Jeremy Steffman and Sun-Ah Jun: JASA Express Letters https://doi.org/10.1121/1.5126107 Published Online 20 September 2019

Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception[a)]

Jeremy Steffman[b)] and Sun-Ah Jun
Department of Linguistics, University of California Los Angeles, 3125 Campbell Hall, Los Angeles, California 90095-1543, USA
jsteffman@ucla.edu, jun@humnet.ucla.edu

Abstract: Two experiments explored how pitch influences perception of vowel duration as a cue to voicing in light of (1) psychoacoustic interactions between pitch and duration and (2) predicted compensatory effects based on the patterning of pitch and duration in the accentual/prominence-marking system of English. Listeners categorized a "coat"–"code" vowel duration continuum with pitch height on the vowel manipulated. In experiment 1 the expected psychoacoustic effect was observed. In experiment 2 the continuum was altered, highlighting pitch as a prosodic property, resulting in predicted compensatory effects. Results thus indicate prosodic patterns can mediate the perception of durational cues in isolated words.

[MG]
Date Received: June 9, 2019 **Date Accepted:** August 28, 2019

1. Introduction

It is well established that acoustic and articulatory properties of speech segments are systematically modulated by prosodic factors [see Cho (2016) for an overview]. However, the ways in which listeners are sensitive to prosodically driven variation in perception remains an open question (Mitterer *et al.*, 2016). Previous studies investigating perceptual compensation for prosodic patterns have tested boundary phenomena, e.g., initial strengthening and phrase-final lengthening. In the present study we investigate listeners' perception of acoustic correlates of prominence marking in American English. Specifically, we ask how listeners' perception of duration and pitch is influenced both by psychoacoustic factors, and by the patterning of pitch and duration as correlates of accentedness (i.e., prominence marking). We test if listeners incorporate their expectations about the lengthening or shortening due to accentedness (or a lack thereof) in their perception of durational cues in isolated words.

Listeners categorized a "coat"–"code" continuum manipulated to differ in vowel duration. In English (among other languages) vowels before voiced obstruents are longer than those before voiceless obstruents and this is a robust cue to voicing for listeners (e.g., Raphael, 1972). Pitch height on the vowel in the target word was also manipulated to have one of two levels, HIGH and LOW (described in Sec. 2.1). Using this continuum, we tested how listeners' perception of durational cues is influenced by changes in pitch height. Two different predictions are considered: psychoacoustic predictions informed by perceptual interactions between duration and pitch, and linguistic/prosodic predictions informed by the patterning of duration and pitch as correlates of accentedness.

Various psychoacoustic interactions between pitch and duration have been documented in the literature (e.g., Lehiste, 1976; Gruenenfelder and Pisoni, 1980). We consider just one interactive aspect: the influence of pitch height on the perception of duration. Higher pitch increases perceived duration (e.g., Yu *et al.*, 2014). This is argued to have a domain general auditory basis in light of the fact that it is observed with non-speech (Brigner, 1988) and patterns similarly for speakers of different languages (Šimko *et al.*, 2016). Previous studies find evidence of this in listeners' numerical ratings of duration (Yu *et al.*, 2014), as well as comparison of two stimuli (Šimko *et al.*, 2016). In our case, if listeners perceive increased pitch as increased vowel duration, they would be predicted to shift categorization such that a vowel with LOW pitch

[a)]A previous version of this work was presented at the 93th Annual Meeting of the Linguistics Society of America.
[b)]Author to whom correspondence should be addressed.

is perceived as shorter (relative to a vowel with HIGH pitch), and thus less likely to be categorized as code, effectively *decreasing* code responses when pitch is LOW.[1]

