9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Anticipatory shortening: Articulation rate, phrase length, and lookahead in speech production

Jason Bishop[1,2], *Boram Kim*[2]

[1]College of Staten Island, City University of New York
[2]The Graduate Center, City University of New York

jason.bishop@csi.cuny.edu, bkim@gradcenter.cuny.edu

Abstract

The present study investigated "anticipatory shortening", the durational compression of syllables in longer prosodic phrases. Our primary motivations were related to this phenomenon's relevance to speech production planning; these durational adjustments depend on upcoming material, and as such, are generally assumed to be indicative of speakers' lookahead. Applying simple correlational analysis to a corpus of (American English) read speech, we asked whether articulation rate (defined as average syllable durations) was most closely related to the length of an associated intermediate phrase, Intonational Phrase, or inter-pause interval. We found that—when final lengthening is removed—the two larger prosodic domains had by far the stronger relationship with articulation rate. We interpret our basic findings as consistent with claims that speakers plan their speech in relatively large chunks, corresponding to at least one Intonational Phrase.

Index Terms: speech production planning, prosodic phrasing, speech rate, articulation rate

1. Introduction

1.1. Phrase length and average syllable duration

Average syllable duration (ASD) is a measure commonly used to assess speech tempo, or "articulation rate" (i.e., the rate of actual speech, excluding silent pause intervals). One of the most consistent findings in studies of articulation rate in recent years is that ASD decreases as the length of the stretch of speech containing it increases. This phenomenon has variably been referred to as "anticipatory shortening", "compensatory shortening", and "medial shortening" (for a very comprehensive review, see [1]).[1] In the present study, we focus on this durational effect when the stretch of speech in question is a sentence-level/phrasal prosodic constituent, defined phonologically within the Autosegmental Metrical framework ([10],[11],[12]).

Our primary interests here involve what the effect of phrase length on ASD tells us about speech production planning. Terms like "anticipatory shortening" reflect the assumption that adjustments to ASD driven by phrase length indicate lookahead on the part of the speaker, who can only make the proper adjustments if overall phrase length is to some extent planned in advance ([13]; see also [14] for detailed discussion). This is consistent with the notion that speakers plan considerably further ahead than one or two words ([15]), and instead plan in chunks defined by prosodic structure ([16]). Recent work on what level of the prosodic hierarchy these chunks might reflect has suggested variability across speakers and speaking conditions (see, e.g., [17],[18],[19]), but evidence from timing at phrase edges (i.e., initial and final lengthening) has been widely interpreted as indicating that speakers plan a basic outline of utterance material in at least Intonational Phrase-sized chunks ([20]; see also [21] for recent review and discussion). Research on speech production planning thus leads to the prediction that phrase length effects on ASD—i.e., anticipatory shortening—will be most apparent when considering the length of the Intonational Phrase containing the syllables, or perhaps the length of longer stretches that include more than one Intonational Phrase (e.g., inter-pause intervals, or "pause groups"). On the other hand, ASD should be less sensitive to the length of lower-level (and, on average, shorter) phrasal constituents, such as the intermediate phrase. However, to our knowledge, this has not been demonstrated previously.[2]

Whether ASD is best predicted by the length of an intermediate phrase (ip), Intonational Phrase (IP), or pause group (PG) is the basic question we asked in the present study, making use of a corpus of speech prosodically annotated for phrase structure. Before going on, it is necessary to point out an additional timing effect related to phrase structure—but orthogonal to the effect of phrase length—that must be addressed. In particular, duration is also adjusted at phrase edges (see [22] and [23], among many others), especially at the ends of phrases, where syllables are known to be considerably lengthened (for a recent review, see [24]). In nearly all previous work on articulation rate (and speech rate more generally), final lengthening has been incorporated into ASD values (e.g., see [25],[26],[27], and references therein). From the perspective of assessing anticipatory shortening, this is possibly problematic, since final syllables (in addition to bearing edge-based lengthening), are more weakly correlated with phrase length than are non-final syllables—and may possibly even have the opposite relationship to it ([28]; but see [29]). Therefore, measures of ASD that include orthogonal

[1] It has long been observed that the compression effects we discuss here cannot be explained as word-level "polysyllabic shortening" (e.g., [2], [3], [4], [5], [6], [7]). However the relationship between compressions associated with higher and lower levels of the prosodic hierarchy remains somewhat unclear ([8]; see also [9] and [1] for more recent discussion).

