Check for updates

# Effects of phonological and phonetic information of vowels on perception of prosodic prominence in English

Suyeon Im*

*Department of English Language and Literature, Soongsil University, Seoul, Korea*

## Abstract

This study investigates how the phonological and phonetic information of vowels influences prosodic prominence among linguistically untrained listeners using public speech in American English. We first examined the speech material's phonetic realization of vowels (i.e., maximum F0, F0 range, phone rate [as a measure of duration considering the speech rate of the utterance], and mean intensity). Results showed that the high vowels /i/ and /u/ likely had the highest max F0, while the low vowels /æ/ and /ɑ/ tended to have the highest mean intensity. Both high and low vowels had similarly high phone rates. Next, we examined the effects of the vowels' phonological and phonetic information on listeners' perceptions of prosodic prominence. The results showed that vowels significantly affected the likelihood of perceived prominence independent of acoustic cues. The high and low vowels affected probability of perceived prominence less than the mid vowels /ɛ/ and /ʌ/, although the former two were more likely to be phonetically enhanced in the speech than the latter. Overall, these results suggest that perceptions of prosodic prominence in English are not directly influenced by signal-driven factors (i.e., vowels' acoustic information) but are mediated by expectation-driven factors (e.g., vowels' phonological information).

**Keywords:** prosodic prominence, vowel identity, acoustic cues, public speech, English

## 1. Introduction

In English, the vowels which convey prosodic prominence are higher in pitch, longer in duration, greater in intensity, steeper in spectral slope, and hyperarticulated compared with the ones which do not (Beckman, 1986; Breen et al., 2010; Cole et al., 2007; Kochanski et al., 2005; Sluijter & van Heuven, 1996; Turk & White, 1999, among others). There is consensus that duration and intensity play important roles in perceptions of prosodic prominence, while existing research varies on the role of F0. Kochanski et al. (2005)

examined a spontaneous speech in the corpus of British and Irish English. They found that loudness and phone duration strongly correlated with prosodic prominence, while other acoustic measures, including F0 measures and spectral slope, did not. Cole et al. (2019) investigated a spontaneous speech in the Buckeye corpus of American English and found that phone rate, intensity, and max F0 strongly correlated with prosodic prominence. Among the acoustic correlates, phone rate was the strongest correlate of prosodic prominence.

Prosodic prominence is not only associated with acoustic cues,

but also related to other linguistic and paralinguistic factors. Prosodic prominence may encode the information status of a word, rhythm, speech style, speech mode, and so on (Calhoun, 2010; Chodroff & Cole, 2018; Hirschberg, 1993, Im et al., 2023; Pierrehumbert & Hirschberg, 1990; Vogel et al., 1995, among others). Linguistic factors' effects on perceptions of prosodic prominence have been investigated (Aylett & Turk, 2004; Baumann & Winter, 2018; Bishop et al., 2020; Breen et al., 2010; Cole et al., 2010, 2019; Im et al., 2023; Turnbull et al., 2017; Watson et al., 2008, among others). Cole et al. (2010) identified two types of linguistic factors based on the cognitive processes of perception: (1) the expectation-driven factors that listeners have prior experiences with the language use or discourse context, and (2) the signal-driven factors that listeners have real time information while listening to the speech. They examined the effects of expectation-driven factors (e.g., lexical repetition, lexical frequency) and signal-driven factors (e.g., acoustic cues) on perceptions of prosodic prominence in the Buckeye corpus of American English. Results showed that (a) prosodic prominence was partly influenced by both expectation-driven factors and signal-driven factors, and (b) expectation-driven factors contributed more to perceptions of prosodic prominence than did signal-driven factors. In a similar vein, Im et al. (2023) examined the effects of information status and pitch accents, as expectation-driven factors, and acoustic cues, as signal-driven factors, on perceptions of prosodic prominence in a public speech of American English. They found that expectation-driven factors (i.e., information status, pitch accents) mediated the effects of signal-driven factors (i.e., acoustic cues) in perceptions of prosodic prominence.

One expectation-driven factor that may play an important role is the identity of the vowel. Previous experimental evidence shows that vowels have intrinsic differences (Fahey & Diehl, 1996; Heffner, 1937; House, 1961; House & Fairbanks, 1953; Kingston, 1992; Lehiste & Peterson, 1959, 1961; Peterson & Barney, 1952; Peterson & Lehiste, 1960; Whalen & Levitt, 1995; Whalen et al., 1999; Young et al., 2001, among others). For monophthongs, high vowels are higher in F0, shorter in duration, and lower in amplitude than low vowels. Lehiste (1970) examined the loudness of vowels in English. Although amplitude was actually higher for low vowels than for high vowels, listeners considered the vowels with more articulatory effort (i.e., high vowels) louder than the vowels with less articulatory effort (i.e., low vowels). This indicates that listeners' perceptions reflect not merely acoustic cues but multiple factors, including vowel identity and its associated articulatory effort in the speaker's production. For diphthongs, the initial vowel tends to determine the F0 and intensity of the entire vowel (Lehiste, 1970). Considering that vowels are the anchors of prosodic prominence, it is possible that vowels' intrinsic differences influence perceptions of prosodic prominence. Despite vowels' potential effect on prosodic prominence, this issue has not been fully addressed in previous research on perceptions of prosodic prominence in relation to expectation-driven and signal-driven factors (e.g., Cole et al., 2010; Im et al., 2023).