This psychoacoustic effect can be contrasted with predicted compensatory effects guided by listeners' interpretation of pitch as a correlate of prosodic accentuation. We first consider some of the structural properties of English prosody related to accentedness. Most accented syllables in English are marked with high (H*) pitch accents (Dainora, 2006), while unaccented syllables tend not to have tonal targets (e.g., Pierrehumbert, 1980). It is well established that, in general, accented syllables and vowels undergo systematic lengthening (e.g., Turk and Sawusch, 1997). These structural properties of the prosodic/intonational system of English engender a very general acoustic consequence: accented syllables have increased duration and pitch relative to unaccented syllables (e.g., Greenberg *et al.*, 2003; Kochanski *et al.*, 2005). In this broad sense, increased pitch and duration can be considered acoustic correlates of accentedness in English. Additionally, increased pitch and duration have been shown to be contributing factors to listeners' perception of prominence in speech (e.g., Ladd and Morton, 1997; Mo, 2011). We can consider this general acoustic correlation in light of prosodically driven compensatory effects (following, e.g., Mitterer *et al.*, 2016). Given that increased pitch correlates with accentedness and contributes to listeners' perception of prominence, we predict that the HIGH pitch condition may give a percept of prominence and accentedness. Because of accentual lengthening, listeners may *expect* longer vowel durations when pitch is HIGH and shorter vowel durations when pitch is LOW. Listeners may therefore compensatorily adjust categorization of the continuum: if an unaccented target (cued by LOW pitch) is expected to be shorter, it should more readily be categorized as code. This predicts that the LOW pitch condition should *increase* code responses, the opposite of the effect predicted by the psychoacoustic influences outlined above.

2. Experiment 1

To test these predictions, we implemented a two-alternative forced choice task. Listeners categorized a stimulus as one of two English words, coat or code. These two words were chosen to be closely matched for lexical frequency from the $\text{SUBTLEX}_{\text{US}}$ corpus (Brysbaert and New, 2009).

2.1 Materials

The stimuli were created from the resynthesized speech of a Tones and Break Indices-trained male American English speaker. The speaker was recorded at 44.1 kHz (32 bit) using SM10A ShureTM microphone and headset in a sound attenuated room. Manipulation of pitch and duration was carried out using PSOLA resynthesis. The starting point for the manipulations was a nuclear (H*) pitched-accented code, produced in the carrier phrase "I'll say code now," where the unaltered target word had a vowel duration of approximately 160 ms. This target word was excised from the carrier phrase and audible voicing after closure was removed to make the coda stop ambiguous in voicing. Pitch on this isolated word was then manipulated to create two conditions. The *f*0 values from the original nuclear pitch-accented target word code will be referred to as the HIGH pitch condition (onset = 135 Hz; offset = 129 Hz). Pitch values for the LOW condition, which we resynthesized onto the target word, were taken from a target word in another carrier sentence in which it was post-focus, thus unaccented, following a contrastively focused "say": "I'll SAY code now" (onset = 112 Hz, offset = 103 Hz). Two vowel length continua were resynthesized from these HIGH and LOW target words, which varied only in pitch. The continua ranged from 60 to 150 ms (note all stimuli were resynthesized in manipulating duration, the unaltered original was not used). The resulting continuum had 15 ms step intervals, with 7 continuum steps in each pitch condition, for a total of 14 unique stimuli. By using pitch values from the prosodic contexts outlined above, we ensure that they are an instantiation of the intended prosodic context under investigation, giving representative pitch for an accented (HIGH pitch) and unaccented (LOW pitch) syllable. Listeners categorized these target words in isolation (with no carrier phrase).

2.2 Participants and procedure

Thirty participants were recruited for experiment 1. Participants were self-reported native American English-speaking adults with normal hearing. Participants were students at UCLA and received course credit. No participant responses were excluded from analysis. Testing was carried out in a sound-attenuated room with participants seated in front of a desktop computer. Stimuli were presented binaurally via a Peltor$^{\text{TM}}$ 3 M$^{\text{TM}}$ (Maplewood, MN) listen-only headset, adjusted to a comfortable

listening level. During trials, participants heard a stimulus and were presented visually with coat and code, one on each side of the screen. Participants indicated their choice via a keypress on the computer keyboard: an "f" keypress indicated the left side choice, and a "j" keypress indicated the right side choice. The side of the screen on which each word appeared was counterbalanced. The inter-trial-interval was 250 ms. Participants categorized 16 instances of each of unique stimulus, for a total of 224 (16×14) randomized trials in the experiment.

