Contextual prominence in vowel perception

Citation for published version:
Steffman, J 2021, 'Contextual prominence in vowel perception: Testing listener sensitivity to sonority expansion and hyperarticulation', *JASA Express Letters*, vol. 1, no. 4, pp. 045203.
https://doi.org/10.1121/10.0003984

Digital Object Identifier (DOI):
10.1121/10.0003984

Published In:
JASA Express Letters

Contextual prominence in vowel perception: Testing listener sensitivity to sonority expansion and hyperarticulation

Cite as: JASA Express Lett. **1**, 045203 (2021); https://doi.org/10.1121/10.0003984
Submitted: 20 January 2021 • Accepted: 15 March 2021 • Published Online: 15 April 2021

Jeremy Steffman

JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Contextual prominence in vowel perception: Testing listener sensitivity to sonority expansion and hyperarticulation

Jeremy Steffman[a)]
Department of Linguistics, Northwestern University, Evanston, Illinois 60208, USA
jeremy.steffman@northwestern.edu

Abstract: High front vowels have been shown to undergo two competing forms of acoustic (and articulatory) modulation due to prosodic prominence—(1) hyperarticulation: more extreme high/front articulations under prominence and (2) sonority expansion: more open articulations, allowing more energy to radiate from the mouth. This study explores how these effects translate into listeners' perception of the contrast between the vowels /i/ and /ɪ/. Results show that listeners uniformly expect a hyperarticulated vowel (acoustically) under prominence, and adjust categorization of an F1/F2 continuum accordingly. Results are discussed in relation to production findings and possible accounts of why listeners favor hyperarticulation in perception.

[Editor: Douglas D O'Shaughnessy] https://doi.org/10.1121/10.0003984

Received: 20 January 2021 **Accepted:** 15 March 2021 **Published Online:** 15 April 2021

1. Introduction and study goals

Prosodic prominence has been shown to modulate how speakers articulate segmental material, with so-called *prominence strengthening* effects that can be described as paradigmatic contrast enhancement (e.g., Cho, 2005). This study tests how such effects translate into speech perception. Consider an example of prominence strengthening: cues to stop voicing contrasts such as VOT and stop closure duration are enhanced under prominence such that prominent stops show greater differentiation in these cues, as compared to non-prominent stops (Cole *et al.*, 2007), enhancing the voicing contrast. With respect to vowels, a comparable example comes from de Jong (1995) who finds that speakers produced prominent rounded vowels (/ʊ/) with increased lip protrusion and rounding, enhancing the [+round] feature of the vowel and its contrast with other vowels. Following de Jong (1995), these effects can be described as localized *hyperarticulation*.

Another influence of prominence strengthening in vowels, generally framed in terms of syntagmatic enhancement (i.e., enhancing distinctions between adjacent units in the speech stream) is so-called *sonority expansion* where sonority is defined as the general openness of the vocal tract (e.g., de Jong *et al.*, 1993). Generally speaking, prominent vowel articulations show increased amplitude of jaw opening and in some cases lingual backing, all serving to produce a more open articulation with more energy radiating from the mouth (e.g., de Jong *et al.*, 1993 and Erickson, 2002). Both hyperarticulation and sonority expansion have been shown to modulate vowel formant structure in line with the articulatory adjustments they entail (Cho, 2005, Kim *et al.*, 2016, and van Summers, 1987). This study thus asks if listeners adjust their perception of formant cues on the basis of prosodic prominence, that is, in line with how formant structure is altered by prominence in speech production.

