A Unified Account of Prominence Effects in an Optimization-Based Model of Speech Timing

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]firstname.lastname@uni-bielefeld.de, [2]juraj.simko@helsinki.fi

Abstract

We show how our optimization-based model of speech timing reproduces three effects of prosodic prominence on suprasegmental timing patterns in speech: (1), the durational interaction between lexical stress and pitch accent, (2), polysyllabic shortening in pitch-accented words and (3), differential behavior of prominent and non-prominent syllables under speaking rate variation. We review the literature and present model simulations that replicate reported phenomena. Results underline the capacity of our model to provide a unified account of the temporal organization of speech.

Index Terms: Speech timing, computational modeling, prominence, optimization

1. Introduction

In this paper, we study effects of prosodic prominence on the temporal organization of speech in our optimization-based model of speech timing [1]. Specifically, we demonstrate how the model reproduces three temporal effects of prominence, (1), the durational interaction between lexical stress and accent, (2), polysyllabic shortening in accented words and (3; more tentatively), interactions of prominence and speaking rate. More detailed discussion of the empirical phenomena will be provided below. Results indicate that our model provides a promising explanatory platform for the phenomena under study, grounding them in a cognitively plausible architecture.

We define prosodic prominence as the perceived salience of a syllable or a larger prosodic unit relative to its context [2]. Previous research shows that it is perceived on a gradual scale [3, 2]. Prominence is manifested in the values of acoustic parameters such as fundamental frequency, intensity, various spectral characteristics, and, crucially for the present study, duration. All of these may be enhanced in prominent syllables [4, 3, 5, 6, 7, 8]. Many languages employ prominence distinctions for linguistic functions. We will look at two of them in particular: the first, *lexical stress*, denotes the greater prominence of a syllable relative to other syllables within the same word [2]. The second, for which we use the general term *accent*, refers to the relative prominence of words within a prosodic phrase or utterance [9]. We shall employ this as a general definition and do not attempt to introduce further distinctions, such as between *phrasal stress* and other types of accent for the present purpose.

From a functional perspective, enhancing prominence may be understood as a strategy employed by speakers in order to emphasize important units in the speech signal so as to draw listeners' attention to these units. For example, lexical stress tends to fall on root morphemes in many languages [10], and it has been shown to play an important role in word recognition and segmentation [11, 12], sometimes being the only cue for distinguishing between otherwise identical words, such as *OBject* and *obJECT* in English. Accent, in turn, is used to mark words in an utterance which are semantically very important, often coinciding with information that is new in discourse. Changing the accent pattern of an utterance typically results in major changes in its interpretation [13, 14, 9].

This functional perspective on prominence lends itself well to interpretation within the framework of Hyper- and Hypoarticulation (H&H) theory [15]. H&H theory assumes that speech patterns are shaped by trade-offs between conflicting demands related to minimization of effort and maximization of communicative success on part of the speaker. On this account, it may be assumed that prominent syllables and words are those which are particularly critical for communicative success. Their greater prominence in relation to their environment would then be a consequence of locally shifting the balance in favor of perceptual clarity, so as to ensure that communication be successful [9]. Under this view, prosodic prominence can be interpreted as "localized hyperarticulation" [16, 17].

In this paper, we provide support for this view, by demonstrating how several temporal effects of prominence emerge automatically from the formalization of H&H-inspired assumptions in an optimization-based model of speech timing. We discuss the implementation of prominence as localized hyperarticulation and show how the above-mentioned prominence effects on timing are replicated by the model. Our results thus add to previous findings on the capacity of the model to account for empirically observed phenomena [1, 18]. The rest of the paper is structured as follows: In Section 2, we introduce the model architecture, paying special attention to the modeling of prominence. In Section 3, we discuss evidence pertaining to the timing phenomena under study and report on model simulations demonstrating their replication. Implications of these results and perspectives for further work are discussed in Section 4.

