THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Vowel-internal cues to vowel quality and prominence in speech perception

OPEN ACCESS

Research Article

Jeremy Steffman*

Vowel-internal cues to vowel quality and prominence in speech perception

Abstract: This study examines how variation in F0 and intensity impacts the perception of American English vowels. Both properties vary intrinsically as a function of vowel features in the speech production literature, raising the question of the perceptual impact of each. In addition to considering listeners' interpretation of either cue as an intrinsic property of the vowel, the possible prominence-marking function of each is considered. Two patterns of prominence strengthening in vowels, sonority expansion and hyperarticulation, are tested in light of recent findings that contextual prominence impacts vowel perception in line with these effects (i.e. a prominent vowel is expected by listeners to be realized as if it had undergone prominence strengthening). Across four vowel contrasts with different height and frontness features, listeners categorized phonetic continua with variation in formants, F0 and intensity. Results show that variation in level F0 height is interpreted as an intrinsic cue by listeners. Higher F0 cues a higher vowel, following intrinsic F0 effects in the production literature. In comparison, intensity is interpreted as a prominence-lending cue, for which effect directionality is dependent on vowel height. Higher intensity high vowels undergo perceptual re-calibration in line with (acoustic) hyperarticulation, whereas higher intensity non-high vowels undergo perceptual re-calibration in line with sonority expansion.

Keywords: F0; intensity; prominence; speech perception; vowel perception

1 Introduction

The notion of suprasegmental parameters in speech (conventionally F0, duration, and intensity) as simply "overlaid" on speech segments has long been known to be

*Corresponding author: Jeremy Steffman**, Department of Linguistics and English Language, The University of Edinburgh, Dugald Stewart Building, EH8 9AD, Edinburgh, UK,
E-mail: jeremy.steffman@ed.ac.uk. https://orcid.org/0000-0003-3675-910X

somewhat problematic (e.g., Fletcher 2010; Lehiste 1970). One reason is the "intrinsic" variation of these parameter across segments. For example, higher vowels generally have higher F0 than lower vowels, all else being equal (e.g., Chen et al. 2021; Hillenbrand et al. 1995). In American English, tense vowels also tend to be longer and higher intensity than lax vowels, and low vowels tend to be longer and higher intensity than high vowels (e.g., Fairbanks et al. 1950; Hillenbrand et al. 1995). Moreover, perception research has shown that listeners use vowel duration as a cue to vowel contrasts in American English (e.g., Hillenbrand et al. 2000; Kondaurova and Francis 2008). One question that arises is the following: do listeners reliably exploit variation in suprasegmental parameters such as F0 and intensity as cues to segmental categories? The present study examines this question as applied to vowel categories, building on several recent studies.

This question becomes more interesting if we consider recent work that shows that variation in prosodic structure, cued by suprasegmental parameters, guides segmental perception (e.g., Mitterer et al. 2016, 2019). Most relevant to this study, contextual variation in prominence-lending suprasegmental parameters has recently been shown to impact listeners' perception of vowels (Steffman 2021a, 2021b). These effects are in line with the way that vowel articulations and acoustics vary under prominence, in so called "prominence strengthening" (e.g., Cho 2005; de Jong 1995), described in Section 2.2. In other words, variation in suprasegmental parameters, if interpreted as variation in prominence, may cause listeners to recalibrate vowel perception in a manner that is distinct from their interpretation as intrinsic cues.

The present study examines how two suprasegmental cues, F0 and intensity, impact perception of contrastive vowel categories in isolated words, in American English. Four different vowel contrasts which differ in their prominence strengthening effects are tested in four complementary experiments.

2 Background

2.1 Intrinsic variation in vowel F0 and intensity

As noted above, F0 generally varies as a function of vowel height. All else being equal, vowels that are higher in the vowel space tend to have higher F0, evident across large scale studies of American English (Hillenbrand et al. 1995; Peterson and Barney 1952). A biomechanical explanation for this pattern that has been pursued in the literature is the so-called "tongue pull hypothesis" (e.g., Hoole and Honda 2011; Ladefoged 1968; Ohala 1973). As described originally by Ladefoged "[…] the tongue is attached to the superior part of the hyoid bone, and some of the laryngeal muscles are attached to

the inferior part. When the tongue is raised these laryngeal muscles are stretched, and the tension of the vocal cords is increased" (p. 41), leading to increases in F0. In line with this account, Honda (1983) presented electromyographic data showing that the hyoid bone is indeed pulled forward via contraction the posterior genioglossus during the production of high vowels. Chen et al. (2021) recently showed that an additional "jaw push" mechanism leads to differences in intrinsic F0, particularly in non-high vowels which are not well explained by the tongue pull hypothesis (where low vowels have lower F0 than mid vowels). Importantly, though these biome-chanical accounts provide an articulatory basis for intrinsic F0, it has also been noted that intrinsic F0 differences may be enhanced in production as a phonologization of the biomechanically-based effect (Honda and Fujimura 1991).

Intensity has also been documented to vary across vowels: the pattern can be described in two ways. First, in American English specifically, so-called high "tense" vowels are higher intensity than their "lax" counterparts. Relevant to the present study, Fairbanks et al. (1950), with data from ten speakers, documented that /i/ was higher in intensity than /ɪ/, and that /u/ was higher in intensity than /ʊ/. They additionally show, for other vowels, that the general pattern is one whereby lower vowels have relatively higher intensity (this comports with Lehiste and Peterson 1959, though they report data from just one speaker). This latter effect can be understood through the lens of vowel sonority as defined in e.g., Beckman et al. (1992), whereby lower, more open vowel articulations allow more energy to radiate from the mouth, with increased intensity.

The present study focuses on the perception of four vowel contrasts. Two of these are the aforementioned high vowel tense/lax pairs: /i/ versus /ɪ/ and /u/ versus /ʊ/. As noted above: /i/ and /u/ have both higher F0 and intensity than /ɪ/ and /ʊ/ respectively (Fairbanks et al. 1950; Hillenbrand et al. 1995). The additional two pairs are non-high front vowels /ɛ/ versus /æ/, and non-high non-front vowels /ʌ/ versus /ɑ/. For these non-high vowel pairs, the higher vowel in the pair (/ɛ/ and /ʌ/), has both higher F0 and lower intensity, in line with both of the patterns sketched above (Chen et al. 2021; Fairbanks et al. 1950).

2.2 "Extrinsic" variation in vowel F0 and intensity in prominence marking

Intensity and F0 also vary based on the prosodic organization of an utterance. One clear predictor of both intensity and F0 is prosodic prominence. In the prominence-marking system of American English, prominence can described as phonologically manifested by the placement of a pitch accent on a metrically strong syllable (Beckman and Pierrehumbert 1986), with different pitch accents conveying different levels of prominence (Bishop et al. 2020; Cole et al. 2019). More generally, if

prominence is understood as the property of "standing out" in the speech stream (Baumann and Cangemi 2020; Baumann and Winter 2018) both increases in F0 and intensity map to greater perceived prominence, particularly for words with a high pitch accent (Bishop et al. 2020; Cole et al. 2010, 2019).[1] On the basis of these studies, there is a general expectation that more extreme F0 and higher intensity should lead to increases in perceived prominence.

2.2.1 Prominence effects in segmental perception

Why then should perceptual prominence matter for vowel perception? Recent work has shown that listeners' perception of segmental material in speech is related to their perception of prosodic features such as phrasing and prominence. The relatedness of these two domains becomes apparent if we consider the sizable body of speech production research which shows that segmental articulations and acoustics are fine-tuned by the prosodic organization of an utterance (e.g., Cho 2004, 2016; Keating 2006). For example, in American English, voice onset time (VOT) in stops at the beginning of a prosodic phrase is longer (Kim et al. 2018b), and perception data has suggested that listeners use prosodic phrasing information in their perception of VOT as a cue to stop voicing contrasts (Kim and Cho 2013; Mitterer et al. 2016). This data and subsequent work (Kim et al. 2018b; Mitterer et al. 2019) has been taken to support a model of spoken language recognition that entails parallel extraction of segmental and prosodic parses of the speech signal, where prosodic parsing involves computing a representation of phrasing and prominence (Cho et al. 2007; McQueen and Dilley 2020). One model that captures this is the Prosody Analyzer (Cho et al. 2007), which posits parallel processing of prosodic and segmental information, and integration of prosodic structure with lexical candidates in lexical competition.

