Dynamic Specification of Vowels in Hijazi Arabic

Wael Almurashi,[1] Jalal Al-Tamimi,[2] and Ghada Khattab[3]

[1]*Department of Languages and Translations, Taibah University, Madinah 344, Saudi Arabia*

[2]*Université de Paris, CNRS, Laboratoire de linguistique formelle, F-75013 Paris, France*

[3]*Department of Speech and Language Sciences, School of Education, Communication and Language Sciences, Newcastle University, Newcastle NE1 7RU, United Kingdom*

ABSTRACT

Research on various languages shows that dynamic approaches to vowel acoustics—in particular Vowel-Inherent Spectral Change (VISC)—can play a vital role in characterising and classifying monophthongal vowels compared with a static model. This study's aim was to investigate whether dynamic cues also allow for better description and classification of the Hijazi Arabic (HA) vowel system, a phonological system based on both temporal and spectral distinctions. Along with static and dynamic F1 and F2 patterns, we evaluated the extent to which vowel duration, F0, and F3 contribute to increased/decreased discriminability among vowels. Data were collected from 20 native HA speakers (10 females and 10 males) producing eight HA monophthongal vowels in a word list with varied consonantal contexts. Results showed that dynamic cues provide further insights regarding HA vowels that are not normally gleaned from static measures alone. Using discriminant analysis, the dynamic cues (particularly the seven-point model) had relatively higher classification rates, and vowel duration was found to play a significant role as an additional cue. Our results are in line with dynamic approaches and highlight the importance of looking beyond static cues and beyond the first two formants for further insights into the description and classification of vowel systems.

1 Introduction

Research on the acoustic patterning of vowels has become increasingly prominent in descriptions of monophthongal vowel systems in various languages. A large part of this work, however, remains focussed on static first (F1) and second (F2) formant measures, typically at the vowel's mid-point. This section explores work focussing on dynamic cues, particularly Vowel-Inherent Spectral Change (VISC; e.g., Nearey and Assmann 1986; Hillenbrand et al. 1995; 1999; 2001; Morrison and Assmann 2013, just to name a few) and their roles in several areas, such as production and perception. This type of investigation (e.g., VISC) has been lacking in the acoustic field and more specifically, in the Arabic context, with the majority of studies focusing on a static approach. This approach is extensively followed because it is believed that measuring the vowel's midpoint, where shifts in formant values are typically minimal, yields the target position a speaker tries to reach when they produce vowels (Peterson and Barney 1952). Therefore, it is thought to represent the best acoustic characteristic of vowels.

We take a closer look at the study of Peterson and Barney (1952). They collected their data by asking participants to produce target vowels in an /hVd/ frame in American English and reported on the vowels' F1 and F2 obtained at the vowel's midpoint. The result showed great variability in formant frequencies in the first and second formant measurements in the scatter plot. Then, 70 listeners who had no knowledge about phonetics were asked to recognise the /hVd/ vocalic elements. They were required to circle 1 of 10 keywords corresponding to the monophthong vowels /ɪ i ɛ ɑ æ ɔ ʊ u ɚ and ʌ/. The listening test was simple, and the signals were recognised by the participants with 94% accuracy. The obvious question that arises is thus the following: How do listeners come to identify the vowels despite the variability observed in the data from Peterson and Barney (1952). Such crucial observations led many researchers to assume that listeners must use other features (e.g., dynamic specification model in particular, VISC) as well as other additional cues (in addition to the first two formants) such as multiple vocalic cues (e.g., fundamental frequency [F0] and third formant frequency [F3]) and vowel duration (Morrison and Assmann 2013). After conducting a considerable amount of VISC research, many researchers (e.g., Nearey and Assmann 1986, Hillenbrand 2013, among others) have found that the cues to vowel identification are not, indeed, expressible in one time slice and that transitional movements within the vowels (including additional cues) perform significant functions in identifying and describing monophthongal vowels. These are explored in more detail in the next section.

1.1 Dynamic approaches to vowel identification and classification using VISC

The term VISC was devised by Nearey and Assmann (1986; Nearey 2013) and defined as the "relatively slowly varying changes in formant frequencies associated with vowels themselves, even in the absence of consonantal context". This is based on the assumption that the formant trajectories of the studied vowels can be characterised by shifts in frequency, typically measured between two locations over the duration of the vowel: one around the vowel's onset (at around 20%) and the other near the vowel's offset (at around 80%). This is because the VISC approach aims to evaluate inherent vowel variation along the vowel target after eliminating the effects of surrounding consonants. VISC has three primary accounts, which reported to perform significant functions in terms of describing and classifying monophthongal vowels. The first model is onset + offset: this is known as the offset model. Many studies have used this model to capture the amount of vowel inherent dynamics. For example, Jin and Liu (2013) found speech dynamics are greater for speakers of languages that have a sparse vowel system (e.g., Chinese, which has six monophthongs) than for those who have a dense vowel system (e.g., Korean and English, which have 10 and 12 monophthongs), potentially due to speakers of low-density languages having more freedom and space to produce their vowels compared to high-density languages (e.g., Manuel 1990; Meunier et al. 2003; Al-Tamimi and Ferragne 2005; Mok 2013; Almurashi et al. 2020, among others). The second model is onset + slope, or the slope model: this is used to reflect the average pace of spectral changes, with a higher value of spectral rate of shift (e.g., rising/positive) suggesting fast dynamic movement over the vowel's duration and a lower value (e.g., falling/negative) suggesting a slower movement (e.g., Fox and Jacewicz 2009; Farrington et al. 2018; Almurashi et al. 2020, among others). The third model is onset + direction, or the direction model: this is used to track the direction of spectral changes (e.g., Nearey and Assmann 1986; Gottfried et al. 1993; Morrison and Nearey 2007; Morrison and Assmann 2013). To note, a considerable amount of research has investigated the direction model using not only two points [20% and 80%] (e.g., Watson and Harrington 1999; Slifka 2003; Chladkova and Hamann 2011), but also three [20%, 50%, and 80%] (e.g., Huang 1992; Zahorian and Jagharghi 1993; Harrington and Cassidy 1994; Hillenbrand et al. 1995; Ferguson and Kewley-Port 2002; Yuan 2013, among others), and multiple points [more than three locations] (e.g., Fox 1983; Van Son and Pols 1992; Adank, Van Hout and Smits 2004; McDougall 2006; McDougall and Nolan 2007; Al-Tamimi 2007a,b; Fox and Jacewicz 2009, among others). Research applying the direction model using multiple measurements has taken the VISC research to an advanced level and demonstrated that such a combined technique can represent detailed information and truer representation of the entire formant trajectories regarding formant spectral movements, potentially revealing dialect-specific patterns which might remain unnoticed when formant values are taken from few locations (Fox and Jacewicz 2009; Darcy and Mora 2015).

In terms of classification accuracy, many acoustic studies (e.g., Hillenbrand and colleagues 1995; 1999; 2001; Arnaud et al. 2011; Almurashi et al. 2020, among others) have used discriminant analysis (e.g., quadratic discriminant analysis [QDA]), to evaluate the role of static and dynamic models (in particular, the direction model) in identifying monophthong vowels. The QDA is considered a conceptual framework that resembles perceptual assimilation processes, as a classification tool (Hillenbrand et al. 1995; 2001). In details, it evaluates the robustness in the observed differences between vowels by looking at the combination of predictors used. The analysis involves a multivariate analysis of variance on the combination of predictors and creates discriminant functions used to separate the vowels. These discriminant functions can be either positively or negatively correlated with each of the predictors. Then, the discriminant analysis tries to separate the vowels into multiple groupings to arrive at an optimal separation between the categories. By using the discriminant analysis, a considerable amount of research has found evidence to support the two-point model, and such a model leads to higher correct classification rates than using a single point (static model) (e.g., Hillenbrand and colleagues 1999; 2001; Arnaud et al. 2011; Almurashi et al. 2020). Other studies found evidence to support the three-point model and that monophthong vowels can have more accurate vowel separation compared with the midpoint model or two-point model (e.g., Huang 1992; Zahorian and Jagharghi 1993; Harrington and Cassidy 1994; Hillenbrand et al. 1995; Ferguson and Kewley-Port 2002, Yuan 2013, among others). Another line of research on dynamic cues reported that vowel identification is not, indeed, expressible in one or even in few time slices, deducting that transitional movements from multiple points (e.g., more than three locations) perform significant functions in classifying monophthongal vowels (e.g., Neel 2004).

Along with VISC measurements, the aforementioned VISC studies note that despite the efficiency of the F1 and F2 values is indisputable, adding additional cues such as multiple vocalic cues (e.g., F0, F3) and vowel duration are beneficial and can aid in providing a more detailed view and understanding. This understanding is crucial for identifying monophthong vowels. For example, Hillenbrand et al. (1995) run QDA on various metrics—namely, F0, F1, F2, and F3 from spectral properties sampled across vowel duration three times, at 20% (onset), 50% (midpoint), and 80% (offset); twice (at 20% and 80%); and once (at 50%). The QDA results showed that extracting such additional cues (in addition to F1 and F2) from dynamic patterns across the vowel duration led to consistent yet fairly modest improvements in category separability. Taken together, merging both approaches (e.g., the use of VISC as a tool to analyse a dynamic aspect in vowel production, and the use of multiple cues (vowel duration, F0, F3) in addition to F1 and F2 would effectively separate vowel categories and provide adequate description, more phonetic details and deeper understanding of the features involved in monophthongal vowels (Morrison and Assmann 2013).

1.2 Dynamic approaches to vowel identification and classification in Arabic

In work on Arabic, the majority of first language (L1) studies have concentrated on static acoustic features of vowels and only two studies have examined the role of dynamic properties in describing and classifying monophthongal vowels. The first study was by Al-Tamimi (2007a,b) who examined the role of dynamic specification of vowel systems in the Jordanian Arabic (/i iː eː a aː oː u and uː/) and Moroccan Arabic (/iː aː ʊ uː and ə/) dialects and French in both production and perception. In production, dynamic correlates were quantified by modelling the transition (onset to midpoint) through regression analyses (linear and polynomial). The results showed that dynamic correlates allowed for a fine-tuned distinction, whereby vowels were clearly separated between and within dialects. In terms of classification accuracy, Al-Tamimi (2007a,b) found a clear advantage to the dynamic stylisation of transition in classification; an increase in classification accuracy in discriminating the two Arabic dialects (e.g., Jordanian and Moroccan) and French, by around 10-30% (depending on the consonants' place of articulation and comparison), was observed (Al-Tamimi, 2007a). Dynamic correlates of vowels further allowed clear separation between and within the two Arabic dialects; rates of 85.68% were obtained for Moroccan Arabic and 88.6% for Jordanian Arabic (using dynamic specification), with an improvement of classification accuracy by 5-8% (Al-Tamimi, 2007b).

The second study was conducted by Almurashi et al. (2020) who investigated VISC models (e.g., offset, slope, and direction models) from two points for the F1, F2, and F3 of Hijazi Arabic (HA) vowels. HA, the dialect which is the focus of this study, is considered one of the main spoken dialectal varieties in the Kingdom of Saudi Arabia and spoken in several cities, such as Jeddah, Taif, Makkah, and Medina (Alzaidi 2014). The HA vowel system contains the MSA/Classical Arabic long vowels /iː aː uː/ and three short vowels /i a u/. Moreover, it contains the two long mid vowels /eː/ and /oː/ that evolved from MSA/Classical Arabic diphthong vowels /aw/ and /aj/ (Abdoh 2011). Almurashi et al. (2020) investigated all HA vowels in /hVd/ syllables that were included in a carrier sentence. The results showed the following: in terms of the offset model, HA vowels had great spectral shifts (up to 200 Hz for F1, up to 600 Hz for F2, and up to 400 Hz for F3), as has been noted in studies on low-density languages (e.g., Jin and Liu 2013; Mok 2013, among others), suggesting that their speakers have more space and freedom to produce their vowels compared with high-density languages. In terms of the slope model, Almurashi et al. (2020) found that using the slope model revealed significant variation across the vowels. For example, the data displayed that the F2 of the low and back vowels had rising slopes, unlike the front vowels, which had falling slopes. In terms of the direction model, Almurashi et al. (2020) found that using the direction was useful in the disambiguation of tense/lax vowels in HA. For instance, the F1 direction of long vowels showed a significantly different spectral change compared with their short counterparts. This finding provided evidence for the existence of a tense/lax distinction in Arabic vowel contrasts which were otherwise thought to be based on length; this issue is still in debate despite mounting evidence indicating a difference in both quality and quantity (e.g., Rosner et al. 1994; Khattab 2007; Al-Tamimi 2007a,b; Khattab and Al-Tamimi 2008; Almbark and Hellmuth 2015; Almurashi et al. 2020; Al-Mazrouei et al. 2023). Almurashi et al. (2020) ran the discriminant analysis on their /hVd/ data, and the results revealed that the three-point model with the first three formants (with and without the duration) resulted in the highest classification accuracy for all eight HA vowels (the average classification rate was 95.5% for the three-point model), followed by the two-point model (the average classification rate was 94.25%), and then the static model (the average classification rate was 93.5%). They concluded that looking at the internal transition behaviour of vowels can be useful in providing a better overview and the three-point approach is the best and most accurate for classifying HA vowels and highlighted the role of vowel duration more than F3 as an additional cue for the classification accuracy of HA vowels.