2. Model Architecture

In our model, we use a computational optimization procedure in order to simulate trade-offs between the hypothesized goals of minimizing effort and maximizing perceptual clarity in suprasegmental speech timing. The model architecture derives from an embodied optimization model of articulatory timing [19, 20]. Input consists of specifications of sequences of syllables, representing speech utterances. Given an input sequence, an optimization algorithm computes the vector S of syllable durations that minimizes the composite cost function C. C is a weighted sum of component functions that represent production and perception constraints on constituent durations.

The basic architecture of the model includes three components, D_S, T and P_S, whose relative influence is controlled by the scalar weighting factors α_D, α_T and α_P, as shown in Equation 1 below. The current model abstracts away from many details of speech production and conceptualizes effort mainly in the sense of time as a "shared resource", rather than physical articulatory effort. This is implemented on a global and a local scale: globally, the durational cost component T captures the overall duration of a whole utterance, i.e., the time used for conveying the message encoded in it. On a local scale, D_S is proportional to individual syllable durations, based on the assumption that the syllable is a basic unit of information which speakers strive to transmit in an efficient manner [17, 21]. The weighting factors α_D and α_T allow for globally imposing premiums on these components, encompassing requirements regarding efficient information transmission (α_D) and global speaking rate (α_T) throughout an utterance.

Of special importance for the present work is component P_S, representing a tendency to maximize perceptual clarity. P_S decreases with syllable duration, based on the reasoning that long durations should facilitate perception. Crucially, P_S is non-linear, being modeled by imposing costs on the reciprocal of syllable durations. Thus, P_S initially decreases rapidly with increasingly longer durations, but eventually flattens out. This technique has an intuitive appeal if one interprets P_S as the *inverse of the probability of recognition* of a syllable. One may assume that this probability grows with syllabic duration up to a point where perfect recognition is reached. Increasing syllabic duration beyond this point will make for little or no improvement in recognizability. Direct evidence for this modeling decision comes from gating studies, where subjects have to identify phonemes from acoustic syllable fragments of varying duration [22, 23]. The weighting factor α_P allows for simulating global constraints with regard to perceptual clarity.

In keeping with the concept of localized hyperarticulation, we model syllabic prominence by using two additional weighting factors, ψ_S and δ_S, which simultaneously boost P_S and decrease D_S for individual syllables, rather than for a whole utterance. This implements the assumption that speakers prioritize clarity over efficiency in prominent constituents. As this mechanism applies to individual syllables, it is used to simulate lexical stress in the model.

Accent is hypothesized to enhance the prominence of whole words, rather than individual syllables. Accentual lengthening also seems to affect all syllables in an accented word, at least in some languages [24, 25]. In order to capture this phenomenon, we implemented an additional cost function, P_W. P_W is basically a copy of P_S that operates at the word level, imposing costs on the reciprocal of the summed durations of all syllables in an accented word. P_W thus provides an impetus to increase the sum of the durations of all syllables within this word. Since the model is agnostic towards the propositional content of simulated utterances, we simply define words as arbitrary non-overlapping sub-sequences of S, with the restriction that a word may include at most one stressed syllable. An additional weighting factor, ψ_W, is used to control the strength of accentual lengthening. Formally, the model is thus defined as

$$C = \alpha_D \sum_S \delta_S D_S + \alpha_P \sum_S \psi_S P_S + \alpha_T T + \psi_W P_W \quad (1)$$

Figure 1 visualizes the architecture of the model for a hypothetical utterance with the medial word being accented. Note that P_W is defined for this word only, assuming that speakers consciously manipulate the prominence of only the accented word.