2.3 Results and discussion

Results were assessed with a linear mixed-effect model with a logistic linking function, using the lme4 package in R (Bates *et al.*, 2015). Pitch was contrast coded (HIGH was mapped to -1 and LOW mapped to 1). The random effect structure consisted of by-subject random intercepts, with maximal random slopes. Results from experiment 1 are visualized in Fig. 1 (left panel). Table 1 shows the model output.

Pitch, the predictor of interest, showed a significant effect ($\beta = -0.35$, $z = -2.73$, $p < 0.01$), whereby overall LOW pitch significantly decreased code responses relative to HIGH pitch. As outlined above, this effect is expected if HIGH pitch increased perceived vowel duration, which, as a cue to voicing, would effectively generate increased code responses in comparison to LOW pitch. In this sense the main effect of pitch observed in experiment 1 is consistent with the psychoacoustic integration predictions outlined above and concurs with previous studies (e.g., Brigner, 1988; Yu *et al.*, 2014). An interaction between duration and pitch was also observed in the model ($\beta = 0.33$, $z = 6.87$, $p < 0.001$). *Post hoc* comparison of contrasts with emmeans (Lenth *et al.*, 2018) shows pitch has no effect at the three lowest steps of the continuum, and at higher steps the effect increases in magnitude as vowel duration increases (Table 1).

These results overall suggest that variation in pitch influences perceived duration as a cue to voicing, such that increased pitch increases perceived duration. However, the presence of the interaction in the model highlights that this effect is contingent on vowel duration and is only observed with longer continuum steps (greater than 90 ms, see Table 1). Yu *et al.* (2014) found that shorter vowels in their stimuli showed reduced pitch-height effects, in line with our finding that our shortest continuum steps lacked any influence of pitch height. It can also be noted that previously cited studies which find this effect use stimuli which are substantially longer (e.g., 150–250 ms in Yu *et al.*, 2014, 150–450 ms in Šimko *et al.*, 2016) than our own continuum (60–150 ms). Taken together this suggests that these psychoacoustic effects may be limited to longer vowels, though further research is needed to confirm this.

The restricted nature of the effect in experiment 1 indicates that the influence of pitch is contingent on the durational values used. In light of this issue, we return to the question of how pitch and duration pattern as correlates of accentuation in English. Following the logic that compensatory processes related to prosodic structure are learned from acoustic patterns in the language, we consider the duration of vowels observed in spoken corpora of English with the goal of seeing how these durational values compare to the stimuli in experiment 1. Previous corpus studies show that

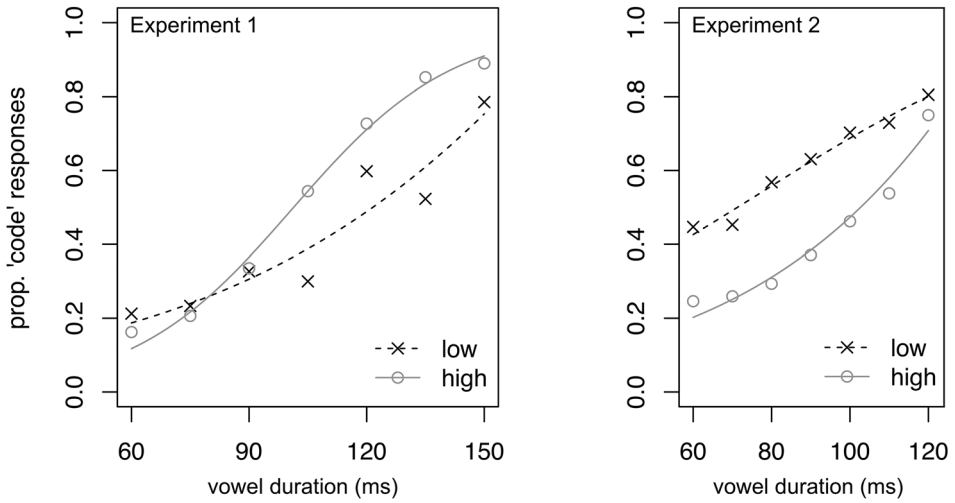

Fig. 1. Categorization along the continuum split by pitch condition for experiment 1 (at left) and experiment 2 (at right). The *x* axis shows vowel duration values. Points show the raw proportion of code responses in each pitch condition. Lines are psychometric curves which are fit to show a smoothed categorization trend.

