

Perceptual Tuning to Structure: Integrating the Phonetic Detail of Coarticulatory Vowel Nasalization With Prosodic and Information Structure

Language and Speech

1–16

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00238309261453071

journals.sagepub.com/home/las

Sahyang Kim¹, Holger Mitterer² ,
and Taehong Cho^{3,4} 

Abstract

This study explores prosodically driven perceptual adjustment of a non-contrastive coarticulatory feature. Specifically, it examines how prosodic prominence and information structure influence the perception of anticipatory vowel nasalization in CVN in American English as a cue to an upcoming nasal consonant. Using a two-alternative forced-choice task, we tested whether the same acoustic degree of nasalization is interpreted differently depending on prosodic and information-structural contexts, and whether these effects are consistent across native English speakers and Korean learners of English. Listeners identified target words (*Bob* or *bomb*) from a nasalization continuum embedded in carrier sentences (“(No,) Riley wrote Bob/bomb slowly”), where the coda consonant was masked by noise. Prosodic prominence was manipulated by adjusting pitch and amplitude of surrounding words while keeping the target constant; and information structure was manipulated by including or omitting *No*, signaling contrastive focus. For a given stimulus, listeners gave more nasal responses (*bomb*) when the target was prosodically prominent or preceded by *No*, implying that they compensate for reduced vowel nasality in these contexts. The fact that *No* influenced perception independent of prosody suggests that information structure can modulate perceptual compensation beyond acoustic-prosodic cues, reflecting a top-down effect of discourse meaning. Finally, although Korean listeners gave fewer nasal responses overall—possibly due to reliance on acoustic detail or perceptual hypercorrection—both groups showed similar sensitivity to prosodic and information-structural cues. These findings taken together point toward a shared perceptual mechanism in both native and non-native perception that integrates fine-grained phonetic detail with prosodic and information-structural contexts.

¹Department of English Education, Hongik University, Seoul, South Korea

²Department of Cognitive Science, University of Malta, Msida, Malta

³Hanyang Institute for Phonetics and Cognitive Sciences of Language, Hanyang University, Seoul, South Korea

⁴Department of English Language and Literature, Hanyang University, Seoul, South Korea

Corresponding author:

Taehong Cho, Department of English Language and Literature, Hanyang University, Seoul 04763, South Korea.

Email: tcho@hanyang.ac.kr

Keywords

vowel nasalization, denasalization, prosodic boundary, gender-related variation, Seoul Korean

1. Introduction

Understanding spoken language requires decoding multiple layers of information embedded in the speech signal, including segmental, lexical, prosodic, and information-structural elements. These layers are encoded through both segmental and suprasegmental dimensions, whose fine-grained phonetic details are systematically shaped and coordinated by prosodic structure—an area central to the phonetics–prosody interface (Byrd & Krivokapić, 2021; Cho, 2016; Cho et al., 2026; Keating et al., 2003; Mücke et al., 2014; Sugahara & Turk, 2004). In line with this perspective—particularly the view that prosodic structure is signaled by the intricate coordination of both segmental and suprasegmental features—perceptual studies have shown that prosodic structure, especially its two key components, boundary and prominence relations conveyed in large part through suprasegmental cues, plays a significant role in modulating segmental perception (e.g., Kim & Cho, 2013; Kim et al., 2018; Mitterer et al., 2016; Steffman, 2019, 2021a; Steffman & Jun, 2019).

For instance, suprasegmental features such as phrase-final lengthening and F0 modulation, which signal prosodic boundaries, may shape listeners' interpretation of segmental content. It has been shown that in English, listeners require a longer VOT to perceive a voiceless stop when it follows a contextual word with final lengthening, compared to when no such lengthening occurs (Kim & Cho, 2013; Mitterer et al., 2016). This perceptual pattern reflects prosodic boundary-related variation in production—specifically, pre-boundary lengthening and post-boundary (domain-initial) strengthening. That is, the final syllable of the preceding word is lengthened, and the onset of the following word is strengthened, with voiceless aspirated stops, for example, typically produced with longer VOTs (e.g., Cho & Keating, 2009). Similarly, F0 cues alone can affect perception. In Korean, an F0 rise preceding a phrase boundary influences how listeners interpret the lenis-to-tense stop phonological alternation: Without an F0 rise, listeners often reinterpret a phonetically tense stop as lenis, whereas a boundary-marking F0 rise blocks this reinterpretation, leading the same phonetic input to be perceived as tense (Kim et al., 2018; see also Steffman & Katsuda, 2021, for related findings in Japanese).

Prosodic prominence, another key component of prosodic structure, also shapes segmental perception. In American English, prominence-related f0 increases modulate how listeners use vowel duration to infer coda voicing. For instance, when categorizing items along a “coat–code” continuum, when the F0 contexts indicate that the target vowel is prominent, listeners require longer vowels to perceive *code* (or they give more *coat* responses for a given stimulus), compensating for prominence-induced lengthening (Steffman & Jun, 2019). Similarly, prominence contextually cued by F0, duration, and amplitude in surrounding words affects the categorization of ambiguous vowels between /e/ and /æ/ (Steffman, 2020, 2021a, 2021b)—that is, when the target syllable is perceived prominent, listeners require a higher F1 and lower F2 (i.e., a more open and retracted vowel) to identify the vowel as /æ/. These perceptual adjustments mirror production patterns of hyperarticulation, in which /æ/ is indeed realized as lower and more retracted under prominence.