In the particular test case adopted here, sonority expansion and hyperarticulation in vowel articulations and their consequences on formant structure can conflict: this is for high front vowels such as American English /i/ and /ɪ/. A hyperarticulated variant of these vowels should be produced with a more closed articulation, that is, an articulation more closely approaching a high/front target (acoustically with lowered F1, and raised F2). In comparison, sonority expansion would entail a more open vowel articulation, not as closely approximating a high front target (acoustically with raised F1, and possibly lowered F2).[1] Empirically, various speech production (cinefluorographic, EMA, acoustic) studies of American English /i/ and /ɪ/ (primarily /i/) show variation. Some suggest sonority expansion (Houde, 1967), others hyperarticulation (Kent and Netsell, 1971), and others some intermediate patterns (Cho, 2005). Also, apparent is within-study inter-speaker variation (Cho, 2005). Taking this literature as a whole, we can conclude that speakers appear to variably prioritize these two prominence strengthening patterns in high front vowels. One outstanding question is thus how listeners relate

[a)]ORCID: 0000-0003-3675-910X.

contextual prominence information to how a prominent high front vowel should be realized, and if listeners, like speakers, are variable in this regard.[2]

Recently, Steffman (2021b) showed that listeners adjust their perception of non-high front vowels /ɛ/ and /æ/ (cued by F1 and F2) based on prominence in line with sonority expansion. However, these vowels, unlike high front vowels do not show the same tension between sonority expansion and hyperarticulation. We thus have a reason to expect that listeners will adjust their perception of formant cues on the basis of contextual prominence, however, what precisely this adjustment might be for high front vowels remains an open question. The present experiment addresses this in testing if and how listeners adjust their categorization of the contrast between /i/ and /ɪ/, in line with acoustic prominence strengthening effects. In testing this we will explore a possible way in which prominence influences speech comprehension, and how listeners tune into and prioritize conflicting prominence strengthening patterns in perception.

2. Methods

The study implemented a two-alternative forced choice (2AFC) task in which listeners categorized a continuum ranging between two vowel endpoints as one of two words, "sit" (/ɪ/) or "seat" (/i/). These target words were placed in a carrier phrase which manipulated contextual prominence, signaling the target as prominent, or non-prominent in relation to context.

2.1 Materials

Stimuli were created by re-synthesizing the speech of a ToBI-trained American English speaker. The speech material was recorded in a sound-attenuated booth in the UCLA Phonetics Lab, using an SM10A Shure™ microphone and headset. Recordings were digitized at 32 bits and a 44.1 kHz sampling rate. Supplementary images showing visual representations of both the continuum manipulation and prominence manipulation, as well as the stimuli, can be found in supplementary materials for the article hosted on the Open Science Foundation (Steffman, 2021a).

The carrier sentence frames were designed to convey a perceived difference in prominence. They are shown schematically with ToBI labels in (1) and (2) below, where *x* is the target (categorized as "sit" or "seat").

(1) I'll say *x* now
H* H* L-L%
(2) I'll SAY *x* now
L + H* L-L%

In (1), the target is prominent, bearing the nuclear pitch accent in the phrase which contains a standard declarative tune. In contrast, in (2) the target follows narrow focus marking, realized with a rising L + H* accent on the word "say"; the target is therefore unaccented and non-prominent. The two prominence conditions in the present study, which will be referred to as the nuclear pitch accent (NPA) and post-focus conditions were created to manipulate only the context surrounding the target (with the target identical across conditions), in such a way that listeners' perception of target prominence was roughly equivalent to the (phonological) prominence distinction exemplified in (1) and (2). These stimuli accordingly present a fairly conservative manipulation, changing only context to ensure that properties of the target sound itself did not shift listeners' perception. Any differences observed across conditions in the experiments that follow can only be attributed to context. The starting point for the creation of these frames was (1) above. The two frames were created by cross-splicing and PSOLA method resynthesis. The NPA condition, in which the target is contextually prominent, was created simply by using the frame in (1), from which the target sound was excised. To create the post-focus condition, the vowel in "say" from (2), with narrow focus, was spliced into the frame, replacing the vowel in "say" from (1). The vowel in "say" in the post-focus condition therefore contained increased amplitude and duration relative to "say" in (1). Pitch on the preceding word "I'll" was re-synthesized to match the pitch values of this word in (2): low-dipping pitch realizing the low target of the L + H* accent. Pitch on "I'll" in the NPA condition was *also* resynthesized, overlaid with nearly identical pitch from another production of (1), to ensure that that both "I'lls" underwent an equal amount of resynthesis. The post-target material was identical across conditions, as produced in (1), which was highly similar to its production in (2). It was unaccented and phrase-final with a low (L-L%) boundary tone. The resulting stimuli thus contained differences in the pre-target pitch contour, and the duration, overall amplitude and envelope of the preceding vowel /eɪ/ in "say."

