DE GRUYTER

Phonetica
International Journal of Phonetic Science

Dynamic Specification of Vowels in Hijazi Arabic

Journal:	*Phonetica*
Manuscript ID	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Almurashi, Wael; Taibah University, Al-Tamimi, Jalal; Université Paris Cité Khattab, Ghada; Newcastle University
Keywords:	Static cues, Dynamic cues, Discriminant analysis, Hijazi Arabic

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]*Laboratoire de Linguistique Formelle, LLF, Université Paris Cité, Paris 75013, 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 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 (50% female) 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.

I. 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 & Assmann 1986; Hillenbrand et al. 1995; 1999; 2001; Morrison & 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 first language (L1) 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 & Barney 1952). Therefore, it is thought to represent the best acoustic characteristic of vowels. Nevertheless, many researchers (e.g., Nearey & Assmann 1986, Hillenbrand 2013, among others) have since found the static model to have important limitations such as not providing sufficient information for describing and discrimination monophthongal vowels, while dynamic correlates were found to be more accurate. These are explored in more detail in the next section.

A. Dynamic approaches to vowel identification and classification using VISC

The term VISC was devised by Nearey and Assmann (1986; Nearey 2013, p. 49) 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. A considerable amount of research has investigated VISC using not only two points [20% and 80%]), but also three [20%, 50%, and 80%] (e.g., Huang 1992; Zahorian & Jagharghi 1993; Harrington & Cassidy 1994; Hillenbrand et al. 1995; Ferguson & Kewley-Port 2002; Yuan 2013, among others), and multiple points [more than three locations] (e.g., Fox 1983; Van Son & Pols 1992; Adank et al. 2004; McDougall 2006; McDougall & Nolan 2007; Al-Tamimi 2007a,b; Fox & Jacewicz 2009, among others), each reported to perform significant functions in terms of describing and classifying monophthongal vowels.

Apart from number of measurements, VISC has three primary accounts, namely: 1) onset + offset: this is known as the offset model. For example, many studies have used the two-point model to capture the amount of vowel inherent dynamics using the offset model and have reported that speech dynamics are greater for languages with a sparse vowel system than for those with a crowded vowel system, 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 & Ferragne 2005; Jin & Liu 2013; Mok 2013; Almurashi et al. 2019; 2020, among others; 2) 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 (i.e. rising/positive) suggesting fast dynamic movement over the vowel's duration and a lower value (i.e. falling/negative) suggesting a slower movement (e.g., Fox & Jacewicz 2009; Farrington et al. 2018; Almurashi et al. 2019; 2020, among others), and 3) onset + direction, or the direction model: this is used to track the direction of spectral changes (Nearey & Assmann 1986; Gottfried et al. 1993; Morrison & Nearey 2007; Morrison & Assmann 2013). Using this model, many researchers note that when formant trajectories are made at many locations rather than extracted from one single point, vowels be characterised more effectively (e.g., Watson & Harrington 1999; Slifka 2003; Chladkova & Hamann 2011; Almurashi et al. 2019; 2020, among others). Research applying the direction model using multiple measurements has taken 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 a few locations (Fox & Jacewicz 2009; Darcy & Mora 2015).

In terms of classification accuracy, many acoustic studies have used discriminant analysis[1] as a classification tool to evaluate the role of static and dynamic models, as well as the role of vowel duration, fundamental frequency (F0), and third formant frequency (F3) as additional cues in identifying monophthong vowels. Some studies have 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. 2019; 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 & Jagharghi 1993; Harrington & Cassidy 1994; Hillenbrand et al. 1995; Ferguson & Kewley-Port 2002, Yuan 2013, among others). Another line of dynamic cues

[1] The discriminant analysis is a statistical method that many studies have used in predicting listeners' categorisation patterns (e.g., Hillenbrand and colleagues 1995; 2001).

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 identifying monophthongal vowels (e.g., Neel 2004). All lines of dynamic studies have also reported that using various additional cues such as the vowel duration, F0, and F3 can aid in the vowel classification accuracy.

B. Dynamic approaches to vowel identification and classification in Arabic

In work on Arabic, the majority of 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 and Moroccan Arabic 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. The second study was conducted by Almurashi et al. (2019; 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—namely /i iː eː a aː oː u and uː/—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 & 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. (2019; 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. (2019; 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. Both Arabic studies concluded that looking at the internal transition behaviour of vowels can be useful in providing a better overview and information of the vowels' properties.

In terms of classification accuracy, both dynamic studies in Arabic (Al-Tamimi 2007a,b; Almurashi et al. 2019; 2020) examined their data by using discriminant analysis to evaluates the role of static versus dynamic cues. 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). Similarly, Almurashi et al. (2019; 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 the three-point approach is the best and most accurate for classifying HA vowels and highlighted that vowel duration is the most important additional cue for the classification accuracy of HA vowels, more than F3.

II. THE CURRENT STUDY

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. (2019; 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). Additionally, they 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 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 & 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. Therefore, this research expands on Almurashi et al.'s (2019; 2020) study by investigating HA vowels more deeply in various phonetic environments, which is recommended by many researchers (e.g., Hillenbrand et al. 1995; Watson & 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 presents 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 vowel production of HA vowels; we also explore to what extent vowel duration, F0, and F3 act as additional cues to classification accuracy.

III. METHODOLOGY

A. Subjects and material

The participants were 20 HA speakers of both genders who were born and raised in Hijaz in the north-west of Saudi Arabia. They 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/, which means '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). Together, the HA stimuli consisted of 8 vowels × 2 words × 3 different consonantal contexts × 3 repetitions × 20 HA participants = 2,880 items.

B. Acoustic analysis

Acoustic analysis was conducted using Praat (Boersma & Weenink 1992–2022). The sound files were manually labelled for each token. The onset and offset of the vocalic segment were manually labelled for each monosyllabic and disyllabic syllable using wide-band and waveform spectrograms 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 (e.g., Nearey 1989; Adank et al. 2004) was run on the midpoints of the first two formant values[2] (on a speaker-by-speaker basis) in RStudio (version 1.4.1103; 2022) and R Core Team (version 4.0.4; 2022).

