Linguistic Research 43(2): 535-562
DOI: 10.17250/khisli.43.2.202606.008

Duration and vowel dynamics in the classification of tense-lax pairs of English high vowels*

Hyun Jin Hwangbo
(Pukyong National University)

Hwangbo, Hyun Jin. 2026. Duration and vowel dynamics in the classification of tense-lax pairs of English high vowels. *Linguistic Research* 43(2): 535-562. This study examines which acoustic features best classify English high tense–lax vowel pairs, /i/–/ɪ/ and /u/–/ʊ/, in the Buckeye Corpus of conversational speech. The effectiveness of vowel duration and spectral cues is compared using random forest models, assessing both static midpoint measurements and dynamic formant trajectories. Results indicate that trajectory models outperform static models overall. When dynamic spectral information is unavailable, including duration significantly improves classification accuracy. Notably, duration and spectral dynamics contribute differently to classification depending on vowel pair: for the front pair (/i/–/ɪ/), the static model with duration is most efficient, whereas for the back pair (/u/–/ʊ/), dynamic spectral information is more important. These results demonstrate that classifying high tense–lax vowel pairs in English conversational speech relies on vowel-specific cue patterns. **(Pukyong National University)**

Keywords tense-lax vowels, vowel classification, conversational speech, duration, vowel dynamics

1. Introduction

English tense and lax vowel pairs, such as /i/-/ɪ/ and /u/-/ʊ/, are described as contrasting across multiple dimensions. These include phonological distribution, articulatory description, and phonetic realization (Davenport and Hannahs 2005; Fromkin et al. 2014; Ladefoged and Johnson 2014). Phonologically, tense vowels appear in a wider variety of syllabic contexts. Lax vowels are more limited and are considered unfavorable in open syllables. Phonetically, tense and lax vowels differ in duration

* This work was supported by the Pukyong National University Research Fund in 2023(202315330001).

and spectral properties. In duration-based descriptions, /i/ and /u/ are longer than their lax counterparts. In spectral descriptions, static measurements in the F1–F2 space usually show lax vowels as more centralized relative to tense vowels. These patterns capture strong tendencies in controlled speech styles. However, it remains open how reliably duration and static vowel space identify tense-lax categories in conversational speech. In this context, spectral overlap and temporal reduction can weaken canonical distinctions.

The role of vowel duration in distinguishing tense and lax vowels has been extensively documented in production studies (see Hillenbrand 2013 for review). In controlled speech materials, tense vowels are consistently produced with longer durations than their lax counterparts among speaker groups (Peterson and Lehiste 1960; Hillenbrand et al. 1995; Yang 2008). For example, in CVC syllables, the high front tense vowel /i/ and the high back tense vowel /u/ have mean durations of 207 ms and 235 ms, respectively, whereas the corresponding lax vowels /ɪ/ and /ʊ/ show shorter durations of 161 ms and 163 ms (Peterson and Lehiste 1960). Similar patterns are observed in connected speech elicited through reading tasks. According to van Santen (1992), the intrinsic durations of the high tense vowels /i/ and /u/ are 153 ms and 156 ms; the corresponding lax vowels /ɪ/ and /ʊ/ are shorter, at 111 ms and 127 ms. These studies show a systematic difference in duration between tense and lax vowels in controlled conditions. However, it remains to be investigated whether these durational differences serve as a reliable phonetic correlate of the tense–lax distinction in conversational speech.

Perceptual studies demonstrate mixed results regarding the role of duration in vowel identification. Shortening vowel duration can bias listeners toward lax responses, but the influence varies as a function of the vowel category. Duration interacts with spectral properties, and low vowels are more affected by duration manipulation than non-low vowels. Hillenbrand et al. (2000) examined perceptual consequences of duration manipulation using resynthesized /hVd/ tokens from Hillenbrand et al. (1995). When vowel duration was altered while spectral structure was preserved, vowel identification accuracy decreased modestly for both shortened and lengthened vowels. This suggests duration alone does not serve as a direct cue to vowel identity. Importantly, shortening disproportionately affected certain mid and low vowels, most notably /æ/ and /ɑ/. These vowels were often identified as /ɛ/ and /ʌ/, respectively. In contrast, high vowel pairs like /i-ɪ/ and /u-ʊ/ showed little change in identification under duration manipulation. These findings indicate that the role of duration in vowel identification is not uniform across vowel categories. Instead, it varies with each vowel's acoustic characteristics. High tense–lax pairs display distinctive spectral features, including formant trajectory patterns. This motivates a closer look at the spectral properties that distinguish tense and lax vowels.

Formant frequencies in the F1–F2 vowel space are measured at the static midpoint. This practice has traditionally been used to capture spectral differences between vowel categories. Lax vowels are described as more centralized, while tense vowels are considered more peripheral in the vowel space. For high front vowels, the tense vowel /i/ has lower F1 and higher F2 than /ɪ/. For high back vowels, the tense vowel /u/ shows both lower F1 and lower F2 than /ʊ/ in citation forms like /hVd/ and /kVd/ (Hillenbrand et al. 1995; Leung et al. 2016). Similar patterns are reported in connected speech during reading tasks (Leung et al. 2016; Sandoval et al. 2019). Under these conditions, lax vowels /ɪ/ and /ʊ/, as well as tense vowel /u/, are less peripheral in F2 than in citation forms. The high front tense vowel /i/ remains relatively stable across speech styles (Leung et al. 2016).

Speaking style affects both temporal and spectral properties of vowels. These properties characterize tense and lax vowels in English and are reduced in more conversational contexts. Both vowel duration and formant values are strongly influenced by speaking style. Compared to citation forms, vowel durations in conversational speech are much shorter. They may even be reduced to the point of syllable deletion (Johnson 2004). Spectral features also vary with speech style. In conversational speech, vowel spaces centralize, and formant undershoot is more common. In citation forms or clear speech, vowels are more likely to reach reference targets (Moon and Lindblom 1994). These studies suggest that temporal and spectral distinctions between vowels are weaker in conversational speech. As a result, tense and lax vowels overlap more in the F1–F2 vowel space than in controlled speaking styles.

When static midpoint measurements fail to separate vowel categories clearly, other acoustic cues, such as duration and vowel dynamics, play an important role in distinguishing vowels. In such cases, speakers may rely on temporal properties or formant trajectory across the vowel, rather than on a static measurement. Empirical evidence supports this idea in many sociophonetic and acoustic studies. Labov and Baranowski (2006) report that in the Northern Cities Shift, the short vowels /e/ and

/ɑ/ overlap substantially at the midpoint but remain distinct due to a 50 ms duration difference. This indicates that temporal cues can presesrve vowel contrasts even when spectral separation at a single point is minimal. Similarly, Stanley (2020) shows that ‘thought’ and ‘lot’ in Cowlitz County, Washington, have almost identical midpoint F1 and F2 values. They differ, however, in trajectory patterns, which distinguish vowels despite substantial overlap at the midpoint. Although static measurements effectively capture key spectral properties, duration and vowel dynamics are crucial for realizing vowel contrasts, especially when static midpoint measurements are insufficient.