In creating the target word, F1 and F2 were manipulated orthogonally by LPC decomposition and resynthesis using the Burg method via a PRAAT script (Winn, 2016). The starting point for the target was a production of "seat," produced with nuclear prominence as in (1). The pitch and intensity of this target sound was set to be the average of these two parameters between the "seat" production in (1) and (2) (a nuclear accented and unaccented target), such that the target itself was ambiguous in terms of prominence, and appropriate for both frames (1) and (2). The formant values for each endpoint were based on model sound productions of "sit" and "seat." Model sound endpoint values were slightly modified from the model speaker's productions of /i/ and /ɪ/, with the F1 dimension expanded slightly to make both dimensions span an equal range in Bark units (Traunmüller, 1990). The goal in making each dimension equally distributed in Bark space was to ensure that any asymmetrical effects of prominence on F1 and F2 were not due to these dimensions spanning a different extent of perceptual space. The resynthesis process estimated source and filter for the starting model sound from the "seat" model. The filter model's F1 and F2 were then adjusted to match those of a model "sit" production. From these two filter models, 2 intermediate filter steps in each F1 and F2 dimensions were created by interpolating between these model endpoint values in Bark. Phase-locked higher frequencies from the starting base /i/ model that were lost in the process of LPC resynthesis were restored to all continuum steps, improving the naturalness of the continuum. The results was four continuum steps varying orthogonally in each dimension, for a total of 16 steps on the continuum, ranging from 3 to 4.8 Bark in F1 space, and 12.4 to 14.2 Bark in F2 space. Each of the 16 target steps was then spliced into the two carrier phrase frames, creating a total of 32 unique stimuli (4 F1 steps × 4 F2 steps × 2 frames).

2.2 Participants and procedure

Listeners were asked to categorize the target sound as the English word "sit" or "seat" by key press ("f" and "j" keys). Participants were presented with orthographic representations of each target word on a computer monitor, one centered on each half of the screen. The side of the screen on which each target appeared was counterbalanced across participants. 38 native speakers of American English were recruited for the present study. All participants completed the study remotely, accessing the experiment online. All were instructed to complete the experiment while wearing headphones in a quiet space. The experiment consisted of 256 randomized trials (8 presentations of 32 unique stimuli), with a self-paced break halfway through the trials. All responses were analyzed.

2.3 Predictions

Given the structure of the stimuli we can consider two predictions, corresponding to (1) a perceptual expectation of hyperarticulation and (2) a perceptual expectation of sonority expansion, in terms of formant structure. First following the hyperarticulation account, listeners might expect a prominent vowel to be realized with lowered F1 and raised F2, reflecting a more fronted, and raised vowel articulation. If listeners compensate accordingly, when the target vowel is prominent (in the NPA condition), they should require lower F1 and higher F2 to categorize a target as /i/. This predicts, overall, relatively *decreased* /i/ responses in the prominent NPA condition. In comparison, following sonority expansion, sonority expanding gestures under prominence would lead to more open articulations overall, resulting in, acoustically, a vowel with higher F1 and potentially lower F2. If listeners compensate accordingly, when the target vowel is prominent (in the NPA condition), they should expect higher F1 and lower F2 in categorizing a target as /i/. This predicts, overall, relatively *increased* /i/ responses in the prominent NPA condition.