Taken together, these studies suggest that listeners compute the prosodic structure from boundary and prominence cues conveyed across both segmental and suprasegmental dimensions, and that they compensate for these effects during segmental perception. This perspective aligns with the view that prosodic structure is derived through the integration of segmental and suprasegmental cues in parallel with segmental analysis—that is, through their simultaneous operation—with speech perception emerging from the interaction of these processing streams (Cho et al., 2007;

Kim et al., 2018; Salverda et al., 2003; Steffman, 2020, 2021a, 2021b; Steffman et al., 2022; see also Warner, 2023). In this line of research, some scholars have suggested that prosodic analysis may operate as an independent process, with its outcome subsequently influencing segmental interpretation (e.g., Poeppel, 2014; Salverda et al., 2003). However, subsequent work (e.g., Cho et al., 2007; Kim et al., 2018; McQueen & Dilley, 2021; Steffman, 2020, 2021b; Steffman et al., 2022; see also Warner, 2023) proposes or discusses a theoretical framework in which listeners may construct prosodic and segmental representations simultaneously by drawing on all available acoustic cues, thereby modulating perception in a contextually appropriate manner—context, in this case, referring to prosodic context. This is consistent with the refined “Prosody Analyzer” framework proposed by Cho et al. (2007) (see also McQueen & Dilley, 2021), which specifically argues for interconnected rather than independent processing of prosodic and segmental cues, as is assumed in some neuroscientific inspired models (Poeppel et al., 2008). The present study extends this view by examining a new case of prosodically driven perceptual adjustment: the fine-grained phonetic realization of coarticulatory vowel nasalization in English.

Extensive phonetic research on prosodic strengthening demonstrates that prosodic boundaries and prominence systematically enhance segmental articulation (and, in turn, its acoustic consequences) (e.g., Cho, 2011, 2016; Cho & Keating, 2001, 2009; Fletcher, 2010; Foucheron, 1999; Foucheron & Keating, 1997; Keating et al., 2003).¹ This heightens phonetic clarity of articulatory features, which may further enhance either syntagmatic contrast—especially in conjunction with boundary-related strengthening—or paradigmatic (phonemic) contrast, particularly in connection with prominence-related strengthening (e.g., Cho, 2016; Foucheron, 1999). This strengthening is also evident in coarticulatory processes: In American English, vowels in CVN words exhibit reduced nasalization under focus-induced prominence (Cho et al., 2017), reflecting coarticulatory resistance that enhances vowel orality. However, it is unclear whether this facilitates word recognition. Nasalized vowels serve as a robust acoustic and perceptual cue for anticipating an upcoming nasal coda. In fact, increased vowel nasalization has been observed in listener-oriented contexts with communicative demands, such as high lexical neighborhood density (Scarborough, 2013) or careful speech (Beddor, 2009). Some perceptual studies (Ali et al., 1971; Beddor et al., 2013; Lahiri & Marslen-Wilson, 1991) have indeed shown that surface vowel nasality directly contributes to segmental prediction, aiding in the identification of nasal codas. More nasalized vowels, therefore, are more likely to be interpreted as CVN words, facilitating lexical access (Scarborough & Zellou, 2013; Zellou, 2022).

These seemingly divergent production patterns—reduced nasality under focus-induced prominence versus increased nasality driven by communicative or phonological demands—raise an important question: How do listeners interpret reduced vowel nasalization in prosodically prominent contexts when predicting an upcoming nasal consonant? The key issue is whether listeners adjust their expectations based on prosodic structure and therefore require less nasalization in prominent contexts than in non-prominent ones to anticipate a following nasal consonant. Alternatively, it can be argued that listeners might require *more* nasalization under prominence to anticipate a nasal consonant, assuming that prosodic salience enhances articulatory cues such as nasality, in line with patterns shown to facilitate lexical processing (Scarborough & Zellou, 2013; Zellou, 2022).

To address this question, the present study examines how contextually induced prominence influences the listeners’ interpretation of non-contrastive coarticulatory vowel nasalization in English. We constructed a *Bob–bomb* continuum by varying nasalization on the vowel while masking the final coda and embedded these stimuli in carrier sentences such as “Riley wrote [Bob/bomb] slowly.” The pitch and amplitude of the surrounding context (“Riley wrote . . .” and “slowly”) were manipulated to create conditions of relative prominence on the target word, which

itself remained acoustically constant. Listeners judged whether they heard *Bob* or *bomb*, relying solely on the degree of vowel nasalization. If listeners incorporate the production pattern in which prominence reduces nasality for phonetic clarity, they may treat a given degree of nasalization as a *stronger* cue in prominent contexts—leading to more nasal (*bomb*) responses. Alternatively, if prominence increases the perceptual expectation for nasalization—as in other listener-oriented contexts—then the same degree of nasalization may be perceived as insufficient, resulting in fewer nasal responses when the word is prominent.

Another question addressed in the present study is whether a focus-related information-structural cue further modulates listeners' perception of fine phonetic details. It has been reported that information-structural context influences listeners' perception at the suprasegmental level (Bishop, 2012). In that study, listeners heard acoustically identical sentences (e.g., *I bought a motorcycle*), presented as answers to questions designed to induce different focus contexts (e.g., *What happened?* for broad focus, *What did you do?* for VP focus, and *What did you buy?* for narrow focus on objects). After listening to each question–answer pair, participants were asked to rate the level of prominence of the verb and object in the answer sentence. The results showed that in the narrow focus condition (in which the object was focused), listeners rated the object as significantly more prominent and the preceding verb as less prominent than when the same acoustic recording of the sentence was situated within broad and VP focus contexts. This effect was stronger when the object received narrow focus (e.g., in response to *What did you buy?*) than when it received contrastive focus (e.g., when correcting a sentence such as *Why's your wife mad? . . . because you bought a car?*). Given that the acoustic signal of the target sentence remained unchanged across conditions, Bishop (2012) provides evidence that suprasegmental (prosodic) perception of relative prominence is influenced by listeners' top-down expectations derived from higher-level contextual knowledge—in this case, the contextual information provided by the preceding question. In a similar vein, the present study examines whether the contrastive information structure, this time, lexically implied by a negative element, modulates listeners' interpretation of fine segmental details, focusing on the role of the negative marker *No* as a cue to contrastiveness in the perception of coarticulatory nasalization.

To test this, a subset of the stimuli included *No*, as in “No, Riley wrote Bob/bomb slowly,” with no change to the acoustic realization of the target word or its carrier sentence. *No* is expected to function as a pragmatic cue signaling contrastiveness, likely invoking an implicit or prior alternative in discourse. This contrastive use is assumed to create an information-structural context resembling focus-induced prominence, where greater vowel nasal reduction has been observed (Cho et al., 2017). If so, listeners may expect less nasal coarticulation in sentences with *No* than without, leading to more nasal responses in the former. This design thus allows us to test whether information-structural cues can modulate the perception of a coarticulatory feature independent of prosody and the speech signal itself (cf. Mücke & Grice, 2014).