In vowel classification studies, vowel dynamics, often referred to as Vowel Inherent Spectral Change (VISC), have been shown to enhance vowel identification and classification relative to static midpoint measures (e.g., Nearey and Assmann 1986; Zahorian and Jagharghi 1993; Hillenbrand et al. 1995; Morrison 2013; Hong 2021, 2023; Hwangbo 2025). That is, incorporating dynamic spectral patterns over time generally improves vowel classification compared to relying on a single static measurement in the discriminant analyses. In addition to formant dynamics, including vowel duration further improved classification, particularly for models based on a single static time point. At the same time, the role of duration in vowel classification is inconsistent across studies. In Zahorian and Jagharghi (1993), duration was found to play a minimal role in classifying vowels produced in CVC syllables. Although duration alone yielded performance above chance level, its contribution was limited, which showed only modest improvements when duration was added. On this basis, Zahorian and Jagharghi (1993) suggested that duration may be more informative in natural connected speech than in carefully produced citation forms, highlighting the importance of considering speech style when evaluating the contribution of temporal cues.

Trajectory-based models and duration play compensatory roles in classifying monophthongs. Watson and Harrington (1999) examined Australian English vowels produced in citation forms such as /hV/ and /hVd/ and classified them using a Gaussian classification method. Overall, the results showed that models incorporating formant dynamics together with duration achieved higher classification accuracy than static models. When duration was not included as a predictor, trajectory-based models still classified monophthongs significantly better than static models, especially for specific vowel pairs such as /i, ɪ/ and /a, ʌ/. This pattern suggests that, for these monophthongal contrasts, dynamic spectral information can compensate for the absence of duration. These results indicate that duration and formant dynamics exhibit a compensatory relationship, rather than contributing independently, particularly for monophthongs.

In addition to studies based on citation forms, support for trajectory-based representations has also been reported in analyses of connected speech. Sandoval et al. (2019) examined vowel classification using the TIMIT corpus and an ensemble-based tree classifier, comparing a static midpoint model with trajectory-based models that incorporated multiple temporal sampling points and vowel duration. The results showed that trajectory-based models achieved approximately 10% higher classification accuracy than static models, and that including duration further improved classification performance, indicating that both dynamic spectral information and temporal cues contribute to vowel classification in connected speech. Evidence from spontaneous conversational speech further supports this pattern. Hwangbo (2025) conducted a classification study using the Buckeye Corpus and demonstrated that trajectory-based models, including F0, outperformed static models in classifying vowels. When vowel duration was included as a predictor, classification accuracy increased overall, with substantial gains observed for static models. In addition, the effect of duration was not uniform across vowels, with specific vowel categories, such as /ɪ/ and /u/, showing greater sensitivity to duration than the others. These studies demonstrate that incorporating duration generally improves vowel classification in connected speech, but that its contribution interacts with the availability of dynamic spectral information and varies across vowel categories.

These studies show that trajectory-based representations generally improve vowel identification and classification relative to static midpoint models. In addition to dynamic spectral information, vowel duration has also been shown to facilitate vowel classification, particularly when spectral cues are limited to static points. However, much of the existing work has relied on citation-form materials or carefully controlled speech, with comparatively fewer studies examining how these cues operate in spontaneous conversational speech. In conversational speech, vowels are expected to be less distinctive in terms of static spectral properties, resulting in increased overlap in the F1–F2 vowel space. Under such conditions, vowel contrasts may be maintained through other acoustic cues, such as spectral dynamics and temporal properties. Importantly, duration and formants have been identified as central characteristics distinguishing tense and lax vowels in English. Given that English high tense and lax vowel pairs in conversational speech are likely to exhibit substantial spectral overlap, the present study investigates which acoustic variables, or combinations of variables, most effectively classify these vowel contrasts. Specifically, duration and formant dynamics are expected to play a greater role when spectral differences are insufficient for reliable categorization.

To address these questions, the present study quantifies spectral differences between high tense–lax vowels and evaluates their classification using machine learning models. Spectral overlap and separation between vowel categories were assessed using Bhattacharyya's affinity (Bhattacharyya 1943) and Euclidean distance, calculated at multiple temporal locations within the vowel, specifically at the onset, midpoint, and offset. Together, these measures provide complementary characterizations of acoustic similarity in the F1–F2 space. Vowel classification was conducted using random forest models (Breiman 2001), including both static models based on midpoint spectral information and trajectory models incorporating formant measurements across time, with and without vowel duration as a predictor. By jointly examining acoustic overlap metrics and classification performance, the study evaluates the relative contributions of duration and spectral dynamics to vowel classification. Particular attention is given to whether these cues operate similarly or differently for front and back high tense–lax vowel pairs.

2. Methods

Tense and lax vowels in conversational speech were collected from the Buckeye Corpus of conversational speech (Pitt et al. 2005). The Buckeye Corpus comprises spontaneous speech from 20 male and 20 female speakers in Central Ohio, all middle-class Caucasians. The corpus contains approximately 300,000 tokens with phonemic labels, and the overall transcription agreement is approximately 80%. From the corpus, the high tense and lax vowels, 'iy, ih, uw,' and 'uh', which correspond to the IPA as /i, ɪ, u, ʊ/, were collected. Both stressed and unstressed vowels at word and phrase initial, medial, and final were included. The data for analysis did not control for any consonantal context.

Vowel duration and the spectral values of F1 and F2 of each vowel were collected by Praat (Boersma and Weenink 2024) with a script from Yoon (2021). Raw formant values in Hertz were used without normalization, as emprical evidence indicates that such transformations do not significantly imporve classification accuracy (Hillenbrand and Gayvert 1993; Hillenbrand et al. 1995; Hong 2021). Specifically, comparative analyses of linear frequencies versus non-linear auditory transforms, including Bark, Mel, and Koenig scales, have shown no clear advantage in in vowel classification over raw frequency data (Hillenbrand and Gayvert 1993; Hong 2021). Moreover, Hillenbrand et al. (1995) observed that incorporating vowel duration and spectral dynamics resulted in high identification rates, suggesting that accurate recognition can be achieved without normalization, provided that temporal and dynamic cues are preserved. Spectral values of F1 and F2 of each vowel were measured at 20%, 50%, and 80% of the vowel duration. To assess whether vowel duration differed systematically between tense and lax vowels, duration was analyzed using a linear mixed-effects model implemented in the 'lme4' package (Bates et al. 2015) in R (R Core Team 2024), with vowel as a fixed effect and speaker included as a random intercept.

As front and back vowel categories are highly separable along the F2 dimension, analyses of the two vowel pairs were conducted separately: the front vowel pair includes the tense 'iy' and the lax 'ih', and the back vowel pair includes the tense 'uw' and the lax 'uh'. Acoustic overlap of the tense and lax vowel pairs was quantified using Bhattacharyya's affinity (BA) (Bhattacharyya 1943), computed in R with the 'adehabitatHR' package (Calenge 2024). Euclidean distance (ED) between the members of the vowel pair was also calculated as a complementary measure of acoustic separation. BA quantifies the degree of overlap between probability distributions and is therefore well-suited for assessing vowel overlap in acoustic space (Warren 2018). BA is represented by numbers between 0 and 1, where 1 indicates total overlap and 0 indicates a distinction between the distributions. In the present study, BA and ED were measured at 20%, 50%, and 80% of vowel duration to examine how acoustic overlap evolves over time. Both BA and ED were evaluated using both raw and normalized scales. The overall pattern for BA and ED remained highly consistent across these analytical methods. Variations in BA values were minimal and did not significantly influence the overall patterns and trends observed. Similarly, ED demonstrated the same pattern, with differences mainly due to scaling. Therefore, the results are presented in raw scales.

