

# Phrase-final lengthening modulates listeners' perception of vowel duration as a cue to coda stop voicing<sup>a)</sup>

Jeremy Steffman

*Department of Linguistics, University of California Los Angeles, 3125 Campbell Hall,  
Los Angeles, California 90095-1543, USA  
jsteffman@ucla.edu*

**Abstract:** The present study addresses how listeners may be sensitive to temporal regularities associated with prosody in their perception of durational segmental cues. Specifically, this study tests whether expectations about phrase-final lengthening mediate listeners' categorization of a “coat”~“code” vowel duration continuum, where vowel duration cues stop voicing. Prosodic position of the target was manipulated such that it was either medial, or final in a carrier phrase. Results indicate that this prosodic manipulation influences categorization such that listeners effectively require longer vowel duration for a “code” response when the target is phrase-final, suggesting that prosodic patterns can modulate listeners' processing of temporal cues.

© 2019 Acoustical Society of America  
[MG]

**Date Received:** March 11, 2019    **Date Accepted:** May 27, 2019

## 1. Introduction

It is well known that the phonetic realization of speech sounds varies systematically as a function of their position in the prosodic configuration of an utterance, conceptualized as the phonetic encoding of prosodic structure [e.g., [Cho \(2016\)](#)]. Though the effects of prosody in speech production are well established, the influence of prosodic patterns in listeners' perception of segmental categories is not well understood and open questions remain ([Mitterer et al., 2016](#)). The present study addresses one such question, exploring how listeners may be sensitive to phrase-final lengthening in their perception of vowel duration as a cue to coda stop voicing.

One aspect of prosody is its organization of the temporal structure of speech. For example, [Turk and White \(1999\)](#) discuss prosody as “hierarchical structure [that] influences the domain and distribution of durational effects” (p. 71). Phrase-final lengthening is one such durational effect whereby, generally speaking, sounds are lengthened in (Intonational) phrase-final position [see [Cho \(2016\)](#) for an overview]. In terms of speech perception, some previous research suggests that listeners may indeed be sensitive to phrasal lengthening in perception of durational contrasts. [Nooteboom and Doodeman \(1980\)](#) tested perception of a Dutch vowel length contrast, and found vowels in (syntactic) phrase and utterance final position required longer duration to be categorized as a phonemically long vowel, suggesting phrasal position may mediate listener expectations of vowel lengthening in perception. This result is highly suggestive, though [Nooteboom and Doodeman](#) do not consider possible differences in categorization originating from durational context and perceptual normalization for changes therein. The present study therefore presents an extension, testing final lengthening effects in perception, while controlling for possible confounds introduced by speech rate normalization which takes on crucial relevance given recent research, discussed below.

Listeners adjust categorization of temporal contrasts based on changes in adjacent/proximal speech rate (e.g., [Miller and Liberman, 1979](#)), and this is typically viewed as originating from a domain-general auditory mechanism ([Wade and Holt, 2005](#)). Proximal rate effects have been shown to occur on the basis of preceding [e.g., [Wade and Holt \(2005\)](#)] and following [e.g., [Miller and Liberman \(1979\)](#) and [Newman and Sawusch \(1996\)](#)] changes in speech rate.

Considering normalization for speech rate in exploring listeners' perception of durational cues as a function of (temporal) prosodic patterns is crucial, given that some prosodic patterns predict the same adjustment in categorization as would be

---

<sup>a)</sup>This work was presented at the 175th meeting of the Acoustical Society America, Minneapolis, MN, USA, May 2018 and a previous version appears in *Proc. Mtgs. Acoust.* **33**, 060001 (2018).