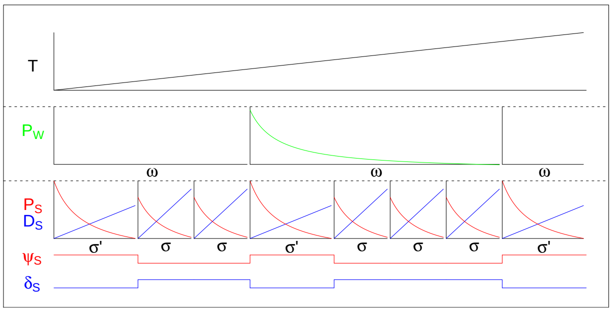

Figure 1: *Model architecture. Cost functions T (utterance level), P_W (word level; ω) and D_S/P_S (syllabic level; σ; apostrophe denotes stresses) as well as parameters δ_S and ψ_S are plotted as a function of respective constituent durations for a hypothetical utterance consisting of a trisyllabic, a tetrasyllabic (accented) and a monosyllabic word. ψ_W is not shown.*

3. Simulation Experiments

3.1. Methodology

The model was implemented in R using the built-in optimization function *optim*. The first experiment (stress-accent interaction) will be reported in Section 3.2.1., the second (polysyllabic shortening) in Section 3.2.2. and the third (speaking rate) in Section 3.2.3. Simulations were run on the syllable sequence depicted in Figure 1, i.e., an "utterance" consisting of a trisyllabic, a tetrasyllabic and a monosyllabic "word", all with initial stress. The only exception to this is Experiment 2 (polysyllabic shortening), where the number of syllables in the accented "word" was varied. Experimentation showed that other modifications of the input, such as adding more words or placing the stressed syllables at different positions within the words, do not affect the qualitative pattern of results. ψ_S was set to 2 for stressed and 1 for unstressed syllables in all simulations. δ_S was set to $1/\psi_S$ in order to reduce the number of free parameters. ψ_W was set to 2. Unless noted otherwise, all other model parameters were set to 1. Crucially, these parameter settings are arbitrary, and no theoretical status is attached to them. Parametric scans revealed that the qualitative pattern of results reported in this paper is stable across a wide range of parameter settings. No attempt was made to model other sources of durational variation, such as syllabic structure or final lengthening.

3.2. Modeling Empirical Results

3.2.1. Interaction of stress and pitch accent

Previous research suggests that accentual lengthening is not distributed uniformly throughout the word. Results from a large-scale corpus study of American English [26] indicate that accentual lengthening is proportionally stronger in stressed than in unstressed vowels once vocalic identity, postvocalic consonant and within-word-position are controlled. Experiments on minimal stress pairs and reiterant syllables in English and Dutch [7, 8] suggest a somewhat more complex picture, indicating that differences diminish in word-final position. For word-initial position, these studies also support proportionally greater accentual lengthening in stressed than in unstressed syllables.

A simulation with the reported parameter settings was run on the test utterance in order to investigate the effect of accent on stressed and unstressed syllables. Figure 2 displays predicted syllable durations. It shows that the model converges and produces meaningful results: there is marked lengthening of stressed compared to unstressed syllables, and also accentual lengthening in both stressed and unstressed syllables.

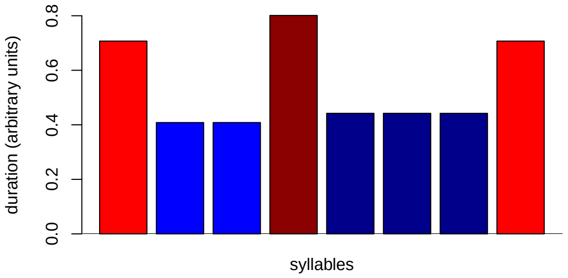

Figure 2: *Syllable durations predicted by the model for test utterance. Light red: +stress -accent; light blue: -stress -accent; dark red: +stress +accent; dark blue: -stress +accent.*