The present study investigates how prosodic prominence is influenced by vowel identity (as an expectation-driven factor) and associated acoustic cues (as signal-driven factors) with linguistically untrained listeners using a public speech in American English. The research questions of the present study are summarized as follows:

(1) How are vowels phonetically realized in a public speech of American English?

(2) How do the phonological and phonetic information of vowels in the speech influence perceptions of prosodic prominence by linguistically untrained listeners?

Given the previous experimental evidence that (a) high and low vowels are associated with higher F0, duration, and intensity than are mid vowels, and (b) perception of prosodic prominence is likely associated with F0, duration, phone rate, and intensity, we expect that high and low vowels may affect perceptions of prosodic prominence more than will mid vowels. For this, we will first examine how the vowels differ from one another in the speech material's phonetic realization (Section 3.1.). Next, we will look at how vowel identity and associated acoustic cues (i.e., max F0, F0 range, phone rate, and mean intensity) influence perceptions of prosodic prominence (Section 3.2.). The schematic representation of the present study, adopted from Cole et al. (2010), is shown in Figure 1.
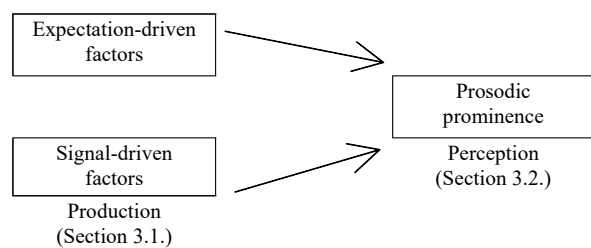


**Figure 1.** The schematic representation of the present study.

Although more linguistic factors are known to be involved with perceptions of prosodic prominence, as shown in the previous research above (e.g., Cole et al., 2010; Im et al., 2023), this study focuses on the effects of vowel entity and associated acoustic cues on prosodic prominence, which previous research has not fully examined. Based on the results, the present study aims to expand our understanding of perception of prosodic prominence as a function of expectation-driven and signal-driven factors.

## 2. Method

### 2.1. Participants and Speech Material

Thirty-five native speakers of American English (12M, 23F, $M_{age}$=24.3) participated in a perception experiment. Most of them were undergraduate or graduate students at a Midwest University in the U.S. The participants were asked to listen to a speech on an online platform called Language Markup and Experimental Design Software (Mahrt, 2013) and select the words they perceived as prominent on a transcript of the speech (Rapid Prosody Transcription; Cole & Shattuck-Hufnagel, 2016). A prominent word was defined as one that sounds higher, longer, and louder, compared to the surrounding words in an utterance. The speech material was obtained from TED Talks and was entitled "Try something new for thirty days" as shown in (3) (www.ted.com/talks/matt_cutts_try_something_new_for_30_days). A male speaker of American English delivered the speech in a clear and lively style. The entire speech was broken into four pieces presented in chronological order during the experiment. The experiment took less than 30 minutes. After the experiment, the participants received monetary compensation.

(3) A few years ago, I felt like I was stuck in a rut, so I decided to follow in the footsteps of the great American philosopher, Morgan Spurlock, and try something new for 30 days. The idea is actually pretty simple. Think about something you've always wanted to add to your life and try it for the next 30 days. (...)

## 2.2. Vowel Identification

The vowels of each word in the entire speech (n=361) were identified based on the CMU Pronouncing Dictionary (Weide, 2005). For most monosyllabic and polysyllabic words (n=293), the vowel with the primary stress was considered as the landing location of pitch accent and was analyzed in the study. Fourteen vowels were observed: /i/ (n=10), /ɪ/ (n=29), /eɪ/ (n=15), /ɛ/ (n=24), /æ/ (n=31), /ɝ/ (n=17), /ʌ/ (n=27), /aɪ/ (n=53), /aʊ/ (n=11), /u/ (n=25), /ʊ/ (n=7), /oʊ/ (n=11), /ɔ/ (n=11), and /ɑ/ (n=22). For some monosyllabic words (n=68), the vowel was considered to have no word-level stress, following the CMU Pronouncing Dictionary. These words were mostly function words (e.g., "a", "the", "to", "is"). Five vowels, /i/, /ɪ/, /ɝ/, /ʌ/, and /ɔ/ were found and were categorized as unstressed vowel (UV) in the analysis (n=68). To summarize, the present study analyzes fifteen vowel categories in total, including those with and without primary stress (i.e., UV, /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɝ/, /ʌ/, /aɪ/, /aʊ/, /u/, /ʊ/, /oʊ/, /ɔ/, and /ɑ/).