With respect to prominence effects specifically, it has been shown that contexual prominence cues mediate vowel perception. Two patterns of prominence-driven changes in vowel production and acoustics are relevant to consider here. First, so-called *sonority expansion*, which refers to the production of prominent vowels with larger amplitude of jaw movement, and lingual modulation to produce the vowel as more open. Sonority is used in this context to refer to increased opening (decreased impedance) allowing more energy to radiate from the mouth (Beckman et al. 1992; Silverman and Pierrehumbert 1990). These effects have been documented clearly for both mid and low vowels in American English (Cho 2005; Erickson 2002), and correspond, acoustically, to both a raising of vowel F1 (first formant) and lowering of vowel

---

1 Bishop et al. (2020) find that maximum F0 does not impact prominence perception in words with low (L*) and down-stepped pitch accents in American English, whereas it has a clear effect for high pitch accents.

F2 (second formant). A different pattern has been documented for American English /i/ and /ʊ/. Cho (2005) finds that /i/ is generally produced with more extreme (higher, more front) lingual articulation under prominence, and a corresponding lowering of F1 and raising of F2. de Jong (1995) also found /ʊ/ was produced with more lingual retraction under prominence, and with increased protrusion of the upper lip, corresponding to a backer and more rounded production of the vowel. Results such as these are not consistent with the sonority expansion model, as they show overall more closed articulations under prominence. Instead, they have been described as localized *hyperarticulation*, which takes place for high vowels in American English. Hyperarticulation entails that a more extreme/precise vowel target is produced under prominence: a more closed articulation for high vowels. Proposed functional motivations for this asymmetry in prominence strengthening effects are overall dispersion in the vowel space and contrast maintenance with mid vowels (Cho 2005). Note that sonority expanding modulations in low vowels specifically can also be considered hyperarticulation, as increased opening constitutes enhancement of their [+low] feature, though the same cannot be said for mid vowels. In comparison, in Tongan, which has a less-crowded vowel space, Garellek and White (2015) show sonority expansion effects even for high vowels, suggesting that these patterns are vowel inventory and language dependent.

Sonority expansion and hyperarticulation patterns have been shown to impact vowel perception. Steffman (2021a, 2021b) showed that listeners make use of contextual prominence information in determining how formant cues are perceived along these lines. In other words, a contextually prominent vowel is expected to be realized as if it were subject to (acoustic) prominence strengthening patterns. In these previous studies, contextual prominence was manipulated as the presence or absence of cues conveying narrow focus preceding the target word in a carrier phrase, making the target vowel less prominent. Testing the perception of the /ɛ/-/æ/ contrast, Steffman (2021b) found that listeners required overall higher F1 and lower F2 to perceive /æ/ in more prominent contexts, reflecting an expectation of sonority expansion, i.e. reflecting the expectation that a contextually prominent vowel would be realized as acoustically lower and backer in the vowel space in terms of F1/F2 (with higher F1, and lower F2). Complementing this result, Steffman (2021a) showed that another pattern is evident for the perception of /i/ versus /ɪ/, whereby listeners required overall lower F1 and higher F2 to perceive /i/, in line with an expectation of acoustic hyperarticulation, i.e. reflecting the expectation that a contextually prominent vowel would be realized as acoustically higher and more front in terms of F1/F2 (with lower F1, and higher F2). These results thus show that vowel-specific strengthening patterns generate perceptual adjustments for listeners. Notably, in these experiments, only the context varied so that the vowel itself was acoustically identical across conditions. One open question for the perception of vowel contrasts is thus if vowel-internal cues will generate the same effect.

3 Research aims and predictions

To examine the questions outlined above, four parallel experiments were run. In each, listeners' perception of a different vowel contrast was tested. The four contrasts were /i/ versus /ɪ/, /u/ versus /ʊ/, /ɛ/ versus /æ/, and /ʌ/ versus /ɑ/. The task used to address the question of listeners' use of F0 and intensity was a simple two alternative forced choice task (2AFC) in which listeners categorized a phonetic continuum (varying in F1 and F2) ranging between the vowel pairs in each experiment. The impact of F0 and intensity on categorization are investigated. In this section we consider various predictions and their implications for the research questions sketched above.

Given the influence of prominence-lending contextual information in vowel perception shown in Steffman (2021b, 2023), a hypothesis grounded in this research is that vowel-internal information will exert the same prominence-based effect. This is taken as a starting point. However, the fact that the present study examines vowel-internal cues, for which intrinsic effects are crucial consideration, presents a clear alternative possibility in some cases, described below. In that sense, the current study is fairly exploratory in nature and considers this alternative possibility fully as well. Predictions for both F0 and intensity are given below. All predictions are schematized in abbreviated fashion in Table 1 for easier reference.

3.1 F0 predictions

Hereafter, the use of F0 as an "intrinsic" cue will refer to interpretation of F0 in line with intrinsic F0 effects in the speech production literature: in other words, the interpretation of F0 as a property of the vowel which is independent of prominence.

Relevant to the contrasts in the present study, the following patterns have been documented for American English. Chen et al. (2021) find significantly higher F0 in /i/ as compared to /ɪ/, significantly higher F0 in /ɛ/ as compared to /æ/, and significantly higher F0 in /ʌ/ as compared to /ɑ/. /u/ versus /ʊ/ was not tested in Chen et al. (2021), however, in line with the predictions that a higher vowel should have higher F0, Hillenbrand et al. (1995) show higher F0 in /u/ (though they do not provide a statistical test of these differences). On the basis of these patterns we can predict that higher F0 should lead to perception of /i/, /u/, /ɛ/, and /ʌ/, as compared to /ɪ/, /ʊ/, /æ/, and /ɑ/, respectively. These effects would constitute a perceptual correlate of the intrinsic F0 differences in vowels documented in the speech production literature.

Table 1: Schematized predictions for of F0 and intensity effects in vowel perception. The top portion of the table summarizes the patterns reviewed in the introduction. The differences are described such that, e.g., /i,u/ > /ɪ,ʊ/ indicates that /i,u/ have greater F0, intensity or duration than /ɪ,ʊ/, according to the speech production data reviewed in the paper. In the predictions section, the directionality of an effect is indicated by text color, where gray text in a cell indicates more responses of the higher vowel in a given pair.

	High vowels	Non-high vowels
Pattern		
F0	/i,u/ > /ɪ,ʊ/	/ɛ,ʌ/ > /æ,ɑ/
Intensity	/i,u/ > /ɪ,ʊ/	/ɛ,ʌ/ < /æ,ɑ/
Duration	/i,u/ > /ɪ,ʊ/	/ɛ,ʌ/ < /æ,ɑ/
Prominence	Hyperarticulation under prominence	Sonority expansion under prominence
Predictions		
F0 as an intrinsic cue	High F0 = more /i,u/ responses	High F0 = more /ɛ,ʌ/ responses
Intensity as an intrinsic cue	High intensity = more /i,u/ responses	High intensity = fewer /ɛ,ʌ/ responses
Perceived duration as a function of F0	High F0 = longer = more /i,u/ responses	High F0 = longer = fewer /ɛ,ʌ/ responses
Perceived duration as a function of intensity	High intensity = longer = more /i,u/ responses	High intensity = longer = fewer /ɛ,ʌ/ responses
F0 as prominence	Hyperarticulated variant expected high F0 = fewer /i,u/ responses	More sonorous variant expected high F0 = more /ɛ,ʌ/ responses
Intensity as prominence	Hyperarticulated variant expected high intensity = fewer /i,u/ responses	More sonorous variant expected high intensity = more /ɛ,ʌ/ responses

Alternatively, we can consider several predictions that would be consistent with F0 being used as a prominence cue in vowel perception. These predictions are drawn directly from the contextual prominence effects shown in Steffman (2021a, 2021b), which tested perception of /i/-/ɪ/ and /ɛ/-/æ/ (with analogous predictions made for the other two contrasts). Specifically, if higher F0 signals prominence, this should lead listeners to expect hyperarticulation of /i/ in their perception of /i/ as compared to /ɪ/. This leads to a requirement of more extreme (/i/-like) F1 and F2 for an /i/ percept, and perceptual re-calibration (overall decreasing /i/ responses with prominence, as found in Steffman 2021a). This prediction is notably the opposite of what is expected if F0 is used as an intrinsic cue. Because hyperarticulation effects are also documented for /u/ versus /ʊ/ (de Jong 1995), we can predict the same general effect for /u/-/ʊ/: higher F0 should lead to decreased /u/ responses. The previously described pattern of sonority expansion motivates another prediction for non-high vowels: under prominence, listeners should expect an acoustically lower/backer realization. This effectively expands the perceptual criterion for

what counts as /ɛ/ (as compared to /æ/) and /ʌ/ (as compared to /ɑ/), predicting a respective increase in /ɛ/ and /ʌ/ responses. This is the same directionality as the predicted intrinsic F0 effect, for the non-high vowels only.