Table 1. Output from the models and comparison of contrasts using emmeans (Lenth *et al.*, 2018). The results from experiment 1 are given at left, the results from experiment 2 are given at right (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). A coat response is the reference level for the dependent variable, and pitch is contrast coded with HIGH mapped to -1 and LOW was mapped to 1. Vowel duration is treated as continuous and centered at zero.

Experiment 1 model output			Experiment 2 model output		
	β(SE)	z-value		β(SE)	z-value
Intercept	−0.12(0.08)	−1.46	Intercept	0.08(0.06)	1.18
pitch	−0.35(0.13)	−2.73**	pitch	0.51(0.15)	3.28**
vdur	1.34(0.11)	11.85***	vdur	0.76(0.07)	10.64***
vdur:pitch	−0.33(0.05)	−6.87***	vdur:pitch	−0.07(0.03)	−2.54*
Experiment 1 comparison of contrasts			Experiment 2 comparison of contrasts		
duration (ms)	Estimate (SE)	z-ratio	duration (ms)	Estimate (SE)	z-ratio
60	−0.28(0.29)	−0.98	60	−1.26(0.33)	−3.83***
75	0.04(0.7)	0.17	70	−1.19(0.32)	−3.68***
90	0.37(0.26)	1.44	80	−1.11(0.32)	−3.49***
105	0.69(0.25)	2.73**	90	−1.04(0.32)	−3.28**
120	1.03(0.26)	3.91***	100	−0.96(0.32)	−3.03**
135	1.25(0.27)	4.89***	110	−0.89(0.32)	−2.76**
150	1.68(0.30)	5.63***	120	−0.82(0.33)	−2.49*

vowels which are analyzed as unstressed (Greenberg *et al.*, 2003; SWITCHBOARD corpus) and perceived by listeners as lacking prominence (Mo, 2011; Buckeye corpus) are both under 100 ms in duration on average and can be much shorter. Our own stimuli in experiment 1 have a maximum duration (150 ms) that aligns fairly well with the durations of accented vowels and a minimum of 60 ms which aligns clearly with the durations of unaccented vowels, according to Greenberg *et al.* (2003). Additionally, Mo (2011, p. 109) finds that vowels with a duration of 150 ms are perceived as prominent by listeners in a Rapid Prosody Transcription task approximately 65% of the time. In contrast, vowels with 90 ms duration are perceived as prominent only 10% of the time. Therefore, longer steps of our continuum in experiment 1 are above the range for unaccented, non-prominent vowels. In light of Mo's results, longer vowel duration steps in experiment 1 may be judged as prominent by listeners on the basis of duration alone, i.e., a longer vowel with LOW pitch is likely uninterpretable as unaccented, given that unaccented vowels in natural speech are much shorter.

This lack of a possible unaccented interpretation may have influenced listeners to disregard pitch as a correlate of accentedness in these stimuli altogether, resulting in the observed psychoacoustic effect. Stated more generally, when durations are longer than typical accented vowels (as in some previous studies), or when durations are longer than typical unaccented vowels (as in the longer steps in experiment 1) listeners simply may not generate expectations about accentual lengthening or shortening on the basis of pitch differences, with the observed psychoacoustic effect emerging as a default.

Following this logic, we predicted that listeners' sensitivity to LOW pitch as a correlate of un-accentedness may be enhanced when a continuum range is restricted to being relatively short, excluding longer steps which cannot be interpreted as unaccented with LOW pitch and which may be judged as prominent on the basis of duration [following Mo (2011)]. To test this hypothesis, we implemented a second experiment, using a continuum spanning shorter vowel durations with the goal of highlighting pitch as a property related to prominence-marking.

3. Experiment 2

3.1 Materials

In experiment 2, two changes were made to the continuum from experiment 1. First, the maximum value was reduced from 150 to 120 ms, creating a new continuum with endpoints of 60 and 120 ms. Second, the step size was reduced from 15 to 10 ms (7 steps total). By reducing the range of the continuum, listeners are exposed to less extreme variability in duration, rendering durational differences less salient, and reducing the possibility of perceiving prominence on the basis of vowel duration alone (Mo, 2011).