## 2.3. Acoustic Measurement

The four acoustic measures of words, max F0 (in semitones), F0 range (in semitones), phone rate, and mean intensity (in dB), were considered based on previous research on perceptions of prosodic prominence in American English (Cole et al., 2019). A word's phone rate is a measure of duration that takes into account the speech rate of the utterance. It was measured with a Praat script from Cole et al. (2019). The three acoustic measures, max F0, F0 range, and mean intensity, were obtained from ProsodyPro (Xu, 2013). The F0 contour of a monophthong and a diphthong in the speech material can be found in the online Supplementary Material of this study. All the acoustic measures went through two steps of normalization following the previous research (Cole et al., 2019; Im et al., 2023): (a) the measures were normalized for the local context to capture how a word is higher in pitch, longer in duration (i.e., slower in speech rate) or louder in intensity, relative to surrounding words in an utterance. Each measure of a word was adjusted by the mean and standard deviation of the surrounding words using the five word-window centered on the target word. Next, (b) the measures were normalized for differences in units (e.g., semitones, dB). Each measure of a word was adjusted by the mean and standard deviation of the measure of the entire words. Increases in max F0, F0 range, and mean intensity are expected to increase the probability of perceived prominence. Increased phone rate (i.e., decreased duration), however, would decrease the likelihood of perceived prominence. To ensure that all the acoustic measures have similar relationships with perceived prominence (i.e., increases in the acoustic measures would increase the probability of perceived prominence), the analysis uses inverse phone rate.

## 2.4. Statistical Analyses

We first examined the phonetic realization of each vowel in the speech. For this, we ran a multivariate multiple regression in R (R Core Team, 2022). The independent variables were fifteen vowels (UV, /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɝ/, /ʌ/, /aɪ/, /aʊ/, /u/, /ʊ/, /oʊ/, /ɔ/, and /ɑ/). The dependent variables were four acoustic measures (max F0, F0 range, inverse phone rate, and mean intensity).

Next, we looked at the effects of vowels' phonological and phonetic information on linguistically untrained listeners' perceptions of words' prosodic prominence. For this, we ran a generalized mixed-effects model using the lme4 package (Bates et al., 2015) in R (R Core Team, 2022). The fixed factors were (1) vowel (UV, /i/, /ɪ/, /eɪ/, /ɛ/, /æ/, /ɝ/, /ʌ/, /aɪ/, /aʊ/, /u/, /ʊ/, /oʊ/, /ɔ/, and /ɑ/), (2) z-normalized max F0, (3) z-normalized F0 range, (4) z-normalized inverse phone rate, (5) z-normalized mean intensity, and (6) interactions between vowel, z-normalized max F0, z-normalized F0 range, z-normalized inverse phone rate, and z-normalized mean intensity. The dependent variable was listeners' binary responses for prosodic prominence (0 as non-prominent, 1 as prominent). The random effect was the intercept of listeners. We also ran a post-hoc pairwise comparison with Tukey method based on the generalized mixed-effects model using the lsmens package (Lenth, 2016).

## 3. Results and Discussion

### 3.1. Production of Vowels

To address the first research question, we examined how the speech phonetically realized intrinsic vowel differences. Table 1 summarizes the multivariate multiple regression. Each cell indicates the estimated effect of a vowel (in row) in reference to UV (set as the intercept in the model) on the phonetic measure of a word (in column). In the multivariate multiple regression, the UV was set as the intercept so that the estimates of the vowels with the primary stress are comparable with each other.

We make a couple of observations. First, among all the vowels, only the diphthongs /aʊ/ and /aɪ/ showed statistical significance for all four phonetic measures, indicating that these vowels, compared to UV, significantly increase in max F0, F0 range, inverse phone rate, and mean intensity. Second, the vowels varied in relationship with phonetic measures. The mid vowels /ɛ/, /ʌ/, and /ɔ/ had significant relationships with all the phonetic measures except F0 range. In other words, these vowels, compared with UV, significantly increased in max F0, inverse phone rate, and mean intensity, while they did not extend F0 range. The low vowels /ɑ/ and /æ/ showed significant relationships with inverse phone rate and mean intensity but not with max F0 and F0 range, indicating that these vowels' inherent characteristics, compared with UV, manifest in the non-F0 measures only. The high vowels /i/ and /u/ had significant relationships with max F0 and inverse phone rate but not with mean intensity, suggesting that high vowels are encoded by the increased max F0 and inverse phone rate. These results in the present study are broadly in line with those in previous studies (e.g., Fahey & Diehl, 1996; House & Fairbanks, 1953; Kingston, 1992; Lehiste & Peterson, 1959, 1961; Whalen & Levitt, 1995; Whalen et al., 1999; Young et al., 2001) in that high and low vowels are associated with F0 and intensity, respectively.