expected based on speech rate normalization. Such a case is illustrated well by two recent studies (Kim and Cho, 2013; Mitterer *et al.*, 2016). Kim and Cho (2013) explored if listeners would compensate for initial strengthening in categorizing a VOT continuum, given that VOT has been shown to be robustly longer in intonational phrase (IP)-initial position [e.g., Cho (2016)]. A target sound (categorized as /pa/ or /ba/) was placed in the carrier phrase “let’s hear *x* again” (where *x* is the target). In one condition the target was IP-initial, preceded by an IP boundary, which was marked by lengthening and a low (L-L%) boundary tone on the word “hear.” In another condition, the target was IP-medial and the carrier phrase was a unitary IP. Kim and Cho found that listeners required longer VOT for a /p/ response in the initial condition. This result was taken as supportive of the hypothesis that listeners are sensitive to prosodically conditioned initial strengthening in segmental categorization, whereby preceding boundary cues led listeners to expect longer VOT for a post-boundary stop. However, more recently, Mitterer *et al.* (2016) noted these effects may possibly result from speech rate normalization. This is a plausible alternative explanation because the conditions used by Kim and Cho varied pre-target duration, where the sound preceding the target was relatively long in the IP-initial condition (due to phrase-final lengthening). Preceding lengthening would be expected to shift categorization of a temporal cue like VOT (via speech rate normalization) in the same direction as observed by Kim and Cho [e.g., Summerfield (1981)]. Both explanations of the effect crucially predict the same perceptual adjustment, though they attribute it to fundamentally different causes: on one hand a low-level auditory effect (speech rate normalization) and on the other a sensitivity to the prosodically driven temporal organization of speech.

Steffman (2019) extended these studies, with findings that suggest speech rate normalization incorporates awareness of intonational structure. However this critically does not address the question of how prosodic patterns might *independently* mediate listeners’ perception of durational cues as proposed by Kim and Cho.

Investigating if patterns associated with prosody shift categorization in a way that is distinct from normalization for speech rate will therefore inform our understanding of how prosodic systems mediate listeners’ interpretation of segmental cues in speech perception.

### 1.1 The present study

The segmental cue chosen as a test case is vowel duration as a cue to coda obstruent voicing. In English, vowels are generally longer preceding voiced, as compared to voiceless obstruents [e.g., Chen (1970)], and this is a robust rate-dependent cue for listeners [e.g., Heffner *et al.* (2017)]. The present study asks how perception of this voicing contrast (cued by duration) varies as a function of a target word’s position in a carrier phrase. As an illustration, consider the two carrier phrases given in (1) and (2) below, which represent the conditions used in the present experiment. Here the target sound *x* is a word from a vowel duration continuum, categorized as “coat” or “code,” with coda /t/ and coda /d/, respectively. Both (1) and (2) are a unitary intonational phrase (IP).

- (1) I’ll say *x* now (*x* = medial),
- (2) I’ll say *x* (*x* = final).

First, given research showing later-occurring material influences speech rate normalization [e.g., Miller and Liberman (1979) and Newman and Sawusch (1996)], the presence of following “now” in the MEDIAL condition would be expected to influence categorization of the preceding target. In the present experiment, “now” is lengthened relative to the target, being approximately 320 ms in duration (45 ms for /n/ and 275 ms for /aʊ/), while the target vowel ranges from 80 to 240 ms (described below). Miller and Liberman (1979) in fact showed the presence versus absence of a following syllable can shift categorization of a durational contrast. In their study, an isolated syllable was categorized as /ba/ or /wa/, where transition duration into the vowel was manipulated to make a continuum (longer = /w/, shorter = /b/). In another condition a second *lengthened* syllable /da/ followed the target, creating a /bada/ to /wada/ continuum. The authors found the presence of a following lengthened syllable (as compared to an isolated monosyllable) shifted categorization such that /w/ responses decreased when the following syllable was present. In comparing these results to the present conditions, a speech rate normalization based prediction would be simply that the presence of a following lengthened “now” [in the MEDIAL condition (1) above] would make the target sound relatively short to listeners. As Miller and Liberman showed, this shift can occur in comparison to a syllable without a following sound [analogous to the FINAL condition (2) above]. Concretely, this would predict that in the MEDIAL condition the

target will be perceived as relatively short, *decreasing* “code” responses overall in the MEDIAL (versus FINAL) condition.

In contrast, if listeners are sensitive to phrase-final lengthening in their perception of vowel duration as a cue to voicing, they should expect the target to be relatively lengthened in FINAL position. This predicts longer required vowel durations for a “code” response in the FINAL condition: an effective *decrease* in “code” responses when the target is FINAL (versus MEDIAL). Such a pattern of results would suggest that *expectations* about prosodic structure are modulating categorization, even in defiance the otherwise expected speech rate normalization effects. The present design thus offers a way to disambiguate the two possible influences on categorization sketched above [following Mitterer *et al.* (2016)] as each predicts a different shift in categorization.