2 The current study

With the importance of dynamic cues in mind and with the majority of work in this area being restricted to English, more work is required to evaluate the importance of dynamic cues across languages. Emerging works from Al-Tamimi (2007a,b) and Almurashi et al. (2020) labs on Arabic suggest that, while dynamic cues in the spectral properties of vowels improve the identification of Arabic vowels, their classification power is attenuated in this language compared with work on other languages. This fact could open a rich testing area for supplementary studies on cross-language comparisons of L1 research. Languages may be compared with regard to the spectral rate of vowel change (slope model), the direction of vowel shifts (direction model), or the amount of vowel change (offset model) noted in their vowel systems. All of these comparisons can reveal how vowels differ in the nature of their dynamic properties and the extent to which they are different or similar to other vowels in other languages. Most importantly, they would be useful and serve as a reference point for future Arabic studies or other language research.

As stated in the background section, to date, dynamic properties of vowels (particularly VISC) have been researched in only a handful of studies on Arabic. Beyond the restricted /hVd/ environment examined by Almurashi et al. (2020), little information is available regarding VISC's role in other consonantal contexts. Looking at vowels across a set of consonants is different than examining vowels in isolation or the /hVd/, as the /hVd/ syllables do not contain many spectral changes (Oh 2013) unlike the consonantal environments which are known to affect vowel formant values (Hillenbrand et al. 2001). Additionally, different/varied contexts can provide a better overview and additional insights into the characterisation of dynamic cues of HA (e.g., whether HA still exhibits diphthongisation [VISC], whether /eː/ vs /oː/ retained any potential diphthongized patterns or whether they are produced as fully monophthongised, whether HA has a tense/lax distinction, and whether a dynamic representation would yield a better estimation of such a distinction) as well as reveal language or dialect-specific fine-grained phonetic detail that is not gleaned from vowels in isolation or restricted contexts (Clopper and Pisoni 2004; Schwartz 2021). Importantly, we know even less about the role of additional correlates such as F0, F3 and duration in characterising HA vowels within a variety of consonants. As mentioned earlier, combining both approaches, namely, the use of VISC as a tool to analyse a dynamic aspect in vowel production and the use of multiple vocalic cues (e.g., F0, F3) and vowel duration in addition to F1 and F2, was found to be useful and provide further insights into the vowels' characters and how they differ (particularly for vowel pairs that are likely to overlap in their F1 and F2 midpoint values such as /eː/ vs /i/ and /oː/ vs /u).

Taken together, and to fill a gap in the literature, this research expands on Almurashi et al.'s (2020) study by investigating HA vowels (in particular short vs long vowels as well as the vowel pairs /eː/ vs /oː/, /eː/ vs /i/, and /oː/ vs /u/) in various phonetic environments, which is recommended by many researchers (e.g., Hillenbrand et al. 1995; Watson and Harrington 1999). In addition, this current study constitutes the first step into the field of intrinsic dynamic correlates, not only in HA but also in the Arabic language, looking at monophthongal vowels in a variety of consonant environments. The purpose is to present a full acoustic description of HA monophthongs. In doing so, we investigate and evaluate the importance of static and dynamic correlates, particularly VISC, in describing and classifying the production of HA vowels; we also explore to what extent vowel duration, F0, and F3 act as additional cues to classification accuracy.

3 Methodology

3.1 Subjects and material

The participants were 20 HA speakers (10 males and 10 females) who were between 18 and 30 years old (median = 23) and born and raised in Hijaz in the north-west of Saudi Arabia. The participants were undergraduate students at Taibah University and reported no history of speech and/or language disorders. Recordings were made on a Roland Edirol R-09 recorder and Audio Technica Cardioid stereo microphone with a sampling rate of 44,100 Hz and 16-bit amplitude resolution. The subjects were placed in a soundproof room at Taibah University and were asked to produce the target HA vowels (/i iː eː a aː oː u and uː/) within monosyllabic or disyllabic words produced in the phrase of /ktoːb _____ marteːn/, “Write ___ twice”. Each HA vowel was put into six words in three different consonantal contexts namely, bilabial _ alveolar; alveolar _ alveolar; velar _ alveolar (where each consonantal context has 2 words containing the target vowel; the set of target words can be found in the Appendix, Table A1). Together, the HA stimuli consisted of 8 vowels × 2 words × 3 different consonantal contexts × 3 repetitions × 20 HA participants = 2,880 items.

3.2 Acoustic analysis

Acoustic analysis was conducted using Praat (Boersma and Weenink 2022, version 6.2.23). The sound files were manually labelled for each token. The boundaries of the vocalic segment were manually labelled for each monosyllabic and disyllabic word using wideband spectrograms and waveforms in addition to auditory verification (Yang 1996) (see illustration in Figure 1). F0 and all formant tracks were obtained using a 0.025 s window length, 50 Hz pre-emphasis, and a spectrogram view range of 5,000 Hz for males and 5,500 Hz for females. The Lobanov normalisation procedure (Lobanov 1971), which was found to perform considerably better than the majority of other procedures (Adank, Smits and Van Hout 2004; Fabricius et al. 2009), was run on the formant frequencies obtained at the midpoint of the vowel[1] (on a speaker-by-speaker basis) in RStudio (RStudio Team 2022, version 1.4.1103) and R (R Core Team 2022, version 4.0.4).

---

[1] The F1 and F2 midpoints were presented in the result section with and without normalisation (raw data). This was done to represent the whole picture of static representations of the monophthongal vowels. To note, the normalised formant frequencies were used only to plot vowels in the F1×F2 space, and not in any of the statistical tests.

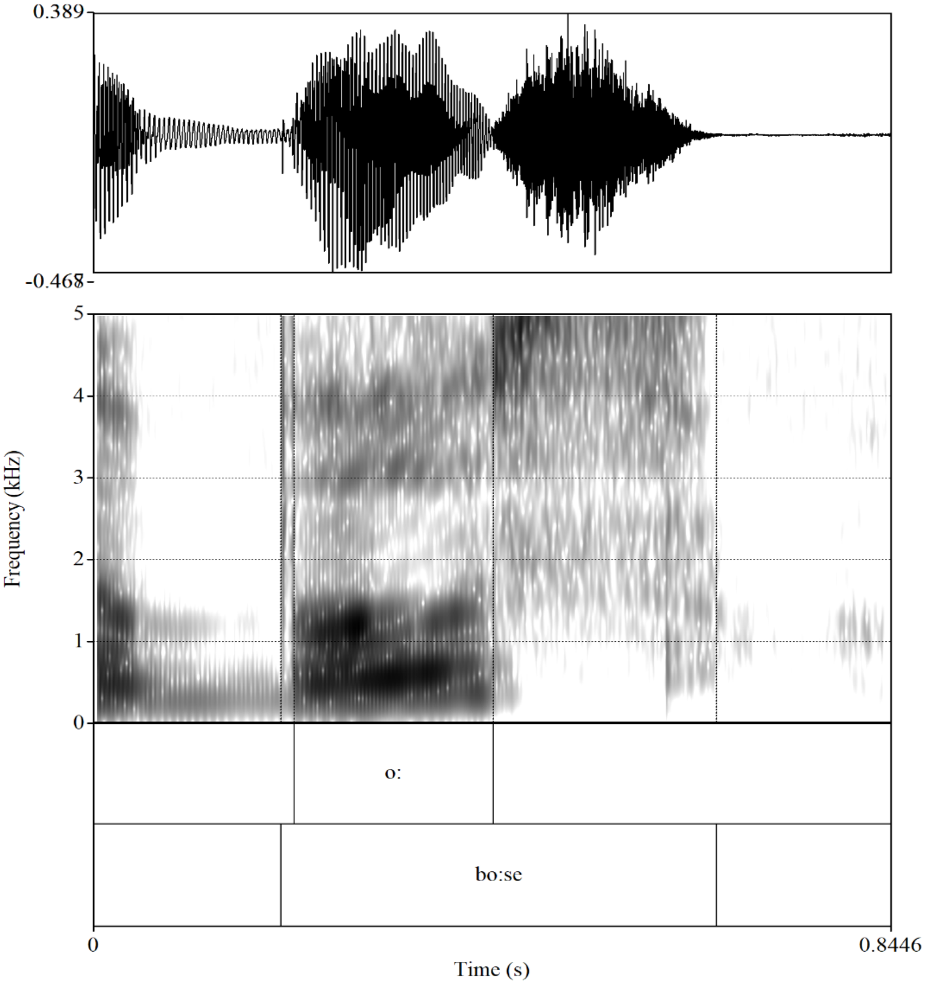


Figure 1: Spectrogram showing formant frequencies of the word /boːse/ ("kiss") as produced by a female HA speaker.

For the purposes of this research, vowel duration (in ms), F0, and the first three formant values were automatically extracted with the aid of a Praat script. The first three formants and F0 values were extracted from one location (50% for the static model), two locations (20% and 80% for the two-point model), three locations (20%, 50%, and 80% for the three-point model), and seven locations (20%, 30%, 40%, 50%, 60%, 70%, and 80% for multiple points[2]) across the vowel

---

[2] Taking more than these measurements for monophthongal vowels would not provide any sudden movements in the vowel trajectories that would justify the use of a large number of measurement points (Cardoso 2015).

duration. For the offset model, we obtained the amount of a vowel's spectral changes by calculating the differences for all three formants and F0 values between the vowel's two measurement locations (in Hertz). For the slope model, we obtained the vowel's rate of change by calculating the differences for all three formant and F0 values between the vowel's two measurement locations and then dividing them by the vowel duration. For the direction model, we obtained the direction of the vowel's spectral shifts by tracking the first three formants and F0 values from two samples (for the two-point model), three samples (for the three-point model), and seven samples (for multiple points). All formant values were manually verified and any errors in formant estimation were corrected by hand. To mitigate Praat measurement error, the Praat script produced a PDF snapshot of each token's spectrogram. These spectrogram PDFs were visually inspected to verify that there were no major formant measurement errors. Additionally, F0, F1, F2, F3, and vowel duration measurements were visually inspected in RStudio (RStudio Team 2022, version 1.4.1103) and R (R Core Team 2022, version 4.0.4) to verify that there were no major measurement errors.

3.3 Statistical analysis

Two types of statistical techniques were used to evaluate the differences in the data—namely, linear mixed-effects modelling (LMM; using the lme4 package (version 1.1.26; Bates et al. 2015) with the afex package (version 0.28-1; Singmann et al. 2018) to select the best fitted/best performing model, followed by pairwise comparisons (Tukey's HSD post-hoc tests) with the multcomp package (version 1.4-16; Hothorn et al. 2016) to determine whether vowels in each pair were significantly different (McDougall 2002; Fox and Jacewicz 2009). We used an alpha level of 0.05, meaning the results would only be considered statistically significant with a p value lower than 0.05. Our outcome was one of the acoustic correlates (F0, F1, F2, and F3 for the static model and for each model of the dynamic cues). Our fixed effects were the vowel identity (with eight levels), consonant (with three levels), and gender (with two levels). Our random effects were the speakers and words to allow for the crossed random effects design to be taken into account (Baayen et al. 2008). For each acoustic correlate, we ran five versions:

```
mdl.1 <- lmer(outcome ~ vowel + consonant + gender + (vowel + consonant | speaker) + (gender | word), data = data)
mdl.2 <- lmer(outcome ~ vowel + consonant + gender + (vowel | speaker) + (gender | word), data = data)
mdl.3 <- lmer(outcome ~ vowel + consonant + gender + (vowel | speaker) + (1 | word), data = data)
mdl.4 <- lmer(outcome ~ vowel + consonant + gender + (1 | speaker) + (1 | word), data = data)
mdl.5 <- lmer(outcome ~ vowel * consonant + gender + (1 | speaker) + (1 | word), data = data)
```

The justification for these models follows from a maximal specification approach (Barr 2013; Barr et al. 2013). First, we decided to include both speakers and words as crossed random effects given the structure of our data. Next, we used gender random slope for words to allow for modelling of any variations with respect to how our males and females produced each word. By vowel and consonant random slopes for speaker were also used to adjust for individual variations. For our fixed effects, we used vowel (variable of interest) in addition to consonant and gender (controlling variables). The controlling variables were used to adjust the coefficients of the fixed and random effects. We used model comparison through Log-Likelihood χ^2 tests and report the results of our optimal model.

The next step was applying the discriminant analysis as a classification tool to evaluate the extent to which the static and dynamic models and other acoustic feature sets (F0, F1, F2, F3, and vowel duration) improve vowel classification. We used the qda function from the MASS package (version 7.3-53.1; Venables and Ripley 2002) to obtain the QDA with a *leave-one-out* cross-validation, or "jackknife" (Hillenbrand et al. 1995; Al-Tamimi, 2007a,b; Almurashi et al. 2020). In detail, this technique divides the data into multiple data sets and then it trains on all of the sets, except one that will be used as a testing data set. It repeats this procedure with each set and then produces the classification accuracy rate. For each of the models (e.g., one-point, two-point, three-point, and seven-point models), we used the vowels as categories to be classified and each of the formant frequencies or each of the formulae and vowel duration outputs as predictors[3]. In detail, we used the production of the full HA vowels as categories and the following predictors as input to each of the discriminant analysis: For the one-point model, we entered the formant values sampled from vowel midpoint at 50%; for the two-point model, we entered the formant values sampled from vowel onset (at 20%) and offset (at 80%); for the three-point model, we entered the formant values sampled from vowel onset (at 20%), midpoint (at 50%), and offset (at 80%); and finally, for the seven-point model, we entered the formant values sampled from seven locations (20%, 30%, 40%, 50%, 60%, 70%, and 80%) across the vowel duration[4]. For each model, we examined various combinations of F0, F1, F2, and F3, with and without the vowel duration. All figures in this paper were created in RStudio (2022) and R Core Team (2022) with the ggplot2 (version 3.3.3; Wickham 2016), dplyr (version 1.0.4; Wickham et al. 2019), tidyverse (version 1.3.0; Wickham 2017), mgcv (version 1.8-34; Wood 2015), and nlme packages (version 3.1-152; Pinheiro et al. 2017).