3.2 Intensity predictions

Similar intrinsic intensity predictions can be taken from Fairbanks et al. (1950). Tense vowels /i/ and /u/ were documented to have higher intensity than lax vowels /ɪ/ and /ʊ/ respectively. For the other vowel contrasts in the experiment, Fairbanks et al. (1950) show that /æ/ has higher intensity than /ɛ/, and that /ɑ/ has higher intensity than /ʌ/. On the basis of these patterns, higher intensity, if used as an intrinsic cue to the contrast, should lead to increased perception of /i/, /u/, /æ/ and /ɑ/, as compared to /ɪ/, /ʊ/, /ɛ/ and /ʌ/, respectively. Prominence-based predictions for intensity follow from those for F0. If higher intensity signals prominence and induces re-calibration in line with hyperarticulation and sonority expansion patterns, it should lead to *increased* increased /ɪ/, /ʊ/, /ɛ/ and /ʌ/ responses, all the intrinsically lower-intensity vowel in each pair. The intensity predictions are thus in the opposite directionality for the intrinsic versus prominence account for all four contrasts.

3.3 Considering possible effects on perceived duration

Finally, for completeness, some additional predictions are forwarded here, which are motivated by previous work showing that the perception of the duration of acoustic events (in both speech and non-speech) is influenced by F0 and intensity. These effects can be understood through the lens of perceptual integrality in the sense of Garner (1974), whereby two stimulus dimensions are perceived as a whole unit, "integrally".

Raised F0 (and more dynamic F0) has generally been found to increase the perceived duration of acoustic events, even with actual (veridical) duration controlled (e.g., Brigner 1988; Yu et al. 2014).[2] If this is the case in the present stimuli, raised F0 (if perceived as longer duration) would cue /i/ and /u/ (as compared to /ɪ/ and /ʊ/) as both of these vowels are longer than their lax counterparts (Hillenbrand et al.

---

2 Steffman and Jun (2019) find that these effects are dependent on the actual duration of a vowel, whereby the effect is only present at longer vowel durations. This shows the effect is "flexible" and may not always occur. The stimuli used in the present study contained vowel durations which were shorter than those that showed the integral F0 effect in Steffman and Jun (2019), with the exception of the continuum testing /ɛ/ versus /æ/, which may be the most likely to show this integral effect for that reason.

1995). The directionality of this effect is notably the same as that predicted from F0 as an intrinsic cue to vowel quality. For the non-high vowel contrasts, higher F0, cuing longer duration, would be predicted to increase /æ/ and /ɑ/ responses (both being longer than /ɛ/ and /ʌ/, respectively). Unlike high vowel contrasts, this effect is the *opposite* of that predicted from F0 as an intrinsic cue. In this sense, the directionality of an F0 effect for non-high vowels (if present) will allow us to assess if F0 is being used as an integral cue in duration perception.

Higher intensity has also been shown to lead to perception of longer duration (Turk and Sawusch 1996). As with F0, higher intensity (if perceived as longer duration) would cue /i/ and /u/ (both being longer than /ɪ/ and /ʊ/ respectively). In the non-high vowels, higher intensity (if perceived as longer duration), would cue /æ/, and /ɑ/ (both being longer than /ɛ/ and /ʌ/ respectively). For each experiment, this effect is the same directionality as the intrinsic intensity effects described above. Such an effect would thus be ambiguous in terms of whether it is due to intensity as an intrinsic cue, or an integral cue in duration perception. The prominence-related prediction is notably in the opposite direction from both the intrinsic intensity prediction and intensity-as-perceived-duration prediction.

3.4 On the mutual exclusivity of the effects

Finally, it is important to note here that the framing of the predictions is binary, in that F0 is either an intrinsic cue *or* a prominence cue. However, the claim here is not that these functions are mutually exclusive in general. The effects are framed as such because they (sometimes) make distinct directional predictions. For example, the intensity predictions are reversed in the intrinsic versus prominence account. Thus, an effect in one direction supports one account, and an effect in the other direction supports the other. It should be noted though that these predictions pertain to the task at hand: identifying speech sounds (placed in words). It may be the case that the intrinsic account is supported in the present study, but that in another task that involves prominence ratings or judgments with the same stimuli, raised F0 may also signal prominence to listeners. More generally, the implication is not that F0 or intensity serves only one perceptual function, but rather that in the domain of segmental perception, cue usage as intrinsic or prominence marking can be indexed based on directionality, in some cases.

Nevertheless, the present study may offer some window into multiple cue functions as intrinsic and prominence lending. For F0 specifically, the non-high vowels show the same predicted effect for the intrinsic and prominence accounts. In that sense the two effects are in agreement, and the effect of high F0 could be additive in this sense. For high vowels, the accounts make competing predictions. If the effects

do compete with one another, we could predict a *smaller* effect of F0 for high vowels as compared to non-high vowels, irrespective of directionality. If this was observed, it could suggest that both effects simultaneously determine the categorization outcome in this task. The existence of no effect for high vowels (only) could also be taken as evidence for this sort of competing influence, though a null result of this sort should be interpreted with caution. For reasons described below, interpretation of effect magnitudes in this study should also be taken with a grain of salt. The competing influence possibility is returned to in the results.

4 Methods

4.1 Materials

In each of the four experiments here, participants categorized a stimulus as one word in a minimal pair that exemplified the vowel contrast. In the experiment testing perception of /i/ versus /ɪ/ the minimal pair was "seat"/"sit". In the experiment testing perception of /u/ versus /ʊ/ the minimal pair was "suit"/"soot". In the experiment testing perception of /ɛ/ versus /æ/ the minimal pair was "ebb"/"ab". In the experiment testing perception of /ʌ/ versus /ɑ/ the minimal pair was "shut"/"shot".

Stimuli were recorded by a male speaker of American English, using a Blue Yeti Tri-Capsule USB Microphone in a sound attenuated room, with a sampling rate of 44.1 kHz, at 32 bits. The stimuli were produced by the speaker in a carrier phrase, from which the target word was later excised. This was "I'll say X now", where X was the target word. The carrier phrase was produced in one of two ways. In one, the target word was the most prominent in the phrase, receiving the nuclear pitch accent. In the other, The pre-target word received focus (e.g., as a contrastive response to "will you write X now?"). In this second frame, the target itself was post-focus and was thus relatively non prominent in comparison to the other rendition of the carrier phrase (e.g., Eady and Cooper 1986; Xu and Xu 2005). These naturally produced variations in F0 and intensity served as the basis for the values employed in creating F0 and intensity conditions. Target words were excised from each of these frames, and were used as the base files in stimulus creation.

The stimulus manipulation process altered F1 and F2 in the vowel, the F0 contour of the vowel, and the intensity of the word overall. Formant manipulation was carried out via LPC decomposition and resynthesis using the Burg method (Winn 2019), as implemented in Praat (Boersma and Weenink 2020). For each experiment, the formant values for the endpoints of the continuum were based on model productions of each endpoint word. For consistency, these endpoints were always selected from prominent productions of the target word. The resynthesis method
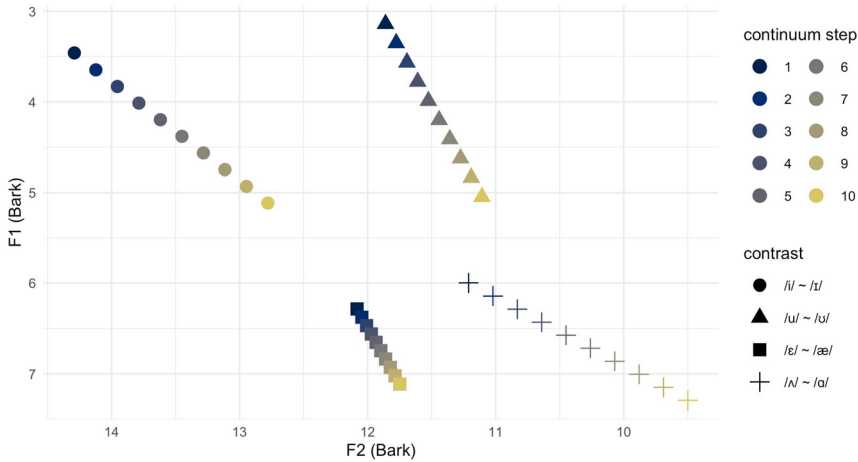
Figure 1: F1 and F2 values (Bark) for the continua used in the four experiments. Each experiment is indicated by the shape of the point, where coloration indicates the continuum step. Note that in each continuum step 1 is endpoint corresponding to the higher vowel of the minimal pair.

started with a natural production of each word, which served as the endpoints of the continuum, and then estimated the source and filter for the starting model vowels. One word was selected as the base, and the filter model F1 and F2 were then varied to match those for the other endpoint of the continuum. Eight intermediate filter steps were created by interpolating between these model endpoint values in Bark space (Traunmüller 1990), with phase-locked higher frequencies from the base model restored to all continuum steps, improving the naturalness of the continuum. The result was a 10 step continuum ranging between each endpoint and varying in F1 and F2 only. The formant values for the continua are shown in Figure 1.