Given this set up, a further prediction can be made. Typical phrase-final vowel durations align more closely with longer continuum steps in the present study [e.g., Turk and Shattuck-Hufnagel (2007)], and it might therefore be expected that the effect of position will be more salient at longer steps, where the target vowel is more plausibly interpreted as phrase-final. Shorter continuum steps are substantially shorter than typical phrase-final vowels, such that listeners may not interpret them as plausibly phrase-final (and therefore subject to lengthening). This predicts an interaction between vowel duration and position such that the effect of position should increase at longer vowel durations. Such an outcome would be supportive of the idea that listeners’ interpretation of prosodic patterns is constrained by language experience, i.e., awareness of typical durations for phrase-final sounds.

## 2. The experiments

Two experiments were implemented, with the only difference being whether the stimuli were blocked in presentation, or randomized, discussed below. The experiments consisted of a 2AFC task in which listeners categorized a stimulus from a vowel duration continuum as of one of two English words, “coat” or “code.” These two words were chosen to be fairly matched for frequency from the SUBTLEX<sub>US</sub> corpus (Brysbaert and New, 2009). Prosodic position was manipulated as represented in (1) and (2) above.

### 2.1 Materials

The stimuli were recorded by a ToBI-trained adult male native English speaker, at 44.1 kHz (32 bit) using a SM10A Shure microphone and headset. The starting point of stimulus manipulation was a naturally produced utterance “I’ll say code now” (with typical declarative intonation: high, H\* pitch accents on “I’ll” and “code,” and a low, L-L% boundary tone). Position was manipulated by removing “now” following the target word, where with “now” removed the target was effectively phrase-final (the FINAL condition). The unaltered carrier phrase constituted the MEDIAL condition. Manipulating position in this way creates the potential confound of listeners being provided with closure duration for the target stop in the MEDIAL condition (where “now” follows the target), but not in the FINAL condition (note there is no release burst for the target stop). This is a potential confound because closure duration could serve as a cue to voicing for the target stop, typically being longer for /t/ as opposed to /d/ (e.g., Flege *et al.*, 1992) though evidence for a robust difference in closure duration across voicing categories is mixed (Luce and Charles-Luce, 1985). If closure duration was short such that it served as a cue for voicing, it would potentially bias responses in the MEDIAL condition in the same way as predicted by phrase-final lengthening, making it impossible to tease these effects apart. Following this logic, closure duration was set to be relatively long, i.e., /t/-biasing (90 ms), based on previous studies (e.g., Flege *et al.*, 1992) and the speaker’s own productions.

The manipulation of target vowel duration was carried out with PSOLA resynthesis. The duration of the original naturally produced vowel was approximately 165 ms. The target word was first excised, and audible voicing after closure was removed to render the stop ambiguous. It should be noted that even though voicing during closure was removed, other acoustic cues to voicing are present on the target vowel itself, for example a lowered/falling first formant and f<sub>0</sub> (Castleman and Diehl, 1996). Though these cues offer the potential to bias responses towards “code” overall, the results of the experiments reported here show that listeners still reliably categorize short vowels as “coat,” in line with previous studies which find vowel duration serves as a robust cue which can adequately distinguish voicing categories for coda obstruents even with other correlates of voicing present (e.g., Raphael, 1972).

The duration of the target vowel was resynthesized to create a continuum ranging from 80 to 240 ms, in 20 ms steps. The initial /k/ in the target word, left

unaltered from its original production had a closure duration 55 ms and VOT of 75 ms. A given resynthesized token was spliced into both the MEDIAL and FINAL frame, creating 18 unique stimuli (2 positional conditions  $\times$  9 continuum steps).

## 2.2 Participants

For each experiment, 32 (different) participants were recruited (BLOCKED experiment: 23 female, mean age = 20.4; RANDOMIZED experiment: 19 female, mean age = 21.6). Participants were students at UCLA and received course credit for participation. All participants were self-reported monolingual American English speakers with normal hearing. No participant responses were excluded in either version of the experiment.