The reduced step size further makes changes in duration less perceptible. A 10 ms step size is quite small for a vowel duration continuum and approaches the just-noticeable difference for continuum steps 100 ms and longer (e.g., Klatt and Cooper 1975). In this sense, vowel duration has become a less reliable cue to voicing for listeners, including potentially perceived duration as a function of pitch. Listeners therefore may be more likely to interpret pitch as prosodic property, compensating for differences in pitch height as originally predicted. If listeners do indeed adjust categorization along these lines, we predict that LOW pitch, if interpreted as cuing a lack of prominence (generating an expectation of shorter target vowels) should *increase* listeners' code responses. This predicts a reversal of the effect of pitch seen in experiment 1.

3.2 Participants and procedure

Thirty (different) participants were recruited. The procedure was identical to experiment 1.

3.3 Results and discussion

The statistical assessment and model fitting procedure were the same as in experiment 1. Results from experiment 2 are visualized in Fig. 1 (right panel). As in experiment 1, increasing vowel duration significantly increased code responses (Table 1). However, the effect was smaller than that in experiment 1, suggesting that, as expected, vowel duration has become a less reliable cue to voicing. In this context of decreased durational influence, pitch showed a significant main effect ($\beta = 0.51$, $z = 3.28$, $p < 0.01$), whereby LOW pitch significantly *increased* code responses relative to HIGH pitch, a reversal of the effect found in experiment 1.

In general, LOW pitch steps in experiment 2 are far more likely to be categorized as code across the continuum compared to those in experiment 1 (see Fig. 1). In contrast, responses to HIGH pitch align fairly closely with categorization of comparable HIGH pitch stimuli in experiment 1, suggesting that the main difference across experiments is listeners' interpretation of LOW pitch. A significant interaction was also observed in the model ($\beta = -0.07$, $z = -2.54$, $p < 0.05$), showing that the magnitude of the effect of pitch decreases systematically as vowel duration increases (see Table 1). The presence of this interaction suggests that listeners are more sensitive to pitch differences at shorter vowel durations, which may also be taken in support of the idea that shorter, and thus more plausibly unaccented vowels enhance listeners' interpretation of LOW pitch as prosodic at these continuum steps.

One question is why HIGH pitch categorization patterns uniformly across experiments, while LOW pitch categorization changes substantially. One possibility is that listeners' experience with co-occurring LOW pitch and shorter vowels would lead them to project an expectation of relative shortening when exposed to LOW pitch, as described above, while uniform HIGH pitch categorization may be explained by a *lack* of HIGH pitch short vowels in listeners' experience. Given that shorter vowels like those in the range of experiment 2 are unlikely to co-occur with higher, accented-like pitch values (recall HIGH pitch values come from a nuclear pitch-accented word), listeners may simply not project prosodic expectations onto shorter HIGH pitch vowels, resulting in unaltered categorization across both experimentexperiments in the HIGH pitch condition. In other words, it is possible that the HIGH pitch stimuli show relatively stable categorization across experiments due to a non-linguistic interpretation of pitch, with default psychoacoustic processing of HIGH pitch and duration as a result. This explanation is speculative and further work will benefit from testing various pitch heights and vowel durations to see if and how listeners experience (or lack thereof) with co-varying prosodic cues engenders sensitivity to particular stimulus properties.

4. General discussion

The present study provides us with a nuanced view of how listeners' experience with prosodic cues may mediate their perception of durational contrasts, and interface with psychoacoustic perceptual processes. experiment 1 suggests that listeners' perception of vowel duration as a cue to voicing can be influenced by pitch height, reflecting psychoacoustic integration of pitch and duration. This aligns with previous literature using explicit judgments to test listeners' perception of duration as a function of pitch height. In experiment 2, which used shorter durations and reduced the perceptibility of durational differences among stimuli, the effect of pitch was reversed entirely. Taken together, these results suggest that compensation for prosodically driven variation in pitch and duration can indeed be observed under the right circumstances, when durational differences are less salient, and also when the durational range of the stimuli

maps onto language-typical patterns for the property of interest, i.e., short vowels are typically unaccented and low pitched.

