic NRU being somewhat longer than vowels from trisyllabic NRU in non-word-final position. This finding may be straightforwardly explained as a secondary stress effect: a word-final unstressed syllable in a trisyllabic NRU is one unstressed syllable removed from the preceding stress ($SU[U]\#$). The assumption of secondary stress assignment would account for the greater duration of vowels in this position relative to word-final vowels in bi- and tetrasyllabic NRU. For non-final unstressed vowels, the situation is reversed: in the case of a trisyllabic NRU, this vowel comes from the syllable directly adjacent to the stressed syllable ($S[U]U\#$), whereas in the tetrasyllabic case, the non-final unstressed category includes observations from the unstressed syllable that is one syllable removed from the initial stressed one ($S[U\mathbf{U}]U\#$), which is a potential site for the putative secondary stress effect. This is consistent with the $3 < 4$ pattern for unstressed vowels in the right panel of Figure 4. The analyses by ISI and word largely mask this effect, but it is visible in the word-final unstressed data by syllable count in the word in the middle panel of Figure 3. As an explanation of this effect, one may invoke the assumption that *eurhythmic* principles play a role in speech production [27]: according to this assumption, languages prefer alternating strong-weak patterns and penalize sequences of prosodically weak syllables.

In contrast to [11], we did not observe an effect of word length even in accented contexts, and no lengthening of unstressed vowels in accented words. This may have to do with the definition of pitch accents in the MARSEC corpus: accent labels in the corpus refer to any salient tonal movement, and not to linguistic categories such as nuclear accent. A tentative interpretation is that the effects observed by [11] may be restricted to words that bear nuclear or contrastive pitch accents. The absence of progressive word-final lengthening in *S -Acc* vowels may suggest that such an effect was simply not detectable in our relatively noisy data, or that the very slight tendency observed by [11] is speaking-style dependent.

5. Conclusions

Our analysis revealed large and reliable lengthening effects of prominence and constituent-final position. There were also "contrastive" rhythmic effects in the direction of preserving alternating long-short duration patterns. PS effects in the word, the ISI, and, crucially, the NRU were shown to be spurious, arising from the distribution of constituent-final syllables in uncontrolled data. Results support the view that prosodic timing in English is restricted to localized lengthening effects, and may be harder to reconcile with models that assume underlying periodicities in speech timing.

6. Acknowledgements

We gratefully acknowledge the present and former staff at the Laboratoire Parole et Langage at the University of Aix-Marseille for making the MARSEC corpus publicly available, Jake Westfall for advice on statistics, and three anonymous reviewers for helpful comments on an earlier draft of this paper. The first author is funded by the Bielefeld graduate school of linguistics and literary studies (LiLi-Kolleg).

7. References

[1] D. Klatt, "Interaction between two factors that influence vowel duration," *The Journal of the Acoustical Society of America*, vol. 54, no. 4, pp. 1102–1104, 1973.

[2] I. Lehiste, "The timing of utterances and linguistic boundaries," *The Journal of the Acoustical Society of America*, vol. 51, no. 6B, pp. 2018–2024, 1972.

[3] R. Port, "Linguistic timing factors in combination," *The Journal of the Acoustical Society of America*, vol. 69, no. 1, pp. 262–274, 1981.

[4] B. Lindblom and K. Rapp, "Some temporal regularities of spoken Swedish," in *Auditory analysis and perception of speech*, G. Fant and M. Tatham, Eds. London: Academic Press, 1975, pp. 387–396.

[5] S. Nooteboom, "Production and perception of vowel duration. a study of durational properties in Dutch," Ph.D. dissertation, University of Utrecht, 1972.

[6] A. Rietveld, "Untersuchung zur Vokaldauer im Deutschen," *Phonetica*, vol. 31, no. 3-4, pp. 248–258, 1975.

[7] D. Abercrombie, *Elements of general Phonetics*. Edinburgh: Edinburgh University Press, 1967.

[8] K. Pike, *The Intonation of American English.* Ann Arbor: University of Michigan Press, 1945.

[9] R. Dauer, "Stress-timing and syllable-timing reanalyzed." *Journal of Phonetics*, vol. 11, no. 1, pp. 51–62, 1983.

[10] A. Eriksson, "Aspects of swedish speech rhythm," Ph.D. dissertation, University of Gothenburg, 1991.

[11] L. White and A. E. Turk, "English words on the procrustean bed: Polysyllabic shortening reconsidered," *Journal of Phonetics*, vol. 38, no. 3, pp. 459–471, 2010.

[12] N. Campbell, "Foot-level shortening in the Spoken English Corpus," in *Proceedings of the 7th FASE Symposium*, Edinburgh, 1988, pp. 489–494.

[13] B. Williams and S. M. Hiller, "The question of randomness in English foot timing: A control experiment," *Journal of Phonetics*, vol. 22, pp. 423–439, 1994.

[14] H. Kim and J. Cole, "The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase," in *Proceedings of Interspeech 2005*, Lisbon, 2005, pp. 2365–2368.

[15] J. Krivokapić, "Rhythm and convergence between speakers of American and Indian English," *Laboratory Phonology*, vol. 4, no. 1, pp. 39–65, 2013.

[16] J. P. Van Santen, "Contextual effects on vowel duration," *Speech Communication*, vol. 11, no. 6, pp. 513–546, 1992.

[17] W. Jassem, *Intonation of Conversational English (educated Southern British).* Nakl. Wroclawskiego Tow. Naukowego; skl. gl.: Dom Ksiazki, 1952, no. 45.

[18] C. Bouzon and D. Hirst, "Isochrony and prosodic structure in British English," in *Proceedings of Speech Prosody 2004*, Nara, Japan, 2004, pp. 223–226.

[19] D. Hirst, "The rhythm of text and the rhythm of utterances: from metrics to models." in *Proceedings of Interspeech 2009*, Brighton, 2009, pp. 1519–1522.

[20] D. Hirst and C. Bouzon, "The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English)," in *Proceedings of Interspeech 2005*, Lisbon, 2005, pp. 29–32.

[21] C. Auran, C. Bouzon, and D. Hirst, "The aix-marsec project: an evolutive database of spoken British English," in *Proceedings of Speech Prosody 2004*, Nara, Japan, 2004, pp. 561–564.

[22] R. Koenker, *Quantile regression*. Cambridge: Cambridge University Press, 2005.

[23] J. M. Chambers and T. Hastie, *Statistical models in S*. London: Chapman & Hall, 1992.

[24] L. White, "English speech timing: a domain and locus approach," Ph.D. dissertation, University of Edinburgh, 2002.

[25] A. E. Turk and S. Shattuck-Hufnagel, "Word-boundary-related duration patterns in English," *Journal of Phonetics*, vol. 28, no. 4, pp. 397–440, 2000.

[26] C. A. Fowler, "Timing control in speech production," Ph.D. dissertation, Indiana University, 1977.

[27] P. Wagner, "Vorhersage und Wahrnehmung deutscher Betonungsmuster," Ph.D. dissertation, University of Bonn, 2002.