[2] But see [29], who reports some statistical analysis of the effects of Intonational Phrase length versus pause group length, although his emphasis (on sociolinguistic implications) differs from ours.

10.21437/SpeechProsody.2018-48

timing effects like final lengthening may obscure a more general relationship between phrase length and ASD for non-final syllables.

In the study presented below, we explored these issues using a speech corpus collected as part of a previous study ([30]), for which phrasal structure was annotated in the Tones and Break Indices (ToBI) annotation system ([31]). As stated above, our basic question was whether the well-documented negative correlation between phrase length and ASD depended significantly on the way "phrase" is defined, and to what extent final syllables affect (or do not affect) the apparent relationship of ASD to phrase length.

2. Experiment

2.1. Methods

2.1.1. Speech corpus

The speech analyzed here was taken from an experiment reported in [30], which investigated cross-speaker variation in prosodic phrasing in read speech. The corpus consisted of speech elicited from 100 native English speakers (42 male/58 female, mostly in their 20s and 30s and from the NYC Metropolitan area), who all read the same moderately-lengthed (156 words) passage. The passage (shown in its entirety in the Appendix) was taken from popular prose writing in [32]. Speakers were digitally recorded (44.1k Hz) reading the passage aloud in a sound-attenuated booth using a head-mounted microphone (Shure SM10); all speakers read the passage twice in order to increase familiarity and fluency, and only the second production was used for the analysis. Recordings were saved as WAV files and set aside for later prosodic annotation that would allow for an analysis of phrase structure.

2.1.2. Prosodic annotation

The read passages were annotated for prosodic phrasing using a modified version of the MAE_ToBI transcription procedure, described in detail in [30]. In brief, two labelers (working together as a pair) identified in each speaker's recording the locations of (a) all disfluencies and (b) all "potential" fluent prosodic boundaries. "Disfluencies" were identified as per the ToBI guidelines and "potential fluent prosodic boundaries" were defined as the locations of any non-disfluent perceived juncture greater than that marking an ordinary word boundary (i.e., anything corresponding to a 2, 3-, 3, 4-, 4, or any uncertainties (e.g., a 2? or 3?). These first-pass "potential" boundary identifications then served as the input to a more conservative second-pass transcription by two additional ToBI-labelers (working independently), who made final decisions about the break values for the first-pass boundary identifications, determining whether they were either word-level boundaries (0,1,1-, 2, or "?-"), ip-level boundaries (3- or 3, or 3?) or IP-level boundary (4-, 4, or 4?). Agreement levels between the two second-pass labelers were generally consistent with what has been reported elsewhere for the ToBI conventions ([33],[34],[35],[36]). Agreement between the two labelers as to the presence of a prosodic break above the word-level was 92.5% (κ = .76), and 100% of the disagreements concerned the presence of an ip boundary versus the absence of a boundary (i.e., there were, unsurprisingly, no cases where one labeler identified a word-level boundary and the other an IP-level boundary). Where both labelers agreed a boundary above the word-level was present, agreement as to its size (i.e., an ip boundary versus an IP boundary) was 94.5% (κ =.88). Because the analysis required definitive decisions about the locations and sizes of boundaries for all speakers, and because we wished to only analyze boundaries about which we had the most certainty, the relatively small proportion of disagreements were settled in the direction of the labeler who marked a smaller degree of juncture. That is, for disagreements where one labeler assumed a word-level boundary and the other an ip-level boundary, the word-level boundary was assumed; in cases where one labeler marked an ip-level boundary and the other an IP-level boundary, the ip-level boundary was assumed. After excluding all cases where a disfluency occurred (if an IP contained a disfluency in any of its ips, all of its ips were excluded), this resulted in a corpus of 3,077 intermediate phrases, 1,900 Intonational Phrases, and 1,124 pause groups for analysis, the basic length statistics of which are shown in Table 1. Important to note is that these are overlapping groups; some IPs consist of only a single ip, and some PGs consist of only a single IP (and, indeed, some of those single-IP PGs also consisted of only one ip). Finally, the connected speech elicited from these passages showed variation—speakers did not all phrase the material in the passage the same way (indeed, as mentioned above, the study in which these passages were elicited aimed to investigate cross-speaker variation in prosodic phrasing).