These results further suggest that prosodically driven compensatory effects for accentual prominence can occur even in isolated words, extending research on the topic which had previously placed words within carrier phrases that varied prosodic context (e.g., Mitterer *et al.*, 2016; Steffman, 2019). The fact that listeners are adjusting categorization of isolated words on the basis of prosodic factors may suggest that they are able to perceptually access prosodic information for words in isolation. One possible lens with which to consider these results is the view that listeners retain phonetically rich representations of sounds in memory as couched in exemplar theories of speech perception (e.g., Pierrehumbert, 2001). That is, prosodic-structural factors may introduce patterned acoustic variability (in duration and pitch), which is encoded and retained by listeners and influences categorization, even in words that are dissociated from an explicit prosodic context. Various previous studies have suggested that listeners retain phonetically rich representations of prosodic information (e.g., Kimball *et al.*, 2015; Schweitzer *et al.*, 2015), and the present study may offer evidence of this in the form of a categorization task. Additionally, it has been argued that effects which can be characterized as compensatory may arise as a natural consequence of exemplar storage (e.g., Johnson, 1997).

The present results also highlight that these processes are not deterministic, rather they are flexible, aligning with previous arguments for such flexibility in the perceptual integration of duration and pitch (e.g., Prince, 2011) which appears to occur in experiment 1, but not experiment 2. This is clearly evidenced by the reversal of the effect where identical stimuli (e.g., the 60 ms step) were categorized as code at markedly different rates based on continuum range and step size. We argue that the effects seen in experiment 1 stem from listeners disregarding pitch as a cue to accentuation for reasons discussed above, such that integration occurs and generates the observed psychoacoustic effects at longer continuum steps. In experiment 2, we argue that listeners are more readily able to interpret pitch as a prosodic cue, especially LOW pitch stimuli, at which point it is possible that experience with prosodic patterns, in the form of stored exemplars, will play a role.

Though an exemplar account may therefore explain the findings in experiment 2, it must crucially be complimented by an explanation of when prosodic patterns will and will not influence listeners' perception. Such a mechanism must necessarily be flexible, in similar fashion to proposed attentional effects on perceptual processing (e.g., Green *et al.*, 1997), which have also been speculated to play a role in prosodic context effects (Steffman, 2019). In the present case, the prosodic effects seen in experiment 2 may occur when attention to pitch as a prosodic cue is enhanced by continuum step size and range, and when acoustic properties in the stimuli map onto stored exemplars. This is speculative, and further research will benefit from testing this idea with varying stimulus presentations or other attentional manipulations (Green *et al.*, 1997; Miller, 1987). Extension of the present results along these lines may prove useful as a way of investigating the perceptual mechanisms underlying these effects.

Further study will therefore better our understanding of how listeners' flexible interpretation of pitch and duration mediates their perception of speech sounds, and under which circumstances language experience with prosodic patterns constrains listeners' interpretation of segmental contrasts.

Acknowledgments

We are grateful to two anonymous reviewers and Marc Garellek for helpful feedback and suggestions. We also thank Adam Royer for recording speech for the stimuli, and Yang Wang and Danielle Bagnas for help with data collection.

References and links

[1]One might wonder if lowered pitch can serve as a micro-prosodic cue for voicing in our stimuli (Kohler, 1985), which would predict that LOW pitch would increase *code* responses. However, as argued in Gruenenfelder and Pisoni (1980), global pitch dynamics over an entire vowel play a central role in the perception of vowel duration (following Lehiste, 1976) in lieu of cuing voicing directly as a micro-prosodic property. Kohler (1985) further shows that a predominant influence of pitch as a direct cue to voicing is in localized pitch changes. Since our manipulations altered overall pitch height, lowered pitch may not be interpretable as a micro-prosodic voicing cue. This is confirmed from the results of experiment 1, which show LOW pitch height is not being interpreted in this way.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.