Finally, this study explores whether prosodically driven perceptual compensation differs between native American English speakers and Korean learners of English. Previous studies have shown that, despite the fact that the two languages employ distinctive prominence systems (e.g., Jun, 2005), speakers of both languages exhibit prominence-induced coarticulatory resistance to vowel nasalization in CVN words in their native languages (Cho et al., 2017; Jang et al., 2018). It has also been shown that Korean learners similarly reduce vowel nasalization under prominence in their L2 English (Jang et al., 2022). Nonetheless, phonetic differences persist: Korean speakers reduce nasal coarticulation even more under prominence than English speakers in both their L1 and L2 (Jang et al., 2018). If such language-specific production patterns inform perception, Korean listeners may assign greater perceptual weight to a given level of vowel nasalization in prominent contexts, resulting in a higher rate of nasal (CVN) responses under prominence compared to native

listeners. Another possibility, if non-native listeners rely more heavily on bottom-up acoustic details—as is often observed in L2 perception (Warner et al., 2022)—they may require greater degrees of nasalization to perceive a nasal segment, resulting in fewer nasal responses overall. Notably, these two patterns are not mutually exclusive: Korean learners may show fewer nasal responses overall but still exhibit a greater sensitivity to prominence-based modulation.

2. Method

2.1. Participants

Sixty participants took part in the study. Twenty native English speakers were visitors or temporary residents in Korea at the time ($M_{age} = 27.7$, range 19–39; 12 women, 8 men). The remaining participants were advanced and intermediate Korean learners of English (20 per group) who were undergraduate and graduate students mostly in their 20s ($M_{age} = 24.3$, range 20–30; 12 women, 8 men for the advanced group; 3 women, 17 men for the intermediate group). The advanced learners of English had an average TOEIC² score of 944 (approximately upper 6%), and the intermediate learners had an average TOEIC score of 770 (approximately upper 36%). All of them self-reported normal hearing and participated for a monetary compensation.

2.2. Stimuli

In creating the speech stimuli, we manipulated the surrounding context rather than the target words themselves, testing how acoustically identical targets could sound relatively more or less prominent depending on their context. This approach follows methods used in previous studies (Steffman, 2020, 2021a, 2021b). Perceived prominence was modulated by adjusting the pitch and amplitude of the surrounding speech within the carrier sentences, while keeping the acoustic properties of the target words constant. Although increased duration strongly influences perceived prominence (Cole et al., 2010; Mo, 2011), we avoided temporal adjustments, as any temporal change may directly affect vowel nasalization in both production and perception (Cho et al., 2017; Solé, 2007; Zellou, 2022).

The stimuli used in the experiment are listed in Table 1. A male native speaker of American English in his 20s recorded the base sentences “(No), Riley wrote Bob/bomb slowly,” placing emphasis (i.e., prominence) on *Bob/bomb*, *Riley*, or both. Target words were *Bob* (oral coda) and *bomb* (nasal coda). Sentences emphasizing either *Bob/bomb* or *Riley* served to examine amplitude differences associated with the presence or absence of focus on target words. Neutrally produced sentences—that is, those produced without reference to information structure and thus bearing pitch accents on both *Riley* and *Bob/bomb*—were used to generate the final experimental stimuli.

The context sentences were generated using PSOLA. In natural sentences, targets were about 6 dB louder and had a pitch that was about 10% higher than the surrounding words. To generate a prominence-absent condition, it was therefore necessary to increase the pitch and amplitude of the context sentence so that the target no longer stood out. To achieve this, the pitch contour of the final prominence-absent stimulus was multiplied by 1.1, resulting in the target having a similar pitch to the surrounding words (see Table 1 for the pitch maxima in each word). Pitch manipulation was done in two steps. First, the pitch contour of an original sentence (with a 10% difference between the target and the context sentence) was multiplied by 1.0488 (i.e., the square root of 1.1), and the sentence was resynthesized. This resynthesized version was then further pitch-manipulated: once by multiplying the existing pitch contour again by 1.0488, and once by multiplying by 1 over 1.0488. The latter manipulation restored the original pitch contour; in this way, both

Table 1. Stimuli in the Four Conditions.

Negation marker “no”	Prominence	Carrier sentence
Absent	Absent	Riley _{104Hz} wrote _{98Hz} [Bob/bomb] _{95Hz,0dB} slowly _{98Hz}
Absent	Present	Riley _{94Hz} wrote _{89Hz} [Bob/bomb] _{95Hz,+6dB} slowly _{88Hz}
Present	Absent	No, Riley _{104Hz} wrote _{98Hz} [Bob/bomb] _{95Hz,0dB} slowly _{98Hz}
Present	Present	No, Riley _{94Hz} wrote _{89Hz} [Bob/bomb] _{95Hz,+6dB} slowly _{88Hz}

Note. Targets words are in the squared brackets. Subscripts indicate pitch maxima for each word and, for the target, the intensity difference to the preceding word.

context sentences ($\pm$ prominence) underwent two cycles of PSOLA manipulation. Finally, the prominence-absent context sentence was multiplied by 2, resulting in a 6 dB gain. Consequently, the target word had an amplitude similar to that of the surrounding words in this version.

Experimental stimuli were created by concatenating a neutrally produced carrier sentence with target words selected from two separate neutral utterances: one with “Bob” and one with “bomb.” From these two endpoints, a continuum ranging from an oral to a nasal stop was generated using the STRAIGHT morphing algorithm (Kawahara et al., 1999) using 10% morphing steps resulting in 11 continuum steps. While the primary experimental manipulation was the creation of this nasalization continuum, the final consonant was also masked, requiring listeners to infer the intended word based solely on vowel nasalization. Previous studies often used gating paradigms with isolated words (Lahiri & Marslen-Wilson, 1991; Ohala & Ohala, 1995), but such methods require the audio to stop and thereby not allow to present the target stimulus in both the preceding and following prosodic structure. Masking was therefore implemented by partially overlaying the target consonant with a masking sound using Praat’s “Concatenate with overlap” function (Boersma, 2001) with a 10-ms overlap, where the consonant was gradually replaced by the masking sound. Masking duration was adjusted across continuum steps because, in the base recordings, the nasal coda (89 ms) was longer than the stop coda (63 ms). For stimuli in the middle of the continuum (Steps 2–10), masking began at vowel offset, ensuring that the entire following consonantal portion was fully masked and leaving vowel nasalization as the only available cue. For the continuum endpoints (Steps 1 and 11), however, consonants were only partially masked to ensure clear identification of “Bob” and “bomb,” thereby providing perceptually unambiguous endpoint baselines for listeners. For the masker sound, we first generated a harmonic complex tone with a base frequency of 400 Hz using the Praat function *Create Sound from tone complex*, which generates a sound in which all harmonics have the same amplitude. This sound was then filtered with a 25-dB/octave roll-off above 2 kHz.