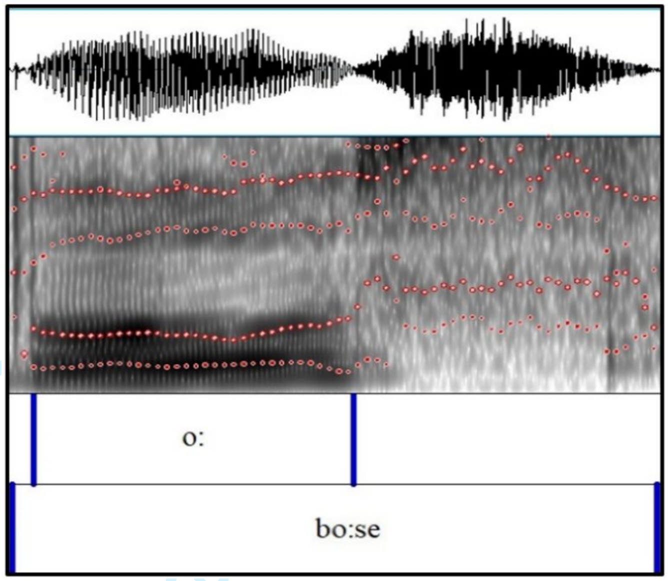


Figure 1: Spectrogram showing formant frequencies of the word ‘boːse’ as produced by a female HA speaker.

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 (Boersma & Weenink 1992–2022). 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[3]) across the vowel 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 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 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

[2] 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.
[3] Taking more than this 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).

(for multiple points). All formant values were manually verified and any errors in formant estimation were corrected by hand.

C. 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 highly significant model, followed by pairwise comparisons (post-hoc tests) with the multcomp package (version 1.4-16; Hothorn et al. 2016) to determine the statistical significance of the study results (McDougall 2002; Fox & 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. 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 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 & Ripley 2002) to obtain the quadratic discriminant analysis (QDA) with a *leave-one-out* cross-validation, or 'jackknife' (Hillenbrand et al. 1995; Almurashi et al. 2020). 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. 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).

IV. 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 I. In addition, full statistical results of the acoustic cues of HA vowels can be found in the Appendix, Table II.

A. Overall patterns of Hijazi Arabic vowels

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 a scatter plot and box plot to characterise the vowels' acoustic features (see Figures 2 with normalisation; and 3 without 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, F0: $\chi^2(2) = 238.2$, $p < 0.0001$; F1: $\chi^2(2) = 87.2$, $p < 0.0001$; F2: $\chi^2(2) = 260.7$, $p < 0.0001$; F3: $\chi^2(2) = 77.2$, $p < 0.0001$. The results of the pairwise comparisons for the /aː/ and /a/ pair showed

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

overall high F1 and low F2 frequencies for /a:/ (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 /i:/ and /i/ pair, the results showed overall low F1 and high F2 frequencies for /i:/ (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 /u:/ vs /u/, the results showed overall low F1 and F2 frequencies for /u:/ (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 /o:/ vs /e:/, the results showed overall high F1 and low F2 frequencies for /o:/ (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.

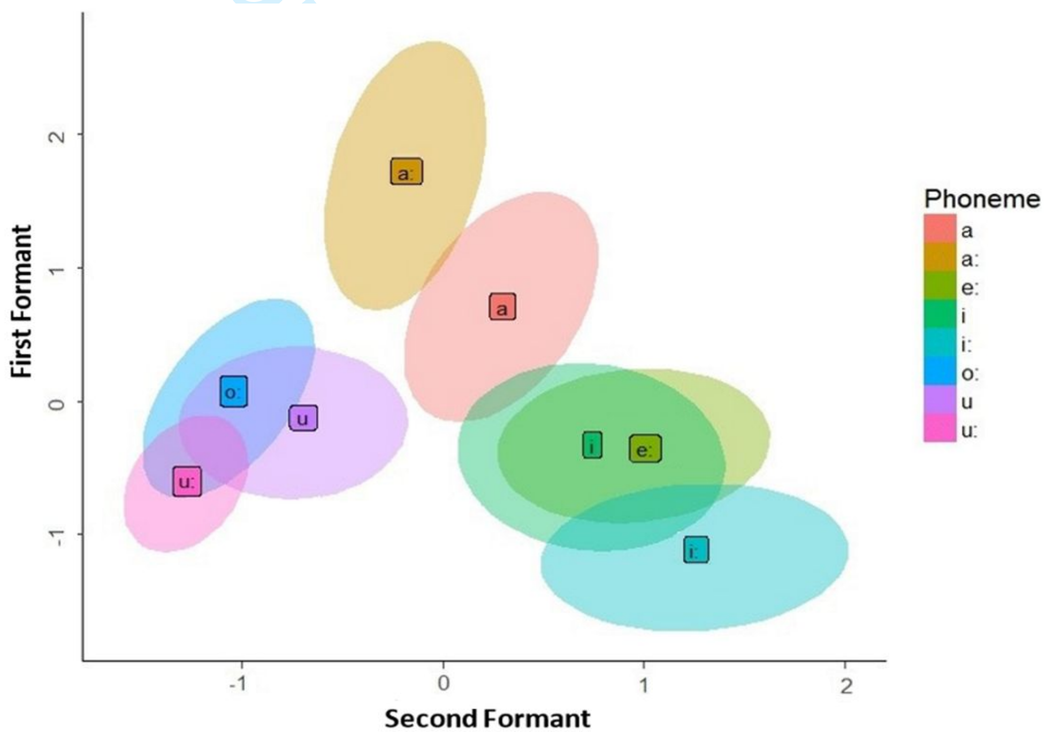


Figure 2: Scatter plot of the normalised midpoints of the first two formant values of the Hijazi Arabic vowels.