---

[3] To note, the offset, slope, and normalised data were not included in the discriminant analysis. Only raw data from static and dynamic model particularly, the direction model.

[4] The same as we applied for dynamic cues' outcomes in LMMs models.

4 Results

This section presents the descriptive and statistical results of the static and dynamic cues of HA monophthongs, accompanied by discriminant analysis. A full summary of the results for the duration, F0, and the first three formant values of HA vowels can be found in the Appendix, Table A2. In addition, full statistical results of the acoustic cues of HA vowels can be found in the Appendix, Table A3.

4.1 Overall patterns of Hijazi Arabic vowels

4.1.1 Static cues

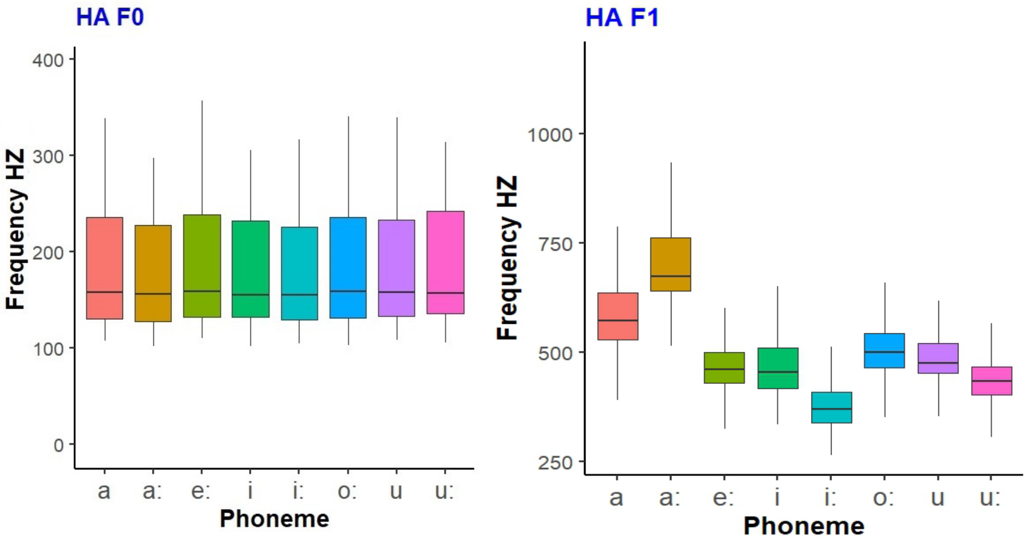
Beginning with the static model, we used the midpoint formant frequencies of the first two formants for all of the HA vowels across different consonant environments in box plots[5] and a scatter plot to characterise the vowels' acoustic features (see Figures 2 without normalisation; and 3 with normalisation). Both Figures show that most of the HA vowels were generally separated. The results of the LMM comparison showed a clear improvement to the model fit when using mdl.2[6], F0: $\chi^2(2) = 238.2$ Hz, $p < 0.0001$; F1: $\chi^2(2) = 87.2$ Hz, $p < 0.0001$; F2: $\chi^2(2) = 260.7$ Hz, $p < 0.0001$; F3: $\chi^2(2) = 77.2$ Hz, $p < 0.0001$. The results of the pairwise comparisons for the /aː/ and /a/ pair showed significantly higher F1 and lower F2 frequencies for /aː/ (for F1, there was a difference of 115.1 Hz, $p < 0.0001$; and for F2, a difference of -235.4 Hz, $p < 0.0001$), with no differences for F0 and F3. For the /iː/ and /i/ pair, the results showed significantly lower F1 and higher F2 frequencies for /iː/ (F1 had a difference of -89.1 Hz, $p < 0.0001$; F2 a difference of 266.6 Hz, $p < 0.0001$), with no differences for F0 and F3. For the pair /uː/ vs /u/, the results showed

---

[5] The box represents the middle '50%' of the data, the lower whisker represents the lower '25%' of the data, and the upper whisker represent the upper '25%' of the data.

[6] Model 2 was the optimal model and the results shown here are those obtained when comparing model 2 to model 1. To note, no significant interactions were found for consonant and for gender in model 2.

significantly lower F1 and F2 frequencies for /uː/ (for F1, there was a difference of -53.1 Hz, $p < 0.0001$; for F2, a difference of -290.3 Hz, $p < 0.0001$), with no differences for F0 and F3. For the pair /oː/ vs /eː/, the results showed significantly higher F1 and lower F2 frequencies for /oː/ (for F1, a difference of 45.7 Hz, $p < 0.0001$; for F2, a difference of -1051.7Hz, $p < 0.0001$; and for F3, a difference of -108.0.7Hz, $p < 0.0005$), with no differences for F0. For the pair /eː/ vs /i/, the results showed significantly lower F2 frequencies for /eː/ (for F2, a difference of -135.6 Hz, $p < 0.0001$), with no differences for F0, F1, and F3. For the pair /oː/ vs /u/, the results showed significantly higher F2 frequencies for /oː/ (a difference of 177.3 Hz, $p < 0.0001$), with no differences for F0, F1 and F3.



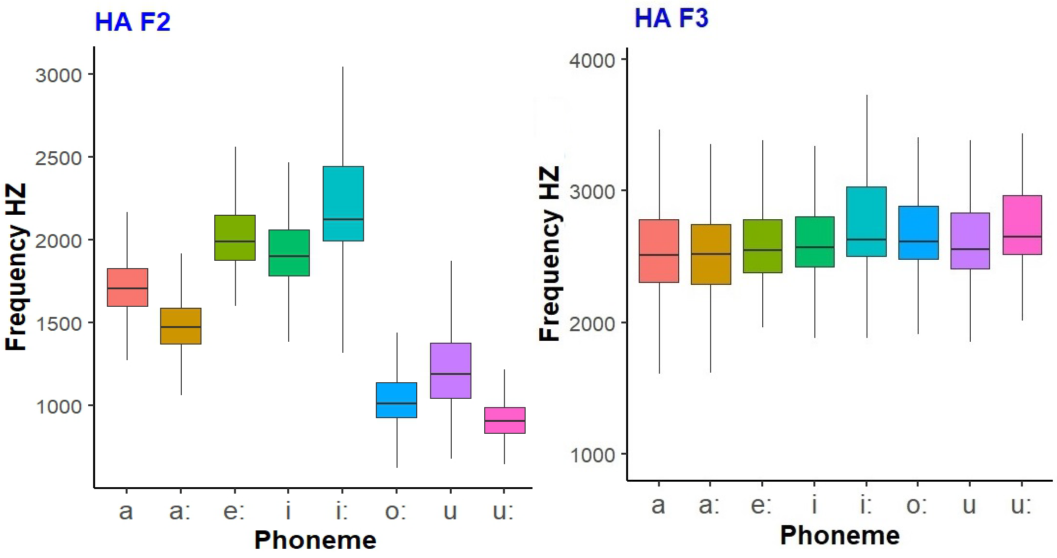


Figure 2: Box plots of the midpoint values of the Hijazi Arabic vowels.

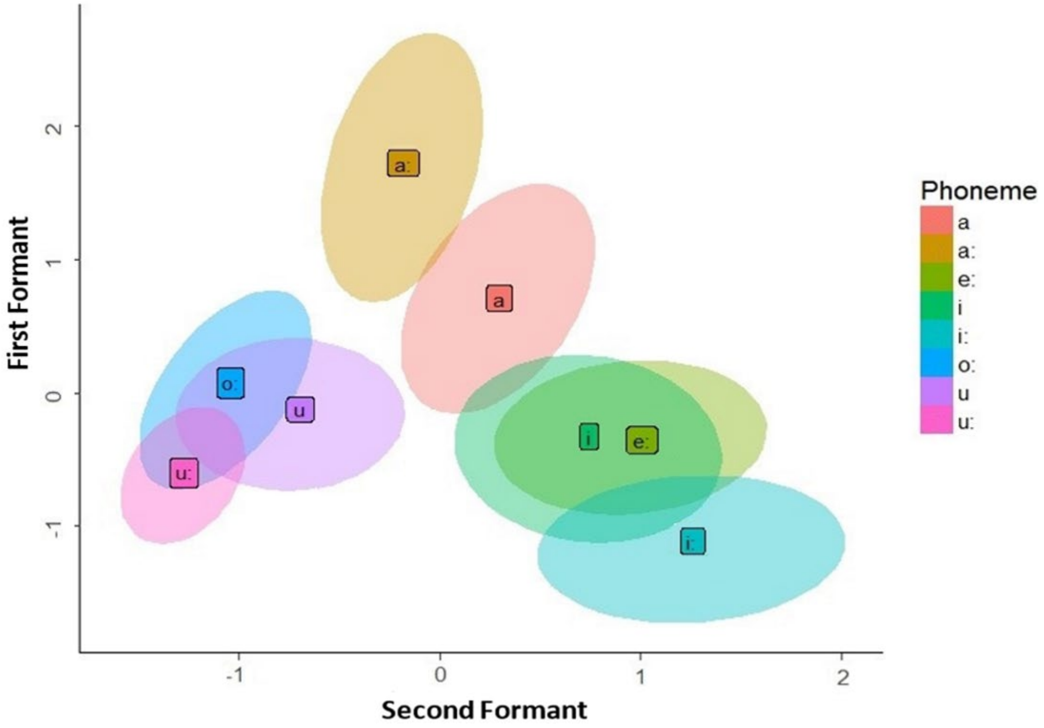


Figure 3: Scatter plot of the normalised midpoints of the first two formant values of the Hijazi Arabic vowels. The ellipses (based on 1.2 SDs) represent the variations occurred in the production of the vowel.

4.1.2 Dynamic cues

We continue with the dynamic models by looking at the offset model using the two-point measurement technique. Figure 4 shows the amount of formant movement changes for each HA vowel, displaying a great amount of spectral movement. The results of the LMM comparison showed a clear improvement to the model fit when using mdl.2[7], F0: $\chi^2(2) = 327.7$ Hz, $p < 0.0001$; F1: $\chi^2(2) = 38.2$ Hz, $p < 0.0001$; F2: $\chi^2(2) = 26.5$ Hz, $p < 0.0001$; F3: $\chi^2(2) = 17.6$ Hz, $p < 0.0001$. Regarding vowel pairs, the results showed that only some pair comparisons were statistically significant. Specifically, for F1, only /aː/ vs /a/ showed a statistically significant difference, with /aː/ having a positive difference of 27.8 Hz, $p < 0.0001$ for F1 and by 81.4 Hz, $p < 0.0001$ for F2; and there were no differences for F0 and F3. Other vowel pairs, such as /iː/ vs /i/ and /u/ vs /uː/, showed no statistical differences between the offset of any of their three formant values or for F0. For the pair /oː/ vs /eː/, the differences were statistically significant for F0 (had a negative difference of -3.9 Hz, $p < 0.0001$), F1 (had a negative difference of -35.9 Hz, $p < 0.0001$), with no differences for F2 and F3. For the pair /eː/ vs /i/, the differences were statistically significant for F0 (had a negative difference of -6.69 Hz, $p < 0.0001$), F1 (had a negative difference of -55.4 Hz, $p < 0.0001$), F2 (had a negative difference of -105.7 Hz, $p < 0.0001$), with no differences for F3. For the pair /oː/ vs /u/, the differences were statistically significant for F1 (had a negative difference of -24.2 Hz, $p < 0.0001$), with no differences for F0, F2 and F3.

---

[7] Model 2 was the optimal model and the results shown here are those obtained when comparing model 2 to model 1. To note, no significant interactions were found for consonant and for gender in model 2.