Random forest classification was employed to evaluate how effectively spectral and temporal cues support vowel classification under varying degrees of overlap. In particular, classification analyses were used to assess the relative contributions of vowel duration and trajectory information to predictive accuracy when spectral overlap between vowel categories is high. For classification, data was then divided into training and test set for random forest classifier using the 'randomForest' package (Liaw and Wiener 2002) in R. Random forest is an ensemble learning method that constructs multiple decision trees, each trained on a bootstrap sample of the data and split using a random subset of features or predictors, which derive its robust model accuracy by aggregating the predictions from all individual trees (Breiman 2001; Gries 2020, 2021; Bernaisch 2022). As such, random forests incorporate randomness at both the observation and predictor levels (Gries 2021). For each model, classification performance is internally estimated using the out-of-bag (OOB) error, defined as the average prediction error across trees trained on the training data (Bernaisch 2022). Model performance was then further assessed by applying the trained classifier to a test set consisting of previously unseen data.

The dataset was partitioned into training and test sets at 70% and 30% of the total tokens, respectively, using the 'caret' package (Kuhn 2008) in R. Table 1 summarizes the number of tokens for each vowel. As the tokens of 'iy' and 'ih' show a large gap, the number of tokens for the vowel 'iy' was upsampled to match the number of 'ih' in the training set to mitigate any bias during the model training. As a result, the front vowel pair in the training set included 26,396 tokens (13,198 tokens for each vowel). In contrast, the test set retained the original proportion to reflect the real-world proportion of the vowel occurrences. The tokens of the back vowel pair exhibit comparable token counts; therefore, no upsampling was applied to the training set, and the test set retained its original proportion. Each model was trained with 500 trees (ntree=500) with 2 variables at each split (mtry=1 for the static without duration model, otherwise mtry=2).

Table 1. The number of vowel tokens in the training and test sets

	iy [i]	ih [ɪ]	uw [u]	uh [ʊ]
Training	4936	13198	1750	1691
Test	2115	5655	749	724
Total	7051	18853	2499	2415

Four random forests were built based on the spectral features and duration. The models based on spectral features were the static and trajectory models. The static model included formants measured at 50%, the midpoint, of the vowel duration, while the trajectory model included formants measured at 20%, 50%, and 80% of the vowel duration. These models either included duration or excluded it, yielding a total of four models, as outlined in Table 2. F1 and F2 indicate the first two formants, and the following numbers after the under bar, 20, 50, and 80, indicate the time point at which the formants were measured. Vowel duration is represented as vowelDur.

Table 2. Random forest models and the predictors of each model

Models	NoDur	Dur
Static	F1_50, F2_50	F1_50, F2_50, vowelDur
Trajectory	F1_20, F2_20, F1_50, F2_50, F1_80, F2_80	F1_20, F2_20, F1_50, F2_50, F1_80, F2_80, vowelDur

In addition to evaluating model performance using out-of-bag (OOB) error rates and predictive accuracy, variable importance analyses were conducted. These analyses were performed to determine the relative importance ranking of predictors within the training data. The variable importance scores reported in this study are based on mean decrease accuracy, which quantifies the extent to which classification accuracy declines when the values of a given predictor are permuted. It is important to note that variable importance rankings in random forest models do not imply causal relevance (Gries 2020, 2021). Instead, these scores reflect the relative contribution of each predictor to the model's classification accuracy: higher importance values indicate a greater contribution to predictive performance. Thus, interpretation should focus on the relative ranking of predictors rather than on their absolute importance values (Strobl et al. 2009). Based on these importance rankings, follow-up random forest models were developed to assess and validate the predictive accuracy associated with the most informative predictors.

3. Results

3.1. Front vowel pair

Formant measurements of the front vowel pair are presented in Table 3 for female and male speakers separately. The data show that the tense vowel 'iy' exhibits lower F1 and higher F2 values than the lax vowel 'ih', consistent with previous findings. Figure 1 illustrates the median formant trajectories of the two vowels in the F1–F2 space, with measurements taken at 20%, 50%, and 80% of the vowel duration. The thick solid lines represent median trajectories across all speakers. Dotted and dashed lines indicate female and male medians, respectively; the label marks the onset (20%), the dot marks the midpoint (50%), and the arrowhead marks the offset (80%) of the vowel duration.

The two vowels display distinct trajectory patterns. The tense vowel 'iy' shows a forward movement along the F2 dimension, corresponding to increased frontness, accompanied by a slight upward movement along F1, corresponding to increased vowel height. In contrast, the lax vowel 'ih' exhibits a shorter trajectory characterized primarily by movement along the F1 dimension, indicating changes in vowel height, with minimal change along F2. Although absolute formant values differ by gender, the overall direction and shape of the trajectories are consistent across male and female speakers.

Table 3. Median values of the first two formants at 20%, 50%, and 80% of the vowel duration for the front vowel pair

Vowels	Gender	F1			F2		
		20%	50%	80%	20%	50%	80%
iy	Female	420	414	407	2252	2385	2328
	Male	358	351	343	1868	1969	1930
ih	Female	467	482	458	1922	1938	1912
	Male	400	410	391	1648	1676	1647

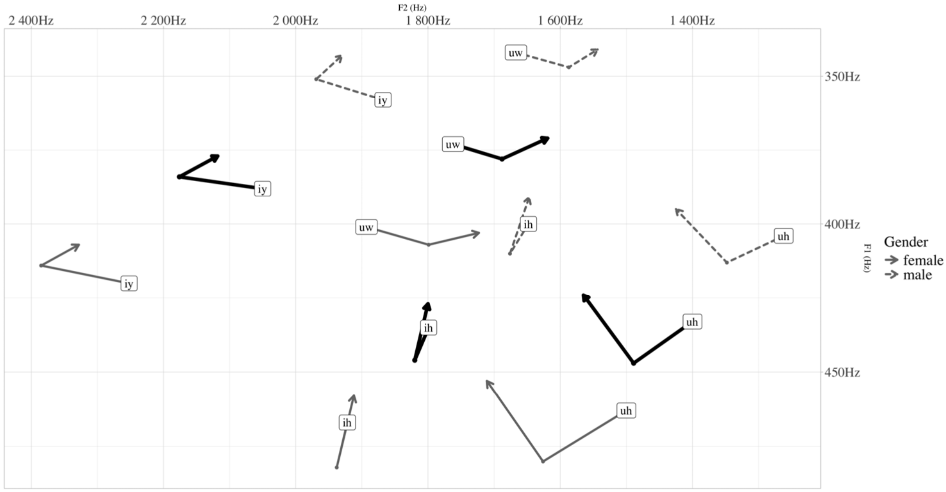

Figure 1. Median formant trajectories segmented by gender. The thick solid line represents the combined median for both genders. Labels, points, and arrowheads indicate the onset (20%), midpoint (50%), and offset (80%) of vowel duration.