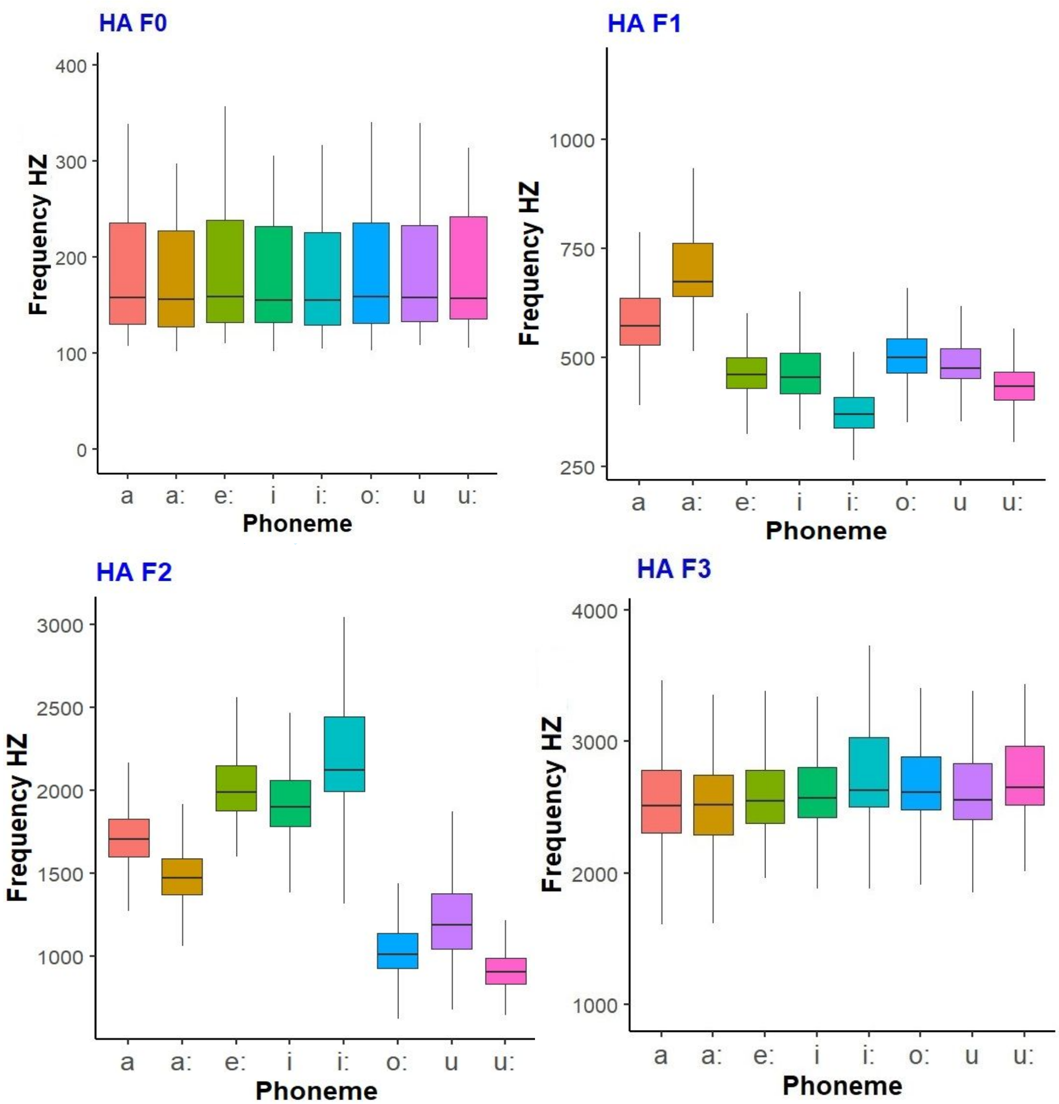


Figure 3: Box plot of the midpoint values of the Hijazi Arabic vowels.

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, F0: $\chi^2(2) = 327.7$, $p < 0.0001$; F1: $\chi^2(2) = 38.2$, $p < 0.0001$; F2: $\chi^2(2) = 26.5$, $p < 0.0001$; F3: $\chi^2(2) = 17.6$, $p < 0.0001$. 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, $p < 0.0001$; for F2, /aː/ had a positive difference of 81.4, $p < 0.0001$;

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, $p < 0.0001$), F1 (had a negative difference of -35.9, $p < 0.0001$), with no differences for F2 and F3. The largest spectral shift across all eight HA vowels (in F0 and F1) was found for /eː/.