## 2.3 Procedure

Testing was carried out in a sound-attenuated booth in the UCLA Phonetics Lab. Stimuli were presented binaurally over Sony MDR-V200 stereo headphones adjusted to a comfortable listening volume. Participants were presented with written instructions on the computer screen that informed them they would hear an English speaker say “I’ll say \_\_\_” and “I’ll say \_\_\_ now,” and that their task was to select if the speaker was saying the word “coat” or “code.” During trials, participants saw the words “coat” and “code” on the screen as they heard the audio stimulus. One word was on the left side of the screen, and the other was on the right (counterbalanced across participants). Participants indicated their response by key press: an “f” key press indicated the word on the left side of the screen, while a “j” key press indicated the word on the right. After each key press, the next trial began automatically. The inter-trial-interval was 250 ms. The system logged key press responses only *after* the offset of the entire stimulus, and participants were instructed to wait until the entire sound had played before pressing a key. This ensures that participants heard the crucial difference between the MEDIAL and FINAL condition (a following “now”) before they responded.

The only difference between the two experiments was the blocking of the stimuli in presentation. In one, the stimuli were entirely randomized. In the other, the stimuli were organized into blocks by position. These will be referred to as the RANDOMIZED and BLOCKED experiments, respectively. In both experiments, participants took a short self-paced break halfway through the trials. In the BLOCKED experiment, stimuli were blocked by position such that each block consisted of four repetitions of all steps on the continuum in a given position condition (randomized within block). In each half of the blocked experiment, participants heard two such FINAL blocks and two MEDIAL blocks, with the order of presentation of these blocks randomized. In both experiments, each participant therefore categorized a total of 16 instances of each unique stimulus (288 trials).

Differences between blocked and randomized presentation have been shown for context effects in previous speech perception research [e.g., Green *et al.* (1997)], where, for example, blocking of stimuli by rate condition decreases speech rate effects in magnitude, though some rate effects do notably persist in blocked presentation (Green *et al.*, 1997). Green *et al.* attribute these sorts of presentation effects to an attentional mechanism wherein variability draws attention to a stimulus dimension and increases its perceptual weight. This same logic can be applied to the present experiment, where variable prosodic position may increase the perceptual weight of prosodic-contextual factors in listeners’ classification of stimuli. The basic empirical prediction is that the effect of position will be larger when stimuli are randomized, such that listeners are exposed to continual variation in across trials. Given that this is a relatively unexplored research area, implementing two experiments along these lines also allows for basic replication in addition to testing possible stimulus presentation effects.

## 2.4 Results

Results were assessed by a linear mixed-effect model with logistic linking function, using the lme4 package in R (Bates *et al.*, 2015). Prosodic position was contrast-coded (MEDIAL mapped to  $-1$ , FINAL mapped to  $1$ ). The random effect structure consisted of by-subject random intercepts, with maximally specified random slopes. EMMEANS (Lenth *et al.*, 2018) was used for *post hoc* testing of interactions. The model outputs, and comparison of contrasts with EMMEANS are shown in Table 1. Categorization plots for each experiment are shown in Fig. 1.

Both experiments showed similar main effects of position whereby overall FINAL position significantly *decreased* “code” responses ( $p=0.013$  in the BLOCKED experiment,  $p<0.001$  the RANDOMIZED experiment). The directionality of the effect

Table 1. Model outputs (at top) and comparison of contrasts with EMMEANS (at bottom) for both blocked and randomized experiments (blocked at left). Values are rounded.