Once these 10 steps were created, F0 was manipulated using the PSOLA method, also implemented in Praat (Boersma and Weenink 2020; Moulines and Charpentier 1990). The goal was to create just two conditions, one in which F0 was relatively high, and one in which it was relatively low. This was accomplished by re-synthesizing the F0 contour of the word to match a prominent production, and non-prominent production. Even though the base file was a prominent rendition of the target word, F0 was resynthesized by overlaying a different prominent production's F0 over the word, such that both conditions were created by resynthesis. In each experiment, this F0 condition will be referred to as the "high F0" condition. The "low F0" condition was created by re-synthesizing the f0 contour from a non-prominent production on the stimulus. The speaker's productions evidenced quite level F0 which did not change dynamically over the vowel. During the resynthesis, any slight rises and falls

Table 2: Mean F0 and intensity values during the target vowel for each experiment (columns) and each F0 and intensity condition (rows).

	/i/-/ɪ/	/u/-/ʊ/	/ɛ/-/æ/	/ʌ/-/ɑ/
High F0 value (Hz)	165	172	132	156
Low F0 value (Hz)	134	126	102	121
F0 *difference* (Hz)	31	46	30	35
High intensity value (dB)	74	77	72	77
Low intensity value (dB)	66	71	64	70
Intensity *difference* (dB)	8	6	8	7

were made level, such that this manipulation is strictly one of F0 height (with a level F0 value). The F0 manipulation resulted in 20 unique stimuli, with 10 continuum steps in each F0 condition. Intensity was subsequently manipulated to be one of two values in each experiment. The "high intensity" condition was created by rendering the target word with the intensity value of a prominent production, while the "low intensity" condition was created by rendering the target word the intensity value of a non-prominent production. This resulted in 40 unique stimuli per experiment, crossing a two-level F0 manipulation with a two-level intensity manipulation at each of the 10 continuum steps. The F0 and intensity values in each experiment are shown in Table 2. Importantly, the goal of these manipulations was to create *relative* differences in F0 and intensity within an experiment, all of which were natural values for the speaker who produced the stimuli. Because of the different values used for each condition across the four experiments, direct comparisons of effect size will be difficult to interpret. However, comparison of effect *directionality* (i.e. does higher F0 lead to the percept of a higher vowel) can be considered across experiments.

4.2 Participants and procedure

For each of the four contrasts tested, 34 self-reported native speakers of American English with normal hearing were recruited from the University of California, Los Angeles (136 participants total). No participant was tested on more than one contrast. The contrasts were tested in a between-participant design as a practical consideration. Given the number of different factor levels (10 continuum steps, two F0 levels, two intensity levels), and the desire for several repetitions of each stimulus, this would create a prohibitively large number of trials in a within-participant design. Participants were undergraduate students, and received course credit for completing the experiment.

Data was collected remotely due to the COVID-19 pandemic, with participants instructed to complete the experiment in a quiet location while using over-the-ear headphones, or earbuds.[3] At the beginning of the experiment, participants were instructed to set the volume to a comfortable level following the practice trials, and not to change the volume once they had set it. There were 8 practice trials in which the two endpoints of the continuum were played in each of the four F0 and intensity conditions.

During a trial, participants heard a stimulus and categorized it as one of two English words which constituted the endpoints of the continuum. They were instructed that their task was to decide what word they had heard, and indicate this decision by key press. The two words were displayed orthographically on the computer screen, each centered in each half of the computer monitor. Participants were instructed to press the 'j' key on the key board to select the choice on left side of the screen, and to press the 'f' key to select the choice on the right side of the screen. The side of the screen on which each word appeared was counterbalanced across participants. After a key press response was registered, the next trial began automatically, with an delay of 500 ms. Each of the 40 unique stimuli were presented a total of 6 times in completely randomized (different randomization for each participant). There were thus 240 trials in an experiment, and all trials were analyzed for a total of 8,160 responses in each experiment (34 speakers by 240 trials).

4.3 Statistical modeling

Results were assessed statistically using a mixed-effects logistic regression model, implemented in the Bayesian framework using the package *brms* (Bürkner 2017). All models were fit with the same general structure. The models were run using R version 4.1.2 (R Core Team 2021) in the RStudio environment (RStudio Team 2021). Weakly informative normally distributed priors were employed for both the intercept and fixed effects.[4] Each model was fit to draw 4,000 samples from the posterior in each of four Markov chains, with a burn-in period of 1,000 iterations in each chain.

---

3 It is important to acknowledge that variation in participation setting and audio hardware is one clear limitation of the study, which introduces some uncontrolled noise into the data collection process. Future studies might benefit from replicating the present results with in-person data collection.

4 All priors except for those for continuum step were set as Normal(0,1) in log-odds space. This captures no prior expectation of an effect of F0 or intensity. The prior for (scaled) F1/F2 continuum step was set to Normal(−3,1), capturing a prior expectation that changes in the F1/F2 continuum (in each experiment going from higher to lower vowel acoustics along the continuum as the continuum step increases numerically), should lead to a decrease in listeners' log-odds of a higher vowel response.

R and Bulk and Tail ESS were inspected to confirm convergence and adequate sampling. For all models, the adapt delta parameter was set to 0.99.

Listeners' categorization response was the dependent variable. It was coded with the higher vowel in a pair mapped to 1, and the lower vowel mapped to 0; e.g., /u/ to 1, /ʊ/ to 0. Responses were predicted as a function of F0 (contrast-coded with high mapped to 0.5, low mapped to −0.5), intensity (contrast-coded with high mapped to 0.5, low mapped to −0.5), and the vowel formant continuum, centered and scaled. All interactions between these effects were additionally included. Random effects were specified as by-participant random intercepts with all fixed effects specified as by-participants random slopes.[5]

In reporting the results from the model, several metrics characterizing the posterior estimate for each effect are presented. The median posterior estimate and 95 % credible intervals (CrI) are given in the full model summaries which are contained in Tables 3 and 4 in the Appendix. This represents the effect (in log-odds), and characterizes the distribution/certainty around that estimate. When 95 % credible intervals *exclude* 0, this suggests a consistent directionality for a non-zero effect, and therefore a reliable, or "credible" effect, analogous to a significant effect in a frequentist model. An additional metric is reported: the "probability of direction" (henceforth pd), computed with *bayestestR* package (Makowski et al. 2019). This metric is useful in that it corresponds more intuitively to a frequentist model's *p*-value. pd indexes the percentage of a posterior distribution which shows a given sign. A posterior centered precisely on zero (i.e., no effect), will have a pd of 50, while a posterior with a strongly skewed negative or positive distribution will have pd that approaches 100. A pd value of 97.5 % corresponds to 95 % credible intervals excluding the value of zero, and hence a credible effect.

All of the data, code for the modeling analysis, and code for generating the figures is contained on an open-access repository hosted on the OSF at: https://osf.io/cew8k/.

5 Results

Results are presented visually in two ways. Figure 2 plots the results for each contrast, showing the categorization functions for the continuum split by both F0 and intensity conditions, as indicated by line coloration and whether the line is dashed or solid. Figure 3 shows just the effect of F0 and intensity conditions, by collapsing across the steps of the continuum and plotting responses as a function of the F0 and intensity conditions for each contrast.