Table 4 summarizes the mean, median, and standard deviation of vowel duration for the front vowel pair. The tense vowel ‘iy’ is longer than the lax vowel ‘ih’. Because the data are drawn from conversational speech, overall vowel durations are shorter than those reported in studies based on citation forms; nevertheless, the pattern of longer tense vowels relative to lax vowels is preserved. To assess whether this difference was statistically significant, vowel duration was analyzed using a linear mixed-effects model with vowel category as a fixed effect and speaker as a random intercept. Results show that the lax vowel ‘ih’ was significantly shorter than the tense vowel ‘iy’, with an estimated difference of 26.79 ms (β=−26.79, SE=0.48, t=−55.28).

Table 4. Mean, median, and standard deviation (SD) of vowel duration for the front vowel pair

Vowel	Mean	SD	Median
iy	90.4	43.6	81
ih	62.9	32.4	57

Table 5 presents BA and ED values for the front vowel pair at each measurement point. BA indicates that acoustic overlap between ‘iy’ and ‘ih’ is highest at the onset (20%), lowest at the midpoint (50%), and increases again toward the offset (80%).[1]

ED shows a complementary pattern, with the greatest separation between the two vowels at the midpoint. These distributional patterns are further illustrated in Figure 2, which visualizes the density distributions of 'iy' and 'ih' in the F1–F2 space at each time point. At 20% of the vowel duration, the density contours of the two vowels display substantial overlap. By 50%, the density peaks become more distinctly separated. At 80%, the distributions exhibit increased overlap once again. The overlap metrics and density visualizations indicate that the front tense–lax contrast is most distinctive at the midpoint.

Table 5. Bhattacharyya's affinity (BA) and Euclidean distance (ED) values at onset, midpoint, and offset of vowel duration for the front vowel pair

	20%	50%	80%
BA	0.907	0.818	0.870
ED	264.04	350.09	313.93

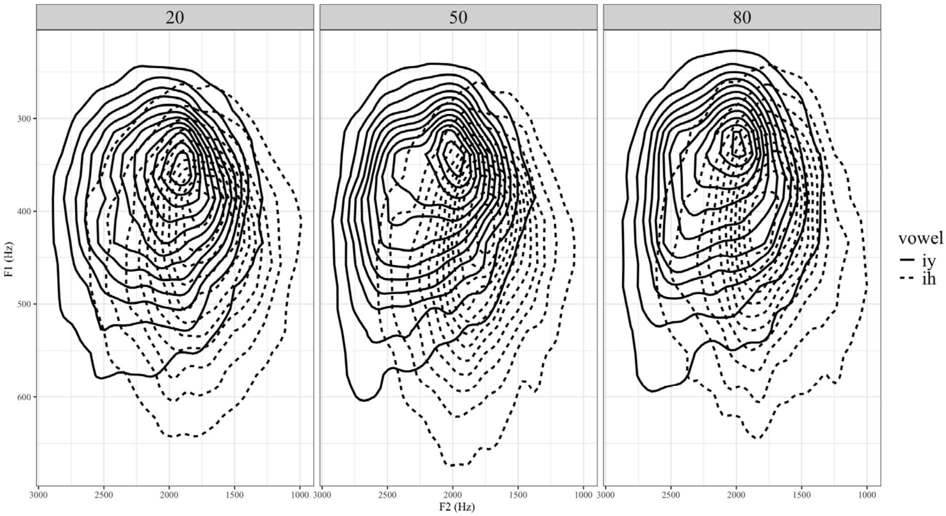

Figure 2. Overlap of the front vowel pair, represented with a density plot

1 Although BA is high for both front and back vowel pairs, indicating significant overlap, the trajectory in Figure 1 shows that the vowel pairs are separated. BA is based on vowel distribution, while the trajectory pattern is derived from the median of the vowels. Therefore, it is possible for vowel pairs to exhibit high overlap in BA while their trajectories are separated.

Random forest models were constructed using the predictors listed in Table 2. Separate models were built for the front and back vowel pairs, using two model types, static and trajectory, each evaluated with and without vowel duration. OOB error rates and test-set prediction accuracy for the front vowel pair are reported in Table 6. All models achieved performance above chance level.

Table 6. Results of random forests for the front vowel pair

Models	OOB		Prediction Accuracy	
	NoDur	Dur	NoDur	Dur
Static	11.55	7.5	76.54	82.02
Trajectory	6.69	5.32	82.69	84.83

Random forest classification results for the front vowel pair are summarized in Table 6. Overall, models incorporating spectral trajectory information consistently outperformed static models. In the NoDur condition, adding trajectory information reduced the OOB error rate by 4.86 percentage points, from 11.55% to 6.69%, and increased prediction accuracy by 6.15 percentage points, from 76.54% to 82.69%. In the Dur condition, trajectory information yielded smaller but still positive gains, reducing the OOB error rate by 2.18 percentage points, from 7.50% to 5.32%, and increasing test accuracy by 2.81 percentage points, from 82.02% to 84.83%.

Including vowel duration as a predictor consistently improved classification performance across model types. In the static model, adding duration reduced the OOB error rate from 11.55% to 7.50%, corresponding to a 4.05 percentage-point decrease, and increased test prediction accuracy from 76.54% to 82.02%, an improvement of 5.48 percentage points. In the trajectory model, adding duration reduced the OOB error rate from 6.69% to 5.32%, a 1.37 percentage-point decrease, and increased test accuracy from 82.69% to 84.83%, a 2.14 percentage-point increase. Notably, the static model with duration achieved test prediction accuracy comparable to that of the trajectory model without duration (82.02% vs. 82.69%), indicating that duration and spectral trajectory information produce comparable improvements in classification accuracy for the front tense–lax vowel pair.

Variable importance scores were extracted from the trajectory model with duration to assess the relative contribution of individual predictors to classification performance. Figure 3 presents the variable importance rankings for the trajectory model with duration. Vowel duration emerged as the highest-ranked predictor, indicating that temporal information plays a prominent role in distinguishing the tense and lax vowels. The following most informative predictors were F1_50 and F2_50, corresponding to spectral information at the vowel midpoint. Notably, midpoint formant measures ranked higher than formant values at earlier or later portions of the vowel, consistent with the overlap analyses showing reduced acoustic overlap at the midpoint. These rankings indicate that classification performance for the front vowel pair relies most strongly on a combination of duration and midpoint spectral information, while dynamic spectral measures at non-midpoint time points contribute less to predictive accuracy in the presence of these cues.