Table 1. *Mean lengths and standard deviations (in σ) and number of observations for the intermediate phrases (ip), Intonational Phrases (IP) and pause groups (PG) analyzed.*

	Length (σ)	S.D. (σ)	Observations
ip	6.0	2.5	3,077
IP	9.4	4.1	1,900
PG	12.7	6.5	1,124

2.1.3. Calculating average syllable duration

The phrasal structure of the speakers' productions of the passage having been determined, we then calculated ASD for each type of "phrase" (we will refer to pause groups, i.e., inter-pause intervals, as a type of "phrase", although we do not assume a phonological status for it). ASD was defined as the total duration of the phrase containing it (extracted automatically in Praat) divided by the number of syllables in that unit. This resulted in six different measures of ASD, depending on whether the prosodic phrase was an ip, an IP, or a PG, and (for each) whether or not the final syllable in that phrase was included in the calculation of ASD. Although in English some effects of final lengthening can be seen before the final syllable, these effects are quite small (for reviews, see [37] and [1]; see also recent findings by [38]), and so we regard the exclusion of final syllables as a strong control for the effects of final lengthening on ASD.

2.2. Results

Unsurprisingly, Pearson correlations showed the relationship between phrase length and ASD to be both inverse and highly significant no matter how phrases were defined, or how ASD was calculated ($p < .001$ in all cases). Table 2 therefore displays the correlations in terms of the more-interpretable corresponding R^2 values, and we focus our attention on the relative strengths of the relationships given different ways of defining "phrase", and of calculating ASD.

To assess the relative strengths of the correlations in Table 2, we carried out a series of statistical comparisons using [39]'s (see also [40]) modification of the Fisher z-test (appropriate for comparing dependent overlapping and dependent non-overlapping correlations), generated using the *cocor* package ([41]) in R statistics ([42]).[3] Considering first differences depending on the definition of "phrase" (i.e., comparing horizontally in Table 2) when final syllables were included, comparisons indicated that the strength of the relationship between ASD and ip length did not differ significantly from that between ASD and IP length ($z = -1.09$, $p > .10$), although both differed significantly from the strength of the relationship between ASD and PG length (ASD ~ ip length vs. ASD ~ PG length: $z = -3.45$, $p < .001$; ASD ~ IP length vs. ASD ~ PG length: $z = -5.32$, $p < .001$). Thus, somewhat in contrast to the predictions based on production planning, variation in ASD was not more strongly related to the length of high-level prosodic phrases (IP or PG) generally; rather, it was equivalently related to the length of lower-level ips and higher-level IPs, and significantly less strongly related to the largest domain, PGs.

The pattern was quite different, and more in line with our planning-based predictions, when considering ASD calculations that excluded final syllables. Here, comparisons indicated that the strength of the relationship between phrase length increased as the size of the domain considered increased (ASD ~ ip length vs. ASD ~ IP length: $z = 15.19$, $p < .001$; ASD ~ IP length vs. ASD ~ PG length: $z = 5.95$, $p < .001$; ASD ~ ip length vs. ASD ~ PG length: $z = 13.85$, $p < .001$). Although all significant from each other, the absolute difference between the two larger phrase domains was quite small (only about 4%, as seen in Table 2), both accounting for approximately five times as much variance in ASD as did ip length. This suggests that the durational shortening of non-final syllables seems to be considerably better predicted by the length of higher-level prosodic phrases than lower-level ones.

Finally, and unsurprisingly given the patterns just described, the relationship between ASD and phrase length *within* phrase type (i.e., comparing vertically in Table 2) differed significantly depending on whether the ASD calculation included or excluded final syllables. In particular, excluding final syllables lead to a much weaker relationship between ASD and phrase length when the domain was the ip ($z = -30.04$, $p < .001$) and a significantly stronger relationship when the domain was IP ($z = 6.01$, $p < .001$) or PG ($z = 10.23$, $p < .001$). When ips are the domain, excluding final syllables led to phrase length's accounting for 26% less variance in ASD. When IPs or PGs were the domain, however, excluding final syllables led to phrase length accounting for 11% and 27% more variance in ASD, respectively.

In summary, the relationship between ASD and phrase length in this corpus replicated a well-established pattern: ASD decreases as phrase length increases. Further, the relationship is highly robust, being statistically significant whether including or excluding phrase-final (and presumably lengthened) syllables, or whether the length that we consider is that of a lower-level (ip) or higher-level (IP or PG) phrase. Nonetheless, decisions as to how to calculate ASD and how to define "phrase" do matter. The most variance in ASD seems to be accounted for when (a) phrase length refers to that of high-level phrases, and (b) phrase-final syllables are excluded from the calculation of ASD. Researchers' treatment of these two factors can make the difference between whether phrase length accounts for 11% or approximately half of the variance in ASD.