Using this strategy, we created two carrier sentence types that induced contrasting prominence conditions on the target word—either prominence or no prominence. In the prominence condition, we lowered the amplitude and maximum pitch of surrounding non-target words in the carrier sentence, making the target word relatively louder and higher-pitched. Conversely, in the no-prominence condition, non-target words were made louder and higher-pitched than the target word. Specifically, the louder version of the carrier sentence—leading to a lower perceived prominence on the target word—had twice the energy (+6 dB) compared to the softer version of the carrier sentence, leading to a higher perceived prominence of the target word. In addition, as shown in Table 1, carrier sentences varied whether the sentence-initial “No” was heard or not. We selected every second step along the 11-step nasalization continuum, resulting in six stimulus levels. Note that the first and last steps, which include audible portions of the consonant and serve as guiding stimuli, were excluded

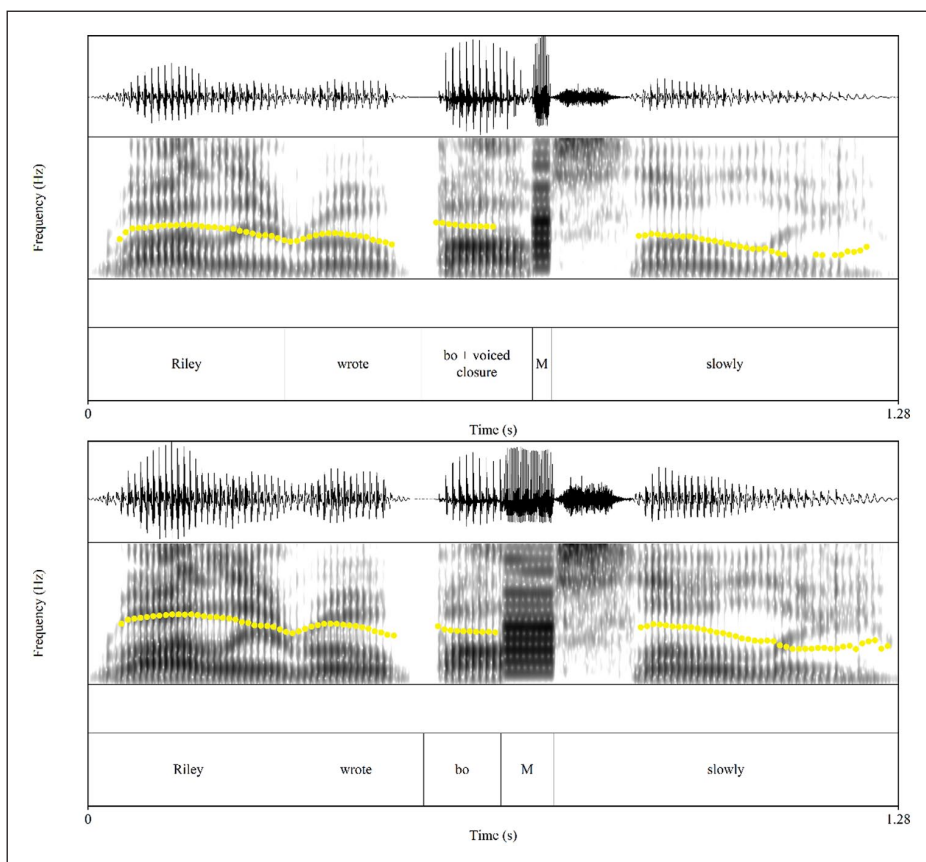


Figure 1. Spectrogram, waveform, and pitch contour (yellow lines) for example stimuli with prominence on the target word with a partially masked coda (upper panel) and without prominence on the target word with a fully masked coda (lower panel).

Note. The endpoint stimuli (Steps 1 and 11) were only partially masked to ensure clear identification of “Bob” and “bomb,” thereby providing perceptually unambiguous endpoint baselines for listeners. In contrast, all stimuli from the middle of the continuum had fully masked consonants, leaving vowel nasalization as the only available cue. “M” indicates the masked portion.

from the analysis. This yielded 24 experimental stimuli: 6 levels of nasalization $\times$ 2 prominence conditions $\times$ 2 versions differing in the presence or absence of sentence-initial *No*.

Figure 1 presents two examples of stimuli. The top panel shows a stimulus from the prominent condition, where the target word is louder and higher-pitched than the surrounding context, and the coda consonant is only partially masked, allowing a short voiced closure to remain audible. The bottom panel illustrates a stimulus from the non-prominent condition, in which the consonant is fully masked, and the non-target words are louder and higher-pitched than the target. The yellow (bright) lines indicate pitch contours, showing a higher pitch on the target word in the prominent condition (top panel) and a relatively lower pitch in the non-prominent condition (bottom panel). All concatenated stimuli were perceptually validated by the tone and break indices (ToBI)-trained authors to confirm that their pitch accent patterns were consistent with the intended prosodic conditions.

2.3. Procedure

The experiments were conducted using ExperimentBuilder on a standard PC. Participants sat in front of a monitor and listened to the stimuli through headphones. Two response options (*Bob* and *bomb*) were displayed on the left and right sides of the screen. After hearing a stimulus, participants responded by pressing the corresponding mouse button. As feedback, the selected option remained on screen while the other disappeared. Each participant heard 24 stimuli repeated 15 times, resulting in 360 trials. After every 50 trials, participants received a progress update and were given the opportunity to take a short break. Stimuli were presented in randomized order within each of 15 blocks.