---

5 In R code: response ~ continuum*intensity*F0 + (1 + continuum*intensity*F0 | participant).
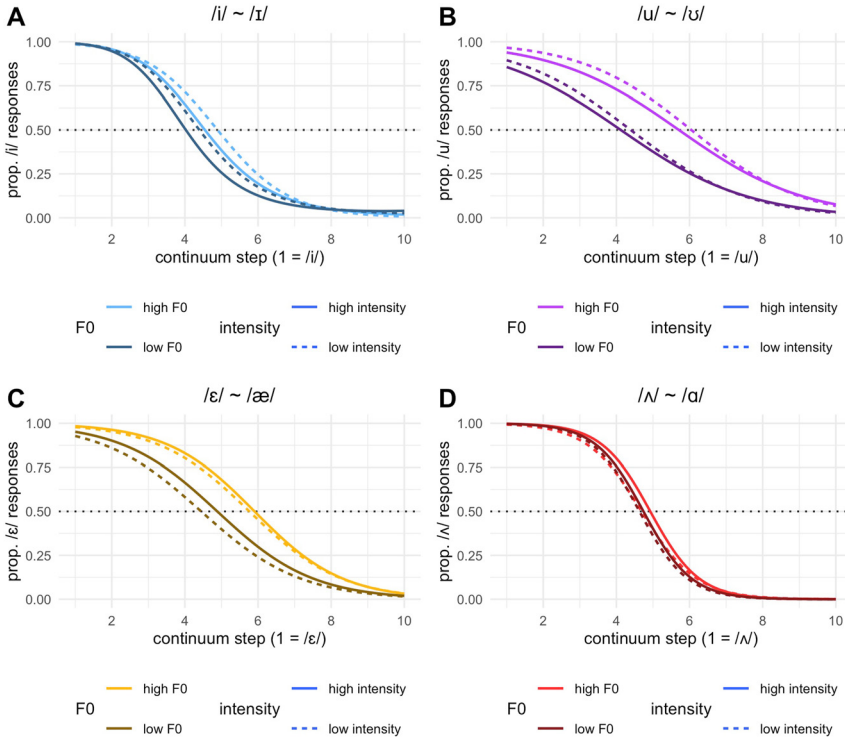
Figure 2: Categorization functions (logistic regression fits) for each experiment, with the continuum step on the *x* axis and listeners' responses on the *y* axis. In each experiment, the proportion of *higher* vowel responses e.g., /i/ relative to /ɪ/ is plotted on the *y* axis.

5.1 Effect of F0

For both /i/-/ɪ/ (β = 0.51, pd = 100) and /u/-/ʊ/ (β = 1.09, pd = 100), F0 exerted a credible effect on categorization, whereby high F0 increased high vowel responses: /i/ and /u/ responses respectively. This result is in line with F0 serving as an intrinsic cue to vowel quality, or as integrated in perception of vowel duration. Notably, following the hyperarticulation-under-prominence account, we would expect the opposite effect as that observed here. The results of both non-high vowel contrasts further agree with this result. For both /ɛ/-/æ/ (β = 1.05, pd = 100) and /ʌ/-/ɑ/ (β = 0.27, pd = 99), higher F0 increased higher vowel responses. This outcome is the opposite as what would be expected if F0 influenced perceived duration in the non-high vowels.

The F0 effect is visible in Figure 2 in the positioning of the brighter colored lines above and to the right of the darker colored lines in each panel. It is more
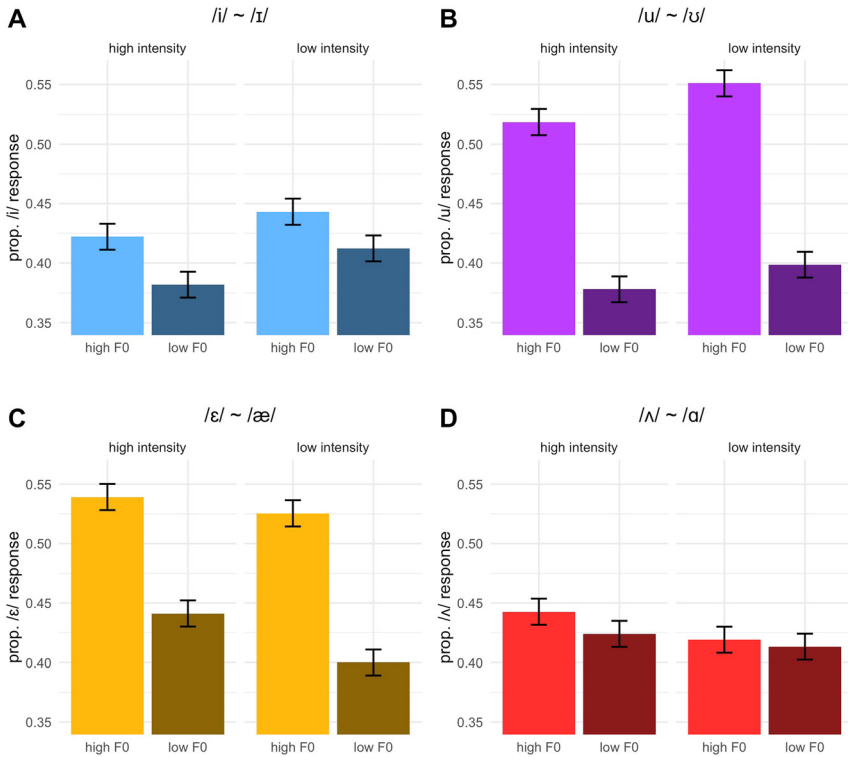
Figure 3: Barplots showing listeners' responses as a function of F0 and intensity conditions (collapsed across the continuum) in each experiment. In each experiment, the proportion of *higher* vowel responses e.g., /i/ relative to /ɪ/ is plotted on the *y* axis. Error bars show one standard error computed from the raw data.

straightforwardly visible in the height of the bars in Figure 3, where in each panel the bars for the high F0 condition show a higher proportion of /i/, /u/, /ɛ/ and /ʌ/ responses than the low F0 condition. The results of F0 overall support the intrinsic cue predictions, i.e., for all experiments higher F0 leads to the percept of a higher vowel. For the two non-high contrasts, the effect is the same directionality as the prominence effect. However, the fact that both high vowel contrasts show a directionality consistent with an intrinsic F0 effect suggests this is the most coherent interpretation of the results.

5.2 Effect of intensity

The intensity results for /i/-/ɪ/ (β = −0.22, pd = 98) and /u/-/ʊ/ (β = −0.20, pd = 100) are similar to one another as well. Higher intensity decreased higher vowel (/i/ and /u/)

responses. Recall that both of the high and tense vowels /i/ and /u/ are higher intensity than their lax counterparts (Fairbanks et al. 1950). The directionality of the intensity effect is therefore the opposite of what would be predicted by its use as an intrinsic cue (see Table 1). Instead, the directionality of the effect is consistent with the prominence account. Higher intensity leads to perceptual re-calibration for hyper-articulation, with a more hyperarticulated variant of the vowel expected under prominence (effectively decreasing higher vowel responses).

A prominence-based interpretation of the intensity results is further supported by examining the effect of intensity in both non-high contrasts. Higher intensity *increased* higher vowel responses for both /ɛ/-/æ/ (β = 0.24, pd = 99) and /ʌ/-/ɑ/ (β = 0.23, pd = 98). This outcome is also the opposite of what would be predicted if intensity served as an intrinsic cue to vowel quality, as lower vowels are generally higher in intensity. It is also not consistent with intensity as an integral cue in duration perception where higher intensity would cue a longer (lower) vowel. Instead, the data supports the prominence account: a higher-intensity non-high vowel, when perceived as prominent, should be realized as lower and backer in the vowel space, with listeners effectively evidencing perceptual re-calibration for sonority expansion for non-high vowels.

There was not credible evidence for an interaction between intensity and F0 in any of the four experiments (see the Appendix for model summaries).

5.3 Combined analysis

Given the differential effects of intensity as a function of vowel height, a combined modeling analysis examined the interaction between vowel height and the effects of F0 and intensity, aggregating data across all four experiments. The purpose of this analysis was to statically confirm the qualitative pattern noted whereby intensity leads listeners to re-calibrate perception differently for high vowels versus for non-high vowels. The model was fit to predict listeners' responses as before, with the higher vowel in each experiment mapped to 1 in codding the model. The fixed effects in the model included vowel height (coded as high/non-high, with high mapped to −0.5, and non-high mapped to 0.5). The other fixed effects in the model were the same as the individual experiment analysis, with the same random effect structure: by participant random intercepts with random slopes for all fixed effects, except for height (as a participant was exposed to only one level of the height variable). A random intercept for experiment was additionally included in the model to account for the differences in overall higher-vowel responses across experiments. The full model summary is included in Appendix. The following questions are central in the combined analysis. First, are F0 effects uniform across vowel heights? Second, is

there statistical evidence for the asymmetrical influence of intensity as a function of vowel height as suggested by the individual experiments analyses?