Table 2. *Relationship between average syllable duration and phrase length (expressed as R^2; all correlations were negative) when phrase length is defined within either intermediate (ip), Intonational (IP) or pause groups (PG), and when phrase-final syllables were either included or excluded from average syllable duration calculations.*

	ip	IP	PG
Final Syllable Included	.37	.41	.29
Final Syllable Excluded	.11	.52	.56

3. Discussion

It is well-known, and well-replicated, that average syllable durations get shorter as more syllables are added to the stretch of speech containing them. The goal of our study was to determine whether this relationship was more or less apparent depending on what stretch of speech is considered. In so doing, we were primarily interested in the implications of this relationship for our understanding of speech production planning; the shortening of syllables based on a larger stretch of speech's length indicates that this larger stretch of speech is to some extent already part of the speaker's plan. As discussed above, work on speech production planning in recent years suggests that speakers' lookahead is quite extensive, likely consisting of one or more Intonational Phrases.

Consistent with this—at least when final syllables were excluded from calculations of ASD—we found that ASD was most strongly associated with the length of higher-level prosodic units. Simple R^2s suggested that more than half of the variance in ASD may be accounted for by the length of the syllables' associated Intonational Phrase or Pause Group. This is compared to the much more modest 11% accounted for by the length of an associated intermediate phrase (and indeed, even some of this explained-variance may largely reflect ips that corresponded to complete IPs and PGs). We propose that the results obtained when excluding final syllables are more informative as to speakers' planning, since final syllables undergo an independent timing effect—i.e., final lengthening. It may be that most of the variation in final syllable durations is predicted by how much or how little speakers need to highlight phrase edges, leaving very little room for other factors to influence final syllable durations. If this is the case, and final syllables simply do not participate in the phrase-length induced shortening that pre-final syllables do (and as noted above, this has in fact been reported previously; [28]), including them in ASD would dampen the correlation between phrase-length and ASD in a way that looks very much like our findings. That is, removing final syllables from the ASD calculation resulted in an improvement to the correlation between ASD and IP length (an increase in R^2 from .41 to

[3] Indeed, we also considered five other similarly-appropriate tests available in the *cocor* package, all of which produced results equivalent to the ones we present here.

.52), but an even bigger improvement for the correlation between ASD and PG length (an increase in R^2 from .29 to .56). This likely reflects the fact that PGs often consist of multiple IPs, and so contain not one, but multiple IP-final syllables within them.

In summary, we interpret our findings as consistent with the ideas that (a) the relation between ASD and phrase length reflects "anticipation" of upcoming material—i.e., planning, and (b) that speakers are planning relatively far in advance—in chunks of one or more IPs. We would like to point out that our study is somewhat preliminary in nature—we have carried out a correlational analysis to identify broad differences between measures of ASD (rather than more complex statistical modeling that takes many different factors into account), and we have done so post-hoc, on a corpus of speech designed for other purposes. Additionally, while most studies have not investigated differences between syllables' position in the phrase at all, we have (at most) distinguished only two (pre-final and "all syllables"), and it is surely the case that anticipatory shortening is not spread across all pre-final syllables evenly (see [43]). Nonetheless, little work has investigated the sensitivity of articulation rate to prosodic phrasing in the context of an explicit phonological model of that structure, such as the Autosegmental Metrical model assumed here. Similarly, although anticipatory shortening is often controlled for in studies of articulation/speech rate by including phrase length (usually PG length) as a factor ([25],[26],[29],[44]), comparatively less attention has been paid to the phenomenon as a window into speakers' planning scope.

4. Conclusions

The study presented above investigated "anticipatory shortening", the apparent durational compression of syllables in longer prosodic phrases. Our primary motivations were related to this phenomenon's relevance to speech production planning; such durational effects based on upcoming length are indicative of speakers' lookahead. Applying simple correlational analysis to a corpus of read speech, we asked whether articulation rate (defined as average syllable durations) was most closely related to the length of an associated intermediate phrase, Intonational Phrase, or pause group (i.e., an inter-pause interval). Our main result was that (when final lengthening is removed) the two larger prosodic domains had the strongest relationship with articulation rate, with approximately half the variation in average syllable duration being accounted for by either Intonational Phrase or pause group length (intermediate phrases accounting for about one fifth as much). We interpret this result as consistent with claims that speakers plan their speech in relatively large chunks, corresponding to at least one Intonational Phrase.