2.4. Statistical Analysis

A statistical analysis was conducted using a generalized linear mixed-effects model. The “guiding” stimuli—Step 1 (*Bob*) and Step 6 (*bomb*)—were excluded, as these included audible portions of the coda consonant. All predictors were contrast-coded. For Prosody, prominent and non-prominent conditions were coded as 0.5 and -0.5 , respectively. For Information Structure, the presence versus absence of *No* was coded as 0.5 and -0.5 . English Competence, with three levels (native speaker, advanced L2, and intermediate L2), was coded with two contrasts: The first compared native speakers (2/3) to all Korean learners ($-1/3$), and the second compared advanced (0.5) to intermediate (-0.5) Korean learners. Nasalization was scaled to range from -0.5 to 0.5 to allow for meaningful comparison of regression weights across predictors. The model included a random intercept for participants and random slopes for Prosody and Nasalization (without correlation terms), representing the maximal random-effects structure that successfully converged. Fixed effects included interactions between Nasalization and Native Language, and a three-way interaction among Native Language, Prosody, and Information Structure. A four-way interaction was excluded to maintain model parsimony and ensure convergence.

3. Results

Figure 2 presents the mean proportions of nasal responses for all three listener groups (different panels) over the nasalization continuum, with different lines for the different conditions of information structure (with or without *no*) and prosody (prominent or not). One of the clearest patterns observed was that as the nasalization continuum changes from Step 1 (*Bob*) to Step 6 (*bomb*), the participants showed a gradual increase in nasal responses. That is, the more nasalized the vowel was, the more nasal the responses were even within the set of stimuli in which the consonant was completely masked (i.e., stimuli two to five of the continuum). As shown in Table 2, this effect of nasalization, observed across all three listener groups, was statistically significant. However, this effect simply indicates that participants were able to use vowel nasalization as a cue in the acoustic signal. More relevant to our research question is how listeners respond to the *same* degree of nasalization across different prominent contexts—namely, the presence or absence of focus-induced prominence (signaled by the surrounding prosodic context), the presence or absence of the utterance-initial *No* (reflecting Information Structure), and how native language background further modulates these responses. To better illustrate these effects, Figure 3 plots the predicted changes of nasal responses depending on Information Structure and Prosody in logOdds for each listener group, based on a generalized linear mixed-effects model that includes information structure and prosody as predictors, holding the degree of vowel nasalization constant.

The results showed that participants gave more nasal responses when the target word was made *contextually* prominent—that is, when the amplitude and pitch of the surrounding context were

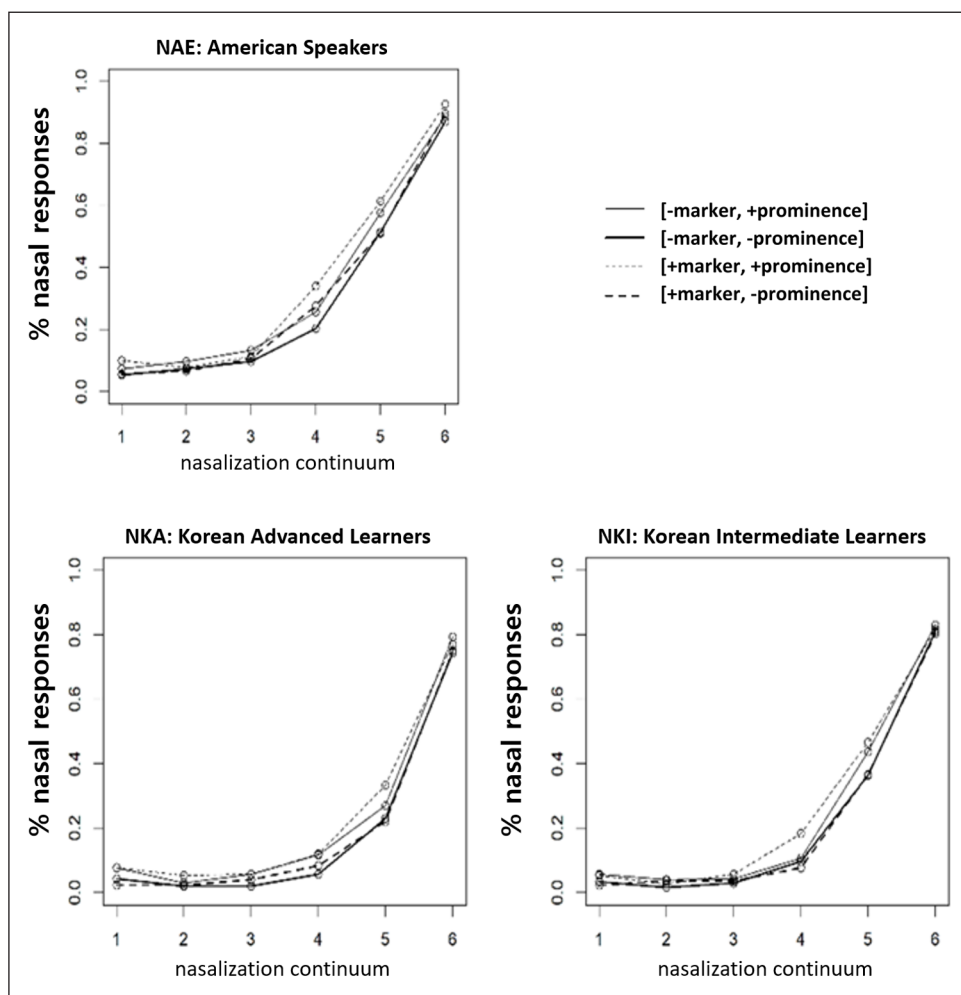


Figure 2. Proportion of “nasal” responses (i.e., trials in which the participant indicated that the intended word was *bomb* rather than *Bob*) for the three listener groups (Native American English speakers, Korean advanced learners of English, and Korean intermediate learners of English).

reduced to highlight the target, as indicated by the gray lines and [+prominence] in Figure 2 compared to the black lines for the [–prominence] condition (see also Figure 3). In addition, more nasal responses were observed in the presence of *No*—that is, when the sentence began with *No*, signaling contrastive focus through information structure, as shown by the dotted lines and [+marker] in Figure 2 (see also Figure 3). Statistical analyses further confirmed these observations, indicating significant effects of both Prosody and Information Structure (Table 2), although the effect sizes were generally small. These effects were not moderated by any interactions among factors.