Figure 3 visualizes results from comparisons between accented and unaccented syllables. As can be seen, the effect of accentuation is greater in absolute as well as proportional terms in stressed than in unstressed syllables, in accordance with published results. This pattern is generated by the interaction between D_S, P_S and P_W: P_W provides an impetus to lengthen all syllables within its scope and thus works in the same direction as P_S. Stressed syllables, which are defined by a higher premium on P_S and lowered D_S, are "more ready" to be lengthened, leading to a stronger effect compared to unstressed ones. A possible interpretation is that in the accented environment, where everything is lengthened, the contrast between stressed and unstressed syllables has to be enhanced to be reliably perceived. This explanation resonates with the idea of accent as a "magnifying lens, i.e. the intensification of phonological contrasts in accented environments. [27, 28, 29]. We hypothesize that deviant results for word-final syllables reported in some studies stem from interactions with word-final lengthening (cf. [8]) and leave this idea open for further research.

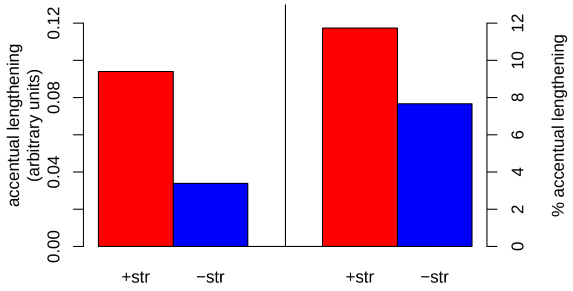

Figure 3: *Absolute (left panel) and proportional (right panel) amount of accentual lengthening in stressed (red) and unstressed (blue) syllables as predicted by the model.*

3.2.2. Polysyllabic shortening in pitch-accented words

Polysyllabic shortening, i.e. an inverse relationship between stressed syllable duration and the number of syllables in the respective word, has been attested in many languages, including English [30], Swedish [31], Dutch [32] and German [33]. Results from more recent investigations, however, suggest that the phenomenon may be confined to pitch-accented words, indicating the distribution of accentual lengthening across the word rather than a genuine compression effect [25, 34].

Polysyllabic shortening was tested in the model by varying the syllable count of the accented medial word, while keeping all parameter settings constant. Figure 4 visualizes stressed (red) and unstressed (blue) syllable durations as a function of the number of syllables in the accented word. The model predicts marked shortening of a stressed syllable as a function of the number of syllables in an accented word, in accordance with the studies mentioned above. As for unstressed syllables, there is a discernible but rather weak shortening effect. This converges with results from Swedish [31] and Dutch [32].

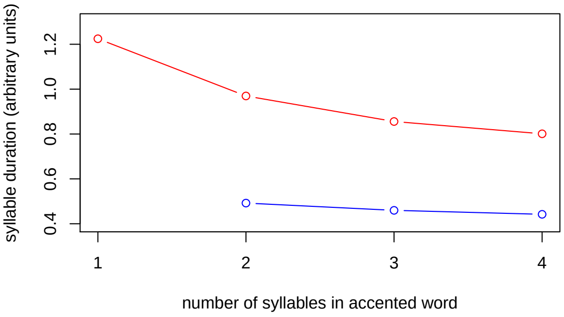

Figure 4: *Polysyllabic shortening in stressed (red) and unstressed (blue) syllables in the accented word.*

The model's prediction for stressed syllables in particular bears close resemblance to empirical results, with the magnitude of shortening gradually decreasing as more syllables are added to the word. The similarity between the stressed trajectory and cost function P_W itself might lead a critical observer to suspect that P_W causes some ad-hoc encoding of the effect, in the fashion of descriptive models that fit rational functions to vowel duration by syllable count in a word [31, 32]. We would like to stress that this is not the case: the effect of P_W, on the contrary, is to *lengthen* all syllables in an accented word, and, crucially, P_W has no access to the number of these syllables.

Rather than being "hardcoded", polysyllabic shortening emerges from the interaction of the individual component cost functions: the interplay of D_S and P_S defines an optimal duration for each syllable in the absence of any higher-level process. P_W, if present, perturbs the balance between D and P_S by providing an impetus to lengthen the summed durations of the syllables within its scope. If the word thus defined contains more syllables, the lengthening evoked by P_W can be shared out among the individual syllables, so that each one of them has to depart less from its optimal duration. This explanation is very much in keeping with [25]'s distributional accent hypothesis.