**Table 1.** The effect of a vowel (in row) on the phonetic realization of words in the speech material (in column)

| Vowel | Max F0 | F0 range | Inverse phone rate | Mean intensity |
|---|---|---|---|---|
| (Intercept) | $\beta=-.40$ $t=-3.48$ $p<.001^{***}$ | $\beta=-.34$ $t=-2.89$ $p<.01^{**}$ | $\beta=-.77$ $t=-6.99$ $p<.001^{***}$ | $\beta=-.62$ $t=-5.52$ $p<.001^{***}$ |
| /i/ | $\beta=1.59$ $t=4.90$ $p<.001^{***}$ | $\beta=1.32$ $t=4.05$ $p<.001^{***}$ | $\beta=1.18$ $t=3.83$ $p<.001^{***}$ | $\beta=.47$ $t=1.50$ n.s. |
| /ɪ/ | $\beta=.36$ $t=1.71$ $p=.09$ | $\beta=.04$ $t=.17$ n.s. | $\beta=.79$ $t=3.92$ $p<.001^{***}$ | $\beta=.32$ $t=1.53$ n.s. |
| /eɪ/ | $\beta=.44$ $t=1.61$ n.s. | $\beta=1.07$ $t=3.90$ $p<.001^{***}$ | $\beta=1.49$ $t=5.74$ $p<.001^{***}$ | $\beta=.61$ $t=2.29$ $p<.05^{*}$ |
| /ɛ/ | $\beta=.57$ $t=2.49$ $p<.05^{*}$ | $\beta=.18$ $t=.81$ n.s. | $\beta=.79$ $t=3.64$ $p<.001^{***}$ | $\beta=.89$ $t=4.03$ $p<.001^{***}$ |
| /æ/ | $\beta=.06$ $t=.30$ n.s. | $\beta=.31$ $t=1.47$ n.s. | $\beta=.61$ $t=3.12$ $p<.01^{**}$ | $\beta=.45$ $t=2.26$ $p<.05^{*}$ |
| /ɚ/ | $\beta=.26$ $t=.99$ n.s. | $\beta=.19$ $t=.74$ n.s. | $\beta=1.09$ $t=4.41$ $p<.001^{***}$ | $\beta=1.04$ $t=4.14$ $p<.001^{***}$ |
| /ʌ/ | $\beta=.48$ $t=2.22$ $p<.05^{*}$ | $\beta=.34$ $t=1.54$ n.s. | $\beta=.91$ $t=4.40$ $p<.001^{***}$ | $\beta=.93$ $t=4.42$ $p<.001^{***}$ |
| /aɪ/ | $\beta=.67$ $t=3.79$ $p<.001^{***}$ | $\beta=.49$ $t=2.78$ $p<.01^{**}$ | $\beta=1.2$ $t=7.20$ $p<.001^{***}$ | $\beta=.95$ $t=5.59$ $p<.001^{***}$ |
| /aʊ/ | $\beta=.81$ $t=2.61$ $p<.01^{**}$ | $\beta=1.23$ $t=3.95$ $p<.001^{***}$ | $\beta=1.17$ $t=3.96$ $p<.001^{***}$ | $\beta=.94$ $t=3.10$ $p<.01^{**}$ |
| /u/ | $\beta=.51$ $t=2.26$ $p<.05^{*}$ | $\beta=.30$ $t=1.35$ $p=$n.s. | $\beta=.63$ $t=2.96$ $p<.01^{**}$ | $\beta=.28$ $t=1.28$ $p=$n.s. |
| /ʊ/ | $\beta=.33$ $t=.87$ n.s. | $\beta=-.07$ $t=-.20$ n.s. | $\beta=-.03$ $t=-.08$ n.s. | $\beta=.75$ $t=2.03$ $p<.05^{*}$ |
| /oʊ/ | $\beta=.39$ $t=1.25$ n.s. | $\beta=.54$ $t=1.75$ $p=.08$ | $\beta=1.09$ $t=3.69$ $p<.001^{***}$ | $\beta=.94$ $t=3.11$ $p<.01^{**}$ |
| /ɔ/ | $\beta=1.04$ $t=3.34$ $p<.001^{***}$ | $\beta=.49$ $t=1.56$ n.s. | $\beta=.94$ $t=3.20$ $p<.01^{**}$ | $\beta=1.02$ $t=3.36$ $p<.001^{***}$ |
| /ɑ/ | $\beta=.23$ $t=.98$ n.s. | $\beta=.32$ $t=1.34$ n.s. | $\beta=1.16$ $t=5.20$ $p<.001^{***}$ | $\beta=1.31$ $t=5.75$ $p<.001^{***}$ |