Finally, the results also revealed a significant effect of Native Language: Native English listeners gave more nasal responses overall than Korean learners of English. There was no interaction between Nasalization and Native Language, indicating that the groups differed in baseline sensitivity rather than in how they used the nasalization continuum. This effect appears to stem from the Korean listeners’ overall lower response rates (i.e., closer to floor), where smaller differences

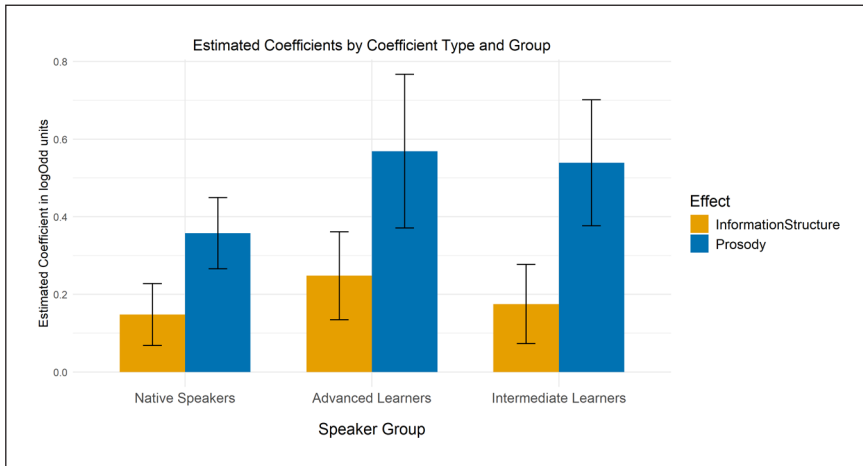


Figure 3. Estimated effects (regression weights) of prosody and information structure for each listener group, based on the generalized linear mixed-effects model.

Note. Error bars indicate standard errors of the estimates.

across the continuum were more strongly compressed by the logistic linking function, which can also be inferred from Figure 2. This suggests that, all else being equal, Korean learners of English as non-native listeners require more nasalization to predict an upcoming nasal coda than native English listeners.³ Importantly, Native Language did not interact with either Prosody or Information Structure, indicating that both native and non-native listeners showed similar effects of these two factors—that is, a given level of vowel nasalization elicited more nasal responses when the target word was prominent or preceded by *No*. There was also no effect of L2 English proficiency (i.e., intermediate vs. advanced), suggesting that Korean listeners' responses were not modulated by their level of English fluency.

4. Discussion

The present study investigated how prosodic prominence and information structure modulate the perception of coarticulatory vowel nasalization in English as a cue to an upcoming nasal consonant. Following previous work grounded in the theoretical framework of the Prosody Analyzer account (e.g., Cho et al., 2007; McQueen & Dilley, 2021) and empirical evidence (Kim & Cho, 2013; Mitterer et al., 2016; Steffman, 2020, 2021a, 2021b; Steffman et al., 2022 Steffman & Jun, 2019), we hypothesized that listeners' computation of prosodic structure—particularly boundaries and prominence—simultaneously adjusts the interpretation of phonetic cues to segmental information by aligning perception with production-based expectations.

Extending this framework, we tested whether such perceptual adjustments apply to anticipatory vowel nasalization—a gradient, non-contrastive feature that nonetheless provides a reliable cue to upcoming nasal segments (Ali et al., 1971; Beddor et al., 2013; Zellou, 2022). Crucially, the production of this vowel nasalization feature varies depending on communicative context: It is enhanced in listener-oriented speech (e.g., careful articulation or high lexical neighborhood density; Beddor, 2009; Scarborough, 2013) but reduced under focus-induced prominence, where coarticulatory resistance enhances vowel orality (Cho et al., 2017). This contrast led us to ask whether listeners would expect reduced nasality under prosodic prominence—mirroring production

Table 2. Model Summary.

Predictor	B (SE)	z	p
Intercept	−3.039 (0.206)	−14.754	<.001
Nasalization	1.488 (0.118)	12.634	<.001
Prosody	0.495 (0.088)	5.644	<.001
InformationStructure	0.19 (0.057)	3.303	<.001
English Competence			
L2Eng_Level	−0.527 (0.506)	−1.042	.297
EngNative	1.744 (0.425)	4.105	<.001
Prosody: InformationStructure	0.141 (0.115)	1.233	.218
Nasalization × English Competence			
Nasalization: L2Eng_Level	−0.318 (0.294)	−1.083	.279
Nasalization: EngNative	−0.172 (0.241)	−0.712	.477
Prosody × English Competence			
L2Eng_Level: Prosody	0.028 (0.226)	0.124	.901
EngNative: Prosody	−0.168 (0.175)	−0.958	.338
InformationStructure × English Competence			
L2Eng_Level: InformationStructure	0.071 (0.152)	0.464	.643
EngNative: InformationStructure	−0.06 (0.111)	−0.547	.584
(Prosody: InformationStructure) × English Competence			
L2Eng_Level: Prosody: InformationStructure	−0.137 (0.305)	−0.449	.654
EngNative: Prosody: InformationStructure	−0.211 (0.221)	−0.954	.34

patterns—or instead anticipate increased nasality, assuming that phonetic cues to coda nasality are generally enhanced in other listener-oriented contexts. Our results supported the former: Listeners incorporated prosodic-structural influences observed in production into their perceptual processing. Participants gave more nasal (*bomb*) responses when the target word appeared in a prosodically prominent context, even though the acoustic signal of the word itself remained unchanged. That is, listeners required less nasalization under prominence to interpret as a cue for an upcoming nasal. Given that prominence was manipulated through contextual prosody while keeping the target word invariant, vowel processing necessarily involved simultaneous computation of the prosodic structure and segmental information related to the upcoming nasal consonant. These findings reinforce the Prosody Analyzer view, which posits the parallel (i.e., synchronous) analysis of segmental and prosodic representations, integrated through mutual reference to all available cues.