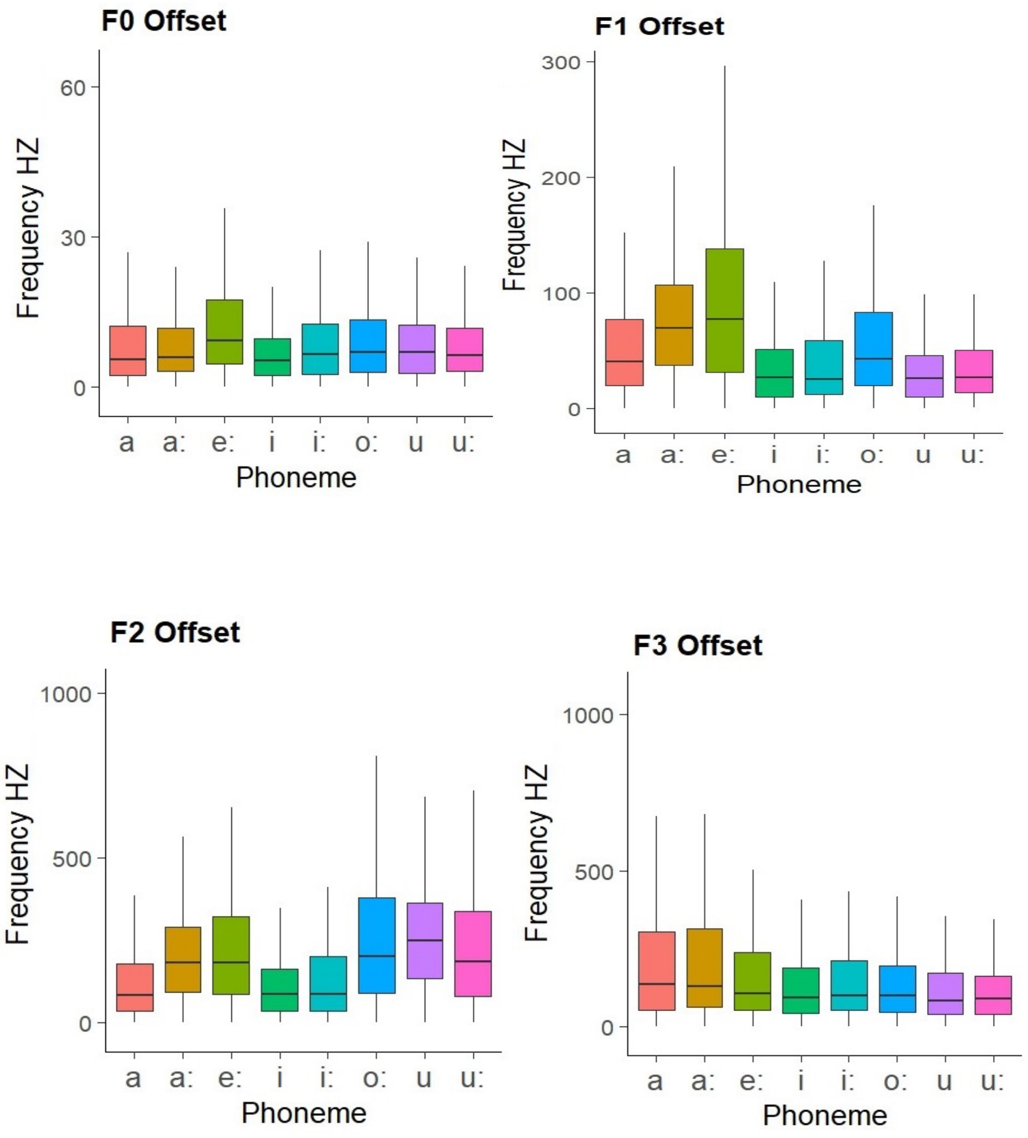


Figure 4: Box plots of the offset model for the Hijazi Arabic vowels.

Regarding the slope of HA from two-point model, Figure 5 shows potential differences among its vowels, with some vowels having their own slope values for each formant. More specifically, the LMM comparison showed clear improvement to the model fit when using mdl.2[8], F0: $\chi^2(2) = 189.9$ Hz, $p < 0.0001$; F1: $\chi^2(2) = 33.4$ Hz, $p < 0.0001$; F2: $\chi^2(2) = 11.3$ Hz, $p < 0.0001$;

---

[8] Model 2 was the optimal model and the results shown here are those obtained when comparing model 2 to model 1. To note, no significant interactions were found for consonant and for gender in model 2.

F3: $\chi^2(2) = 27.0$ Hz, $p < 0.0001$. Comparison of vowel pairs showed that for /aː/ and /a/, the differences were statistically significant for F0 (had a negative difference of -0.05 Hz, $p < 0.0001$), for F1 (had a positive difference of 0.2 Hz, $p < 0.0001$), for F2 (had a negative difference of -0.8 Hz, $p < 0.0001$), and no significant difference for F3. For /iː/ and /i/, the results showed a negative difference in slopes for F1 (difference of -0.19 Hz, $p < 0.0001$), a positive slope for F2 (difference of 0.6 Hz, $p < 0.0001$), and no significant differences for F0 and F3. For /uː/ and /u/, the results showed no significant differences in slopes for F0, F1, F2, and F3. For the pair /oː/ vs /eː/, the results showed a significant slope with overall a positive difference for F1 (difference of 0.39 Hz, $p < 0.0001$) and a negative difference for F2 (difference of -1.3 Hz, $p < 0.0001$), with no significant differences in slopes for F0 and F3. For the pair /eː/ vs /i/, the results showed a significant slope with overall a positive difference for F1 (difference of 0.71 Hz, $p < 0.0001$) and a negative difference for F2 (difference of -1.81 Hz, $p < 0.0001$), with no significant differences in slopes for F0 and F3. For the pair /oː/ vs /u/, the results showed a significant slope with overall a positive difference for F1 (difference of 0.27 Hz, $p < 0.0001$) and for F2 (difference of 1.06 Hz, $p < 0.0001$), with no significant differences in slopes for F0 and F3.

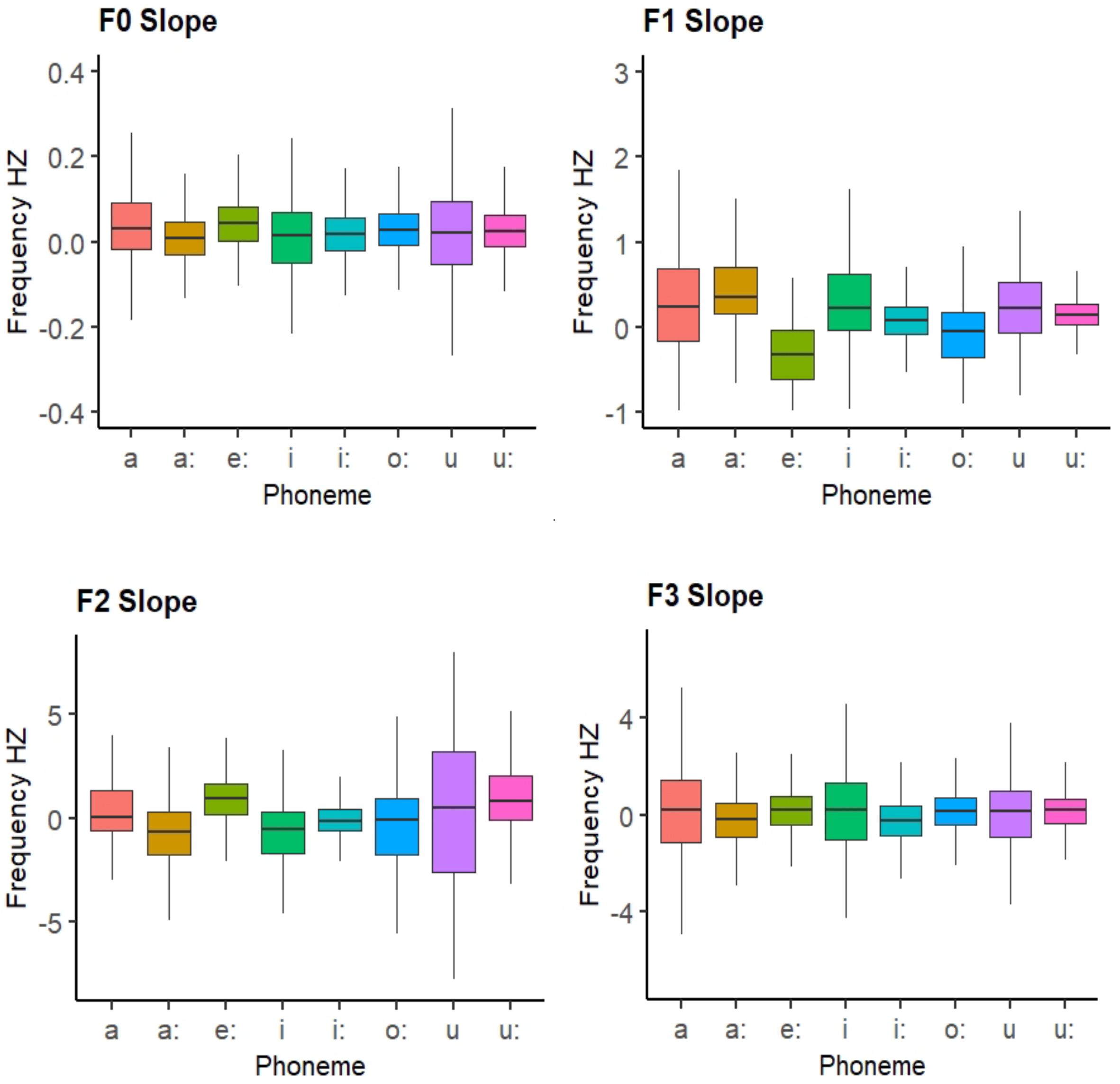


Figure 5: Box plots of the slope model of the Hijazi Arabic vowels.

With respect to the direction of HA using the two-point model, Figure 6 shows variation among HA vowels. According to the results of the LMM comparison, there was a clear improvement to the model fit when using mdl.2[9], F0: $\chi^2(2) = 277.2$ Hz, $p < 0.0001$; F1: $\chi^2(2) =$

---

[9] Model 2 was the optimal model and the results shown here are those obtained when comparing model 2 to model 1. To note, no significant interactions were found for consonant and for gender in model 2.

134.0 Hz, $p < 0.0001$; F2: $\chi^2(2) = 152.5$ Hz, $p < 0.0001$; F3: $\chi^2(2) = 93.1$ Hz, $p < 0.0001$. Comparison of vowel pairs showed that for /a/ and /aː/, there was an overall significantly higher direction related to the transition of /aː/ for F1 (difference of 111.6 Hz, $p < 0.0001$), a significantly higher direction for F2 (difference of 208.2 Hz, $p < 0.0001$), and no differences for F0 and F3. For /i/ and /iː/, the results showed no differences for F0 but significant differences in direction for F1, F2, and F3: high for F1 (difference of 68.8 Hz, $p < 0.0001$), low for F2 (difference of -228.1 Hz, $p < 0.0001$), and low for F3 (difference of -94.8 Hz, $p < 0.0001$). For the pair of /u/ vs /uː/, the results showed overall significant differences in direction for /uː/ in F1, F2, and F3: For F1, the high direction difference amounted to 36.4 Hz, $p < 0.0001$; for F2, the high direction difference was 167.9 Hz, $p < 0.0001$; and for F3, the low direction difference was -79.2 Hz, $p < 0.0001$. There were no differences for F0. For the pair /oː/ vs /eː/, the results showed significant differences in directions with an overall high difference for F1 (a high transition difference of 41.3 Hz, $p < 0.0001$) and low difference for F2 (a low transition difference of -865.8 Hz, $p < 0.0001$), with no significant differences in directions for F0 and F3. For the pairs /eː/ vs /i/, the results showed significant differences in directions with an overall low difference for F2 (a low transition difference of -112.1 Hz, $p < 0.0006$) with no significant differences in directions for F0, F1, and F3. For the pairs /oː/ vs /u/, the results showed significant differences in directions with an overall low difference for F1 (a low transition difference of -25.7 Hz, $p < 0.0001$) and high difference for F2 (a high transition difference of 123.4 Hz, $p < 0.0001$), with no significant differences in directions for F0 and F3.

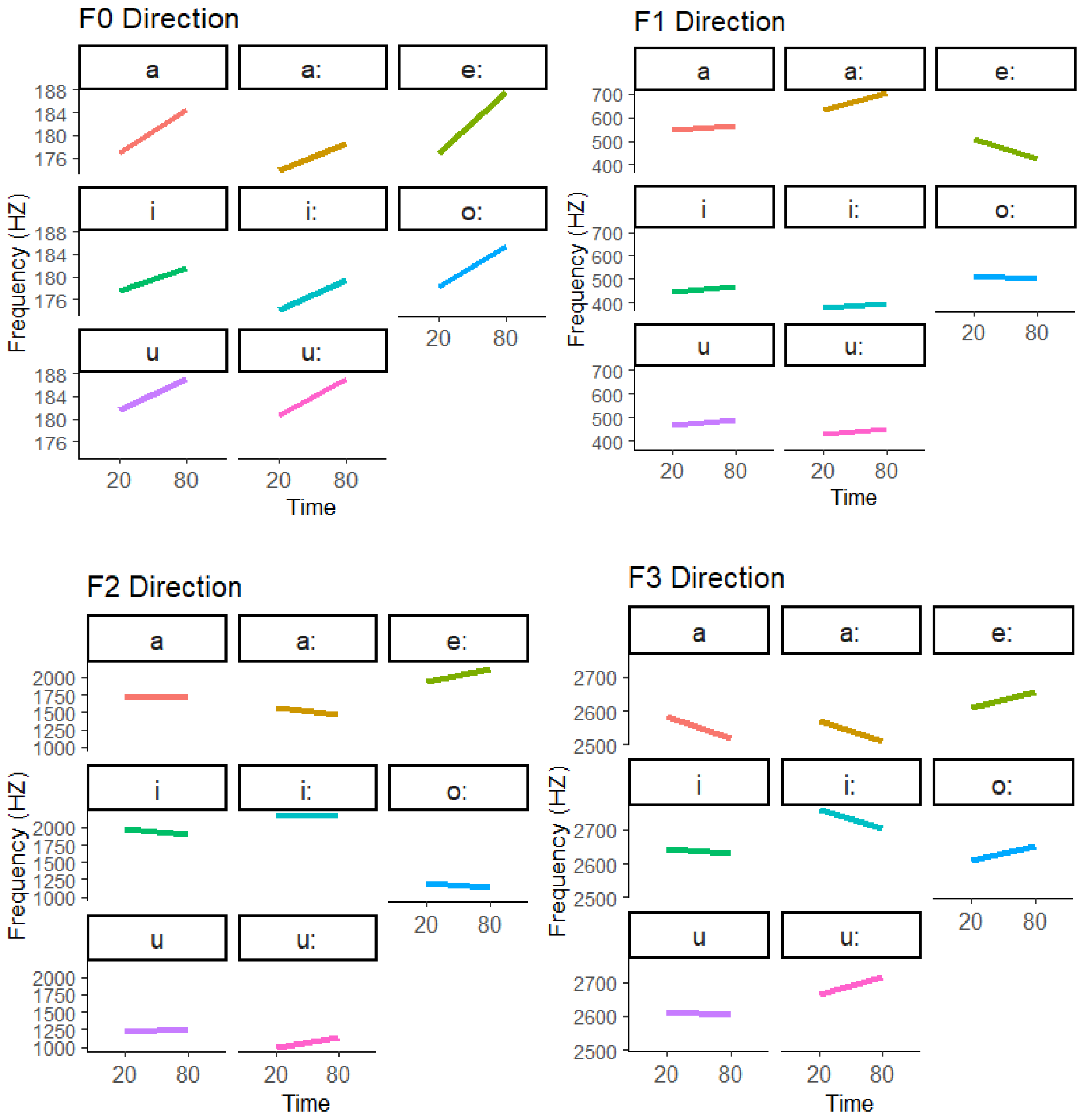


Figure 6: Results of the direction (measured at two points) of the Hijazi Arabic vowels.