Figure 4 shows the key results from the combined modeling analysis, plotting the effects of intensity and F0 as bar plots, split by vowel height (grouping both high vowel contrasts with one another, and both non-high vowel contrasts with one another). The combined analysis finds a credible main effect of F0 (β = 0.72, pd = 100) and no credible evidence for an interaction between F0 and vowel height (pd = 80). In other words, there isn't evidence for systematic height-based differences in the effect of F0, and the effect is the same directionality for both high and non-high vowels (higher F0 favoring perception of the higher vowel in a pair). Here we can also note that the effect is qualitatively larger for high vowels, which is the opposite of what would be predicted if intrinsic F0 and prominence effects exert a competing influence on categorization outcomes, as raised in Section 3.4. Because they predict opposite effects for high vowels only, they might be predicted to "cancel each other out" to some extent. The fact that this is not observed thus does not support this possibility, though it may also be due to differences in each continuum. In that sense, these effects do not rule this possibility out, though they also do not provide definitive support for it.

There was not a main effect of intensity in the combined model (pd = 71), though there was a credible interaction between intensity and vowel height (β = −0.45, pd = 100), statistically confirming the observation that the impact of intensity is different for high vowels versus non-high vowels. The interaction was examined
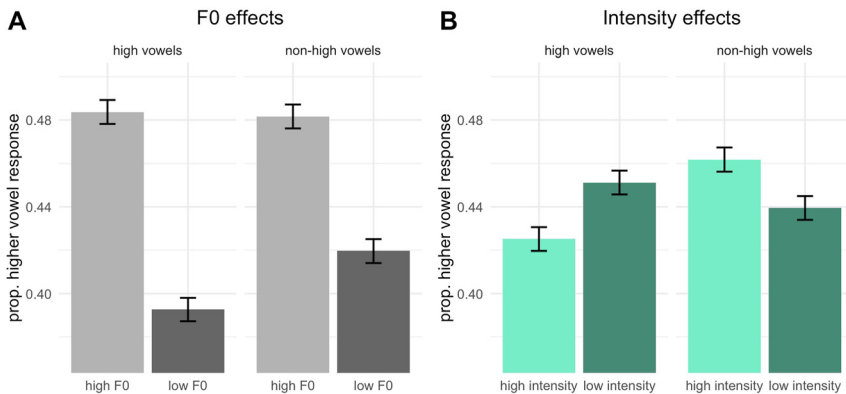
Figure 4: Barplots showing the effects of F0 (panel A), and intensity (panel B), as a function of vowel height, shown as the sub-panels within each panel A and B. Responses are collapsed across the continuum and the cue that is not plotted (i.e., collapsed across intensity in panel A). Note that the *y* axis in each plot shows the proportion of higher vowel responses from each experiment, plotting the proportion of /i/ /u/ /ɛ/ and /ʌ/ responses.

further by testing marginal effects of intensity within each level of height using *emmeans* (Lenth 2021). This assessment finds a credible effect of intensity for high vowels whereby high intensity decreases higher vowel responses (β = −0.20, pd = 100). Showing the opposite effect, high intensity for non-high vowels increases higher vowel responses (β = 0.25, pd = 100). In alignment with the individual modeling results, the combined analysis thus allows us to conclude the intensity exerts an asymmetrical directional influence, dependent on vowel height. This asymmetry can be understood as originating from different patterns of prominence strengthening.

6 Discussion and conclusions

The present study examined the role of F0 and intensity as cues to vowel contrasts in American English. Various potential ways in which F0 and intensity might serve as cues were considered. The first consideration was whether or not these cues would be interpreted as "intrinsic" properties of the vowel, reflecting the documented patterns of F0 and intensity as co-varying with vowel features in the speech production literature. Additionally, the role of F0 and intensity as prominence cues, which might generate an expectation for the realization of a given vowel, was considered.

The F0 results are consistent across the four experiments in showing that listeners used F0 as an intrinsic cue in vowel perception. For all four contrasts considered, higher level F0 leads to the perception of a higher vowel. The uniformity of the direction of these results across contrasts can be taken as indicating the relevance of intrinsic F0 across different regions of the vowel space, which is contrast independent. In this sense, although intrinsic variation in F0 is, in part, physiologically based (though also potentially phonologized), it appears to be a salient perceptual cue that affects perceptual categorization. This is, in essence, a mirror image of the speech production literature on intrinsic F0. One important consideration is the fact that the present F0 manipulation was one that varied overall F0 height with a level F0 pattern. F0 movements that signal prominence may also be dynamic, where for example, a prominence-conveying difference between two pitch accents can include variation in the slope of an F0 rise and other cues, and prominence perception varies as a function of pitch accent category (e.g., Bishop et al. 2020; Cole et al. 2019). Future work will accordingly be needed to extend the present results to examine how different sorts of variation in F0 may signal prominence, or be interpreted as an intrinsic property of the vowel.

The consistency of these results across the four experiments and the link to biomechanical production constraints can additionally be taken to predict that these effects should be consistent across languages, even those with different prosodic

systems, or different patterns of prominence-related vowel modulations. The proposed explanation for why (level) F0 variation is interpreted as an intrinsic cue is that it is linked to consistent patterns in production. Though increases in F0 have generally been shown to signal prominence as noted in the introduction, the fact that prominent (pitch-accented) vowels can have low pitch accents might make F0 a less-consistent vowel-internal prominence cue, or one that is dependent on pitch accent type and context (Bishop et al. 2020, see also Kochanski et al. 2005). This raises an alternative possibility regarding cross-linguistic generalization of the effect: in languages where raised F0 marks prominence more consistently, perceptual prominence effects for F0 may be observed. Cross-linguistic extension of the present results would thus be a beneficial test for the consistency of these effects, and their relation to language-independent biomechanical effects in vowel production, and language-specific patterns of F0 variation. One additional interesting test case would be a language with lexical tone, in which F0 may not be as strong a cue to vowel quality, due to its functionality in conveying lexical contrasts.

The directional uniformity of the F0 effects across the four pairs of vowels is different than the pattern observed for intensity, whereby the directionality of the intensity effect is dependent on vowel height, as shown by the credible interaction in the combined analysis. This difference, as described above, is taken to reflect the ways in which vowels are modulated as a function of prominence strengthening. In other words, a vowel with higher intensity was perceived as if it was subject to prominence strengthening effects. If it was a high vowel, the perceptual responses reflected re-calibration for hyperarticulation (expectation of a higher and more fronted vowel production); if it was a non-high vowel the perceptual responses reflected re-calibration for sonority expansion (expectation of a lower and more backed vowel production). The directionality of the effects also clearly speak against an intrinsic cue account, in the sense that for all four experiments, a higher intensity favored the perception of a vowel category that is described as being intrinsically lower intensity (Fairbanks et al. 1950; Lehiste and Peterson 1959).

The results thus support the claim that intensity variation is interpreted by listeners as a prominence-lending (prosodic) feature, which could be linked to it's relative consistency in signaling prominence, as shown for example in Kochanski et al. (2005). The influence of prominence in the perception of vowel contrasts has been shown in various recent studies, both for phrasal prominence (Steffman 2021a, 2021b), and in local prominence-marking cues such as the presence of vowel-initial glottalization (Steffman 2020, 2023), where glottalization is argued to serve a prominence marking function (Dilley et al. 1996; Garellek 2014). The intensity effects seen here align with the phrase-level prominence effects documented in Steffman (2021a, 2021b), which tested the same front vowel pairs. This set of results, having generated the same shifts in categorization with acoustically very different prominence cues,

can be taken to suggest the present effects are related to more generalized promi-
nence perception, not just higher intensity per se. This predicts we should observe
the same effects for other prominence cues, including those which are acoustically
very different from the intensity manipulations used here (e.g., longer VOT in a
preceding voiceless stop, which co-occurs with prominence; Cho and Keating 2009;
Cole et al. 2007; Kim et al. 2018a). From a cross-linguistic perspective, for languages in
which higher intensity has been shown to mark prominence in speech production,
we can assume that it should cue prominence to listeners as well. An interesting
question then becomes how perceived prominence may interact with language-
specific prominence strengthening effects. For example Garellek and White (2015),
show that lexically stressed syllables in Tongan are produced with higher intensity.
Interestingly, as noted in Section 2.2.1, /i/ and /u/ in Tongan are not hyperarticulated,
and instead are produced as lower in the vowel space (with higher F1), when lexically
stressed. Perceptual categorization responses that comported with this pattern
would thus be in the opposite direction of what we observed here for American
English. The effect of intensity might be predicted to vary across languages then, to
the extent that it cues prominence and induces perceptual effects in line with
language-specific patterns of prominence strengthening.