Another important finding is that the presence of *No* as a contrastive marker within the information structure significantly increased nasal responses. Crucially, this effect was additive and independent of the effect of prosodic prominence, indicating that listeners can modulate segmental interpretation based solely on discourse-level information-structural cues, independent of prosodic structure. This aligns with prior claims that information structure can influence speech perception independent of phonetic realization—not necessarily mediated by phrase-level prominence (e.g., pitch accent in German; Mücke & Grice, 2014). It further suggests that contrastive contexts rooted in information structure may activate expectations about phonetic realization patterns—such as the coarticulatory resistance observed under focus-induced strengthening in production—even when those patterns are not directly cued in the speech signal other than the presence of *No*, reinforcing the view that higher-level structural information may directly inform speech perception.

Finally, while the effects of discourse context and prosody were consistent across listener groups, the groups diverged in baseline sensitivity: Korean listeners produced fewer nasal responses overall than English listeners, regardless of L2 proficiency, indicating that they required a greater degree of nasalization to infer a nasal coda compared to native English listeners. This may reflect a heavier reliance on bottom-up acoustic cues in L2 perception (Warner et al., 2022), or alternatively, a form of perceptual hypercorrection shaped by cross-linguistic awareness. That is, Korean listeners may be attuned to the fact that English permits more extensive vowel nasalization—sometimes even replacing nasal stops entirely (Clumeck, 1976; Cohn, 1990)—and may overcompensate by raising their perceptual threshold for nasality. While this interpretation is plausible, it remains a hypothesis to be tested in future work.

What is more crucial is that, despite this baseline difference, the effects of prominence and information structure were not moderated by native language, indicating that the integration of prosodic and discourse cues into segmental perception operates similarly across native and non-native listeners. This may reflect, to some extent, converging production patterns in both languages. As noted in the introduction, Korean speakers, like English speakers, exhibit coarticulatory resistance to nasalization under focus (Cho et al., 2017; Jang et al., 2018, 2022), and Korean learners of English show similar reductions in their L2 production. However, Korean speakers differ from English speakers in showing greater prominence-induced resistance in their L1, and this pattern appears to carry over, at least in part, into their L2 production. Nonetheless, these language-specific production differences do not appear to shape L2 perception: Korean listeners did not exhibit a stronger expectation of reduced nasalization in prominent contexts than native listeners. Instead, prosodic structure appears to guide the interpretation of non-contrastive phonetic features in similar ways in both L1 and L2. This dissociation between production and perception suggests that, while the phonetic implementation of prosodic and information-structural cues in production remains language-specific—likely reflecting language-specific motor gestures internalized in phonetic grammar (cf. Cho, 2025) and shaped by L1 experience—their integration in perception may rely on a more general cognitive mechanism.

In this sense, listeners—whether native or non-native—appear to make comparable use of available prosodic and information-structural cues in interpreting fine phonetic details. To the extent that human languages exhibit broadly similar interactions between information structure and the speech signal (e.g., prominence or newness being associated with increased acoustic-phonetic salience), this may reflect a general relationship between information structure and its phonetic realization across languages, which listeners exploit in perception. Thus, our findings may be interpreted as indicating that, despite cross-linguistic differences in how prosodic prominence is phonetically implemented, listeners—regardless of their native language background—show similar perceptual responses to prominence and information-structural cues, which in turn may account for the observed cross-linguistic similarity in perceptual patterns. Given the speculative nature of this interpretation, however, further research is needed to determine whether and to what extent it holds across languages and contexts.

Finally, the effect of information structure signaled by “No,” independent of the prosodic realization of the upcoming speech material, aligns with models of speech in which the mapping between pragmatic focus and prosodic realization is probabilistic rather than absolute (Calhoun, 2010). Because multiple constraints jointly shape the realization of information structure, speakers do not invariably rely on prosodic enhancement alone to signal contrast. Consequently, it is adaptive for listeners to leverage discourse cues—such as the negation marker *No*—to guide the interpretation of fine-grained phonetic details. As discussed in the Introduction, higher-level information-structural context can generate top-down expectations that modulate the perception of suprasegmental prominence (Bishop, 2012). The present findings extend this principle, suggesting

that listeners similarly use lexical cues to contrast to form expectations about coarticulatory detail at the segmental level, even in the absence of additional prosodic reinforcement.

In conclusion, this study showed that listeners' interpretation of non-contrastive coarticulatory vowel nasalization is shaped by both prosodic context and information structure in ways that reflect production patterns. Because only contextual prosody varied, requiring simultaneous computation of prosodic structure and segmental content for a given stimulus, the results align with the view of parallel processing. Notably, the effect of information structure (signaled by *No*) emerged without any accompanying phonetic change, indicating that both prosodic and discourse-level cues can directly influence segmental perception. These effects were consistent across native and non-native listeners, suggesting a shared perceptual mechanism for processing fine phonetic details at the interface of prosody and information structure, underscoring the integrated role of prosodic and higher-level information in speech perception. Future research should further test the generality of these effects across other coarticulatory processes, low-level phonetic phenomena, and diverse language pairings.

Acknowledgements

We thank both the native American English listeners and the Korean learners of English for their participation in this experiment.

ORCID iDs

Holger Mitterer  <https://orcid.org/0000-0003-4318-0032>

Tachong Cho  <https://orcid.org/0000-0002-8148-745X>

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A2A03084924).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1. Thus, the term “enhancement” is used here in a general sense, rather than the specific sense defined by Stevens and Keyser (2010). In their framework, a phonological distinction is “enhanced” by heightening the phonetic clarity of a secondary gesture; that is, the defining gestures of a phonological feature are augmented by additional or secondary articulatory gestures (either superimposed or derived from neighboring segments) and their resulting acoustic consequences.
2. Test of English for International Communication. (<https://www.ets.org/toeic>)
3. Although Korean exhibits less coarticulatory vowel nasalization than English (Cho et al., 2017; Jang et al., 2018), Korean learners often overcompensate in L2 production, hyper-articulating nasalization in English CVN contexts (Jang et al., 2022). Our perceptual findings mirror this production pattern: Learners require an exaggerated degree of nasalization to successfully predict an upcoming English nasal coda.