5. **Appendix:** Passage (taken from [32])

Our patch of tangled yard was an exotic foreign country. I had spent so much of my life in dark theatres and dim hotel rooms, where the only thing green was the peeling paint on the walls, that this seemed perfectly natural to me. This was where I had my first bite of mud pie; where I set up a card table and mixed household chemicals, toothpaste, and my mother's face powder, doing what I called "experiments". Now that I was sick I wasn't allowed out of the house, but when I stood on the headboard of my bed I could look through a high window into the backyard and see the concrete wall I used to climb over with my friends. We'd sneak under the cover of the banana trees and light matches I had stolen from my mother's kitchen. I could just reach the big box of wooden matchsticks she kept on the top shelf of the old four-legged gas stove.

6. Acknowledgements

The authors would like to thank Jaekoo Kang, who assisted with forced alignment in part of the analysis of average syllable duration. We also thank Olivia Ayala, Lorenza Colona, Nadia Zaki, and other undergraduate research assistants in the CUNY Prosody Laboratory for their help with recording and processing of sound files.

7. References

[1] J. Fletcher, "The prosody of speech: Timing and rhythm," in *The Handbook of Phonetic Sciences, Second Edition*, 2010, pp. 521–602.

[2] A. Malécot, R. Johnston, and P.-A. Kizziar, "Syllabic rate and utterance length in French," *Phonetica*, vol. 26, no. 4, pp. 235–251, 1972.

[3] D. O'Shaughnessy, "Consonant durations in clusters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 22, no. 4, pp. 282–295, 1974.

[4] I. Lehiste, "Interaction between test word duration and length of utterance," in S. Garnes, I. Lehiste, P. Miller, L. Shockey, & A. Zwicky, Eds, Ohio State Working Papers in Linguistics (No. 17), 1974, pp. 160–169.

[5] I. Lehiste, "Interaction between test word duration and length of utterance", in L.R. Waugh & C.H. Schooneveld, Eds, *The Melody of Language*, Baltimore, MD: University Park Press, 1980, pp. 169–176.

[6] J. J. De Rooij, Speech punctuation: An acoustic and perceptual study of some aspects of speech prosody in Dutch. Rooij, 1979.

[7] L. H. Nakatani, K. D. O'Connor, and C. H. Aston, "Prosodic aspects of American English speech rhythm," *Phonetica*, vol. 38, no. 1–3, pp. 84–105, 1981.

[8] M. E. Beckman and J. Edwards, "Lengthenings and shortenings and the nature of prosodic constituency," in *Papers in Laboratory Phonology*, J. Kingston and M. E. Beckman, Eds. Cambridge: Cambridge University Press, 1990, pp. 152–178.

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

[10] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," PhD Dissertation, Massachusetts Institute of Technology, 1980.

[11] M. E. Beckman and J. B. Pierrehumbert, "Japanese prosodic phrasing and intonation synthesis," in *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, 1986, pp. 173–180.

[12] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 1996.

[13] S.G. Nooteboom and I.H. Slis, "The Phonetic Feature of Vowel Length in Dutch," *Lang Speech*, vol. 15, no. 4, pp. 301–316, Oct. 1972.

[14] S. G. Nooteboom, "Limited lookahead in speech production," in *Producing speech: Contemporary issues—for Katherine Safford Harris*, 1995, pp. 3–18.

[15] W. J. Levelt, *Speaking: From intention to articulation. ACL*. MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, Massachusetts, 1989.

[16] L. Wheeldon and A. Lahiri, "Prosodic units in speech production," *Journal of Memory and Language*, vol. 37, no. 3, pp. 356–381, 1997.

[17] J. Krivokapic, “The planning, production, and perception of prosodic structure,” PhD Dissertation, University of Southern California, 2007.

[18] J. Krivokapic, “Prosodic planning in speech production,” in S. Fuchs, M. Weirich, D. Pape, & P. Perrier, Eds, *Speech planning and dynamics*, Peter Lang Publishing, pp. 157–190, 2012.