Finally, the present results do not offer strong support for either F0 or intensity
as influencing perceived duration to the extent that it impacts vowel perception. F0
effects in the high vowel pairs are ambiguous in this regard: higher F0 favors
perception of tense vowels /i/ and /u/, which are both generally higher in F0 and
longer in duration (Chen et al. 2021; Hillenbrand et al. 1995). The results in the non-
high vowel pairs do not have this same ambiguity however, as higher F0 favors
perception of /ɛ/ and /ʌ/, which have intrinsically higher F0, but are generally shorter
than /æ/ and /ɑ/ respectively. The directionality of the effect in these latter two
experiments thus suggests F0 is not impacting duration perception to the extent that
it overrides intrinsic F0 effects, if at all. The intensity results are similarly unam-
biguous across experiments. In each, higher intensity favors perception of a vowel
that is shorter than its counterpart, i.e. /ɪ/, /ʊ/, /ɛ/ and /ʌ/. The intensity results thus
suggest that again, if the perception of vowel duration is influenced by variation in
intensity, it is not to the extent that it influences perception of these contrasts, for
which the intensity results clearly support a prominence-perception account.

6.1 Some future directions

The present results build on the previous investigation of prominence effects in
vowel perception in being a first test of vowel-internal cues. Previous tests of
prominence effects in vowel perception (Steffman 2020, 2021a, 2021b), and of prosodic

effects in segmental perception more generally (e.g., Kim and Cho 2013; Mitterer et al. 2019), have focused on strictly contextual influences. The present effects show that, perhaps unsurprisingly, segment-internal cues also constitute an important piece to the puzzle. A fuller understanding of these effects in the context of parallel segmental and prosodic processing (Cho et al. 2007; McQueen and Dilley 2020) will come from experiments which consider the influence of both segment-internal and contextual information together. One test which seems promising is the orthogonal manipulation of contextual and vowel-internal F0 cues. As contextual F0 variation has been shown to impact prominence perception, and vowel-internal F0 cues have been shown here to be interpreted as intrinsic to the vowel, these F0 variations can be tested in tandem to see which is prioritized, and the relative weighting among cues when in conflict (i.e., when intrinsic vowel-internal F0 favors one response, and contextual prominence-lending F0 favors another). The same question may be asked in terms of the timecourse of cues' influence in processing. Contextual prominence information has been shown to be processed at multiple timescales using eye-tracking, with the timecourse and dynamics of processing varying depending on the cue (Steffman 2020, 2023). Phrasal prominence shows a relatively delayed effect, while the highly local cue of vowel-initial glottalization is processed rapidly. Vowel-internal suprasegmental information may be expected to be processed rapidly, and evidence a different timescale from phrasal prominence, whether it shows a competing effect (e.g., vowel-internal F0 as an intrinsic cue), or an additive one (e.g., vowel intensity as prominence cue). Putting this prediction to the test will help situate the present results in a more complete understanding of the role(s) of suprasegmental information in speech perception and online processing (as discussed in e.g., McQueen and Dilley 2020).

One further direction should consider how these effects relate to phrasal variation in F0 and its role in the intonational system of a language. Beyond conveying prominence, F0 serves an important role is signaling discourse and pragmatic functions, and conveying information structure (see e.g., Breen et al. 2010; Ladd 2008). The present study could thus be extended to test how F0 variation which conveys intonational meaning (e.g., question vs. statement) is related to the intrinsic effects shown here. For example, at the end of an utterance, rising F0 movements in the final syllable (with overall higher F0) cause listeners to interpret a sentence as seeking information (e.g., Sostarics and Cole 2023). This raises an interesting question: if raised F0 is interpreted as conveying intonational information, will intrinsic F0 effects diminish or disappear? One possibility is that when F0 is serving an "extrinsic" function, e.g., signaling an intonational boundary tune, listeners will not interpret F0 variation as an intrinsic cue. Extending these effects to target words which are placed in varying intonational contexts seems like a promising future direction to test the extent to which intrinsic F0 effects are related to listeners'

interpretation of F0's intonational functions. Extensions along these lines will necessarily include consideration of F0 dynamics and, e.g., slope variation in addition to F0 height, as noted above.

Another pertinent question raised by the present results is what mechanism may account for the development of both the F0 and intensity effects. On promising avenue for future research in this regard might take a distributional learning perspective to this question (e.g., Theodore and Monto 2019; Yoshida et al. 2010). The distributional information which is relevant here is the *co-occurence* of F0 and intensity with variation in formant structure. To test how listeners might track these cues, one test could present biased distributions which either agree with, or compete with the results found in this study. For example, stimuli could be presented to listeners in which higher F0 tends to co-occur with (acoustically) higher vowel F1 and F2 values. Conversely, a competing distributional bias would be one in which higher F0 co-occurs with lower vowel F1 and F2, going against intrinsic F0 effects. Examining how variations of this sort modulate the existence and size of intrinsic F0 effects would offer a lens into the possible importance of distributional, or co-occurrence-based, tracking of suprasegmental and segmental cues, and would be a first step in testing these mechanisms as an explanation for the emergence of these effects.

In sum, the results taken together show the importance of considering suprasegmental cues in vowel perception, both in terms of intrinsic patterns and in terms of prominence-related suprasegmental variation. Most fundamentally, the results speak to the need to understand that these effects are not uniform, either across suprasegmental cues (F0, intensity), or across vowel contrasts.

The present results further call for the continued consideration of suprasegmental information in perception and the role of prosody in segmental and lexical processing (e.g., Kim et al. 2018b; Mitterer et al. 2016, 2019; Steffman 2021a, 2021b). It is hoped that a more complete understanding of the role of suprasegmental cues in perception will come from the future consideration of cross-linguistic influences, combined testing of vowel internal and contextual cues, their relationship to intonational structure, and the exploration of mechanisms underpinning these effects.

Appendix

Table 3: Model summaries for each of the four experiments, with estimates (posterior medians), error, lower (L) and upper (U) credible intervals, and probability of direction (pd) metric.

/i/-/ɪ/	Estimate	Est. error	L-95 % CI	U-95 % CI	pd
intercept	−1.05	0.12	−1.29	−0.82	100
intensity	−0.22	0.10	−0.42	−0.01	98
F0	0.51	0.16	0.19	0.84	100
continuum	−3.50	0.24	−3.99	−3.02	100
intensity:F0	0.04	0.18	−0.30	0.38	59
intensity:continuum	0.20	0.18	−0.15	0.55	87
F0:continuum	−0.01	0.19	−0.36	0.37	54
intensity:F0:continuum	−0.10	0.28	−0.65	0.45	64

/u/-/ʊ/	Estimate	Est. error	L-95 % CI	U-95 % CI	pd
intercept	−0.30	0.11	−0.53	−0.07	99
intensity	−0.20	0.07	−0.34	−0.06	100
F0	1.09	0.19	0.73	1.49	100
continuum	−2.26	0.20	−2.64	−1.87	100
intensity:F0	−0.15	0.15	−0.44	0.13	86
intensity:continuum	0.24	0.11	0.03	0.46	99
F0:continuum	−0.06	0.10	−0.24	0.13	73
intensity:F0:continuum	0.16	0.22	−0.26	0.62	76

/ɛ/-/æ/	Estimate	Est. error	L-95 % CI	U-95 % CI	pd
intercept	−0.22	0.13	−0.48	0.04	95
intensity	0.24	0.10	0.04	0.44	99
F0	1.05	0.18	0.68	1.41	100
continuum	−2.75	0.16	−3.08	−2.44	100
intensity:F0	−0.19	0.20	−0.58	0.20	83
intensity:continuum	−0.20	0.12	−0.42	0.03	95
F0:continuum	−0.33	0.13	−0.59	−0.08	99
intensity:F0:continuum	0.04	0.24	−0.41	0.52	56

/ʌ/-/ɑ/	Estimate	Est. error	L-95 % CI	U-95 % CI	pd
intercept	−1.28	0.13	−1.55	−1.02	100
intensity	0.23	0.11	0.01	0.45	98
F0	0.27	0.11	0.05	0.49	99
continuum	−4.96	0.30	−5.57	−4.37	100
intensity:F0	−0.06	0.28	−0.62	0.46	58
intensity:continuum	−0.24	0.22	−0.26	0.66	86
F0:continuum	0.22	0.23	−0.26	0.66	83
intensity:F0:continuum	−0.56	0.58	−1.73	0.58	83

Table 4: Model summaries for the combined modeling analysis, with estimates (posterior medians), error, lower (L) and upper (U) credible intervals, and probability of direction (pd) metric.