With further focus on the direction model, the three-point model showed a better acoustic characteristic of HA vowels compared with the static and two-point models. Figure 7 presents the F0, F1, F2, and F3 directions of HA vowels, which differed considerably across the vowels. Regarding the statistical results of the three-point model, the LMM comparison showed a clear improvement to the model fit when using mdl.2[10], F0: $\chi^2(2) = 277.6$ Hz, $p < 0.0001$; F1: $\chi^2(2) = 124.8$ Hz, $p < 0.0001$; F2: $\chi^2(2) = 246.7$ Hz, $p < 0.0001$; F3: $\chi^2(2) = 130.7$ Hz, $p < 0.0001$. Comparing vowel pairs showed the following for /a/ and /aː/: a significantly higher direction for F1 (transition difference of 112.8 Hz, $p < 0.0001$), a significantly higher direction for F2 (difference of 217.3 Hz, $p < 0.0001$), and no differences for F0 and F3. For /i/ and /iː/, the results showed no differences for F0 and significant differences in direction for F1, F2, and F3 values: a high direction for F1 (difference of 75.6 Hz, $p < 0.0001$) and low directions for F2 (difference of -240.9 Hz, $p < 0.0001$) and F3 (difference of -99.1 Hz, $p < 0.0001$). For /u/ and /uː/, the results showed no differences for F0 and overall significant differences in direction for F1, F2, and F3 for /u/: for F1, a high direction (difference of 42.0 Hz, $p < 0.0001$); for F2, a high direction (difference of 208.7 Hz, $p < 0.0001$); and for F3, a low direction (difference of -90.4 Hz, $p < 0.0001$). For the pair /oː/ vs /eː/, the results showed significant differences in directions with an overall high difference for F1 (a high transition difference of 42.1 Hz, $p < 0.0001$), and low difference for F2 (a low transition difference of -927.8 Hz, $p < 0.0001$), with no significant differences in directions for F0 and F3. For the pairs /eː/ vs /i/, the results showed significant differences in directions with an overall low difference for F2 (a low transition difference of -117.3 Hz, $p < 0.0001$) with no significant differences in directions for F0, F1, and F3. For the pairs /oː/ vs /u/, the results showed significant differences in directions with an overall low difference for F1 (a low transition difference of -28.6 Hz, $p < 0.0001$) and high difference for F2 (a high transition difference of 153.0 Hz, $p < 0.0001$), with no significant differences in directions for F0 and F3.

---

[10] Model 2 was the optimal model and the results shown here are those obtained when comparing model 2 to model 1. To note, no significant interactions were found for consonant and for gender in model 2.

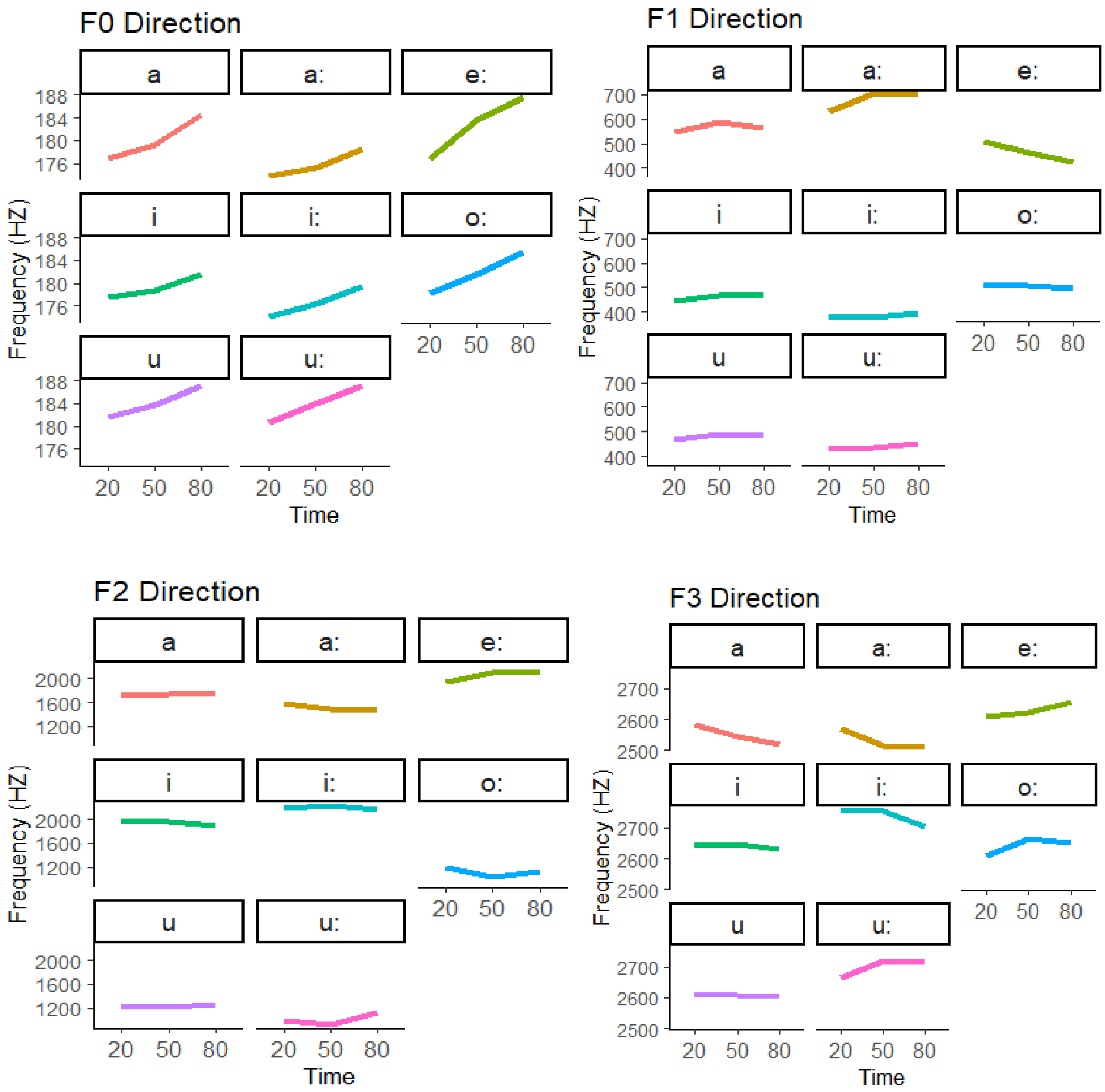


Figure 7: Results of the direction (measured at three points) of the Hijazi Arabic vowels.

Finally, the F0, F1, F2, and F3 directions of HA vowels when using the multiple points, as presented in Figures 8 and 9, differed considerably across the vowels. As can be seen from Figure 8, the formant trajectory plot implies that HA vowels are produced as dynamic vowels, and that /a:/, /u:/, and /e:/ in particular appear to exhibit a great amount of movement in either F1 or F2.

The LMM comparison showed a clear improvement to the model fit when using mdl.2[11], F0: $\chi^2(2) = 262.9$ Hz, $p < 0.0001$; F1: $\chi^2(2) = 118.1$ Hz, $p < 0.0001$; F2: $\chi^2(2) = 188.8$ Hz, $p < 0.0001$; F3: $\chi^2(2) = 139.0$ Hz, $p < 0.0001$. Comparing vowel pairs showed that for /a/ and /aː/, there were significant differences related to /aː/ for F1 and F2, with no differences for F0 and F3 (for F1, the difference was 116.5 Hz, $p < 0.0001$; and for F2, the difference was -224.0 Hz, $p < 0.0001$). For /i/ and /iː/, the results showed overall significant differences in direction for F1, F2, and F3, with no differences for F0 (for F1, the difference was -79.1 Hz, $p < 0.0001$; for F2, the difference was 243.0 Hz, $p < 0.0001$; and for F3, the difference was 107.2 Hz, $p < 0.0001$). For /uː/ and /u/, the results showed significant differences in direction values for F1, F2, and F3, with no differences for F0 (for F1, the difference was -45.1 Hz, $p < 0.0001$; for F2, the difference was -233.5 Hz, $p < 0.0001$; and for F3, the difference was 98.7 Hz, $p < 0.0001$). For the pair /oː/ vs /eː/, the results showed significant differences in directions, with an overall high difference for F1 (a high transition difference of 44.7 Hz, $p < 0.0001$), a low difference for F2 (a low transition difference of -958.1 Hz, $p < 0.0001$), with no significant differences in directions for F0 and F3. For the pair /eː/ vs /i/, the results showed significant differences in direction for F2 (a low transition difference of -120.1 Hz, $p < 0.0001$), with no significant differences in directions for F0, F1, and F3. For the pair /oː/ vs /u/, the results showed significant differences in direction values for F1, F2, and F3, with no differences for F0 (for F1, the difference was -30.2 Hz, $p < 0.0001$; for F2, the difference was 160.9 Hz, $p < 0.0001$; and for F3, the difference was -41.5 Hz, $p < 0.005$).

---

[11] Model 2 was the optimal model and the results shown here are those obtained when comparing model 2 to model 1. To note, no significant interactions were found for consonant and for gender in model 2.

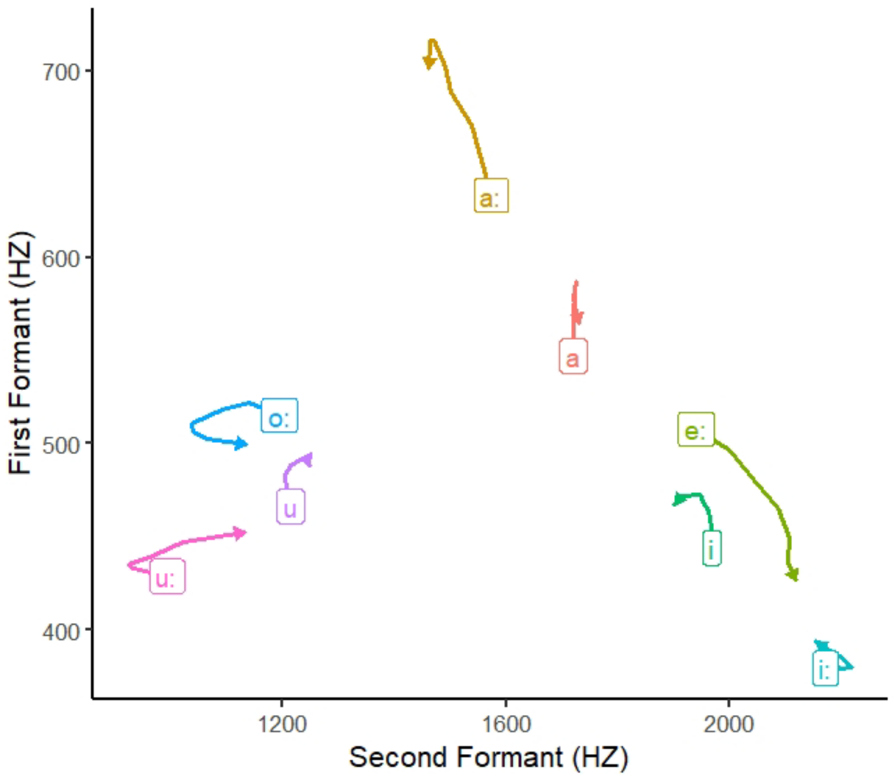
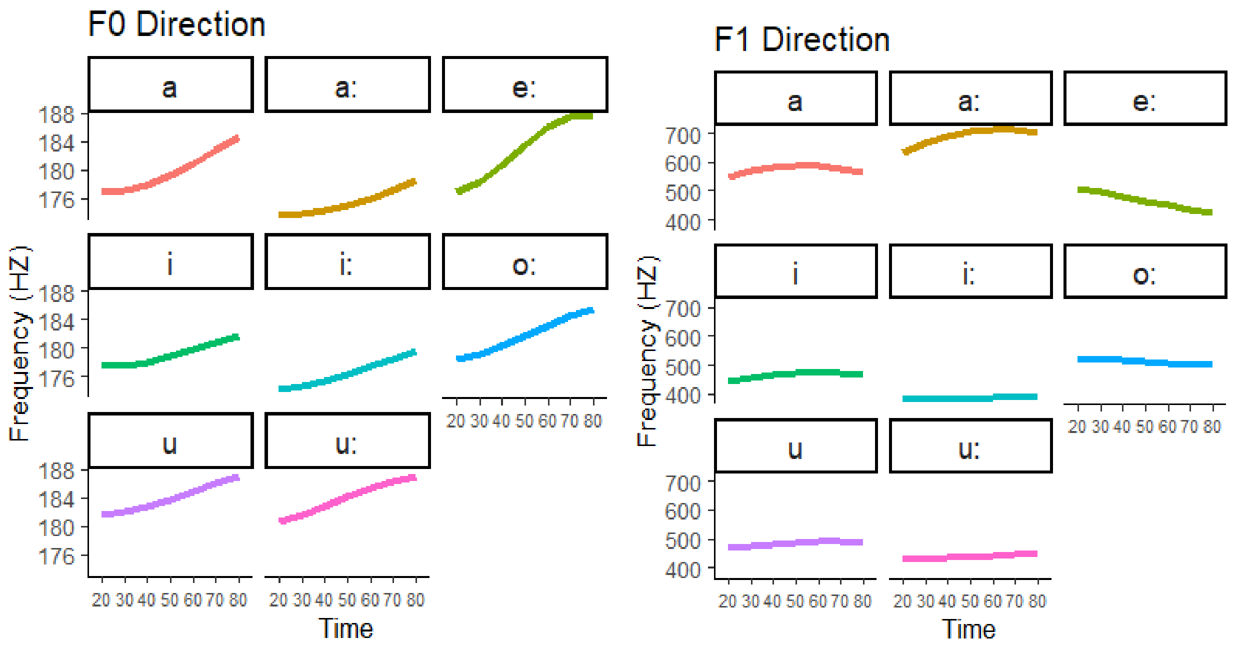