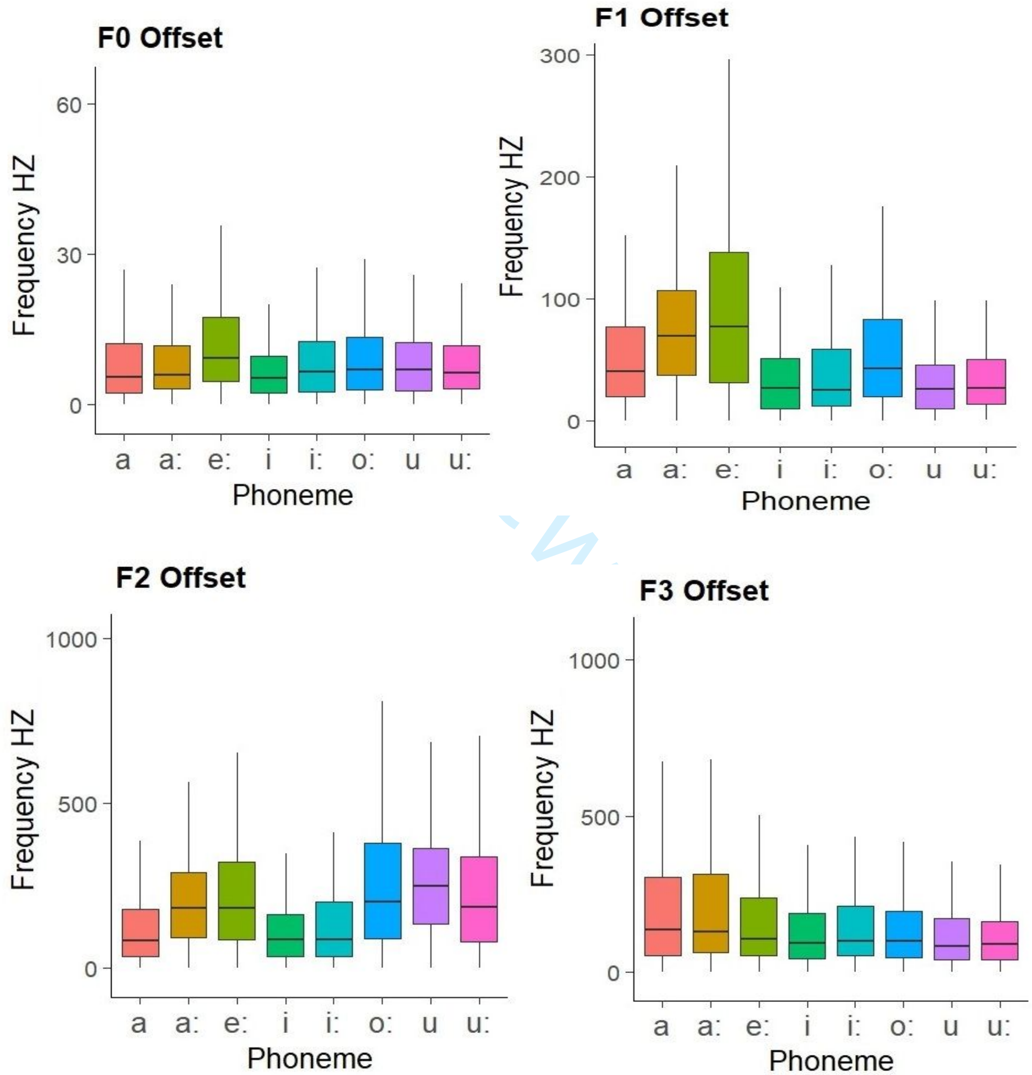
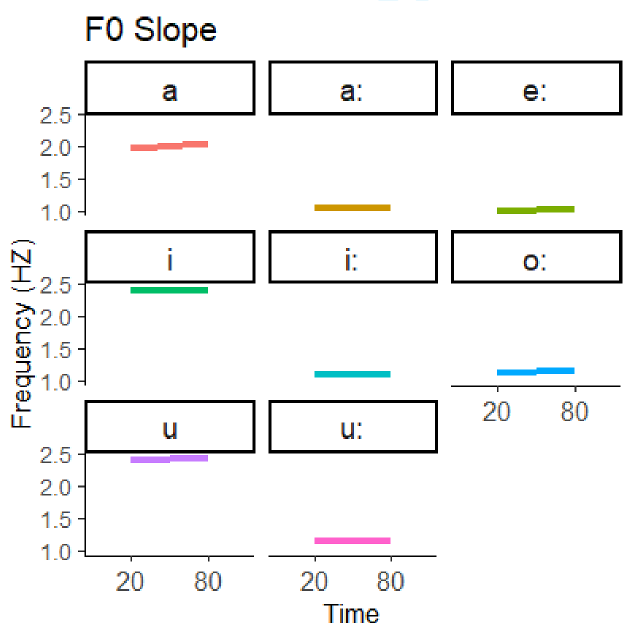
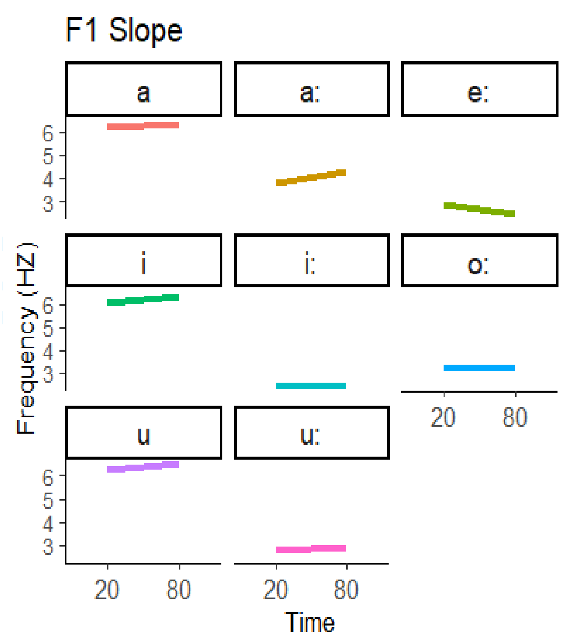


Figure 4: Box plot of the offset model for the Hijazi Arabic vowels.

Regarding the slope of HA from two-point model, Figure 5 shows potential differences among its vowels, with some vowels having their own slope feature for each formant. More specifically, the LMM comparison showed clear improvement to the model fit when using mdl.2, F0: $\chi^2(2) = 189.9$, $p < 0.0001$; F1: $\chi^2(2) = 33.4$, $p < 0.0001$; F2: $\chi^2(2) = 11.3$, $p < 0.0001$; F3: $\chi^2(2) = 27.0$, $p < 0.0001$. Comparison of vowel pairs showed that for /a:/ and /a/, there was an overall negative slope related to the transition of /a:/ for F0 (difference of -0.05, $p < 0.0001$), a positive slope for F1 (difference of 0.2, $p < 0.0001$), a negative slope for F2 (difference of -0.8, $p < 0.0001$), and no difference for F3. For /i:/ and /i/, the results showed a falling slope for F1 (difference of -0.19, $p < 0.0001$), a rising slope for F2 (difference of 0.6, $p < 0.0001$), and no differences for F0 and F3. For /u:/ and /u/, the results showed no significant slopes for F0, F1, F2, and F3. For the pair /o:/ vs /e:/, the results showed a significant slope with overall positive difference for F1 (difference of 0.39, $p < 0.0001$) and a negative transition for F2 (difference of -1.3, $p < 0.0001$), with no significant slopes for F0 and F3.