3. Results

Results were assessed using a Bayesian mixed-effects logistic regression implemented with *brms* in R (Bürkner, 2017). The model was fit with uninformative priors, and predicted listeners' categorization response (an /i/ response mapped to 1, /ɪ/ mapped to 0), as a function of prominence (contrast-coded with NPA mapped to 0.5, post-focus to –0.5) F1, and F2 (both scaled and centered), as well as interactions for these fixed effects. Random effects consisted of by-participant random intercepts with the fixed effect terms and their interactions as by-participant random slopes. Estimates with 95% credible intervals which exclude zero are taken to have a reliable (i.e., credible) impact on listener responses. The model output is shown in Table 1. The data for this experiment and an R script showing the data analysis and visualization can be found in the supplementary materials for the article (Steffman, 2021a).

F1 and F2 evidenced an expected, and credible, influence in listeners' categorization, as shown by panel A of Fig. 1. As F1 increases, /i/ responses decrease, and likewise as F2 increases, /i/ responses increase both in a gradient fashion, as can be seen in panel A. Turning to the effect of prominence, panel B of Fig. 1 plots listeners' responses, arrayed with /i/ responses plotted on the y axis. As can be seen by the differentiation in prominence conveyed by line type, overall listeners show credibly *decreased* /i/ responses in the prominent NPA condition ($\beta = -0.26$, 95%CI = [–0.42, –0.09]). As outlined above, this outcome is consistent with the prediction based on hyperarticulation, that is, listeners expected an acoustically more extreme

Table 1. Effect estimates, error and 95% credible intervals. Credible effects are bolded.

	Estimate	Est. Error	L-95% CI	U-95%CI
Intercept	**−0.55**	**0.15**	**−0.84**	**−0.25**
Prominence	**−0.26**	**0.08**	**−0.42**	**−0.09**
F1	**−1.80**	**0.15**	**−2.09**	**−1.51**
F2	**2.63**	**0.18**	**2.28**	**2.99**
F1:F2	**0.77**	**0.11**	**0.56**	**0.99**
F1:prominence	−0.01	0.10	−0.19	0.19
F2:prominence	0.01	0.11	−0.20	0.22
F1:F2:prominence	−0.01	0.10	−0.21	0.19

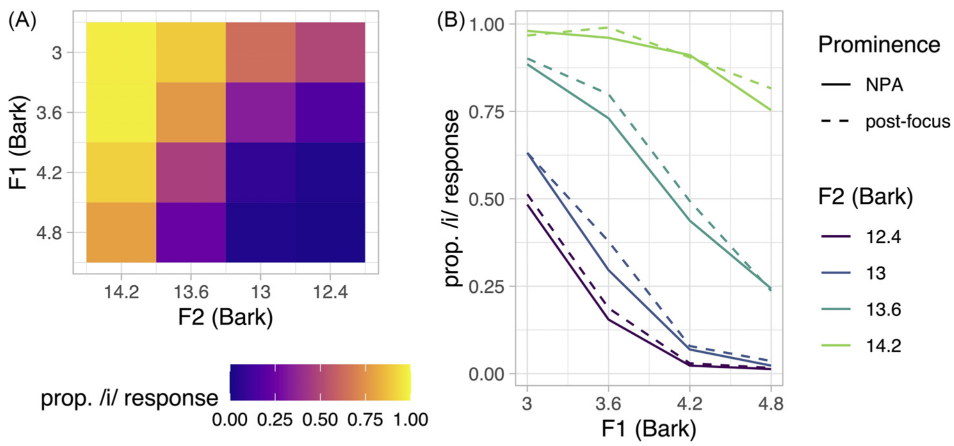

Fig. 1. (A) categorization responses at each continuum step arrayed in a grid, with F1 on the y axis and F2 on *x* axis (note axes are reversed to mirror the vowel space, with /i/ in the top left corner). Responses are represented on a color scale, shown below the plot. (B) /i/ responses (*y* axis) as a function of F1 (*x* axis), F2 (color scale), and prominence (line type).

realization of the vowel under prominence, contra sonority expansion. Two additional estimates in the model were seen to be credible, though they are only discussed briefly here as they are not directly relevant to the questions at hand. First the intercept was seen to be credibly less than zero, suggesting an overall /ɪ/ bias in the continuum, evident in Fig. 1(B).[3] Nevertheless, as is clear from Fig. 1, categorization at continuum endpoints is well anchored. The interaction between F1 and F2 was additionally found to be credible, showing that listeners' sensitivity to a given formant was modulated by the other formant value. This is not discussed further here as it is peripheral to the main point of interest, though a visualization showing the interaction can be found in the supplementary materials (Steffman, 2021a).