In order to examine which vowel is associated with more extreme phonetic value, we visualize the estimates of vowels for each phonetic measure based on the model output shown in Table 1. Figures 2–5 display the relationships between vowels and phonetic measures. Figure 2 shows the relationship between vowel (x-axis) and max F0 (y-axis). We observe that compared to UV, the vowels /i/, /ɛ/, /ʌ/, /aɪ/, /aʊ/, /u/, and /ɔ/ showed substantially higher max F0. These vowels can be rearranged from highest to lowest estimates as follows: /i/>/ɔ/>/aʊ/>/aɪ/>/ɛ/>/u/>/ʌ/. The other vowels, /ɪ/, /eɪ/, /æ/, /ɚ/, /ʊ/, /oʊ/, and /ɑ/ did not significantly differ from UV in max F0.
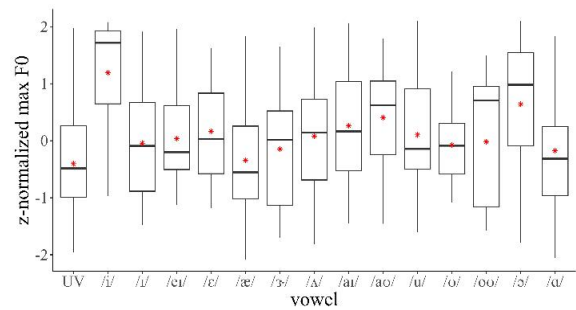


**Figure 2.** The vowel (x-axis) and z-normalized max F0 (y-axis) of words in the speech material.

Figure 3 visualizes the relationship between vowel and F0 range. The vowel /i/ and the diphthongs /eɪ/, /aɪ/, and /aʊ/ had substantially expanded F0 ranges, compared with UV. These vowels can be rearranged in decreasing order of estimates: /i/ > /aʊ/ > /eɪ/ > /aɪ/. Compared with Figure 2, more vowels, /ɪ/, /ɛ/, /æ/, /ɚ/, /ʌ/, /u/, /ʊ/, /oʊ/, /ɔ/, and /ɑ/ in Figure 3 did not significantly differ from UV. From Figures 2 and 3, we confirmed that the high vowels show substantially higher F0 measures than the low vowels in the speech.
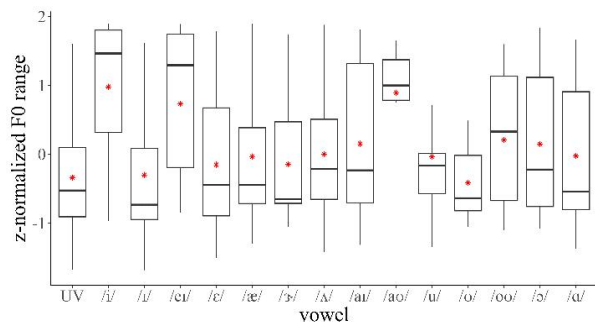


**Figure 3.** The vowel (x-axis) and z-normalized F0 range (y-axis) of words in the speech material.

Figure 4 shows the relationship between vowel and inverse phone rate. All the vowels except /ʊ/ had significantly faster inverse phone rates (i.e., longer duration) than UV. Diphthongs tended to show higher estimates than monophthongs. This might not be surprising given that diphthongs, which combine two vowels, are inherently longer than monophthongs. What is surprising is the higher estimate of the high vowel /i/ than the low vowel /ɑ/. Previous research (Heffner, 1937; House, 1961; House & Fairbanks, 1953; Peterson & Lehiste, 1960) argues that duration should be longer for low than for high vowels due to the articulatory process (i.e., more time needed for larger jaw opening for low vowels than for high ones), which turned out not to be the case in the present study, perhaps due to differences in phonetic measurement (i.e., vowel duration in the previous research vs. word phone rate in the present study). The vowels can be ranked in decreasing order of estimates as follows: /eɪ/>/aɪ/>/i/>/aʊ/>/ɑ/>/ɚ/>/oʊ/>/ɔ/>/ʌ/>/ɪ/>/ɛ/>/u/>/æ/. Only the vowel /ʊ/ was found not to differ significantly from UV in inverse phone rate. In the speech material, /ʊ/ was mostly observed in function words (e.g., "your," "would") and could have been reduced. Perhaps for this reason, /ʊ/ might not have significantly differed from UV in F0 measures and inverse phone rate in this speech.
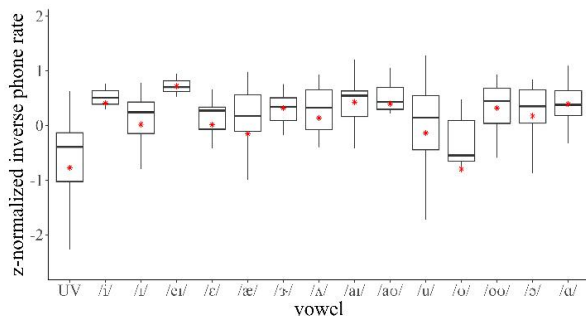
**Figure 4.** The estimated effect of vowel on the likelihood of perceived prominence of words.