Figure 8: Vowel formant trajectories in the F1-F2 space (measured at seven points) of the Hijazi Arabic vowels. Arrows represent the direction of formant movement.



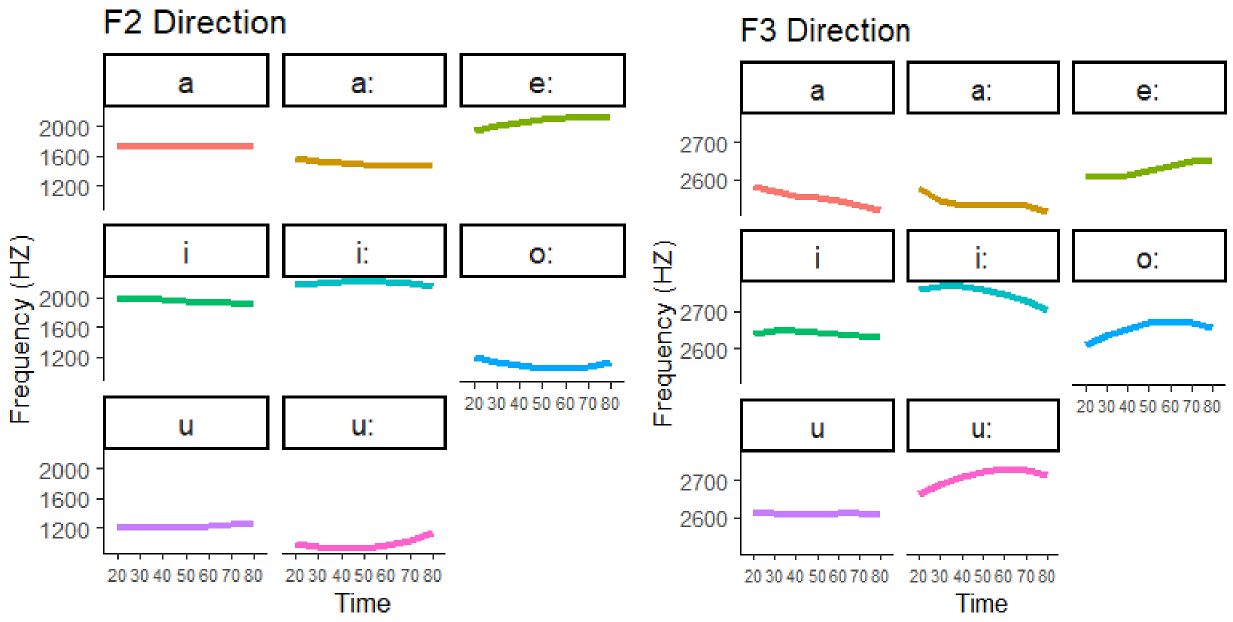


Figure 9: Results of the direction (measured at seven points) of the Hijazi Arabic vowels.

4.2 Discriminant analysis

The QDA results showed that taking seven samples of the vowel duration resulted in the highest classification accuracy (between 77% and 91%, with an average of 85%) for all eight HA vowels, compared to using the other dynamic models, including the three-point model, which came in second place (the correct classification rate being between 69% and 83%, with an average of 76%), and the two-point model, which came in third place (the correct classification rate being between 67% and 83%, with an average of 75%) followed by the static model, which had a classification rate between 61% and 79%, with an average of 71% (see Table 1). However, all four proposed measures obtained their best rates of discrimination accuracy when the combination of F0, F1, F2, and vowel duration was used. The roles of vowel duration, F0, and F3 as additional cues were as follows: The inclusion of the vowel duration with the formant frequencies in any model led to a substantial improvement of 9% to 15% (average of 11%) in vowel separation. On the other hand, the inclusion of F0 in the proposed models improved the discrimination rate of HA vowels by 3% to 5%, or by an average of 4%, whereas with the inclusion of F3, the improvement ranged from 1% to 3%, with an average of 2% overall. Finally, the correct classification rate when using the duration alone was 27%.

	One-point		**Two-point**		**Three-point**		**Seven-point**	
	No Dur	Dur	No dur	Dur	No dur	Dur	No Dur	Dur
F1-F2	61	76	67	79	69	79	77	88
F1-F3	64	78	69	80	70	80	79	89
F0-F2	65	79	72	83	72	83	81	91
F0-F3	66	79	73	83	73	83	82	91

Table 1: Discriminant analysis results showing the percentage in the the classification accuracy of the HA vowels, trained on various combinations of parameters for one-point, two-point, three-point, and seven-point models (F1-F2 indicates F1 and F2; F1-F3 indicates F1, F2, and F3; F0-F2 indicates F0, F1, and F2; F0-F3 indicates F0, F1, F2, and F3).

5. Discussion

5.1 Acoustic correlates

This section discusses the statistical results of the static and dynamic cues of the vowels' production of HA speakers. A table summarizing all significant results can be found in Table 2. As mentioned earlier, full statistical results of the acoustic cues of HA vowels (with *p*-values) can be found in the Appendix, Table A3 .

		Static	**Offset**	**slope**	**Direction 2**	**Direction 3**	**Direction 7**
/a:/ vs /a/	F0	–	–	✓	–	–	–
	F1	✓	✓	✓	✓	✓	✓
	F2	✓	✓	✓	✓	✓	✓
	F3	–	–	–	–	–	–
/u:/ vs /u/	F0	–	–	–	–	–	–
	F1	✓	–	–	✓	✓	✓
	F2	✓	–	–	✓	✓	✓
	F3	✓	–	–	✓	✓	✓
/i:/ vs /i/	F0	–	–	–	–	–	–
	F1	✓	–	✓	✓	✓	✓
	F2	✓	–	✓	✓	✓	✓
	F3	✓	–	–	✓	✓	✓
/o:/ vs /e:/	F0	–	✓	–	–	–	–
	F1	✓	✓	✓	✓	✓	✓
	F2	✓	–	✓	✓	✓	✓
	F3	–	–	–	–	–	–
/e:/ vs /i/	F0	–	✓	–	–	–	–
	F1	–	✓	✓	–	–	–
	F2	✓	✓	✓	✓	✓	✓
	F3	–	–	–	–	–	–
/o:/ vs /u/	F0	–	–	–	–	–	–
	F1	–	✓	✓	✓	✓	✓
	F2	✓	–	✓	✓	✓	✓
	F3	–	–	–	–	–	✓

Table 2: Summary of the statistical results of the acoustic cues of Hijazi Arabic vowels; ticks denote significant results.

5.1.1 Static correlates

The data on the acoustic correlates of HA vowels showed interesting results even when considering static measures alone. For example, the midpoint model showed a significant difference between the HA short and long vowels. The short HA vowels, /i a u/, were centralised compared with their long counterparts, /iː aː uː/, potentially suggesting a lax quality. This result supports other studies (e.g., Rosner et al. 1994; Khattab 2007; Al-Tamimi 2007a,b; Khattab and Al-Tamimi 2008; Almbark and Hellmuth 2015; Almurashi et al. 2020; Al-Mazrouei et al. 2023) that propose long and short Arabic vowels differ in terms of quantity and quality. Such a finding is expected when considering that acoustic duration and length are often interlinked (Almurashi et al. 2020). Although the vowels of HA were separated in the scatter plot (see Figure 3 in the Result section), quite a few variations occurred in the production of some vowels, which was expected because these vowels were produced across a variety of consonant environments rather than a single consonantal context (Hillenbrand et al. 2001; Williams and Escudero 2014; Elvin et al. 2016).

5.1.2 Dynamic correlates

With respect to the offset model, the data revealed that HA monophthongs exhibit a great amount of spectral changes, particularly in the first three formant frequencies, but generally without noticeable differences between HA long and short vowel pairs. Such a result was expected due to the HA vowel system allowing for more variability in production. This finding is in line with those of other researchers, who have noted that speech dynamics are greater for languages with sparse vowel systems (e.g., Manuel 1990; Meunier et al. 2003; Al-Tamimi and Ferragne 2005; Jin and Liu 2013; Mok 2013). Speakers typically fully utilise their phonetic vowel space (Manuel 1990; Meunier et al. 2003). In a dense vowel space less production variability can be tolerated as the speakers have limited freedom to disperse their production of each vowel category in order to avoid overlap between vowels in the phonetic space, which might hamper perception and blur phonological distinctions. In a sparse vowel space, however, speakers have more freedom to disperse their production of vowels without causing considerable blurring of phonetic contrasts that might lead to perceptual confusion (Mok 2013). Further, the amount of spectral movement for HA in this study was found to be greater than the offset results found by Almurashi et al. (2020), who focussed on /hVd/ syllables. This suggests that the properties of vowels within the /hVd/ environment are comparable to their characteristics when produced in isolation (Stevens and House 1963; Oh 2013), while the various consonantal contexts used in this study yielded more spectral movement even within the middle 60% portion of the vowel.

Regarding the slope model, we noticed that HA vowels had positive slopes in most cases, and the higher spectral rate of vowel changes denotes faster spectral movements of HA monophthongal vowels during the vowel duration (Fox and Jacewicz 2009; Farrington et al. 2018). Another important aspect of the slope properties of HA vowels was the different rates of vowel changes between the vowel pairs, particularly the front vowel pairs and in the first two formants; short front vowels had slope values that were different from those of their long front counterparts. This finding suggests that slope models can provide insights into dynamic patterns of realisation for vowel contrasts that are based on temporal as well as spectral contrast (e.g., Fox and Jacewicz 2009; Farrington et al. 2018; Almurashi et al. 2020, among others).

The direction model using two, three, and especially seven points provided the most optimal characterisation of the dynamic patterns of HA vowels production. By way of explanation, the data revealed that the difference between the F1 production of the vowel pair /oː/ vs /u/ was not statistically significant when taking one point located at the steady state of the vowel (e.g., static model). However, in looking at the same vowel pairs using the direction from more than one point (e.g., two, three, and seven points), we found that a significant difference exists. This finding supports the necessity of investigating monophthongal vowels dynamically to represent better and more information about formant spectral movements (e.g., Hillenbrand and colleagues 1995; 2001; Adank, Van Hout and Smits 2004; McDougall 2006; McDougall and Nolan 2007; Almurashi et al. 2020, among others). Importantly, more significant differences were found between the trajectories of the HA vowels using the seven-point direction model than any of the other models looked at here. For example, the F3 production of the vowel pair /o:/ vs /u/ showed no noticeable differences when using the static model or the direction model based on two or three points, whereas extracting multiple points (seven measurements) during the vowel duration revealed a statistically significant difference. Such a result suggests that the more measuring points from the vowel duration, the better the understanding, and the fuller the extent of the vowel spectral changes that might remain unnoticed when formant values are taken from fewer locations (Fox and Jacewicz 2009; Darcy and Mora 2015). The direction model also emphasised some of the same findings as the static model, mainly that the F1 and F2 directions of short vowels are significantly different from those of their long vowel counterparts for HA speakers. This supports findings from other studies on Arabic that short and long Arabic vowels are different not only in terms of their quantity but also their quality (e.g., Khattab 2007; Al-Tamimi 2007a,b; Khattab and Al-Tamimi 2008, among others). Such a result is also in line with acoustic studies (e.g., Watson and Harrington 1999; Slifka 2003; Fox and Jacewicz 2009; Almurashi et al. 2020, among others) that found that using formant trajectories was useful for within-class separation of lax/tense vowels.