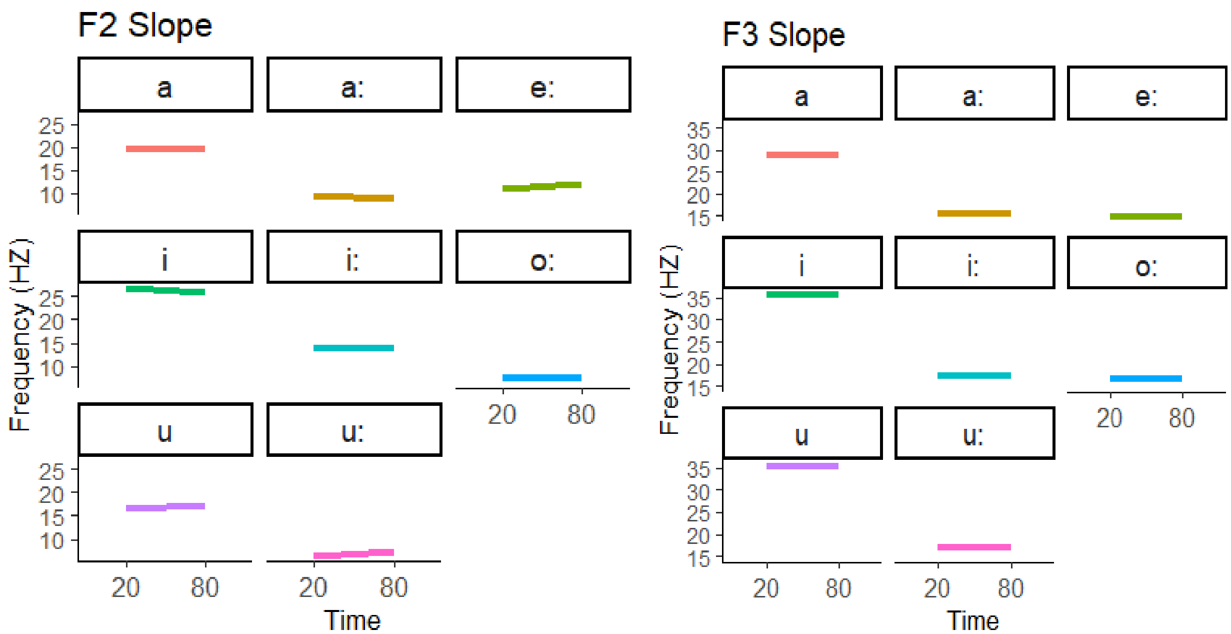


Figure 5: Results of the slope model of the Hijazi Arabic vowels.

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, F0: $\chi^2(2) = 277.2$, $p < 0.0001$; F1: $\chi^2(2) = 134.0$, $p < 0.0001$; F2: $\chi^2(2) = 152.5$, $p < 0.0001$; F3: $\chi^2(2) = 93.1$, $p < 0.0001$. Comparison of vowel pairs showed that for /a/ and /a:/, there was an overall low direction related to the transition of /a:/ for F1 (difference of -111.6, $p < 0.0001$), a high direction for F2 (difference of 208.2, $p < 0.0001$), and no differences for F0 and F3. For /i/ and /i:/, the results showed no differences for F0 but significant differences in direction for F1, F2, and F3: high for F1 (difference of 68.8, $p < 0.0001$), low for F2 (difference of -228.1, $p < 0.0001$), and low for F3 (difference of -94.8, $p < 0.0001$). For the pair of /u/ vs /u:/, the results showed overall significant changes in direction for /u:/ 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 /o:/ vs /e:/, the results showed significant differences in directions with an overall high difference for F1 (a high transition difference of 41.3, $p < 0.0001$) and low difference for F2 (a low transition difference of -865.8, $p < 0.0001$), with no significant changes in directions for F0 and F3.

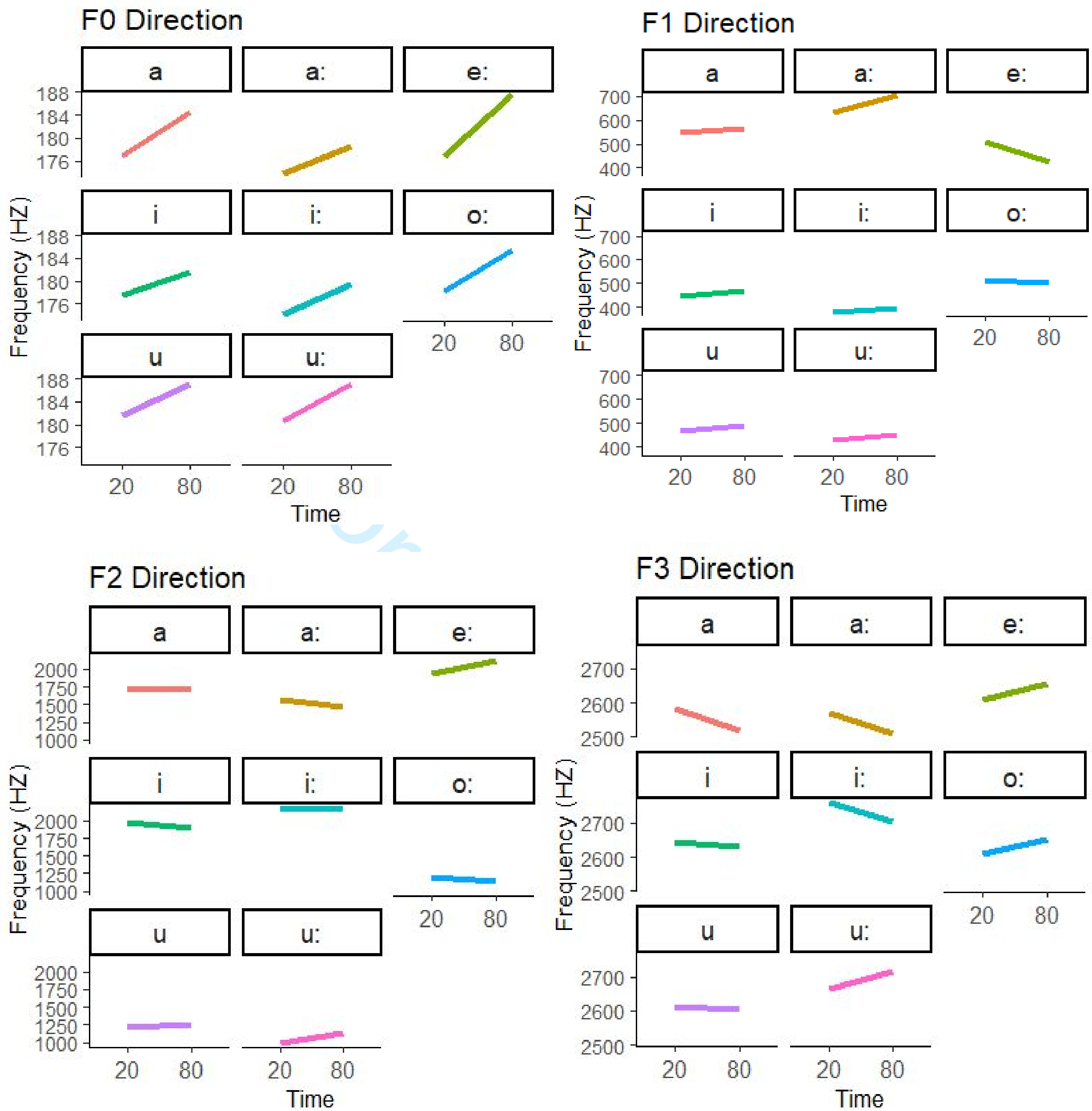
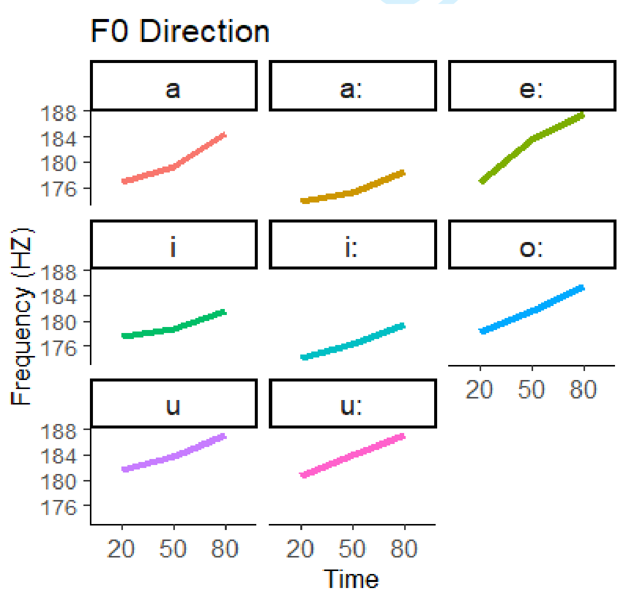
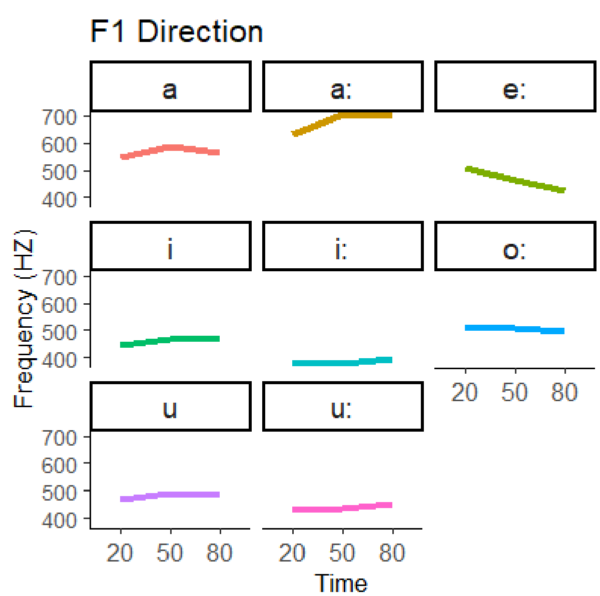