[19] C. Petrone, S. Fuchs, and J. Krivokapić, “Consequences of working memory differences and phrasal length on pause duration and fundamental frequency,” in *Proceedings of the 9th International Seminar on Speech Production (ISSP)*, 2011, pp. 393–400.

[20] P. Keating and S. Shattuck-Hufnagel, “A prosodic view of word form encoding for speech production,” *UCLA working papers in phonetics*, pp. 112–156, 2002.

[21] S. Shattuck-Hufnagel, “Prosodic frames in speech production,” in *The handbook of speech production*, 2015, pp. 419–444.

[22] C. Fougeron and P. A. Keating, “Articulatory strengthening at edges of prosodic domains,” *The journal of the acoustical society of America*, vol. 101, no. 6, pp. 3728–3740, 1997.

[23] P. A. Keating, “Phonetic encoding of prosodic structure,” in J. Harrington & M. Tabain, Eds, *Speech production: Models, phonetic processes, and techniques*, New York: Psychology Press, 2006, pp. 167–186.

[24] A. Turk and S. Shattuck-Hufnagel, “Timing in talking: what is it used for, and how is it controlled?,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130395, Nov. 2014.

[25] H. Quené, “Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo,” *The Journal of the Acoustical Society of America*, vol. 123, no. 2, pp. 1104–1113, 2008.

[26] E. Jacewicz, R. A. Fox, and L. Wei, “Between-speaker and within-speaker variation in speech tempo of American English,” *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 839–850, 2010.

[27] C. G. Clopper and R. Smiljanic, “Regional variation in temporal organization in American English,” *Journal of Phonetics*, vol. 49, pp. 1–15, 2015.

[28] T. Cambier-Langeveld, “The domain of final lengthening in the production of Dutch,” *Linguistics in the Netherlands*, vol. 14, no. 1, pp. 13–24, 1997.

[29] T. Kendall, *Speech rate, pause and sociolinguistic variation: studies in corpus sociophonetics*. Springer, 2013.

[30] J. Bishop, “Does working memory predict explicit prosodic phrasing? (And does explicit prosodic phrasing predict implicit prosodic phrasing?),” Ms, City University of New York, Under review.

[31] M. E. Beckman and G. A. Elam, “Guidelines for ToBI labelling, version 3.0,” *The Ohio State University Research Foundation*, 1997.

[32] A. Alda, *Never Have Your Dog Stuffed: And Other Things I’ve Learned*. Random House, 2009.

[33] T.-J. Yoon, S. Chavarria, J. Cole, and M. Hasegawa-Johnson, “Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI,” in *Eighth International Conference on Spoken Language Processing*, 2004.

[34] J. F. Pitrelli, M. E. Beckman, and J. Hirschberg, “Evaluation of prosodic transcription labeling reliability in the ToBI framework,” in *Third International Conference on Spoken Language Processing*, 1994.

[35] A. K. Syrdal and J. McGory, “Inter-transcriber reliability of ToBI prosodic labeling,” in *Sixth International Conference on Spoken Language Processing*, 2000.

[36] M. Breen, L. C. Dilley, J. Kraemer, and E. Gibson, “Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch),” *Corpus Linguistics and Linguistic Theory*, vol. 8, no. 2, pp. 277–312, 2012.

[37] A. E. Turk and S. Shattuck-Hufnagel, “Multiple targets of phrase-final lengthening in American English words,” *Journal of Phonetics*, vol. 35, no. 4, pp. 445–472, 2007.

[38] T. Cho, J. Kim, and S. Kim, “Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English,” *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. EL384–EL390, 2013.

[39] N. C. Silver, J. B. Hittner, and K. May, “Testing dependent correlations with nonoverlapping variables: A Monte Carlo simulation,” *The Journal of Experimental Education*, vol. 73, no. 1, pp. 53–69, 2004.

[40] O. J. Dunn and V. Clark, “Correlation coefficients measured on the same individuals,” *Journal of the American Statistical Association*, vol. 64, no. 325, pp. 366–377, 1969.

[41] B. Diedenhofen and M. B. Diedenhofen, “Package ‘cocor,’” 2016.

[42] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: Foundation for Statistical Computing, 2017.

[43] J. Yuan, M. Liberman, and C. Cieri, “Towards an integrated understanding of speaking rate in conversation,” in *Ninth International Conference on Spoken Language Processing*, 2006.

[44] H. Quené, “Longitudinal trends in speech tempo: The case of Queen Beatrix,” *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. EL452–EL457, 2013.