Model output						
	Blocked			Randomized		
	$\beta$ (SE)	$z$	$p$	$\beta$ (SE)	$z$	$p$
Intercept	0.57(0.13)	4.26	< 0.001	0.38(0.10)	3.70	< 0.001
Position	-0.25(0.09)	-2.62	0.013	-0.46(0.07)	-6.43	< 0.001
Vowel dur	2.42(0.18)	13.23	< 0.001	1.79 (0.22)	8.05	< 0.001
Vowel dur: position	-0.31(0.08)	-3.70	< 0.001	-0.17(0.06)	-3.03	< 0.01
Comparison of contrasts						
Duration (ms)	Blocked			Randomized		
	Estimate (SE)	$z$ -ratio	$p$	Estimate (SE)	$z$ -ratio	$p$
80	0.50(0.27)	1.87	0.06	-0.40(0.17)	-2.26	0.024
100	0.26(0.22)	1.18	0.24	-0.53(0.15)	-3.48	< 0.01
120	0.02(0.19)	0.12	0.91	-0.66(0.13)	-4.83	< 0.001
140	-0.22(0.18)	-1.24	0.21	-0.79(0.13)	-5.93	< 0.001
160	-0.46(0.18)	-2.49	0.013	-0.92(0.14)	-6.43	< 0.001
180	-0.70(0.21)	-3.28	0.001	-1.05(0.16)	-6.40	< 0.001
200	-0.94(0.26)	-3.67	< 0.001	-1.18(0.19)	-6.13	< 0.001
220	-1.18(0.31)	-3.85	< 0.001	-1.31(0.23)	-5.81	< 0.001
240	-1.42(0.36)	-3.92	< 0.001	-1.44(0.26)	-5.51	< 0.001

therefore aligns with the hypothesis that listeners are adjusting categorization based on an expectation of phrase-final lengthening. The larger magnitude of the effect in the RANDOMIZED experiment also aligns with the prediction that listeners will exhibit greater sensitivity to prosodic context when it varies frequently in (randomized) stimulus presentation as outlined above.

A significant interaction between vowel duration and position was also found in both experiments ( $p < 0.001$  in the BLOCKED experiment,  $p < 0.01$  in the RANDOMIZED experiment), showing the effect of position varied across the continuum. To investigate the interaction, the effect of position at each continuum step was tested. In the BLOCKED experiment a marginally significant effect of position ( $p = 0.06$ ) was observed at the shortest continuum step (80 ms), whereby MEDIAL position decreased “code” responses (reversed directionality from the main effect of position). From 100 to 140 ms on the continuum, there was no effect of position, and at all longer steps, the

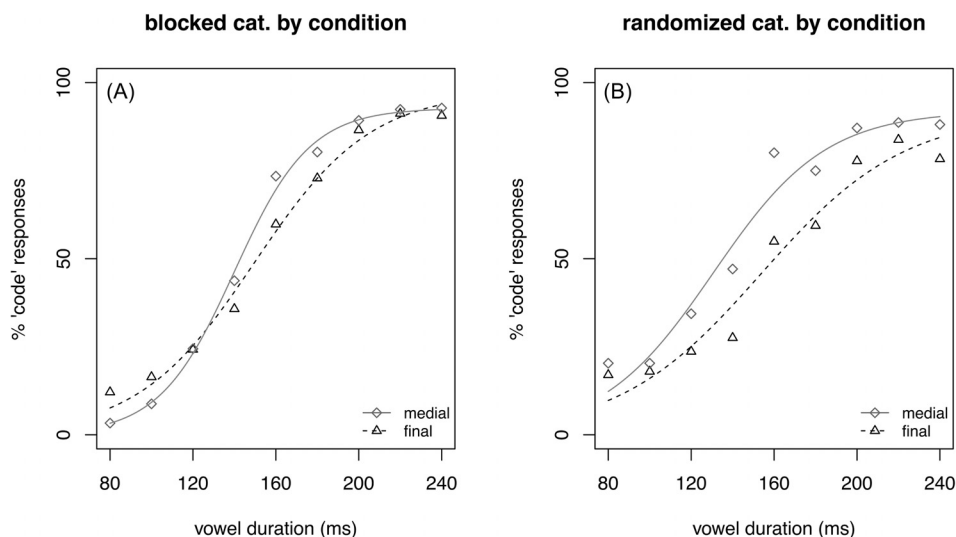


Fig. 1. Categorization split by position condition for both the blocked experiment [(A), left], and randomized experiment [(B), right]. In each plot, the  $x$  axis shows vowel duration values from the continuum, and the  $y$  axis shows the percentage of “code” responses in each condition, at each continuum step. Points represent the raw percentage of code responses, lines are psychometric curves fit to show a smoothed categorization trend.