References

- Ali, L., Gallagher, T., Goldstein, J., & Daniloff, R. (1971). Perception of coarticulated nasality. *The Journal of the Acoustical Society of America*, 49(2B), 538–540.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4), 785–821.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, 133(4), 2350–2366.
- Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning (interface explorations)* (pp. 239–270). Mouton de Gruyter.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Byrd, D., & Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics*, 7, 31–53.
- Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86(1), 1–42. <https://www.jstor.org/stable/40666298>
- Cho, T. (2011). Laboratory phonology. In N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *The continuum companion to phonology* (pp. 343–368). Continuum.
- Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass*, 10(3), 120–141.
- Cho, T. (2025). Advancements of phonetics in the 21st century: Phonetic universals, language variation, and phonetic grammar. *Journal of Phonetics*, 111, 101426. <https://doi.org/10.1016/j.wocn.2025.101426>
- Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190. <https://doi.org/10.1006/jpho.2001.0131>
- Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, 37(4), 466–485. <https://doi.org/10.1016/j.wocn.2009.08.001>
- Cho, T., Kim, D., & Kim, S. (2017). Prosodically conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics*, 64, 71–89.
- Cho, T., Kim, S., & Mitterer, H. (2026). Linguistic and cognitive functions of fine phonetic detail underlying sound systems and sound change. *Journal of Phonetics*, 114, 101470. <https://doi.org/10.1016/j.wocn.2025.101470>
- Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35(2), 210–243. <https://doi.org/10.1016/j.wocn.2006.03.003>
- Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, 4(4), 337–351.
- Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425–452.
- Cohn, A. C. (1990). Phonetic and phonological rules of nasalization. PhD dissertation. *UCLA Working Papers in Phonetics*, 76.
- Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 521–602). Wiley-Blackwell.
- Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics*, 97, 1–74.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728–3740.
- Jang, J., Kim, S., & Cho, T. (2018). Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical Society of America*, 144, EL33–EL39. <https://doi.org/10.1121/1.5044641>
- Jang, J., Kim, S., & Cho, T. (2022). Prosodic structural effects on non-contrastive coarticulatory vowel nasalization in L2 English by Korean learners. *Language and Speech*, 66(2), 381–411. <https://doi.org/10.1177/00238309221108657>
- Jun, S.-A. (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press.

- Kawahara, H., Masuda-Katsuse, I., & de Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Communication*, 27, 187–207.
- Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory phonology VI: Phonetic interpretation* (pp. 145–163). Cambridge University Press.
- Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America*, 134(1), EL19–EL25.
- Kim, S., Mitterer, H., & Cho, T. (2018). A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *Plos One*, 13(8): e0202912, 1–28.
- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245–294.
- McQueen, J. M., & Dilley, L. C. (2021). Prosody and spoken-word recognition. In C. Gussenhoven, & A. Chen (Eds.), *The oxford handbook of language and prosody* (pp. 509–521). Oxford University Press.
- Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics*, 54, 68–79.
- Mo, Y. (2011). *Prosody production and perception with conversational speech*. Unpublished doctoral dissertation. University of Illinois, Urbana-Champaign.
- Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, 44, 47–61.
- Mücke, D., Grice, M., & Cho, T. (2014). More than a magic moment – Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics*, 44, 1–7.
- Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. In J. L. Beckman & P. W. Dickey (Eds.), *Phonology and phonetic evidence* (pp. 41–60). Mouton de Gruyter.
- Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, 24, 142–149. <https://doi.org/10.1016/j.conb.2014.07.005>
- Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 363(1493), 1071–1086. <https://doi.org/10.1098/rstb.2007.2160>
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90(1), 51–89. [https://doi.org/10.1016/S0010-0277\(03\)00139-2](https://doi.org/10.1016/S0010-0277(03)00139-2)
- Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*, 41(6), 491–508.
- Scarborough, R., & Zellou, G. (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, 134(5), 3793–3807.
- Solé, M. J. (2007). Controlled and mechanical properties in speech. In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 302–321). Oxford University Press.
- Steffman, J. (2019). Phrase-final lengthening modulates listeners’ perception of vowel duration as a cue to coda stop voicing. *The Journal of the Acoustical Society of America*, 145(6), EL560–EL566.
- Steffman, J. (2020). *Prosodic prominence in vowel perception and spoken language processing* [Doctoral dissertation]. University of California, Los Angeles.
- Steffman, J. (2021a). Contextual prominence in vowel perception: Testing listener sensitivity to sonority expansion and hyperarticulation. *JASA Express Letters*, 1(4), 045203. <https://doi.org/10.1121/10.0003984>
- Steffman, J. (2021b). Prosodic prominence effects in the processing of spectral cues. *Language, Cognition and Neuroscience*, 36(5), 586–611. <https://doi.org/10.1080/23273798.2020.1862259>
- Steffman, J., & Jun, S.-A. (2019). Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception. *The Journal of the Acoustical Society of America*, 146(3), EL251–EL257.

- Steffman, J., & Katsuda, H. (2021). Intonational structure influences perception of contrastive vowel length: The case of phrase-final lengthening in Tokyo Japanese. *Language and Speech*, 64(4), 839–858.
- Steffman, J., Kim, S., Cho, T., & Jun, S.-A. (2022). Prosodic phrasing mediates listeners' perception of temporal cues: Evidence from the Korean accentual phrase. *Journal of Phonetics*, 94, 101156.
- Stevens, K. N. & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, 38(1), 10–19.
- Sugahara, M., & Turk, A. (2004, March 23–26). Phonetic reflexes of morphological boundaries at a normal speech rate. In B. Bel & I. Marlien (Eds.), *Proceedings of speech prosody 2004* (pp. 353–356). International Speech Communication Association.
- Warner, N. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues of spoken word recognition in phonetic research. *Journal of Phonetics*, 101, 101275. <https://doi.org/10.1016/j.wocn.2023.101275>
- Warner, N., Brenner, D., Tucker, B. V., & Ernestus, M. (2022). Native listeners' use of information in parsing ambiguous casual speech. *Brain Sciences*, 12(7), 930.
- Zellou, G. (2022). *Coarticulation in phonology*. Cambridge University Press.