3.2.3. Interaction between prominence and speaking rate

In the third experiment, we ran various simulations on the test utterance with varying α_T, in order to simulate variation in speaking rate due to time constraints. Higher values of this parameter increase the cost for utterance duration, leading to increased speaking rate. Figure 5 depicts proportional shortening of (unaccented) stressed and unstressed syllables as a function of the rate parameter. As can be seen, proportional shortening is stronger in stressed than in unstressed syllables.

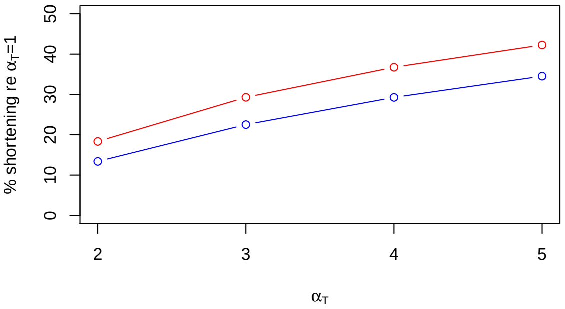

Figure 5: *Percentage shortening of stressed (red) and unstressed (blue) syllables at faster rates ($\alpha_T > 1$) relative to durations at slow rate ($\alpha_T = 1$).*

This prediction receives somewhat mixed support from the literature. For American English, results by [35] and [30] are sometimes cited as indicating the opposite pattern, i.e., stronger shortening of unstressed compared to stressed syllables in fast speech. The methodology in these studies, however, was to compare shortening of a stressed (and pitch-accented) syllable with that of the surrounding carrier sentence as a whole, including the target syllable itself. This is not quite the same as showing that, all else being equal, unstressed syllables shorten more strongly than stressed syllables in fast speech, especially since the phonetic material in the rest of the sentence was of course different from the target syllable in these studies. One more recent experimental study on Dutch [36] directly compares stressed and unstressed syllables and finds proportionally larger shortening in unstressed syllables in fast speech. The majority of studies, however – [37, 38] for French, [39] for Greek and [40] for American English – support our model's prediction, indicating that the proportional magnitude of shortening in fast speech correlates positively with prominence.

Interestingly, [36] also refer to H&H theory as an explanation of their results, arguing that stressed syllables shorten less strongly than unstressed syllables so as to preserve the informationally most important parts of the signal. We would propose an alternative explanation: stressed syllables shorten *more* strongly than unstressed ones because they are longer and, hence, there is "more room" for shortening without marked perceptual loss. This is precisely what follows from the shape of the cost function P_S (cf. Figure 1): a long syllable can undergo substantial shortening with only a slight increase in perceptual cost. For shorter syllables, even a small decrease in duration will lead to markedly higher costs. This explanation hints at the well-attested phenomenon of incompressibility, an idea also expressed by [39] and [40]. Indeed, incompressibility has been shown to emerge from the architecture of our model [1]. Further empirical study is needed to decide between these hypotheses.

4. Discussion and Conclusions

Results show that our model provides a convincing account of effects of prosodic prominence in the temporal domain. The technique of incorporating prominence by locally shifting an H&H continuum in favor of perceptual constraints is theoretically well-founded, as the design of the perception cost functions is directly informed by results from speech perception research. The replication of several temporal effects of prominence demonstrates the empirical adequacy of our modeling approach. Interestingly, the explanations of the effects suggested by our model tend to converge with well-motivated research hypotheses. It is the purpose of computational modeling to demonstrate that theoretically conceived ideas actually work and generate empirically observed patterns once implemented and tested. In our opinion, our model fulfills this task very successfully for the domain of temporal effects of prominence.