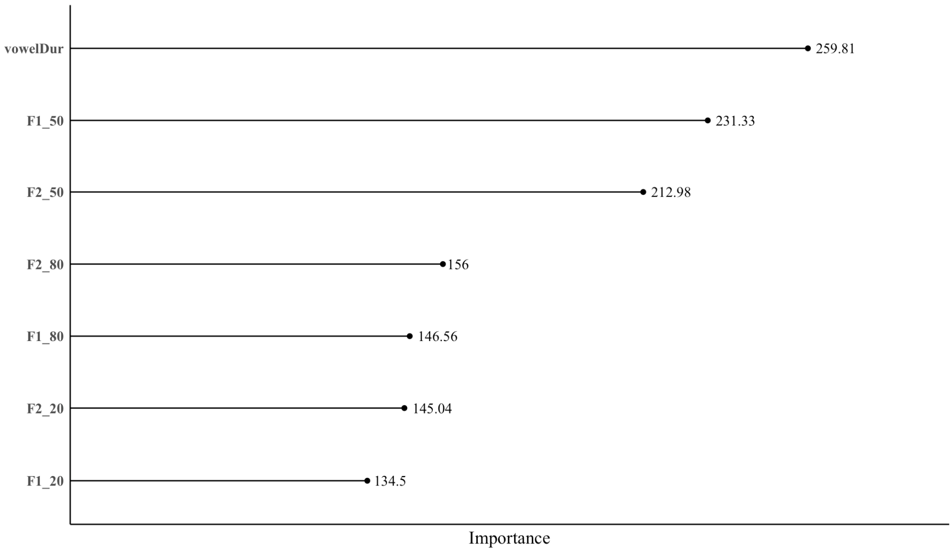

Figure 3. Variable importance scores of the front vowel pair

To further examine how these top-ranked predictors contribute to classification, additional random forest models were constructed by incrementally adding predictors in order of importance, as shown in Table 7. When vowel duration was used as the sole predictor, classification performance was below chance level, with an OOB error rate of 32.47% and a test prediction accuracy of 66.82%. Adding F1_50 as a second predictor improved performance, reducing the OOB error rate to 19.60% and increasing test accuracy to 72.47%, but classification accuracy remained below chance level. When F2_50 was added as a third predictor, classification performance improved substantially, with the OOB error rate decreasing to 7.43% and test prediction accuracy increasing to 81.99%, which exceeded chance level.

These results show that although vowel duration is the most informative single predictor according to variable importance rankings, duration alone is insufficient for classifying the front tense–lax contrast. Reliable classification of ‘iy’ and ‘ih’ emerges only when temporal information is combined with midpoint spectral cues. Notably, this combination of predictors corresponds precisely to the static model with duration, which yielded classification performance comparable to that of the trajectory model without duration.

Table 7. Random forest classification results for the front vowel pair using variable importance rank; asterisks denote statistical significance

	OOB	Prediction Accuracy
vowelDur	32.47%	66.82%
vowelDur+F1_50	19.6%	72.47%
vowelDur+F1_50+F2_50	7.43%	81.99%*

Overall, the results indicate that a static model incorporating vowel duration and midpoint spectral information provides an efficient and effective classification of the front vowel pair. The convergence of multiple analyses supports this conclusion: the overlap measures showed the least acoustic overlap at the vowel midpoint, the variable importance rankings identified vowel duration and midpoint formant values as the most informative predictors, and the incremental random forest analyses demonstrated that reliable classification emerged only when these predictors were combined. Although vowel duration was ranked as the most crucial individual predictor, its predictive value was limited when considered in isolation and increased substantially only when spectral information at the midpoint was present. These results underscore that variable importance reflects the relative predictive contribution within the model rather than causal dominance, and that successful classification of the front tense–lax contrast depends on the joint availability of temporal and spectral cues rather than on any single acoustic dimension.

3.2. Back vowel pair

The formant measurements of the back vowel pair reveal trajectory patterns that differ markedly from those of the front vowel pair. F1 of the tense vowel ‘uw’ remains relatively stable across the vowel duration, whereas the lax vowel ‘uh’ exhibits an increase in F1 toward the midpoint, followed by a decrease toward the offset, indicating greater variation along the height dimension. F2 demonstrates a distinct divergence between the two vowels. F2 of ‘uw’ decreases steadily over time, while F2 of ‘uh’ increases correspondingly, illustrating that the two vowels move in opposite directions along the front–back dimension. These opposing F2 trajectories result in increasing frontness for ‘uh’ and increasing backness for ‘uw’ over the course of the vowel. The trajectories are visualized in Figure 1, which shows that these directional patterns are consistent across male and female speakers.

Table 8. Median values of the first two formants at 20%, 50%, and 80% of vowel duration for the back vowel pair

Vowels	Gender	F1			F2		
		20%	50%	80%	20%	50%	80%
uw	Female	401	407	403	1893	1799	1723
	Male	342	347	341	1667	1587	1544
uh	Female	463	480	453	1500	1626	1711
	Male	404	413	395	1262	1348	1425

Table 9 summarizes the mean, median, and standard deviation of vowel duration for the back vowel pair. The tense vowel ‘uw’ is longer than the lax vowel ‘uh’, aligning with anticipated patterns. A linear mixed-effects model confirmed that ‘uh’ was significantly shorter than ‘uw’, with an estimated difference of 28.07 ms ($\beta=-28.07$, $SE=1.25$, $t=-22.56$), indicating a robust duration contrast between the two vowels.

Table 9. Mean, median, and standard deviation (SD) of vowel duration for the back vowel pair

Vowel	Mean	SD	Median
uw	89.7	51.1	79
uh	63.2	35.4	55

Table 10 presents BA and ED values for the back vowel pair at each measurement point. Generally, the back vowels exhibit significant acoustic overlap; however, the temporal pattern of the overlap diverges from that observed in the front vowel pair. BA is minimal at the onset and increases consistently toward the offset. At the same time, ED decreases substantially over time. These measurements indicate that the two vowels progressively become closer in acoustic space as the vowel unfolds. These patterns are visualized in Figure 4, which shows the density distributions of ‘uw’ and ‘uh’ in the F1–F2 space at each time point. At the onset, the distributions are more clearly separated, particularly along the F2 dimension. Toward the midpoint and offset, the distributions show increasing overlap, reflecting convergence of the two vowels in acoustic space. Together, the overlap metrics and density plots indicate that the back tense–lax contrast is most distinct at the onset and least distinct toward the offset.