FINAL condition significantly decreased “code” responses (aligning with the main effect), as shown in Table 1. In the RANDOMIZED experiment the effect of position shows the same directionality at all vowel duration steps, though it increases in magnitude as vowel duration increases. These interactions therefore suggest that listener expectations of phrase-final lengthening exert a more substantial influence at higher continuum steps, aligning with the prediction that positional effects should be largest at continuum steps which map onto more typical phrase-final durations. In this sense the interactions observed in the both experiments can be taken as supporting the idea that the present results reflect listener expectations about phrase-final lengthening, which is constrained to some extent by learned patterns in the language input. The marginally significant effect of position seen at the shortest continuum step, in the BLOCKED experiment only, exhibits the expected speech rate effect, suggesting that when position is less variable, and durations are shorter, speech rate effects may also occur. These results thus suggest that while prosody can indeed modulate listeners’ interpretation of durational cues, the extent to which these effects occur is flexible, and varies based on stimulus and presentation factors.

### 3. General discussion

The present results showed that target sounds in phrase-final position required significantly longer vowel durations to be categorized as “code.” These results thus suggest temporal regularities associated with prosodic structure (here phrase-final lengthening) are indeed used by listeners in their perception of segmental contrasts, which is also supported by the finding that positional influences increased at longer vowel durations. These findings thus point to an interplay between prosodically driven temporal patterns and listeners’ interpretation of segmental cues, in line with the way prosody is encoded in segmental durations in speech (Cho, 2016; Turk and Shattuck-Hufnagel, 2007). The present results can therefore be taken as supporting the proposal of Kim and Cho (2013) that prosody can mediate listeners’ interpretation of segmental cues.

As outlined in Sec. 1, if contextual duration alone was driving the effect, the expected directionality would be reversed. This prediction was based on previous work showing that categorization of a /ba/-/wa/ continuum shifted when the target was an isolated monosyllable, as compared to being followed by a lengthened second syllable (Miller and Liberman, 1979). One crucial difference between the present study and Miller and Liberman is that here the target was placed in a carrier phrase, in lieu of being an isolated mono-/di-syllable. These results may therefore offer potential insight into the circumstances in which prosodic factors play a role in perception, where listeners do not project expectations about phrasal prosodic structure in isolated mono-/di-syllables [as in Miller and Liberman (1979)], but do so when stimuli more closely resemble expected groupings of words into phrases (as in the present study). This is speculative, and further research is needed to address this point, with the goal of elucidating when prosodic factors do and do not constrain how listeners categorize segmental contrasts.

The relationship between domain-general auditory processes like speech rate normalization and additional higher-level factors in rate-dependent speech perception remains an important unresolved question [e.g., Bosker *et al.* (2017)]. The present results highlight that prosody merits consideration along these lines as well. One promising lens with which to consider prosodic effects in perception is a two-stage model [e.g., Bosker *et al.* (2017)], in which category boundary shifts for durational cues may arise in two distinct stages of processing. In such a model, domain-general speech rate normalization arises early in a first stage of processing (Bosker *et al.*, 2017; Reinisch and Sjerps, 2013). A later stage includes domain-specific cognitive adjustments that, crucially, involve “[...] comparison between a target sound and its expected realization given a certain context” [Bosker *et al.* (2017), p. 175]. Prosodically driven adjustments in perception of the kind explored here would by hypothesis occur at this later stage. One empirical prediction along these lines is that the time-course of speech rate and prosodic effects in perception should differ, with rate effects preceding prosodic effects. By hypothesis, speech rate effects may occur automatically (Bosker *et al.*, 2017, Newman and Sawusch, 2009) but later be modulated by prosodic contextual factors. Using time-course sensitive methods, such as eye tracking [following, e.g., Reinisch and Sjerps (2013)] therefore presents a promising extension of the present results, which may help elucidate the processes underlying these prosodic effects, and their relationship to speech rate normalization.

Further research will accordingly benefit from exploring the conditions under which prosodic structure exerts an influence on segmental categorization and how these

effects play out with online measures. Extending the present results along these lines will help us better understand how prosodic patterns constrain speech categorization and interact with other perceptual processes in listeners' interpretation of durational cues.