Figure 6: Results of the direction (measured at two points) of the Hijazi Arabic vowels.

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, F0: $\chi^2(2) = 277.6$, $p < 0.0001$; F1: $\chi^2(2) = 124.8$, $p < 0.0001$; F2: $\chi^2(2) = 246.7$, $p < 0.0001$; F3: $\chi^2(2) = 130.7$, $p < 0.0001$. Comparing vowel pairs showed the following for /a/ and /a:/: a low direction for F1 (transition difference of -112.8, p <

0.0001), a high direction for F2 (difference of 217.3, $p < 0.0001$), and no differences for F0 and F3. For /i/ and /i:/, 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, $p < 0.0001$) and low directions for F2 (difference of -240.9, $p < 0.0001$) and F3 (difference of -99.1, $p < 0.0001$). For /u/ and /u:/, the results showed no differences for F0 and overall significant changes 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 /o:/ vs /e:/, the results showed significant differences in directions with an overall high difference for F1 (a high transition difference of 42.1, $p < 0.0001$), and low difference for F2 (a low transition difference of -927.8, $p < 0.0001$), with no significant changes in directions for F0 and F3.





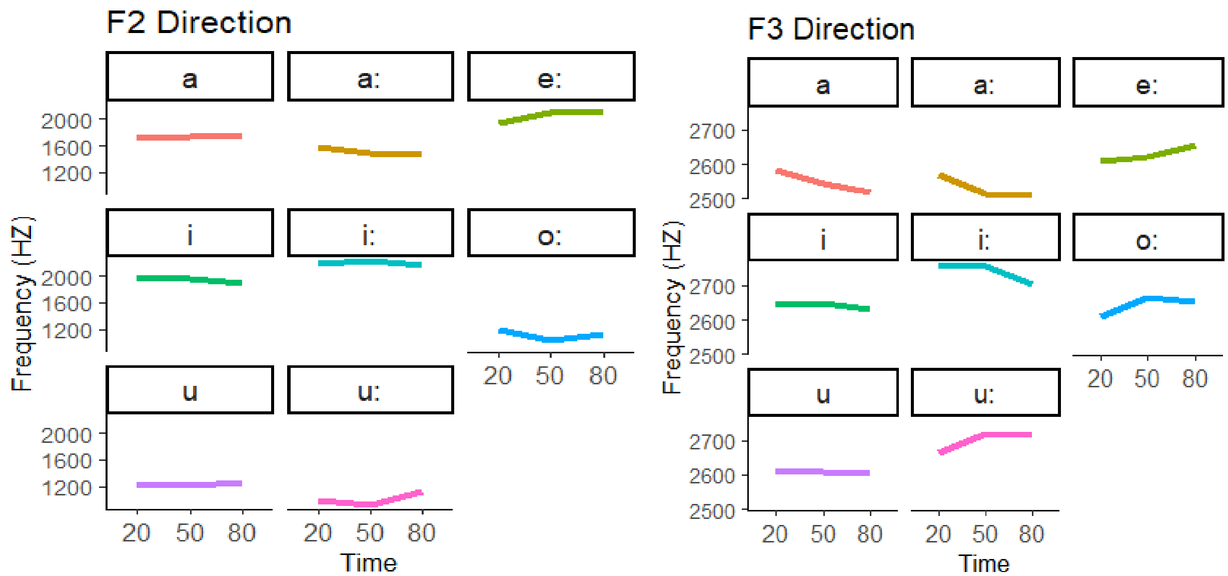


Figure 7: Results of the direction (measured at three points) of the Hijazi Arabic vowels.

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:/ 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, F0: $\chi^2(2) = 262.9$, $p < 0.0001$; F1: $\chi^2(2) = 118.1$, $p < 0.0001$; F2: $\chi^2(2) = 188.8$, $p < 0.0001$; F3: $\chi^2(2) = 139.0$, $p < 0.0001$. Comparing vowel pairs showed that for /a/ and /a:/, there were significant differences related to /a:/ 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 /i:/, 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 /u:/ and /u/, the results showed significant changes 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 /o:/ vs /e:/, the results showed significant differences in directions, with an overall high difference for F1 (a high transition difference of 44.7, $p < 0.0001$), a low difference for F2 (a low transition difference of -958.1, $p < 0.0001$), with no significant changes in directions for F0 and F3.