Brigner, W. L. (**1988**). "Perceived duration as a function of pitch," Percept. Motor Skills **67**(1), 301–302.
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.
Cho, T. (**2016**). "Prosodic boundary strengthening in the phonetics–prosody interface," Lang. Linguist. Compass. **10**(3), 120–141.
Dainora, A. (**2006**). "Modeling intonation in English: A probabilistic approach to phonological competence," in *Laboratory Phonology 8*, edited by L. Goldstein, D. H. Whalen, and C. T. Best (Walter de Gruyter, Berlin, Germany).
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.
Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (**2003**). "Temporal properties of spontaneous speech—A syllable-centric perspective," J. Phonetics **31**(3), 465–485.
Gruenenfelder, T. M., and Pisoni, D. B. (**1980**). "Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production," Percept. Psychophys. **28**(6), 514–520.
Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic Press, San Diego, CA), pp. 145–165.
Kimball, A., Cole, J., Dell, G., and Shattuck-Hufnagel, S. (**2015**). "Categorical vs. episodic memory for pitch accents in American English," in *Proceedings of the 18th International Congress of Phonetic Sciences*, pp. 1–4.
Klatt, D. H., and Cooper, W. E. (**1975**). "Perception of segment duration in sentence contexts," in *Structure and Process in Speech Perception* (Springer, Berlin, Heidelberg), pp. 69–89.
Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). "Loudness predicts prominence: Fundamental frequency lends little," J. Acoust. Soc. Am. **118**(2), 1038–1054.
Kohler, K. J. (**1985**). "F0 in the perception of lenis and fortis plosives," J. Acoust. Soc. Am. **78**(1), 21–32.
Ladd, D. R., and Morton, R. (**1997**). "The perception of intonational emphasis: Continuous or categorical?," J. Phonetics **25**(3), 313–342.
Lehiste, I. (**1976**). "Influence of fundamental frequency pattern on the perception of duration," J. Phonetics **4**(2), 113–117.
Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (**2018**). *emmeans: Estimated Marginal Means, aka Least-Squares Means*, available at https://CRAN.R-project.org/package=emmeans.
Miller, J. L. (**1987**). "Mandatory processing in speech perception: A case study," in *Modularity in Knowledge Representation and Natural-Language Understanding*, edited by Jay L. Garfield (MIT Press, Cambridge, MA), pp. 309–322.
Mitterer, H., Cho, T., and Kim, S. (**2016**). "How does prosody influence speech categorization?," J. Phonetics **54**, 68–79.
Mo, Y. (**2011**). "Prosody production and perception with conversational speech," Doctoral dissertation, University of Illinois at Urbana-Champaign.
Pierrehumbert, J. (**1980**). "The phonology and phonetics of English intonation," Ph.D. thesis, Massachusetts Institute of Technology.
Pierrehumbert, J. B. (**2001**). "Exemplar dynamics: Word frequency lenition and contrast," in *Typological Studies in Language*, edited by J. L. Bybee and P. Hopper (John Benjamins Publishing, Amsterdam, the Netherlands), Vol. 45, pp. 137–158.
Prince, J. B. (**2011**). "The integration of stimulus dimensions in the perception of music," Q. J. Experiment. Psychol. **64**(11), 2125–2152.
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.
Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., and Dogil, G. (**2015**). "Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation," Speech Commun. **66**, 65–81.
Šimko, J., Aalto, D., Lippus, P., Włodarczak, M., and Vainio, M. (**2016**). "Pitch, perceived duration and auditory biases: Comparison among languages," in *18th International Congress of Phonetic Sciences*, Glasgow Scotland.
Steffman, J. (**2019**). "Phrase-final lengthening modulates listeners' perception of vowel duration as a cue to coda stop voicing," J. Acoust. Soc. Am. **145**(6), EL560–EL566.
Turk, A. E., and Sawusch, J. R. (**1997**). "The domain of accentual lengthening in American English," J. Phonetics **25**(1), 25–41.
Yu, A., Hyunjung, L., and Jackson, L. (**2014**). "Variability in perceived duration: Pitch dynamics and vowel quality," in *Proceedings of the 4th International Symposium on Tonal Aspects of Languages*, pp. 41–44.