To further inspect how variable participants are, particularly with respect to the prominence effect, by-participant effect estimates were obtained by summing random slope values with a fixed effect of interest (see, e.g., Politzer-Ahles and Piccinini, 2018). This generates a model-estimated value for each participant's weighting of a given effect. This allows us to observe, in model terms, how participants vary with respect to F1, F2, and prominence. It was observed that participants were remarkably consistent with respect to the prominence effect. All participants' estimates are negative, showing an adjustment in line with hyperarticulation for each of them [smallest magnitude estimate: −0.15, largest magnitude estimate: −0.34; see the R script in the supplementary materials (Steffman, 2021a) for all estimates and their distribution]. This outcome suggests listeners consistently expect a hyperarticulated realization of the target vowel under prominence, and none show an adjustment consistent with sonority expansion (i.e., none show a positive estimate indicating increased log-odds of an /i/ response in the NPA condition).

We can further inspect if these by-participant estimates correlate with one another, as a way of assessing if participants' weighting of formant cues corresponds to their sensitivity to the prominence effect. The absolute value of these by-participant estimates from the model was taken to visualize this.[4] As shown in Fig. 2(A), F1 shows only a weak positive correlation with prominence ($\tau = 0.12$, $p = 0.30$), while in comparison F2 shows a robust positive correlation ($\tau = 0.39$, $p < 0.001$) with prominence (panel B). This result shows that listeners who use F2 more a cue to distinguish the contrast show larger adjustments on the basis of contextual prominence for this contrast. Thus, though by-participant estimates do not reveal variability in the *directionality* of the effect across participants, they do show that participants' cue-weighting correlates with their sensitivity to contextual prominence. This is consistent with previous findings suggesting that the acoustic consequences of hyperarticulation are more impactful in terms of F2 in high front vowels (Cho, 2005; Kim *et al.*, 2016), thus listeners who rely more on this dimension to distinguish the contrast, compensate more for changing F2 as a function of prominence.

4. Conclusions

In summary, these results show that listeners exhibit a general and consistent adjustment in categorization of the contrast between American English /i/ and /ɪ/ under prominence. This shows, most generally, that listeners relate the contextual prominence of a vowel with an expectation of how it should be realized. In this particular test case we see strong evidence for an expectation of (acoustic) hyperarticulation: that is, that prominent high front vowels should be realized with lower F1 and higher F2.[5] The present study manipulated only F1 and F2 in a vowel, which leaves open the question of how other sonority-cuing parameters, beyond F1 (e.g., voice quality, as in Chong *et al.*, 2020), factor into this equation. The possibility of using other cues to signal sonority might serve as an explanation for why hyperarticulation effects are prevalent here: if it is the case that other acoustic parameters (which do not directly impact formant structure) often help to encode sonority in prominent high front vowels, listeners may not expect formant variation in line with sonority