We would also argue that our "localized hyperarticulation" approach provides a more satisfactory account of prosodic prominence than the technique commonly employed in oscillatory models of speech timing, where prominence is incorporated by slowing down a syllabic oscillator for an individual period [41, 42]. This technique could be given some post-hoc perceptual motivation, but it is not clear whether it adds any explanatory value to the model. In contrast, our approach towards incorporating perceptual prominence represents the core of our model's explanatory power, as has been demonstrated by the replication of several timing phenomena within one unified model, based on a mechanism that is directly informed by results from speech perception research.

Importantly, the replication of timing phenomena demonstrated in this paper is an emergent result of the optimization procedure, and there are no explicit mechanisms that would "hardcode" the reported durational patterns in the model. For example, while the lengthening of stressed versus unstressed syllables and accented versus unaccented words is an obvious consequence of the respective parameter settings (although it stems from a well-motivated mechanism), the *interaction* between both effects reported in Section 3.2.1. is a non-trivial outcome of the cost optimization – there is no dedicated model component that would explicitly enforce the observed superadditive combination of stress-induced and accentual lengthening.

Our present model is arguably rather simple and abstract, especially concerning the conceptualization of effort. We would also like to stress that it should not be viewed as a real-time production model. While we claim that the trade-off between the constraints modeled by our cost functions does have psychological reality, we are not endorsing a view of optimization being computed "online" in speech production. The model's abstract conception is intentional, since we believe it to be a necessary requirement for understanding basic processes, before more complex issues can be addressed. We are currently working on a more realistic computational platform that will enable us to consider effort in a more principled way, and to obtain a more complete picture of speech timing phenomena.

5. Acknowledgements

We thank 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.

6. References

[1] A. Windmann, J. Šimko, B. Wrede, and P. Wagner, "Modeling durational incompressibility," in *Proceedings of Interspeech 2013*, Lyon, France, 2013, pp. 1375–1379.

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

[3] G. Fant and A. Kruckenberg, "Preliminaries to the study of Swedish prose reading and reading style," *STL-QPSR*, vol. 2, no. 1989, pp. 1–83, 1989.

[4] D. B. Fry, "Experiments in the perception of stress," *Language and speech*, vol. 1, no. 2, pp. 126–152, 1958.

[5] B. Heuft, T. Portele, P. Wagner, C. Widera, and M. Wolters, "Perceptual prominence," in *Speech and Signals*, W. Sendlmeier, Ed. Frankfurt a. M.: Hector, 2000, pp. 97–115.

[6] B. M. Streefkerk, "Prominence. acoustic and lexical/syntactic correlates," Ph.D. dissertation, University of Amsterdam, 2002.

[7] A. M. C. Sluijter, *Phonetic correlates of stress and accent*. Holland Academic Graphics The Hague, 1995, vol. 15.

[8] A. M. Sluijter and V. J. Van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," *The Journal of the Acoustical society of America*, vol. 100, p. 2471, 1996.

[9] H. Schmitz, *Accentuation and Interpretation*. New York: Palgrave, 2008.

[10] C. H. Echols and E. L. Newport, "The role of stress and position in determining first words," *Language acquisition*, vol. 2, no. 3, pp. 189–220, 1992.

[11] Z. Bond, "Listening to elliptic speech: pay attention to stressed vowels," *Journal of Phonetics*, vol. 9, no. 1, pp. 89–96, 1981.

[12] A. Cutler, "Linguistic rhythm and speech segmentation," in *Music, Language, Speech and Brain*, J. Sundberg, L. Nord, and R. Carlson, Eds. London: Macmillan, 1991, pp. 157–166.

[13] D. L. Bolinger, "A theory of pitch accent in English," *WORD – Journal of the International Linguistic Association*, vol. 14, no. 2-3, pp. 1–149, 1958.

[14] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[15] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech production and speech modeling*, W. Hardcastle and A. Marchal, Eds. Dordrecht: Kluwer, 1990, pp. 403–439.