Figure 5 displays the relationship between vowel and mean intensity. Most vowels, /eɪ/, /ɛ/, /æ/, /ɝ/, /ʌ/, /aɪ/, /aʊ/, /ʊ/, /oʊ/, /ɔ/, and /ɑ/, showed substantially higher mean intensity than the UV. These vowels can be rearranged from highest to lowest estimates in the following order: /ɑ/>/ɝ/>/ɔ/>/aɪ/>/aʊ/>/oʊ/>/ʌ/>/ɛ/>/ʊ/>/eɪ/>/æ/. The three high vowels, /i/, /ɪ/, and /u/, did not significantly differ from UV in mean intensity. These results are consistent with the previous research (Lehiste & Peterson, 1959; Young et al., 2001) that low vowels have higher intensity than do high vowels.
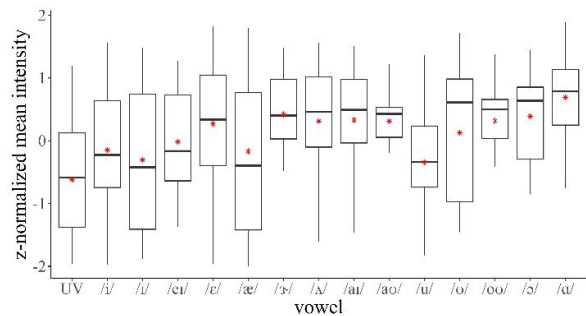


**Figure 5.** The vowel (x-axis) and z-normalized mean intensity (y-axis) of words in the speech material.

In sum, vowels vary in which phonetic cue they are strongly associated with. Compared to UV, the diphthongs are the only vowels that substantially increased in all phonetic measures, max F0, F0 range, inverse phone rate, and mean intensity. The other vowels substantially increased in some, but not all, phonetic measures. The mid vowels /ɛ/, /ʌ/, and /ɔ/ increased in all phonetic measures except F0 range. The low vowels /æ/ and /ɑ/ were associated with higher inverse phone rate and mean intensity compared with the other vowels. The high vowels /i/ and /u/ were related to higher max F0 and inverse phone rate compared with the other vowels. These results from the current study, in alignment with those from previous research, suggest that vowels with more articulatory effort (i.e., high or low vowels) are associated with more extreme phonetic values (i.e., higher max F0, F0 range, inverse phone rate, and mean intensity) than ones with less articulatory effort (i.e., mid vowels).

## 3.2. Perception of Prosodic Prominence

To address the second research question, we examined how the phonological and phonetic information of vowels discussed in Section 3.1. influences linguistically untrained listeners' perceptions

of prosodic prominence. Table 2 summarizes the main effects from the generalized mixed-effects model. The summary of the model's interaction effects can be found in the online Supplementary Material of this study. Each variable (in row) shows its estimated effect on the probability of a word being perceived as prominent. The vowel estimates were calculated in reference to the UV (set as the intercept in the model).

We observe that all the acoustic cues and all the vowels except /ʊ/ significantly affect the likelihood of perceived prominence. This suggests that not only the phonetic but also the phonological information of vowels influence perceptions of words' prosodic prominence. In other words, listeners were likely to rate a word's prosodic prominence, taking vowel identity into account independent of its associated acoustic cues in the speech. Only /ʊ/ did not significantly contribute to perceptions of prosodic prominence.

**Table 2.** The main effect of phonological and acoustic information of vowel (in row) on the probability of perceived prominence by linguistically untrained listeners

| Variable | est. | SE | z | p-value |
|---|---|---|---|---|
| (Intercept) | −3.32 | .16 | −20.25 | <.001*** |
| /i/ | 1.87 | .19 | 9.97 | <.001*** |
| /ɪ/ | 1.37 | .17 | 8.19 | <.001*** |
| /eɪ/ | 1.37 | .18 | 7.59 | <.001*** |
| /ɛ/ | 2.63 | .16 | 16.45 | <.001*** |
| /æ/ | 1.78 | .16 | 11.21 | <.001*** |
| /ɝ/ | 1.07 | .19 | 5.60 | <.001*** |
| /ʌ/ | 2.57 | .16 | 16.52 | <.001*** |
| /aɪ/ | .55 | .16 | 3.36 | <.001*** |
| /aʊ/ | 1.93 | .19 | 10.40 | <.001*** |
| /u/ | .98 | .18 | 5.46 | <.001*** |
| /ʊ/ | −.38 | .44 | −.87 | n.s. |
| /oʊ/ | 1.61 | .23 | 7.14 | <.001*** |
| /ɔ/ | 1.99 | .19 | 10.48 | <.001*** |
| /ɑ/ | 2.15 | .17 | 12.72 | <.001*** |
| Max F0 | .35 | .03 | 10.51 | <.001*** |
| F0 range | .28 | .03 | 9.17 | <.001*** |
| Inverse phone rate | .49 | .05 | 9.70 | <.001*** |
| Mean intensity | .14 | .04 | 3.86 | <.001*** |