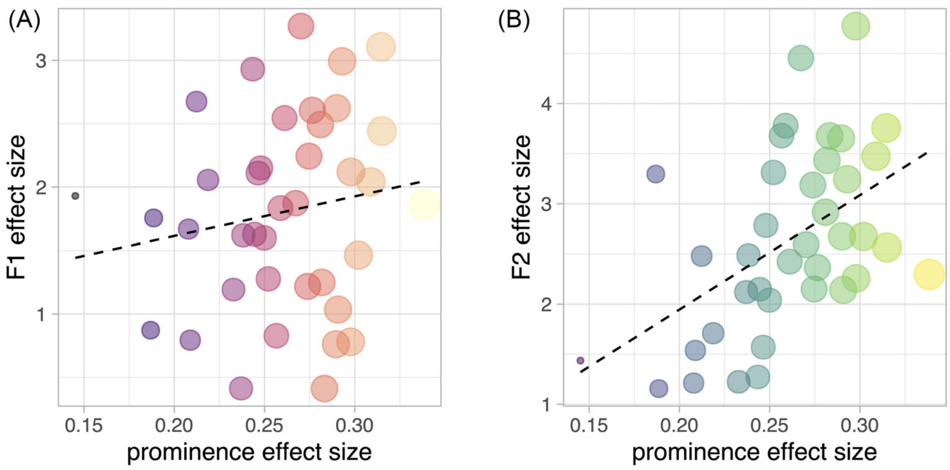

Fig. 2. By-participant effect correlations between the effect of F1 (panel A, left) and F2 (panel B, right), with the effect of prominence. Each individual point represents a given participant. The size of the prominence effect is additionally indicated by dot size and color: larger dots and a lighter color correspond to a larger effect.

expansion, instead tuning into the hyperarticulation pattern. Future work can address this question by adding in more naturalistic co-variation between context and vowel-intrinsic sonority cues, and seeing if this modulates the observed effects (in, e.g., strengthening, the hyperarticulation effects seen here when other cues in the vowel convey sonority).

One outstanding question is why these results are so consistent across participants. Whereas the speech production literature suggests that speakers are variable in which prominence strengthening pattern they prioritize, here we see robust and consistent adjustments in line with hyperarticulation. One possible account for this effect is that perceptual mechanisms at play are subject to a more general constraint which is not operative in the variably attested speech production patterns. Because the American English front vowel space is acoustically crowded, one possibility is that these effects reflect an expectation of perceptual dispersion in that space (consistent with Steffman, 2021b), wherein prominent vowels are expected to show acoustic differentiation from other vowels in the language. For high front vowels, this would entail higher and more-fronted vowel acoustics (hyperarticulation). A test for this account could come from comparing the present study to an analogous test in a language with a less crowded vowel space. For example, in Tongan with the five vowel system /i,e,a,o,u,/, the high front vowel /i/ shows consistent and robust sonority expansion as a function of lexical prominence (Garellek and White, 2015), unlike the variability seen in the American English data. Testing if Tongan listeners therefore show a reversed perceptual pattern as that evidenced here would help us explore the extent to which these influences are dependent on the vowel inventory of the language at hand, and thus offer further insight into the apparent asymmetry in prominence strengthening effects in production as compared to perception.

Future work can help address these questions further in exploring how malleable the present effects are, in testing if, for example, they could be overridden with an exposure phase that evidenced a sonority expansion pattern of prominence strengthening in /i/ and /ɪ/. Given that the present results tested only one speaker's voice in the stimuli, another pertinent further extension of these results would be exploring how well they generalize across different speakers and prominence-lending contexts. This too would offer the possibility of exploring if listeners maintain speaker-specific expectations of prominence strengthening, if for example, they are shown during exposure that different speakers attest different patterns. More generally, exploring how both vowel-intrinsic prominence information (e.g., duration and pitch, cf. Steffman and Jun, 2019), as well as signal extrinsic information which shapes prominence perception (e.g., Bishop, 2012), factor into this equation will help better our understand of prominence effects in perception more holistically.

Acknowledgments

Many thanks are due to Adam Royer for recording materials for the stimuli, and to Sun-Ah Jun, Taehong Cho, Pat Keating, and Megha Sundara for valuable discussion.

References and links

[1]Sonority expansion is generally linked more to F1, however, Erickson (2002) showed front vowels can show lowered F2 under prominence suggesting lingual backing can occur with sonority expansion.