Table 10. Bhattacharyya’s affinity (BA) and Euclidean distance (ED) at onset, midpoint, and offset of vowel duration for the back vowel pair

	20%	50%	80%
BA	0.881	0.897	0.938
ED	317.23	174.47	59.24

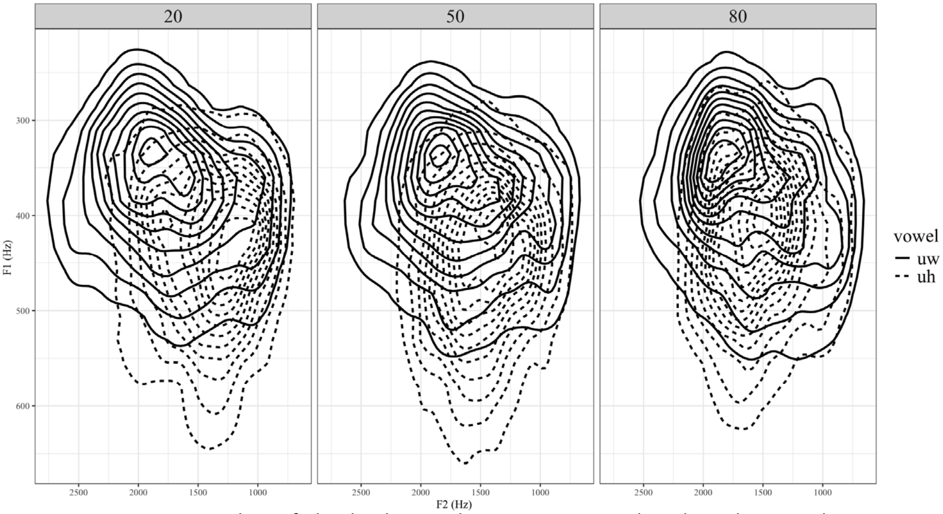

Figure 4. Overlap of the back vowel pair, represented with a density plot

Random forest classification results for the back vowel pair are summarized in Table 11. Overall, trajectory models outperformed static models in both OOB error rates and test prediction accuracy. In the NoDur condition, adding trajectory information reduced the OOB error rate from 31.94% to 21.30%, a decrease of 10.64 percentage points, and increased test prediction accuracy from 67.41% to 80.11%, an improvement of 12.70 percentage points. In the Dur condition, trajectory information reduced the OOB error rate from 24.41% to 18.45%, a decrease of 5.96 percentage points, and increased test accuracy from 76.85% to 81.67%, an improvement of 4.82 percentage points. Adding vowel duration also improved classification performance, with larger effects observed in the static model than in the trajectory model. In the static model, adding duration reduced the OOB error rate from 31.94% to 24.41%, a decrease of 7.53 percentage points, and increased test prediction accuracy from 67.41% to 76.85%, an improvement of 9.44 percentage points. In the trajectory model, duration yielded more minor improvements, reducing the OOB error rate from 21.3% to 18.45%, a decrease of 2.85 percentage points, and increasing test prediction accuracy from 80.11% to 81.67%, an improvement of 1.56 percentage points. These results show that vowel duration has a greater effect on classification when spectral trajectory information is absent; however, trajectory information yields larger improvements than duration for the back vowel pair.

Table 11. Results of random forests for the back vowel pair

Models	OOB		Prediction Accuracy	
	NoDur	Dur	NoDur	Dur
Static	31.94	24.41	67.41	76.85
Trajectory	21.3	18.45	80.11	81.67

Variable importance scores for the back vowel pair are illustrated in Figure 5. The rankings reveal a pattern distinct from that observed for the front vowel pair. The most important predictor is F2_20, followed by F1_50 and vowel duration, with only a slight difference in importance between the latter two predictors.

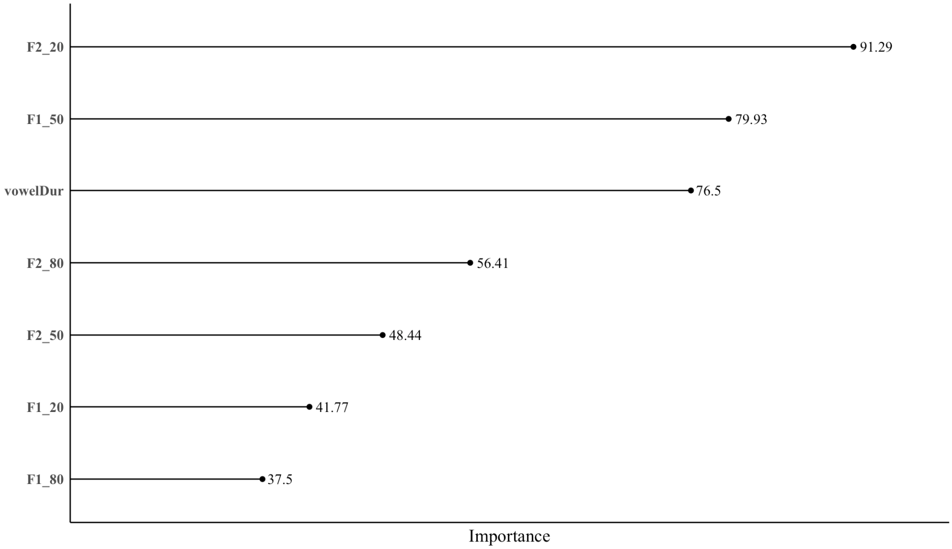

Figure 5. Variable importance scores of the back vowel pair

To further examine how these predictors contribute to classification, additional random forest models were constructed by incrementally adding predictors in order of importance, as shown in Table 12. When F2_20 was used as the sole predictor, classification performance was limited, with an OOB error rate of 41.64% and a test prediction accuracy of 60.96%, which exceeded chance-level performance. Adding F1_50 as a second predictor reduced the OOB error rate by 9.93 percentage points and increased test prediction accuracy by 8.42 percentage points, yielding performance slightly higher than that of the static model without duration. When vowel duration was added as a third predictor, the OOB error rate decreased by a further 8.84 percentage points, and test accuracy increased by 7.81 percentage points, resulting in performance comparable to that of the static model with duration.

Table 12. Random forest classification results for the back vowel pair using variable importance rank; asterisks denote statistical significance

	OOB	Prediction Accuracy
F2_20	41.64%	60.96%*
F2_20+F1_50	31.71%	69.38%*
F2_20+F1_50+vowelDur	22.87%	77.19%*

The results indicate that classification of the back vowel pair relies on spectral information drawn from different temporal portions of the vowel than is the case for the front vowel pair. In particular, F2 at the onset emerges as the most informative predictor, consistent with the overlap measures, which indicate relatively lower overlap at the onset and increasing overlap toward later portions of the vowel. The trajectory analyses, including the overlap measures, show that 'uw' and 'uh' diverge early along the front–back dimension and then converge toward the offset. Notably, the most informative spectral predictors for the back vowel pair are not temporally aligned. That is, F2 at the onset and F1 at the midpoint contribute distinct predictive information. This suggests that classification of 'uw' and 'uh' draws on spectral cues distributed across the vowel rather than concentrated at a single time point.

However, from a perceptual perspective, it is arbitrary to utilize F1 and F2 at different time points. Therefore, a follow-up analysis was conducted using F2_20+F1_20, both with and without duration. Without duration, the OOB error was 32.96%, with a test prediction accuracy of 68.09%. When duration was included, the OOB error decreased to 23.8%, and the prediction accuracy increased to 76.24%. Both sets of results show minimal deviation from those presented in Table 11 and 12. Comparing the classification results of the onset to the static model suggests that spectral cues at the onset, specifically F2, are as informative as those at the midpoint for classifying the back vowel pair.

The results concerning the back vowel pair suggest that although vowel duration enhances classification performance, its impact is most significant in the absence of spectral trajectory information. Once trajectory information is available, duration contributes smaller gains than spectral cues. Among the trajectory information, these results suggest that dynamic spectral information, particularly early F2, plays a central role in classifying the back tense–lax contrast, with duration serving a secondary, supportive role.