### Acknowledgments

Many thanks are due to Megha Sundara, Pat Keating, and Sun-Ah Jun for feedback, and to two anonymous reviewers for valuable comments. Further thanks to Adam Royer for recording speech for the stimuli, and to Yang Wang for help with data collection.

### References and links

- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* **67**(1), 1–48.
- Bosker, H. R., Reinisch, E., and Sjerps, M. J. (2017). "Cognitive load makes speech sound fast, but does not modulate acoustic context effects," *J. Mem. Lang.* **94**, 166–176.
- Brysbaert, M., and New, B. (2009). "Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English," *Behav. Res. Methods* **41**(4), 977–990.
- Castleman, W. A., and Diehl, R. L. (1996). "Effects of fundamental frequency on medial and final [voice] judgments," *J. Phon.* **24**(4), 383–398.
- Chen, M. (1970). "Vowel length variation as a function of the voicing of the consonant environment," *Phonetica* **22**(3), 129–159.
- Cho, T. (2016). "Prosodic boundary strengthening in the phonetics–Prosody interface," *Lang. Ling. Compass* **10**(3), 120–141.
- Flege, J. E., Munro, M. J., and Skelton, L. (1992). "Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish," *J. Acoust. Soc. Am.* **92**(1), 128–143.
- Green, K. P., Tomiak, G. R., and Kuhl, P. K. (1997). "The encoding of rate and talker information during phonetic perception," *Percept. Psychophys.* **59**(5), 675–692.
- Heffner, C. C., Newman, R. S., and Idsardi, W. J. (2017). "Support for context effects on segmentation and segments depends on the context," *Atten. Percept. Psychophys.* **79**(3), 964–988.
- Kim, S., and Cho, T. (2013). "Prosodic boundary information modulates phonetic categorization," *J. Acoust. Soc. Am.* **134**(1), EL19–EL25.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (2018). "emmeans: Estimated marginal means, aka least-squares means," <https://CRAN.R-project.org/package=emmeans> (Last viewed October 26, 2018).
- Luce, P. A., and Charles-Luce, J. (1985). "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," *J. Acoust. Soc. Am.* **78**(6), 1949–1957.
- Miller, J. L., and Liberman, A. M. (1979). "Some effects of later-occurring information on the perception of stop consonant and semivowel," *Percept. Psychophys.* **25**(6), 457–465.
- Mitterer, H., Cho, T., and Kim, S. (2016). "How does prosody influence speech categorization?," *J. Phon.* **54**, 68–79.
- Newman, R. S., and Sawusch, J. R. (1996). "Perceptual normalization for speaking rate: Effects of temporal distance," *Percept. Psychophys.* **58**(4), 540–560.
- Newman, R. S., and Sawusch, J. R. (2009). "Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another," *J. Phon.* **37**(1), 46–65.
- Nooteboom, S. G., and Doodeman, G. J. N. (1980). "Production and perception of vowel length in spoken sentences," *J. Acoust. Soc. Am.* **67**(1), 276–287.
- Raphael, L. J. (1972). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," *J. Acoust. Soc. Am.* **51**(4B), 1296–1303.
- Reinisch, E., and Sjerps, M. J. (2013). "The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context," *J. Phon.* **41**(2), 101–116.
- Steffman, J. (2019). "Intonational structure mediates speech rate normalization in the perception of segmental categories," *J. Phonetics* **74**, 114–129.
- Summerfield, Q. (1981). "Articulatory rate and perceptual constancy in phonetic perception," *J. Exp. Psychol. Hum. Percept. Perform.* **7**(5), 1074–1095.
- Turk, A. E., and Shattuck-Hufnagel, S. (2007). "Multiple targets of phrase-final lengthening in American English words," *J. Phon.* **35**(4), 445–472.
- Turk, A. E., and White, L. (1999). "Structural influences on accentual lengthening in English," *J. Phon.* **27**(2), 171–206.
- Wade, T., and Holt, L. L. (2005). "Perceptual effects of preceding nonspeech rate on temporal properties of speech categories," *Percept. Psychophys.* **67**(6), 939–950.