	Estimate	Est. error	L-95 % CI	U-95 % CI	pd
intercept	−0.62	0.37	−1.31	0.25	94
vowel height	0.03	0.56	−1.18	1.18	53
F0	0.72	0.09	0.55	0.89	100
intensity	0.02	0.04	−0.06	0.11	71
continuum	−3.36	0.13	−3.62	−3.11	100
vowel height:F0	0.14	0.17	−0.20	0.47	80
vowel height:intensity	−0.45	0.08	−0.60	−0.29	100
vowel height:continuum	0.84	0.25	0.36	1.34	100
intensity:F0	−0.08	0.09	−0.25	0.09	83
intensity:continuum	0.03	0.08	−0.12	0.18	64
F0:continuum	−0.06	0.09	−0.22	0.12	76
vowel height:intensity:F0	0.01	0.16	−0.30	0.32	54
vowel height:F0:continuum	0.08	0.13	−0.18	0.33	72
vowel height:intensity:continuum	0.36	0.12	0.13	0.61	100
intensity:F0:continuum	−0.06	0.14	0.33	0.23	67
vowel height:intensity:F0:continuum	0.20	0.24	−0.26	0.68	81

References

Baumann, Stefan & Bodo Winter. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38.

Baumann, Stefan & Francesco Cangemi. 2020. Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics* 81. 100993.

Beckman, Mary E., Jan Edwards & Janet Fletcher. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In Gerard J. Docherty & D. Robert Ladd (eds.), *Gesture, segment, prosody, (Papers in Laboratory Phonology)*, 68–89. Cambridge: Cambridge University Press.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3. 255–309.

Bishop, Jason, Grace Kuo & Boram Kim. 2020. Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from rapid prosody transcription. *Journal of Phonetics* 82. 100977.

Boersma, Paul & David Weenink. 2020. Praat: Doing phonetics by computer (version 6.1.09). Available at: http://www.praat.org.

Breen, Mara, Evelina Fedorenko, Michael Wagner & Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7–9). 1044–1098.

Brigner, Willard L. 1988. Perceived duration as a function of pitch. *Perceptual and Motor Skills* 67(1). 301–302.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28.

Chen, Wei-Rong, Douglas H. Whalen & Mark Tiede. 2021. A dual mechanism for intrinsic f0. *Journal of Phonetics* 87. 101063.

Cho, Taehong. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32(2). 141–176.

Cho, Taehong. 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *The Journal of the Acoustical Society of America* 117(6). 3867–3878.

Cho, Taehong. 2016. Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass* 10(3). 120–141.

Cho, Taehong, James M. McQueen & Ethan A. Cox. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35(2). 210–243.

Cho, Taehong & Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37(4). 466–485.

Cole, Jennifer, Heejin Kim, Hansook Choi & Mark Hasegawa-Johnson. 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, Jennifer, José I. Hualde, Caroline L. Smith, Christopher Eager, Timothy Mahrt & Ricardo Napoleão de Souza. 2019. Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics* 75. 113–147.

Cole, Jennifer, Yoonsook Mo & Mark Hasegawa-Johnson. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2). 425–452.

de Jong, Kenneth. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America* 97(1). 491–504.

Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4). 423–444.

Eady, Stephen J. & William E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America* 80(2). 402–415.

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

Fairbanks, Grant, Arthur S. House & Eugene L. Stevens. 1950. An experimental study of vowel intensities. *The Journal of the Acoustical Society of America* 22(4). 457–459.

Fletcher, Janet. 2010. The prosody of speech: Timing and rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 521–602. Chichester: John Wiley & Sons, Inc.

Garellek, Marc. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics* 45. 106–113.

Garellek, Marc & James White. 2015. Phonetics of Tongan stress. *Journal of the International Phonetic Association* 45(01). 13–34.

Garner, Wendell R. 1974. The stimulus in information processing. In Howard R. Moskowitz, Bertram Scharf & Joseph C. Stevens (eds.), *Sensation and measurement*, 77–90. Dordrecht: Springer.

Hillenbrand, James, Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97(5). 3099–3111.

Hillenbrand, James M., Michael J. Clark & Robert A. Houde. 2000. Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America* 108(6). 3013–3022.

Honda, Kiyoshi. 1983. Relationship between pitch control and vowel articulation. *Haskins Laboratories Status Report on Speech Research* 73. 269–282.

Honda, Kiyoshi & Osamu Fujimura. 1991. Intrinsic vowel F0 and phrase-final F0 lowering: Phonological versus biological explanations. In Jan Gauffin & Britta Hammarberg (eds.), *Phonatory mechanisms: Physiology, acoustics, and assessment*, 149–158. San Diego: Singular Publishing Group.

Hoole, Phil & Kiyoshi Honda. 2011. Automaticity versus feature-enhancement in the control of segmental f0. In G. Nick Clements & Rachid Ridouane (eds.), *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive sound categories*, 131–171. Amsterdam/Philadelphia: John Benjamins BV.

Keating, Patricia. 2006. Phonetic encoding of prosodic structure. In Jonathan Harrington & Marija Tabain (eds.), *Speech production: Models, phonetic processes, and techniques*, 167–186. London: Psychology Press.

Kim, Sahyang, Holger Mitterer & Taehong Cho. 2018b. A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLoS One* 13(8). e0202912.

Kim, Sahyang, Jiseung Kim & Taehong Cho. 2018a. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics* 71. 65–80.

Kim, Sahyang & Taehong Cho. 2013. Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America* 134(1). EL19–EL25.

Kochanski, Greg, Esther Grabe, John Coleman & Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America* 118(2). 1038–1054.

Kondaurova, Maria V. & Alexander L. Francis. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America* 124(6). 3959–3971.

Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1968. *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge: Cambridge University Press.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: Massachusetts Institute of Technology Press.

Lehiste, Ilse & Gordon E. Peterson. 1959. Vowel amplitude and phonemic stress in American English. *The Journal of the Acoustical Society of America* 31(4). 428–435.

Lenth, Russell V. 2021. *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.7.1-1. Available at: https://CRAN.R-project.org/package=emmeans.

Makowski, Dominique, Mattan S. Ben-Shachar & Daniel Lüdecke. 2019. bayestestr: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software* 4(40). 1541.

McQueen, James M. & Laura Dilley. 2020. Prosody and spoken-word recognition. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford handbook of language prosody*, 509–521. Oxford: Oxford University Press.

Mitterer, Holger, Sahyang Kim & Taehong Cho. 2019. The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language* 108. 104034.

Mitterer, Holger, Taehong Cho & Sahyang Kim. 2016. How does prosody influence speech categorization? *Journal of Phonetics* 54. 68–79.

Moulines, Eric & Francis Charpentier. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9(5–6). 453–467.

Ohala, John J. 1973. Explanations for the intrinsic pitch of vowels. *Monthly internal memorandum*, 9–26. Berkeley: Phonology Laboratory, University of California at Berkeley.

Peterson, Gordon E. & Harold L. Barney. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24(2). 175–184.

R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

RStudio Team. 2021. *RStudio: Integrated development environment for R*. Boston, MA: RStudio, PBC.

Silverman, Kim & Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. In Mary E. Beckman & John Kingston (eds.), *Papers in laboratory phonology*, 72–106. Cambridge: Cambridge University Press.

Sostarics, Thomas & Jennifer Cole. 2023. Testing the locus of speech-act meaning in English intonation. *Proceedings of the 20th international congress of phonetic sciences*. Prague: ICPhS 2023.

Steffman, Jeremy. 2020. *Prosodic prominence in vowel perception and spoken language processing*. Los Angeles: University of California PhD thesis.

Steffman, Jeremy. 2021a. Contextual prominence in vowel perception: Testing listener sensitivity to sonority expansion and hyperarticulation. *JASA Express Letters* 1(4). 045203.

Steffman, Jeremy. 2021b. Prosodic prominence effects in the processing of spectral cues. *Language, Cognition and Neuroscience* 36(5). 586–611.

Steffman, Jeremy. 2023. Vowel-initial glottalization as a prominence cue in speech perception and online processing. *Laboratory Phonology* 14(1). https://doi.org/10.16995/labphon.8753.

Steffman, Jeremy & Sun-Ah Jun. 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.

Theodore, Rachel M. & Nicholas R. Monto. 2019. Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review* 26(3). 985–992.

Traunmüller, Hartmut. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88(1). 97–100.

Turk, Alice E. & James R. Sawusch. 1996. The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America* 99(6). 3782–3790.

Winn, Matthew. 2019. Vowel formant continua from modified natural speech (Praat script). Available at: http://www.mattwinn.com/praat.html.

Xu, Yi & Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33(2). 159–197.

Yoshida, Katherine A., Ferran Pons, Jessica Maye & Janet F. Werker. 2010. Distributional phonetic learning at 10 months of age. *Infancy* 15(4). 420–433.

Yu, Alan, Hyunjung Lee & Jackson Lee. 2014. Variability in perceived duration: Pitch dynamics and vowel quality. *Proceedings of the 4th international symposium on tonal aspects of languages*, 41–44. Nijmegen: TAL 2014.