4. Discussion

This paper aimed to determine which acoustic features best distinguish high tense and lax vowels in conversational speech, with a focus on spectral characteristics and vowel duration within a random forest classification approach. The results demonstrate that, for both the front and back vowel pairs, the trajectory models outperform the static models in classification accuracy. Furthermore, duration enhances classification performance when combined with spectral information, both static and dynamic. Importantly, the roles of duration and trajectory differ between the front and back vowel pairs. Specifically, for the front vowel pair, the static model with duration most efficiently classifies the tense and lax vowels, whereas for the back vowel pair, trajectory information, particularly the onset of F2, plays a crucial role.

For the high front tense–lax vowel pair, the classification results indicate that static midpoint with duration performs to a level comparable to that of a trajectory-based model without duration. The comparable performance of the static model with duration and the trajectory model without duration suggests that the relative contribution of duration increases when dynamic spectral information is absent, indicating a compensatory relationship between duration and spectral dynamics. This interpretation is further supported by the fact that adding duration to the trajectory model yields smaller gains. In other words, either duration combined with static midpoint spectral information or explicit spectral trajectories can support near-optimal classification performance for the front vowel pair.

The variable importance analysis reinforces this pattern by ranking duration as the most informative predictor for the front vowels. This aligns with prior research by Hwangbo (2025), indicating that duration significantly influences the classification accuracy of the lax vowel, /ɪ/. Duration and spectral information at the midpoint are necessary to reach ceiling-level performance. In other words, duration enhances classification the most when combined with spectral cues, particularly those at the vowel midpoint. This finding is clarified through the overlap analyses, which demonstrate that although the front vowel pair exhibits substantial overlap throughout the vowel duration, overlap is minimized at the midpoint. That is, midpoint spectral information provides the most informative static representation of the front tense–lax vowel contrast. Although the midpoint shows minimal acoustic overlap between the front vowel pair, the overlap remains relatively high. Therefore, to compensate for spectral overlap, vowel duration enhances classification accuracy primarily when it is combined with the spectral cues at the midpoint. Thus, either trajectory information or duration is crucial for classification prediction accuracy for the front tense–lax vowel pair.

For the high back tense–lax vowel pair, a different pattern emerges. Although adding duration substantially improves classification accuracy in static models, incorporating spectral trajectory information yields even larger improvements, leading trajectory models to outperform static models across duration conditions. This outcome reflects the distinctive dynamic properties of the back vowels, where the tense vowel 'uw' and the lax vowel 'uh' move in opposite directions along the F2 dimension over time. The vowel dynamics contribute largely because the directionality of the vowels is opposite, with the tense vowel 'uw' moving backward and the lax vowel 'uh' moving forward along the F2 dimension. This result also shows that the temporal location of maximal spectral contrast is critical for cue efficiency. The overlap analysis confirms this interpretation by showing that acoustic overlap between the two vowels increases toward the offset, indicating that the vowels are most distinct early in the vowel. As a result, spectral information captured at the onset is particularly informative for classification. Because static models rely on midpoint spectral information, which corresponds to a region of relatively greater overlap for the back vowels, duration plays a larger role in facilitating classification in these models. However, once trajectory information is available, particularly information about early F2 movement, duration contributes relatively minor gains. This pattern indicates that although duration supports classification when spectral cues are limited, dynamic spectral information is the dominant contributor to classification accuracy for the back vowel pair. The variable importance analysis further supports this interpretation by identifying F2 at the onset as the most informative predictor for the back vowel pair. A noticeable spectral separation at the onset diminishes the relative importance of duration, thereby improving classification when static midpoint cues are employed.

Comparing the high front and high back tense–lax vowel pairs reveals that although both pairs exhibit substantial acoustic overlap in conversational speech, the nature of this overlap and the resulting cue efficiency differ systematically across vowel categories. In both cases, high overlap limits the effectiveness of static spectral information alone, necessitating additional cues, such as vowel duration or spectral dynamics, for reliable classification. However, the relative contribution of these cues is conditioned by how and when spectral distinctions are realized over the course of the vowel.

For the front vowel pair, overlap in the F1–F2 space is high throughout the vowel duration but reaches a minimum at the midpoint, where both F1 and F2 are maximally separated. This temporal concentration of spectral contrast makes midpoint spectral information particularly informative in a static representation. Under these conditions, vowel duration plays a prominent role in classification by compensating for residual spectral overlap, especially when dynamic spectral information is not explicitly represented. As a result, the static model, which combines midpoint spectral cues with duration, achieves classification performance comparable to that of a trajectory model without duration. Importantly, while duration is ranked as the most informative predictor for the front vowel pair, it is insufficient on its own and must be combined with informative spectral cues at the midpoint to support robust classification. In contrast, the back vowel pair exhibits a different temporal organization of spectral contrast. Although overall overlap remains high, the tense and lax back vowels diverge early along the F2 dimension and then converge toward the offset, resulting in minimal overlap at the onset and increasing overlap over time. This pattern makes early spectral information, particularly F2 at the onset, highly informative for classification. Consequently, the trajectory models substantially outperform the static models for the back vowel pair, even in the absence of duration. While vowel duration improves classification when only static midpoint information is available, its contribution becomes secondary once dynamic spectral cues are included. The variable importance analysis reinforces this pattern by identifying F2 at the onset and F1 at the midpoint as the most informative predictors, reflecting maximal separation.

In summary, these findings suggest that cue effectiveness in tense–lax vowel classification varies across vowel categories and is influenced by the temporal distribution and directionality of spectral contrast. For the front vowel pair, where spectral separation is temporally localized at the midpoint and trajectory directionality is limited, duration functions as an effective compensatory cue when dynamic information is absent. For the back vowel pair, where trajectory directionality creates early spectral divergence, dynamic spectral information dominates classification, and duration plays a secondary role. Thus, although both duration and spectral dynamics contribute to classification under conditions of high overlap, their relative importance is shaped by vowel-specific patterns of spectral change over time. Tense and lax vowels are often described as differing in duration and centrality within the vowel space, and the present findings support this general characterization while refining it in meaningful ways. Across both front and back vowel pairs, static spectral information alone is insufficient for robust classification in conversational speech. Instead, either vowel duration or spectral trajectory information is required to achieve high classification accuracy. For the front vowel pair, duration and spectral dynamics play complementary roles, whereas for the back vowel pair, trajectory directionality, particularly early front–back movement, provides the primary basis for classification. This contrast highlights that tense–lax distinctions are not realized through a uniform set of acoustic cues across vowel categories. Rather, cue efficiency reflects vowel-specific patterns of spectral organization over time, with different vowels exploiting different temporal regions to maintain contrast.

5. Conclusion

The purpose of this study is to identify the acoustic predictors that most effectively classify high tense and lax vowels in conversational speech, focusing on spectral properties and vowel duration within a random forest classification framework. Because conversational speech is characterized by substantial acoustic overlap between vowel categories, spectral distinction in the vowel space alone is often insufficient for reliable classification, necessitating consideration of additional cues, such as vowel duration, spectral dynamics, or a combination of both. The results demonstrate that although duration is a robust feature distinguishing tense and lax vowels, its role in classification differs systematically between the high front and high back tense–lax vowel pairs. Notably, the results show that these vowel pairs rely on different combinations of acoustic predictors for optimal classification. Specifically, a static model incorporating midpoint spectral information and duration provides the most economical classification for the high front tense–lax pair, whereas dynamic spectral information, particularly F2 at the onset, is crucial for classifying the high back tense–lax pair.