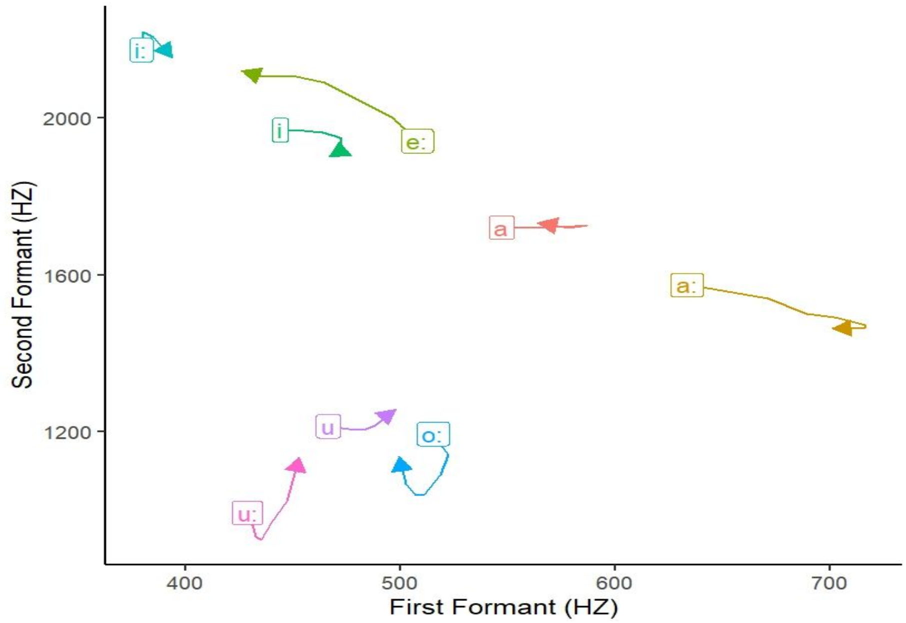
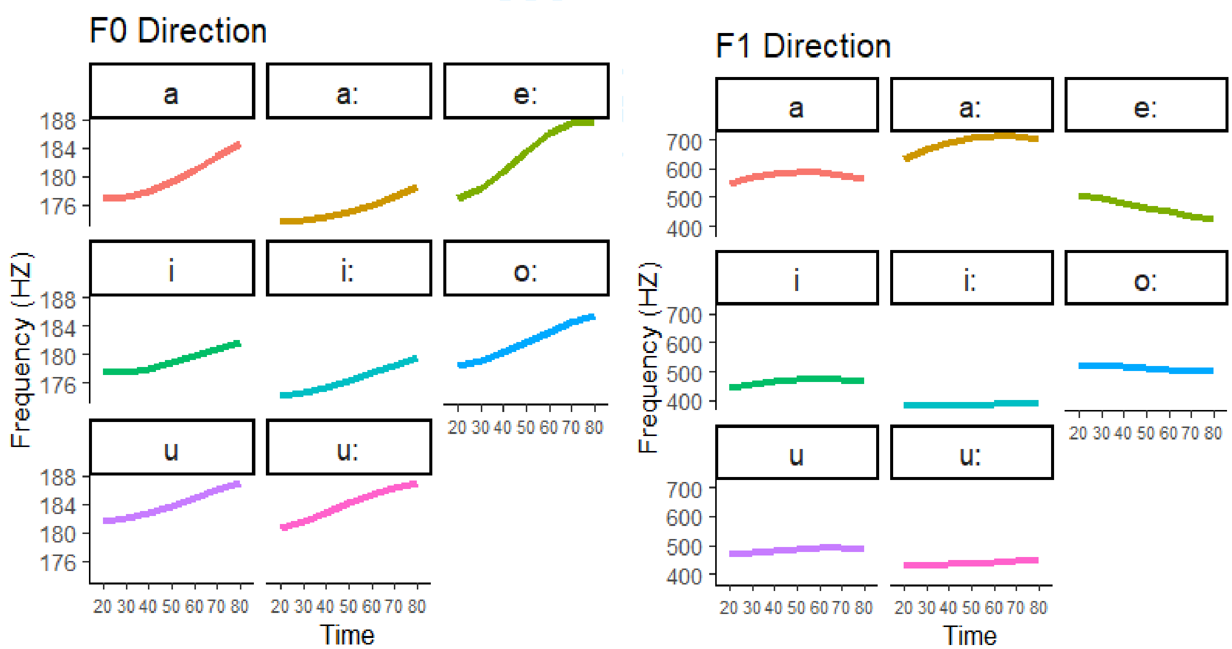


Figure 8: Vowel formant trajectories in the F1-F2 space (measured at seven points) of the Hijazi Arabic vowels. Arrows represent the direction of formant movement.