Interestingly, the direction results showed another difference among HA vowels where some long vowels such as /e: a: u: and o:/ had a greater amount of diphthongization in production (see Figure 8 in the result section). Such a result for /e:/ and /o:/ was expected since they are derived from the underlying diphthong /aj/ and /aw/ (respectively) in Arabic phonology. The diphthongal trajectories for long /uː/ and /aː/, on the other hand, are considered an intriguing finding and indicate that some monophthongs are characterised by VISC between the vowels' two targets, in much the same way found for diphthongs, and such a finding might be crucial for their perceptual identification.

5.2 Discriminant analysis

The data demonstrate that measuring more than three points (e.g., seven-point model) is the best and most accurate for classifying HA vowels in comparison to the other models. The three-point model came second in terms of performance, followed by the two-point model and finally the static model, which yielded the least accurate classification rate. These results are in line with studies on other languages (e.g., Nearey and Assmann 1986; Huang 1992; Zahorian and Jagharghi 1993; Harrington and Cassidy 1994; Hillenbrand et al. 1995; Hillenbrand and Nearey 1999; Hillenbrand et al. 2001; Neel 2004; Ferguson and Kewley-Port 2002; Arnaud et al. 2011; Yuan 2013; Almurashi et al. 2020). The comparatively low classification rate of the static model suggests that the cues to vowel identification cannot all be revealed from a one-time slice and that the spectral movements perform significant functions in identifying the vowel identity (e.g., Nearey and Assmann 1986; Harrington and Cassidy 1994; Hillenbrand et al. 1995; Hillenbrand and Nearey 1999; Hillenbrand et al. 2001, among others). However, it is worth pointing out that although the static model came last in terms of classification performance, the data still yielded an acceptable classification accuracy.

The QDA results of HA in this study generally yielded relatively lower accuracy rates than those found in Almurashi. et al.'s (2020) for the same vowels in an /hVd/ environment (74.5% for the three-point model, 73.75% for the two-point model, and 69.75% for the static model[12]). The relatively higher averages in Almurashi et al.'s (2020) research may be due to the minimal and more uniform effect of the consonants in the /hVd/ environment. These findings highlight the importance of recognizing the effect of various consonantal contexts on whole vowel trajectories (Hillenbrand et al. 2001; Oh 2013) and to include these in experiments rather than generalizing from results from vowels in isolation or in the /hVd/ context[13].

Despite the efficiency of the F1 and F2 values in identifying vowels, F0 was found to play an important role in classifying HA vowels. F3, on the other hand, had little influence on accurately classifying HA vowels, which is in agreement with other studies (e.g., Hillenbrand et al. 2001; Almurashi et al. 2020), and this may be due to the fact that F3 is a better index for lip rounding and speaker physiology than inherent vowel identity[14]. Importantly, this study highlights that vowel duration has a vital role in accurately classifying HA vowels, which is expected for a language like Arabic with a quantitative vowel contrast (e.g., Almurashi et al. 2020). Including vowel duration increased the separation of vowels when using a discriminant analysis more than is typically found for languages with qualitative vowel contrasts such as English (e.g., Hillenbrand et al. 1995; 2001; Watson and Harrington 1999). This can be explained by considering the phonological role of vowel duration as a cue to distinguishing short and long vowels in HA vowels.

6 Conclusion

---

[12] To make this comparison more reliable, we calculated the average of the HA QDA results in this study based on the F1, F2, and F3 (without the F0) as Almurashi el al. (2020) did in their paper.

[13] To note, these findings regarding the influence of the various consonantal contexts on vowels are primarily based on the QDA classification not from LMM tests performed for the present investigation.

[14] Although the vowel pair /e:/ vs /o:/ is presumably distinct in terms of rounding, the result showed no statistical differences for F3. Hence, further studies are recommended to examine such a pair in more complex consonant environments to provide an in-depth analysis of the role of F3.

The main purpose of this research was to evaluate the role of static versus dynamic F1/F2 cues in describing and classifying HA monophthongal vowels, along with examining the role of vowel duration, F0, and F3 as additional cues. Taken together, both classification and description results showed that the cues to vowel identification improved when the method used went beyond measuring a single steady portion and that inherent vowel variations perform significant functions in terms of describing and classifying monophthongal vowels. According to Tiffany (1953), this single-point target is nearly and undoubtedly very simplistic. Our findings are in line with dynamic approaches and highlight the importance of looking beyond static cues and beyond the first two formants for a comprehensive profiling of the vowels in a given phonological system and for improved representation of cross-linguistic and cross-dialectal differences.

Acknowledgments

This work was supported by Taibah University to the first author (WA) and partially supported by a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex Empirical Foundations of Linguistics – EFL to the second author (JA). We thank all of our subjects who participated in this study.

Ethical approval

Ethical approval to collect this study was obtained from Newcastle University Ethics Committee (Ref: 2427/2017).

Author contributions

The authors confirm contribution to the paper as follows:

WA: Data collection; WA and JAT: Data analysis tools (e.g., PRAAT, R and RStudio); WA, JA, and GK: Made a substantial contribution to the conceptualisation of the article, the analysis and interpretation of data, revising the article critically for important intellectual content, and approving the version to be published.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Appendix

TABLE A1: The set of target words that were used for the HA.

HA Vowels				
HA vowel	**Place of articulation**	**IPA**	**HA word**	**English gloss**
/uː/	Bilabial_ Alveolar	/buːsi/	بُوسي[15]	A female name
		/buːz/	بُوز	Mouth
	Alveolar_Alveolar	/duːd/	دُود	Worms
		/tuːt/	تُوت	Blueberry
	Velar_Alveolar	/kuːsa/	كُوسة	Zucchini
		/kuːra/	كُورة	Ball
/u/	Bilabial_ Alveolar	/burj/	بُرج	Tower
		/burr/	بُر	Wheat
	Alveolar_Alveolar	/duss/	دُس	Hide
		/durj/	دُرج	Drawer
	Velar_Alveolar	/kull/	كُل	Eat
		/guddaːm/	قُدام	Deal
/iː/	Bilabial_ Alveolar	/biːsaːn/	بِيسان	A female name
		/biːr/	بِير	Well
	Alveolar_Alveolar	/ʒadiːd/	جَديد	New
		/diːdaːn/	دِيدان	Worms
	Velar_Alveolar	/kiːs/	كِيس	Bag
		/giːss/	قِيس	Measure
/i/	Bilabial_ Alveolar	/biss/	بِس	Cat
		/bilaːl/	بِلال	A male name
	Alveolar_Alveolar	/diss/	دِس	Hide
		/dirham/	دِرهم	Dirham (Currency)

---

[15] In the Arabic script, ħarakāt (“diacritics”) are used to indicate the short vowels and placed below or above the root consonants.

	Velar_Alveolar	/kidd/	كِد	To work hard
		/kilma/	كِلمة	Word
/a:/	Bilabial_ Alveolar	/ba:ss/	باس	Kissed
		/ba:t/	بات	Slept
	Alveolar_Alveolar	/da:s/	داس	Step
		/miħta:s/	مِحتاس	Messy
	Velar_Alveolar	/ka:s/	كاس	Cup
		/ka:sir/	كاسر	Breaker
/a/	Bilabial_ Alveolar	/bass/	بَسْ	Enough
		/bard/	بَرد	Cold
	Alveolar_Alveolar	/dall/	دَل	Guide
		/dass/	دَس	Hid
	Velar_Alveolar	/kadd/	كَد	Worked hard
		/katt/	كَت	Threw something (Liquid)
/o:/	Bilabial_ Alveolar	/bo:se/	بَوس	Kiss
		/bo:t/	بَوت	Football boot
	Alveolar_Alveolar	/do:la/	دَولة	Country
		/do:ri:/	دَوري	League
	Velar_Alveolar	/ko:t/	كَوت	Jacket
		/ko:la/	كولا	Cola
/e:/	Bilabial_ Alveolar	/be:t/	بَيت	House
		/be:z/	بَيز	Oven mitts
	Alveolar_Alveolar	/de:sam/	دَيسم	A male name
		/te:ss/	تَيس	Male-goat
	Velar_Alveolar	/ge:d/	قَيد	Constraint
		/ke:d/	كَيد	Cunning

TABLE A2: Average of the formant frequencies (at 20%, 30%, 40%, 50, 60%, 70%, and 80%) and vowel duration for each Hijazi Arabic vowel.

		F0 (Hz)	**F1 (Hz)**	**F2 (Hz)**	**F3 (Hz)**	**Duration (ms)**
/u:/	20%	180	428	992	2663	169
	30%	181	431	954	2689	
	40%	182	432	932	2709	
	50%	184	435	924	2720	
	60%	185	440	966	2732	
	70%	186	446	1021	2729	
	80%	186	452	1133	2714	

/**i**ː/	20%	174	379	2173	2757	169
	30%	174	381	2193	2770	
	40%	175	379	2197	2763	
	50%	176	380	2220	2756	
	60%	177	384	2206	2751	
	70%	178	390	2173	2723	
	80%	179	393	2153	2704	
/**a**ː/	20%	173	633	1573	2571	175
	30%	173	670	1538	2548	
	40%	174	688	1500	2533	
	50%	175	702	1491	2514	
	60%	175	716	1471	2538	
	70%	176	716	1464	2538	
	80%	178	700	1462	2510	
/**e**ː/	20%	176	507	1941	2610	187
	30%	178	496	1999	2605	
	40%	180	480	2046	2610	
	50%	183	464	2089	2622	
	60%	186	449	2107	2639	
	70%	187	436	2105	2645	
	80%	187	426	2121	2654	
/**o**ː/	20%	178	515	1194	2608	172
	30%	179	522	1139	2629	
	40%	180	518	1090	2658	
	50%	181	510	1037	2663	
	60%	183	506	1040	2669	
	70%	184	502	1065	2674	

	80%	185	499	1136	2653	
/u/	20%	181	466	1213	2614	80
	30%	181	476	1206	2613	
	40%	182	482	1203	2606	
	50%	183	488	1214	2607	
	60%	184	491	1230	2611	
	70%	186	491	1243	2610	
	80%	186	487	1249	2604	
/i/	20%	177	444	1969	2642	79
	30%	177	455	1968	2644	
	40%	177	463	1962	2644	
	50%	178	469	1953	2648	
	60%	179	472	1947	2640	
	70%	180	471	1915	2627	
	80%	181	467	1901	2630	
/a/	20%	176	547	1720	2582	92
	30%	177	568	1721	2565	
	40%	177	581	1723	2559	
	50%	179	586	1727	2544	
	60%	180	586	1725	2546	
	70%	182	577	1720	2528	
	80%	184	563	1731	2517	

Table A3: The statistical results of the acoustic cues of Hijazi Arabic vowels; grey cells denote non-significant results.

	F0	**F1**	**F2**	**F3**

		Diff	*P* <	Diff	*P* <	Diff	*P* <	Diff	*P* <
/aː/ vs /a/	Static model	-4.03	0.9832	115.1	0.0001	-235.4	0.0001	-29.7	0.9392
	Offset model	0.78	0.9964	27.8	0.0001	81.4	0.0001	32.3	0.5303
	Slope model	-0.05	0.0001	0.28	0.0001	-0.8	0.0001	0.36	0.5565
	Direction model (two-point)	-4.53	0.9689	111.6	0.0001	-208.2	0.0001	-9.03	0.9999
	Direction model (three-point)	-4.36	0.7543	112.8	0.0001	-217.3	0.0001	-15.9	0.9940
	Direction model (seven-point)	-4.44	0.0686	116.5	0.0001	-224.0	0.0001	-12.7	0.9249
/uː/ vs /u/	Static model	0.20	1.0000	-53.1	0.0001	-290.3	0.0001	-112.9	0.0002
	Offset model	-0.63	0.9991	1.02	0.9999	-43.6	0.0679	-12.2	0.9960
	Slope model	-0.01	0.9922	-0.07	0.5969	0.30	0.7266	0.27	0.8425
	Direction model (two-point)	-0.46	1.0000	-36.4	0.0001	-167.9	0.0001	-79.2	0.0001
	Direction model (three-point)	-0.24	1.0000	-42.0	0.0001	-208.7	0.0001	-90.4	0.0001
	Direction model (seven-point)	-0.03	1.0000	-45.1	0.0003	-233.5	0.0001	-98.7	0.0001
/iː/ vs /i/	Static model	-2.45	0.9992	-89.1	0.0001	-266.6	0.0001	-108.0	0.0005
	Offset model	1.68	0.7876	2.78	0.9954	26.8	0.5976	4.23	0.9999
	Slope model	-0.01	0.9890	-0.19	0.0001	0.6	0.0001	-0.15	0.9933
	Direction model (two-point)	-2.74	0.9982	-68.8	0.0001	-228.1	0.0001	-94.8	0.0001
	Direction model (three-point)	-2.64	0.9742	-75.6	0.0001	-240.9	0.0001	-99.1	0.0001

	Direction model (seven-point)	-2.60	0.5824	-79.1	0.0001	-243.0	0.0001	-107.2	0.0001
/oː/ vs /eː/	Static model	-1.88	0.9998	45.7	0.0001	-1051.7	0.0001	40.9	0.7404
	Offset model	-3.9	0.0001	-35.9	0.0001	19.2	0.8929	-13.4	0.9931
	Slope model	-0.01	0.8251	0.39	0.0001	-1.3	0.0001	0.01	1.0000
	Direction model (two-point)	-0.46	1.0000	41.3	0.0001	-865.8	0.0001	79.2	0.3511
	Direction model (three-point)	-0.82	0.9999	42.1	0.0001	-927.8	0.0001	12.7	0.9985
	Direction model (seven-point)	-1.08	0.9935	44.7	0.0001	-958.1	0.0001	24.4	0.3007
/eː/ vs /i/	Static model	-4.74	0.9584	4.31	0.9942	-135.6	0.0001	26.2	0.9694
	Offset model	-6.69	0.0001	-55.4	0.0001	-105.7	0.0001	-33.3	0.4902
	Slope model	-0.02	0.1979	0.71	0.0001	-1.81	0.0001	-0.42	0.3351
	Direction model (two-point)	-2.57	0.9988	-11.4	0.9998	-112.1	0.0006	3.91	1.0000
	Direction model (three-point)	-3.29	0.9225	-6.17	0.9999	-117.3	0.0001	11.3	0.9992
	Direction model (seven-point)	-3.73	0.1567	-2.50	0.9999	-120.1	0.0001	12.8	0.9195
/oː/ vs /u/	Static model	2.08	0.9997	-19.4	0.0831	177.3	0.0001	-56.1	0.3436
	Offset model	-0.90	0.9916	-24.2	0.0001	15.9	0.9586	-30.2	0.6174
	Slope model	0.01	1.0000	0.27	0.0001	1.06	0.0001	-0.33	0.6564
	Direction model (two-point)	2.35	0.9993	-25.7	0.0001	123.4	0.0001	-21.4	0.9964
	Direction model (three-point)	2.26	0.9892	-28.6	0.0001	153.0	0.0001	-33.0	0.7804

Direction model (seven-point)	-0.03	1.0000	-30.2	0.0001	160.9	0.0001	-41.5	0.0056

References

Abdoh, Eman. 2011. A study of the phonological structure and representation of first words in Arabic. PhD thesis, University of Leicester, Leicester, UK.