[16] K. J. De Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *The journal of the acoustical society of America*, vol. 97, no. 1, pp. 491–504, 1995.

[17] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and Speech*, vol. 47, no. 1, pp. 31–56, 2004.

[18] A. Windmann, J. Šimko, and P. Wagner, "Probing theories of speech timing using optimization modeling," in *Proceedings of Speech Prosody 2014*, Dublin, Ireland, 2014, pp. 346–350.

[19] J. Šimko and F. Cummins, "Embodied task dynamics." *Psychological review*, vol. 117, no. 4, pp. 1229–1246, 2010.

[20] J. Šimko and F. Cummins, "Sequencing and optimization within an embodied task dynamic model," *Cognitive Science*, vol. 35, no. 3, pp. 527–562, 2011.

[21] R. J. Van Son and J. P. Van Santen, "Duration and spectral balance of intervocalic consonants: A case for efficient communication," *Speech Communication*, vol. 47, no. 1, pp. 100–123, 2005.

[22] W. Grimm, "Perception of segments of English-spoken consonant-vowel syllables," *The Journal of the Acoustical Society of America*, vol. 40, no. 6, pp. 1454–1461, 1966.

[23] M. Tekieli and W. Cullinan, "The perception of temporally segmented vowels and consonant-vowel syllables," *Journal of Speech, Language and Hearing Research*, vol. 22, no. 1, p. 103, 1979.

[24] T. Cambier-Langeveld and A. Turk, "A cross-linguistic study of accentual lengthening: Dutch vs. English," *Journal of Phonetics*, vol. 27, no. 3, pp. 255–280, 1999.

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

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

[27] K. De Jong and B. Zawaydeh, "Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration," *Journal of Phonetics*, vol. 30, no. 1, pp. 53–75, 2002.

[28] K. De Jong, "Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration," *Journal of Phonetics*, vol. 32, no. 4, pp. 493–516, 2004.

[29] M. Ortega-Llebaria, "Comparing the magnifying lens effect of stress to that of contrastive focus in Spanish," in *3rd Conference on Laboratory Approaches to Spanish Phonology, Somerville, MA*, 2008.

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

[31] 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.

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

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

[34] J. Siddins, J. Harrington, F. Kleber, and U. Reubold, "The influence of accentuation and polysyllabicity on compensatory shortening in German," in *Proceedings of Interspeech 2013*, Lyon, France, 2013, pp. 1002–1006.

[35] G. Peterson and I. Lehiste, "Duration of syllable nuclei in English," *The Journal of the Acoustical Society of America*, vol. 32, no. 6, pp. 693–703, 1960.

[36] E. Janse, S. Nooteboom, and H. Quené, "Word-level intelligibility of time-compressed speech: prosodic and segmental factors," *Speech Communication*, vol. 41, no. 2, pp. 287–301, 2003.

[37] D. Duez, "Effects of articulation rate on duration in read French speech," in *Proceedings of Eurospeech*, Budapest, 1999, pp. 715–718.

[38] V. Pasdeloup, R. Espesser, and M. Faraj, "Rate sensitivity of syllables in French: a perceptual illusion?" in *Proceedings of Speech Prosody 2006*, Dresden, 2006, p. 216.

[39] M. Fourakis, A. Botinis, and M. Katsaiti, "Acoustic characteristics of Greek vowels," *Phonetica*, vol. 56, no. 1-2, pp. 28–43, 1999.

[40] M. Fourakis, "Tempo, stress, and vowel reduction in American English," *The Journal of the Acoustical Society of America*, vol. 90, p. 1816, 1991.

[41] M. O'Dell and T. Nieminen, "Coupled oscillator model of speech rhythm," in *Proceedings of ICPhS 1999*, San Francisco, 1999, pp. 1075–1078.

[42] E. Saltzman, H. Nam, J. Krivokapic, and L. Goldstein, "A task-dynamic toolkit for modeling the effects of prosodic structure on articulation," in *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 2008, pp. 175–184.