[2]Though hyperarticulation and sonority expansion can generate conflicting patterns of formant variation and lingual articulation, it should be remarked that other sonority cues, which do not directly impact formant structure (e.g., voice quality as in Chong *et al.*, 2020), might help to strengthen vowel sonority, even when lingual articulations manifest hyperarticulation. For example, Cho (2005) found increased lip aperture in prominent /i/, suggesting a sonority enhancing articulation, even though this vowel was generally hyperarticulated in terms of

lingual articulation and vowel acoustics in his study.

[3]The intercept estimate is obtained at scaled values of 0 in the centered variables for both formants. This is halfway between 3.6 and 4.2 Bark in F1, and between 13 and 13.6 Bark in F2. If one approximates where the categorization functions in Fig. 1(B) would fall at these values, one can see this point is below 0.50 prop. /i/ responses, evidencing the negative intercept estimate.

[4]Recall that all by-participant estimates for the prominence effect were negative (decreased log-odds of an /i/ response). All by-participant estimates were negative for F1, and positive for F2.

[5]This overall pattern differs from what Steffman (2021b) observed with non-high vowels /ɛ/ and /æ/, which in a comparable experiment were expected to be realized as more open, or sonorous, in terms of F1/F2 when prominent. Perceptual prominence strengthening effects therefore appear to vary based on vowel features.

Bishop, J. (**2012**). "Information structural expectations in the perception of prosodic prominence," in *Prosody Meaning* (Walter de Gruyter, Berlin), pp. 239–269.

Bürkner, P.-C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Stat. Softw. **80**(1), 1–28.

Cho, T. (**2005**). "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English," J. Acoust. Soc. Am. **117**(6), 3867–3878.

Chong, A. J., Risdal, M., Aly, A., Zymet, J., and Keating, P. (**2020**). "Effects of consonantal constrictions on voice quality," J. Acoust. Soc. Am. **148**(1), EL65–EL71.

Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (**2007**). "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech," J. Phon. **35**(2), 180–209.

de Jong, K. (**1995**). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," J. Acoust. Soc. Am. **97**(1), 491–504.

de Jong, K., Beckman, M. E., and Edwards, J. (**1993**). "The interplay between prosodic structure and coarticulation," Lang. Speech **36**(2-3), 197–212.

Erickson, D. (**2002**). "Articulation of extreme formant patterns for emphasized vowels," Phonetica **59**(2-3), 134–149.

Garellek, M., and White, J. (**2015**). "Phonetics of Tongan stress," J. Int. Phon. Assoc. **45**(01), 13–34.

Houde, R. (**1967**). "A study of tongue body movement during selected speech sounds," Ph.D. thesis, University of Michigan, Ann Arbor, MI.

Kent, R. D., and Netsell, R. (**1971**). "Effects of stress contrasts on certain articulatory parameters," Phonetica **24**(1), 23–44.

Kim, S., Choi, J., and Cho, T. (**2016**). "Linguistic contrast enhancement under prosodic strengthening in L1 and L2 speech," poster presented at the *15th Conference on Laboratory Phonology*, Ithaca, NY.

Politzer-Ahles, S., and Piccinini, P. (**2018**). "On visualizing phonetic data from repeated measures experiments with multiple random effects," J. Phon. **70**, 56–69.

Steffman, J. (**2021a**). "Materials for: 'Contextual prominence in vowel perception: Testing listener sensitivity to sonority expansion and hyperarticulation,' " https://osf.io/ykbjf/ (Last viewed 3/29/2021).

Steffman, J. (**2021b**). "Prosodic prominence effects in the processing of spectral cues," Lang. Cogn. Neurosci. (published online), 1–26.

Steffman, J., and Jun, S.-A. (**2019**). "Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception," J. Acoust. Soc. Am. **146**(3), EL251–EL257.

Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**(1), 97–100.

van Summers, W. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**(3), 847–863.

Winn, M. (**2016**). "Vowel formant continua from modified natural speech (PRAAT script)," http://www.mattwinn.com/praat/Make_Formant_Continuum_v38.txt, version 38 (Last viewed 3/30/2021).