Figure 6 visualizes the effects of vowels (x-axis) on the likelihood of perceived prominence (y-axis) based on Table 2. The vowels can be rearranged from highest to lowest estimates as follows: /ɛ/>/ʌ/>/ɑ/>/ɔ/>/aʊ/>/i/>/æ/>/oʊ/>/eɪ/>/ɪ/>/ɝ/>/u/>/aɪ/>/ʊ/>/ə/. The mid vowels /ɛ/ and /ʌ/ tended to have higher effects on probability of perceived prominence than did high and low vowels. This is surprising given that mid vowels are not considered to have effortful articulatory processes and extreme acoustic values (c.f. high and low vowels). Taken together, these results suggest that the acoustic information of vowels in the speech might not directly influence perceptions of prosodic prominence and may be mediated by other expectation-driven factors, including vowels' phonological information. If acoustic cues related directly to perceived prominence, we should have observed that high and low vowels with more extreme acoustic realizations were perceived as more prominent than mid vowels, which turned out not to be the case in the present study.
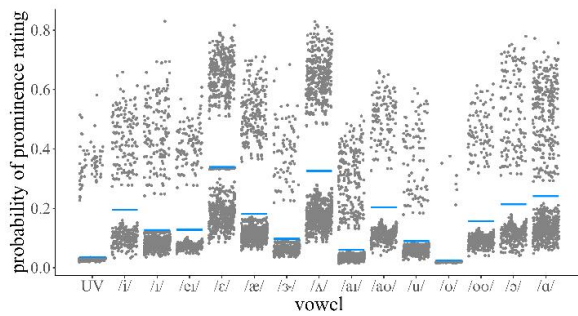
**Figure 6.** The estimated effect of vowel on the likelihood of perceived prominence of words.

A post-hoc pairwise comparison shows that vowels can be categorized into four classes, as shown in Table 3. Vowels' estimated effects on perceived prominence decrease from Class 1 to Class 4 (i.e., high effects for Class 1 and low effects for Class 4). Class 1 includes the mid vowels /ɛ/ and /ʌ/. From the post-hoc pairwise comparison, there was no significant difference between /ɛ/ and /ʌ/ ($\beta=-.06$, $z=-.55$, n.s.) while /ɛ/ significantly differed from /ɑ/ ($\beta=-.48$, $z=-3.85$, $p<.05^*$).

Class 2 consists of the high vowel /i/; the low vowels /æ/, /ɔ/, and /ɑ/; and the diphthongs with the high back vowel, /oʊ/ and /aʊ/. The post-hoc pairwise comparison showed no significant difference between /ɑ/ and /ɔ/ ($\beta=.16$, $z=.95$, n.s.); /ɑ/ and /aʊ/ ($\beta=.22$, $z=1.40$, n.s.); /ɑ/ and /i/ ($\beta=.27$, $z=1.65$, n.s.); /ɑ/ and /æ/ ($\beta=.36$, $z=2.88$, n.s.); and /ɑ/ and /oʊ/ ($\beta=.54$, $z=2.64$, n.s.). It, however, revealed that /ɑ/ significantly differed from /eɪ/ ($\beta=.77$, $z=5.16$, $p<.001^{***}$).

Class 3 includes the high vowels /ɪ/ and /u/; the mid vowel /ɚ/; and the diphthong with the high front vowel /eɪ/. From the post-hoc pairwise comparison, there was no significant difference between /eɪ/ and /ɪ/ ($\beta=-.50$, $z=-2.94$, n.s.); /eɪ/ and /ɚ/ ($\beta=-.30$, $z=-1.70$, n.s.); /eɪ/ and /u/ ($\beta=.39$, $z=2.39$, n.s.), while /eɪ/ significantly differed from /aɪ/ ($\beta=-.82$, $z=-5.69$, $p<.001^{***}$).

Finally, Class 4 consists of the high vowel /ʊ/; the diphthong with the high front vowel /aɪ/; and UV. The post-hoc pairwise comparison showed no significant difference between /aɪ/ and /ʊ/ ($\beta=.94$, $z=2.18$, n.s.); and /aɪ/ and UV ($\beta=-.55$, $z=-3.36$, n.s.).

**Table 3.** The four classes of vowels based on their estimated effects on probability of perceived prominence as shown in Table 2

| Class | Vowel |
|---|---|
| 1 | /ɛ/, /ʌ/ |
| 2 | /ɑ/, /ɔ/, /aʊ/, /i/, /æ/, /oʊ/ |
| 3 | /eɪ/, /ɪ/, /ɚ/, /u/ |
| 4 | /aɪ/, /ʊ/, UV |

UV, unstressed vowel.

In sum, our results show that vowels' phonological and acoustic information influences perceptions of words' prosodic prominence. Increases in all the acoustic measures of vowels yield increased likelihood of words' perceived prominence. Among the acoustic measures, phone rate affected probability of perceived prominence most, followed by the F0 measures. Mean intensity affected likelihood of perceived prominence least. Also, the vowels affected probability of perceived prominence differently. The mid vowels, /ɛ/ and /ʌ/, which were phonetically encoded by max F0, phone rate, and intensity but not by F0 range in the speech, affected perceptions

of prosodic prominence most. High vowels, low vowels, and diphthongs, which substantially differed from the other vowels' phonetic measures in the speech, showed lower effects on perceived prosodic prominence than did mid vowels. Taken together, these results suggest that signal-driven factors (i.e., vowels' phonetic information) do not directly influence perceptions of prosodic prominence and are mediated by expectation-driven factors (i.e., vowels' phonological information) in the speech.