Despite its contributions, this study is subject to several limitations. First, the analysis is confined to high tense–lax vowel pairs, which constrains the generalizability of the findings. High vowels provide a clear context for examining spectral overlap and cue efficiency in conversational speech. However, other English tense–lax contrasts, such as mid and low vowels, may exhibit different patterns of spectral overlap and distinct temporal distributions of spectral contrasts. Accordingly, the cue efficiency patterns identified here should not be assumed to generalize to the entier English vowel system without further investigation. A second limitation concerns F3, which is closely associated with lip rounding. Although the high back vowel pair examined in this study involves lip rounding, and thus affects apparent vocal tract length, the role of F3 in classifying this back high tense–lax vowel pair was not tested. Future research should include F3 as a predictor to more fully assess its contribution to the classification of the back high tense–lax contrast.

In conclusion, this study demonstrates that cue efficiency in classifying high tense–lax vowels in conversational speech is influenced not only by the extent of spectral overlap but also by the temporal distribution and directionality of spectral change. Moreover, these factors differ systematically across vowel categories. While tense and lax vowels have traditionally been characterized as being distinguished by duration and centrality in the vowel space, the present results show that these cues are not uniformly informative across all high tense–lax vowels. For high front vowels, duration plays a prominent role in classification when dynamic spectral information is absent, but only in combination with informative midpoint spectral cues. In contrast, for high back vowels, dynamic spectral information, particularly early front–back movement reflected in F2 at the onset, provides the most efficient basis for classification, with duration serving a more limited, secondary role. The present study implies that the relative importance of duration and spectral properties is vowel-specific rather than universal.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

Bernaisch, Tobias. 2022. Comparing generalised linear mixed-effects models, generalised linear mixed-effects model trees and random forests: Filled and unfilled pauses in varieties of English. In Ole Schützler and Julia Schlüter (eds.), *Data and methods in corpus linguistics*, 163-193. Cambridge: Cambridge University Press.

Bhattacharyya, Anil. 1943. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society* 35: 99-109.

Boersma, Paul and David Weenink. 2024. Praat: Doing phonetics by computer.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1): 5-32.

Calenge, Clement. 2024. adehabitatHR: Home range estimation.

Davenport, Michael and Stephen J. Hannahs. 2005. *Introducing phonetics & phonology 2nd edn*. London: Hodder Arnold.

DiCanio, Christian, Hosung Nam, Jonathan D. Amith, Rey Castillo García, and D. H. Whalen. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics* 48: 45-59.

Fromkin, Victoria, Robert Rodman, and Nina Hyams. 2014. *An introduction to language 10th edn.* Melbourne: Cengage Learning.

Gries, Stefan Th. 2020. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. C*orpus Linguistics and Linguistic Theory* 16(3): 617-647.

Gries, Stefan Th. 2021. *Statistics for linguistics with R: A practical introduction. Mouton Textbook 3rd edn*. Berlin: De Gruyter Mouton.

Hillenbrand, James M. and Robert T. Gayvert. 1993. Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech, Language, and Hearing Research* 36(4): 694-700.

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

Hillenbrand, James M., Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of Acoustics Society of America* 97(5): 3099-3111.

Hong, Soonhyun. 2021. Roles of temporal patterns of vowel-intrinsic cues in model identification of Korean vowels in spontaneous speech. *Studies in Phonetics, Phonology, and Morphology* 27(2): 321-351.

Hong, Soonhyun. 2023. Roles of dynamic patterns of lower formants, vowel identity, and gender in predicting postvocalic consonant place in Korean spontaneous speech. *Studies in Phonetics, Phonology, and Morphology* 29(2): 211-246.

Hwangbo, Hyun Jin. 2025. Dynamic cues in vowel classification: A discriminant analysis of conversational speech corpus. *Korea Journal of English Language and Linguistics* 25: 289-310.

Jacewicz, Ewa and Robert Allen Fox. 2012. The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *The Journal of the Acoustical Society of America* 131(2): 1413-1433.

Johnson, Keith. 2004. Massive reduction in conversational American English. In Kiyoko Yoneyama and Kikuo Maekawa (eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*. 29-54. Tokyo, Japan: The National International Institute for Japanese Language.

Kuhn, Max. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5): 1-26.

Labov, William and Maciej Baranowski. 2006. 50 msec. *Language Variation and Change* 18(3): 223-240.

Ladefoged, Peter and Keith Johnson. 2014. *A course in phonetics 7th edn.* Stamford, CONN: Cengage Learning.

Leung, Keith K. W., Allard Jongman, Yue Wang, and Joan A. Sereno. 2016. Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America* 140(1): 45-58.

Liaw, Andy and Matthew Wiener. 2002. Classification and regression by randomForest. *R News* 2(3): 18-22.

Moon, Seung-Jae and Björn Lindblom. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America* 96(1): 40-55.

Morrison, Geoffrey Stewart. 2013. Theories of vowel inherent spectral change. In Geoffrey Stewart Morrison and Peter F. Assmann (eds.), *Vowel inherent spectral change*, 31-48. Berlin: Springer.

Nearey, Terrance M. and Peter F. Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America* 80(5): 1297-1308.

Neel, Amy T. 2008. Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research* 51(3): 574-585.

Peterson, Gordon E. and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6): 693-703.

Pitt, Mark A., Keith Johnson, Elizabeth Hume, Scott Kiesling, and William Raymond. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45(1): 89-95.

R Core Team. 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Sandoval, Steven, Rene L. Utianski, and Heike Lehnert-LeHouillier. 2019. Average formant trajectories. *Perspectives of the ASHA Special Interest Groups* 4(4): 719-732.

Stanley, Joseph A. 2020. *Vowel dynamics of the elsewhere shift: A sociophonetic analysis of English in Cowlitz county, Washington.* PhD Dissertation. University of Georgia.

Strobl, Carolin, James Malley, and Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4): 323-348.

van Santen, Jan P. H. 1992. Contextual effects on vowel duration. *Speech Communication* 11(6): 513-546.

Warren, Paul. 2018. Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association* 48(3): 305-330.

Watson, Catherine I. and Jonathan Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America* 106(1): 458-468.

Yang, Byung-Gon. 2008. An acoustical comparison of English tense and lax vowels produced

by Korean and American males. *Speech Sciences* 15(4): 19-27.
Yoon, Kyuchul. 2021. *Praat & scripting*. Daegu: Book Kyul.
Zahorian, Stephen A. and Amir Jalali Jagharghi. 1993. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America* 94(4): 1966-1982.

Hyun Jin Hwangbo
Assistant Professor
Division of English Language and Literature
Pukyong National University
45 Yongso-ro, Nam-gu,
Busan, 48513, Korea
E-mail: hjhwangbo@pknu.ac.kr

Received: 2026. 02. 13.
Revised: 2026. 03. 18.
Accepted: 2026. 05. 01.