Adank, Patti, Roel Smits & Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5), 3099-3107. https://doi.org/10.1121/1.1795335

Adank, Patti, Roeland van Hout & Roel Smits. 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America*, 116(3), 1729-1738. https://doi.org/10.1121/1.1779271

Al-Mazrouei, Aisha, Aisha Negm & Vladimir Kulikov. 2023. The vowel system of Qatari Arabic: Evidence for Peripheral/Non-Peripheral distinction between long and short vowels. *Journal of the International Phonetic Association*, 1-19. https://doi.org/10.1017/S0025100323000117

Almbark, Rana, and Sam Hellmuth. 2015. Acoustic analysis of the Syrian vowel system. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, UK

Almurashi, Wael, Jalal Al-Tamimi & Ghada Khattab. 2020. Static and dynamic cues in vowel production in Hijazi Arabic. *The Journal of the Acoustical Society of America*, 147(4), 2917-2927. https://doi.org/10.1121/10.0001004

Al-Tamimi, Jalal. 2007a. *Indices dynamiques et perception des voyelles: étude translinguistique en arabe dialectal et en français* (Dynamic indices and vowel perception: translinguistic study in Arabic and in French dialects). PhD research, University Lyon, Lyon, France, (accessible here in French :http://theses.univ-lyon2.fr/documents/lyon2/2007/al-tamimi_je).

Al-Tamimi, Jalal. 2007b. Static and dynamic cues in vowel production: a cross dialectal study in Jordanian and Moroccan Arabic. In *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, 541-544.

Al-Tamimi, Jalal & Emmanuel Ferragne. 2005. Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2465-2468.

Alzaidi, Muhammad. 2014. Information structure and intonation in Hijazi Arabic. PhD thesis, University of Essex, Colchester, UK.

Arnaud, Vincent, Caroline Sigouina & Johanna-Pascale Roy. 2011. Acoustic description of Quebec French high vowels: First results. In *Proceedings of the 17th ICPhS*, Hong Kong, China, 244-247.

Baayen, Harald, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4)*,* 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, Dale. 2013. Random effects structure for testing interactions in linear mixed-effects models. *Quantitative Psychology and Measurement*, 4, 1–2. https://doi.org/10.3389/fpsyg.2013.00328

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily .2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3)*,* 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. R package version 1.1.26. https://CRAN.R-project.org/package=lme4

Boersma, Paul & David Weenink Boersma. 2022. Praat: Doing phonetics by computer. http://www.praat.org

Cardoso, Amanda. 2015. Dialectology, phonology, diachrony: Liverpool English realisations of PRICE and MOUTH. PhD research, University of Edinburgh, Edinburgh, UK.

Chladkova, Kateřina & Silke Hamann. 2011. High vowels in Southern British English: /u/-fronting does not result in merger. In *Proceedings of the 17th ICPhS*, Hong Kong, China, 476-479.

Clopper, Cynthia & David Pisoni. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32(1), 111-140. https://doi.org/10.1016/S0095-4470(03)00009-3

Darcy, Isabelle & Joan Mora. 2015. Tongue movement in a second language: the case of Spanish/ei/-/e/ for English learners of Spanish. In *Proceedings of the 18th ICPhS*, Glasgow, UK.

Elvin, Jaydene, Daniel Williams & Paola Escudero. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America*, 140(1), 576-581. https://doi.org/10.1121/1.4952387

Fabricius, Anne, Dominic Watt & Daniel Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change*, 21(3), 413-435. https://doi.org/10.1017/S0954394509990160

Farrington, Charlie, Tyler Kendall & Valerie Fridland. 2018. Vowel dynamics in the Southern vowel shift. *American Speech: A Quarterly of Linguistic Usage*, 93(2), 186-222. https://doi.org/10.1215/00031283-6926157

Ferguson, Sarah & Diane Kewley-Port. 2002. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 112(1), 259-271. https://doi.org/10.1121/1.1482078

Fox, Robert. 1983. Perceptual structure of monophthongs and diphthongs in English. *Language and Speech*, 26(1), 21-60. https://doi.org/10.1177/002383098302600103

Fox, Robert & Ewa Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, 126(5), 2603-2618. https://doi.org/10.1121/1.3212921

Gottfried, Michael, James Miller & Donald Meyer. 1993. Three approaches to the classification of American English vowels. *Journal of Phonetics*, 21(3), 205-229. https://doi.org/10.1016/S0095-4470(19)31337-3

Harrington, Jonathan & Stephen Cassidy. 1994. Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language and Speech*, 37(4), 357-373. https://doi.org/10.1177/002383099403700402

Hillenbrand, James. 2013. Static and dynamic approaches to vowel perception. In Geoffrey Morrison & Peter Assmann (Ed.), *Vowel inherent spectral change* (9-30). Springer, Berlin.

Hillenbrand, James & Terrance Nearey. 1999. Identification of resynthesized /hvd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, 105(6), 3509-3523. https://doi.org/10.1121/1.424676

Hillenbrand, James, Michael Clark & Terrance Nearey. 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, 109(2), 748-763. https://doi.org/10.1121/1.1337959

Hillenbrand, James, Laura Getty; Michael Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5), 3099-3111. https://doi.org/10.1121/1.411872

Hothorn, Torsten, Frank Bretz, Peter Westfall, Richard Heiberger, Andre Schuetzenmeister & Susan Scheibe .2016. Package "multcomp": Simultaneous inference in general parametric models. *Project for Statistical Computing*, Vienna, Austria. R package version 1.4-16. http://cran.stat.sfu.ca/web/packages/multcomp/multcomp.pdf

Huang, Caroline. 1992. Modelling human vowel identification using aspects of formant trajectory and context. In Yoh'ichi Tohkura, Eric Vatikiotis-Bateson & Yoshinori Sagisaka (Ed.), *Speech perception, production and linguistic structure* (43-61). IOS Press, Amsterdam.

Jin, Su-Hyun & Chang Liu. 2013. The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America*, 133(5), 363-369. https://doi.org/10.1121/1.4798620

Khattab, Ghada. 2007. A phonetic study of gemination in Lebanese Arabic. In *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, 153-158.

Khattab, Ghada & Jalal Al-Tamimi. 2008. Durational cues for gemination in Lebanese Arabic. *Language and Linguistics*, 11(22), 39-55.

Lobanov, Boris. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America,* 49, 606–608. https://doi.org/10.1121/1.1912396

Manuel, Sharon. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, 88(3), 1286-1298. https://doi.org/10.1121/1.399705

McDougall, Kirsty. 2002. Speaker-characterising properties of formant dynamics: a case study. In *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, 403-408.

McDougall, Kirsty. 2006. Dynamic features of speech and the characterisation of speakers: Towards a new approach using formant frequencies. *International Journal of Speech Language and the Law*, 13(1), 89-126. https://doi.org/10.1558/sll.2006.13.1.89

McDougall, Kirsty & Francis Nolan. 2007. Discrimination of speakers using the formant dynamics of /uː/ in British English. In *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, 1825-1828.

Meunier, Christine, Cheryl Frenck-Mestre, Taissia Lelekov-Boissard & Martine Le Besnerais. 2003. Production and perception of vowels: Does the density of the system play a role? In *Proceedings of the 15th ICPhS*, Barcelona, Spain, 723-726.

Mok, Peggy. 2013. Does vowel inventory density affect vowel-to-vowel coarticulation?. *Language and Speech*, 56(2), 191-209. https://doi.org/10.1177/0023830912443948

Morrison, Geoffrey & Terrance Nearey .2007. Testing theories of vowel inherent spectral change. *The Journal of the Acoustical Society of America*, 122(1), 15-22. https://doi.org/10.1121/1.2739111

Morrison, Geoffrey & Peter Assmann. 2013. *Vowel inherent spectral change*. Springer Science and Business Media. https://doi.org/10.1007/978-3-642-14209-3

Nearey, Terrance. 2013. Vowel inherent spectral change in vowels in North American English. In Geoffrey Morrison & Peter Assmann (Ed.), *Vowel inherent spectral change* (49-85). Springer, Berlin.

Nearey, Terrance & Peter Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, 80(5), 1297-1308. https://doi.org/10.1121/1.394433

Neel, Amy. 2004. Formant detail needed for vowel identification. *Acoustics Research Letters Online*, 5(4), 125-131. https://doi.org/10.1121/1.1764452

Oh, Eunjin. 2013. Dynamic spectral patterns of American English front monophthong vowels produced by Korean-English bilingual speakers and Korean late learners of English. *Linguistic Research*, 30(2), 293-312. https://doi.org/10.17250/khisli.30.2.201308.007

Pinheiro, José, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, Siem Heisterkamp & Bert Van Willigen. 2017. Package "nlme". Linear and nonlinear mixed effects models. R package version 3.1-152. https://CRAN.R-project.org/package=nlme

Peterson, Gordon & Harold Barney. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175-184. https://doi.org/10.1121/1.1906875

R Core Team. 2022. R: A language and environment for statistical computing (version 4.0.4). Vienna, Austria: R Foundation for Statistical Computing. [Software Resource]. ISBN 3-900051-07-0. https://www.R-project.org/

Rosner, Burton & John Pickering. 1994. *Vowel perception and production*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198521389.001.0001

RStudio. 2022. Rstudio: Integrated development environment for R (version 1.4.1103). Boston, MA. [Software Resource]. https://rstudio.com/

Schwartz, Geoffrey. 2021. The phonology of vowel VISC-osity–acoustic evidence and representational implications. *Glossa: A Journal of General Linguistics*, 6(1), 1-30. http://doi.org/10.5334/gjgl.1182

Singmann, Henrik, Ben Bolker, Jake Westfall, Frederik Aust, Mattan Ben-Shachar, Søren Højsgaard, John Fox, Michael Lawrence, Ulf Mertens, Jonathon Love, Russell Lenth & Rune Christensen. 2018. afex: Analysis of factorial experiments. R Package Version 0.28-1. https://CRAN.R-project.org/package=afex.

Slifka, Janet. 2003. Tense/lax vowel classification using dynamic spectral cues. In *Proceedings of the 15th ICPhS*, Barcelona, Spain, 921-924.

Stevens, Kenneth & Arthur S. House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6(2), 111-128. https://doi.org/10.1044/jshr.0602.111

Tiffany, William. 1953. Vowel recognition as a function of duration, frequency modulation and phonetic context. *The Journal of Speech and Hearing Disorders*, 18(3), 289-301. https://doi.org/10.1044/jshd.1803.289

Van Son, Rob & Louis Pols. 1992. Formant movements of Dutch vowels in a text, read at normal and fast rate. *The Journal of the Acoustical Society of America*, 92(1), 121-127. https://doi.org/10.1121/1.404277

Venables, Bill & Brian Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. R package version 7.3-53.1

Watson, Catherine & 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. https://doi.org/10.1121/1.427069

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York. R package version 3.3.3. https://ggplot2.tidyverse.org

Wickham, Hadley. 2017. Package tidyverse: Easily install and load the "tidyverse". R package version 1.3.0. https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf

Wickham, Hadley, Romain François, Lionel Henry & Kirill Müller. 2019. Dplyr: A grammar of data manipulation. R package version 1.0.4. https://CRAN.R-project.org/package=dplyr

Williams, Daniel & Paola Escudero. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America*, 136(5), 2751-2761. https://doi.org/10.1121/1.4896471

Wood, Simon. 2015. Package "mgcv". R package version 1.8-34 http://cran.r-project.org/web/packages/mgcv/mgcv.pdf

Yang, Byunggon. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24(2), 245-261. https://doi.org/10.1006/jpho.1996.0013

Yuan, jiahong. 2013. The spectral dynamics of vowels in Mandarin Chinese. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 1193-1197.

Zahorian, Stephen & Amir Jagharghi. 1993. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, 94(4), 1966-1982. https://doi.org/10.1121/1.407520