## 4. Conclusion

This study has investigated how intrinsic differences in vowels influence perceptions of prosodic prominence with linguistically untrained listeners using a public speech in American English. We first examined how the speech phonetically realized vowels. The high or low vowels were associated with higher max F0, F0 range, inverse phone rate, or mean intensity than the mid vowels, indicating that vowels with more articulatory effort manifest in more extreme phonetic values than ones with less articulatory effort. Next, we investigated how the phonological and acoustic information of vowels in the speech influences listeners' perceptions of prosodic prominence. All the vowels except /ʊ/ made significant effects on perceived prominence, independent of the acoustic measures. Surprisingly, the mid vowels affected perceived prominence more than the high vowels, low vowels, and diphthongs with more effortful articulatory processes and extreme acoustic cues. These results can be taken as evidence that signal-driven factors (i.e., vowels' phonetic information) do not directly affect perceptions of prosodic prominence and are mediated by expectation-driven factors (i.e., vowels' phonological information) in the speech. Overall, the present study expands our understanding of perceptions of prosodic prominence in relation to expectation-driven and signal-driven factors in a spontaneous speech in American English. The results can potentially be applied to the areas of language processing and acquisition.

## Supplementary Materials

For those who might be interested, (1) the F0 contour of a monophthong and a diphthong (Section 2.3.) and (2) the complete model output of the generalized mixed-effects model (Section 3.2.) can be found in the online repository, https://osf.io/a7fc6.

## References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31-56.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20-38.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, The Netherlands: Fortis.

Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and

signal-extrinsic factors in the perception of prosodic prominence: Evidence from rapid prosody transcription. *Journal of Phonetics*, *82*, 100977.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7-9), 1044-1098.

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*(1), 1-42.

Chodroff, E., & Cole, J. (2018, September). Information structure, affect, and prenuclear prominence in American English. *Proceedings of the Interspeech 2018* (pp. 1848-1852). Hyderabad, India.

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, *75*, 113-147.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *Journal of Phonetics*, *35*(2), 180-209.

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.

Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology*, *7*(1), 8.

Fahey, R. P., & Diehl, R. L. (1996). The missing fundamental in vowel height perception. *Perception & Psychophysics*, *58*(5), 725-733.

Heffner, R. M. S. (1937). Notes on the length of vowels. *American Speech*, *12*(2), 128-134.

Hirschberg, J. (1993). Pitch accent in context predicting intonational prominence from text. *Artificial Intelligence*, *63*(1-2), 305-340.

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America, 33*(9), 1174-1178.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *25*, 105-113.

Im, S., Cole, J., & Baumann, S. (2023). Standing out in context: Prominence in the production and perception of public speech, *Laboratory Phonology*, *14*(1).

Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech*, *35*(1-2), 99-113.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*, 1038-1054.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, *31*, 428-435.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, *33*, 419-425.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1-33.

Mahrt, T. (2013). Language markup and experimental design software [Computer software]. Retrieved from www.timmahrt.com/lmeds.html

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271-311). Cambridge, MA: MIT Press.

R Core Team. (2022). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org/

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*(4), 2471-2485.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*(2), 171-206.

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience*, *32*(8), 1017-1033.

Vogel, I., Bunnell, H. T., & Hoskins, S. (1995). The phonology and phonetics of the rhythm rule. In B. Connell, & A. Arvaniti (Eds.), *Phonology and phonetic evidence* (pp. 111-127). Cambridge, UK: Cambridge University Press.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*(3), 1548-1557.

Weide, R. (2005). The Carnegie Mellon University pronouncing dictionary. Retrieved from www.speech.cs.cmu.edu/cgi-bin/cmudict

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1999). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics*, *27*(2), 125-142.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349-366.

Xu, Y. (2013, August). ProsodyPro – A tool for large-scale systematic prosody analysis. *Proceedings of the Tools and Resources for the Analysis of Speech Prosody* (pp. 7-10). Aix-en-Provence, France.

Young, L. H., Zajac, D. J., Mayo, R., & Hooper, C. R. (2001). Effects of vowel height and vocal intensity on anticipatory nasal airflow in individuals with normal speech. *Journal of Speech, Language, and Hearing Research*, *44*(1), 52-60.

• **Suyeon Im,** Corresponding author
Assistant Professor, Dept. of English Language and Literature
Soongsil University
369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea
Tel: +82-2-820-0338
Email: sim@ssu.ac.kr
Fields of interest: Laboratory Phonology, Phonetics, L2 acquisition