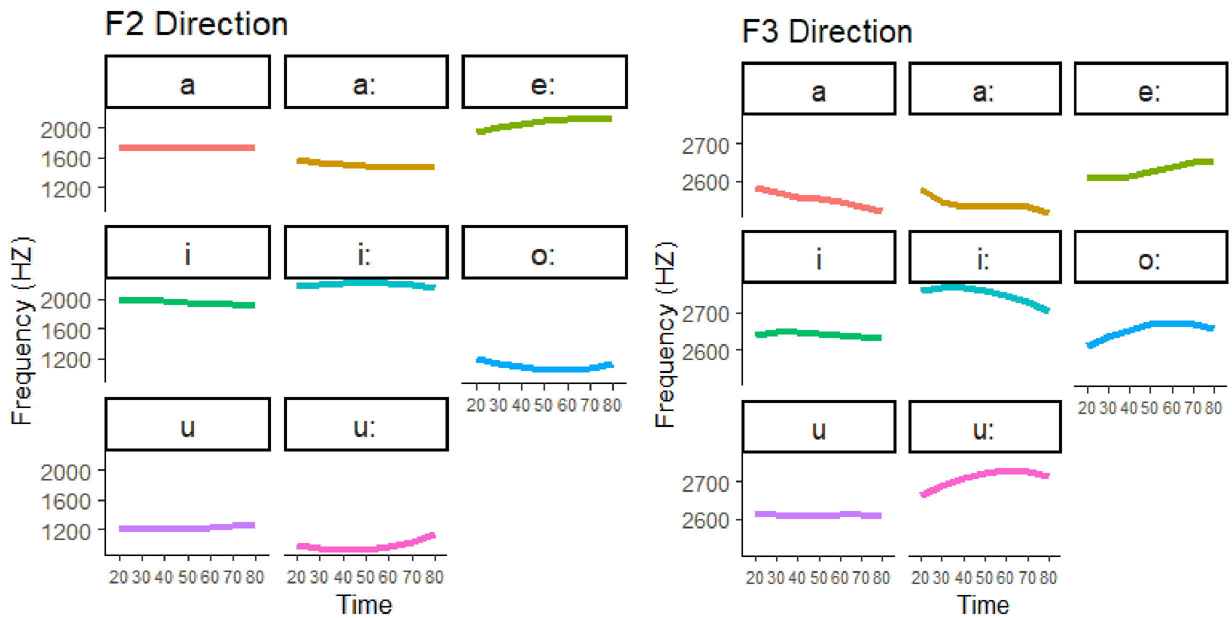


Figure 9: Results of the direction (measured at seven points) of the Hijazi Arabic vowels.

B. 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, F3, 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 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).

V. DISCUSSION

A. Acoustic correlates

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 & Al-Tamimi 2008; Almbark & Hellmuth 2015; Almurashi et al. 2019; 2020) that propose long and short Arabic vowels have differences differ in terms of quantity and quality. Such a finding is expected when considering that articulatory duration and tenseness are often interlinked (Almurashi et al. 2020). Although the vowels of HA were separated in the scatter plot, 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 & Escudero 2014; Elvin et al. 2016).

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 & Ferragne 2005; Jin & 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. (2019; 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 & 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 was exhibited in faster spectral changes of HA monophthongal vowels during the vowel duration (Fox & 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 & Jacewicz 2009; Farrington et al. 2018; Almurashi et al. 2019; 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. 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, suggesting that the full extent of a vowel's spectral change can only be obtained when more samples are extracted across a vowel's duration, s (e.g., Hillenbrand and colleagues 1995; 2001; Adank et al. 2004; McDougall 2006; McDougall & Nolan 2007; Almurashi et al. 2019; 2020, among others). Such a result supports the necessity of investigating monophthongal vowels dynamically to represent better and more information about formant spectral movements that might be dialect specific and that might remain unnoticed when formant values are taken from fewer locations (Fox & Jacewicz 2009; Darcy & 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 & Al-Tamimi 2008, among others). Such a result is also in line with acoustic studies (e.g., Watson & Harrington 1999; Slifka 2003; Fox & Jacewicz 2009; Almurashi et al. 2019; 2020, among others) that found that using formant trajectories was useful for within-class separation of lax/tense vowels.

B. 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 & Assmann 1986; Huang 1992; Zahorian & Jagharghi 1993; Harrington & Cassidy 1994; Hillenbrand et al. 1995; Hillenbrand & Nearey 1999; Hillenbrand et al. 2001; Neel 2004; Ferguson & Kewley-Port 2002; Arnaud et al. 2011; Yuan 2013; Almurashi et al. 2019; 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 & Assmann 1986; Harrington & Cassidy 1994; Hillenbrand et al. 1995; Hillenbrand & 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 (2019; 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[5]). The relatively higher averages in Almurashi et al.'s (2019; 2020) research may be due to

[5] 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. (2019; 2020) did in their paper.

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 the /hVd/.

Despite the efficiency of the F1 and F2 values in identifying vowels, this study highlights that vowel duration is the most important additional cue for accurately classifying HA vowels, which is expected for a language like Arabic with a quantitative vowel contrast (e.g., Almurashi et al. 2019; 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 English (e.g., Hillenbrand et al. 1995; 2001; Watson & 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. Although F0 was not considered the most useful additional cue in classifying HA vowels, it was found to play an important role, as it came in second place after vowel duration. F3 had little influence on accurately classifying HA vowels, which is in agreement with other studies (e.g., Hillenbrand et al. 2001; Almurashi et al. 2019; 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.

VI. CONCLUSION

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 the French Investissements d'Avenir—Labex EFL program (ANR-10-LABX-0083), contributing to the IdEx Université Paris Cité—ANR-18-IDEX-0001 to the second author (JA). We thank all of our subjects who participated in this study.

APPENDIX

TABLE I: 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%	180.6498	428.7802	992.8039	2663.897	169.5198
	30%	181.7039	431.8196	954.1682	2689.587	
	40%	182.8093	432.7631	932.9327	2709.881	
	50%	184.0104	435.1163	924.4994	2720.124	
	60%	185.2902	440.102	966.5947	2732.844	
	70%	186.3553	446.9516	1021.7249	2729.116	
	80%	186.9829	452.6375	1133.9975	2714.066	
/**i**:/	20%	174.283	379.5073	2173.5059	2757.674	169.405
	30%	174.7038	381.7072	2193.6776	2770.5	
	40%	175.3231	379.6633	2197.5666	2763.658	
	50%	176.3967	380.1656	2220.2814	2756.57	
	60%	177.3359	384.7984	2206.1459	2751.82	
	70%	178.52	390.3901	2173.9366	2723.554	
	80%	179.5327	393.8791	2153.8847	2704.766	
/**a**:/	20%	173.717	633.2845	1573.9027	2571.411	175.2373
	30%	173.7319	670.6459	1538.9286	2548.519	
	40%	174.2721	688.8972	1500.0218	2533.305	
	50%	175.1598	702.0196	1491.6184	2514.625	
	60%	175.9754	716.3627	1471.2804	2538.406	
	70%	176.8994	716.5109	1464.8424	2538.246	
	80%	178.6306	700.9156	1462.072	2510.018	
/**e**:/	20%	176.9545	507.8262	1941.2462	2610.413	187.1906
	30%	178.2568	496.2518	1999.8325	2605.394	
	40%	180.6544	480.3074	2046.0682	2610.061	
	50%	183.5993	464.9599	2089.294	2622.35	
	60%	186.1766	449.8387	2107.9941	2639.256	
	70%	187.3743	436.1	2105.7021	2645.579	
	80%	187.4905	426.1607	2121.5285	2654.821	
	20%	178.3292	515.2156	1194.0021	2608.857	

/o:/	30%	179.2004	522.0503	1139.7825	2629.837	172.403
	40%	180.3707	518.424	1090.5806	2658.942	
	50%	181.7193	510.7365	1037.54	2663.347	
	60%	183.1565	506.4708	1040.1466	2669.499	
	70%	184.5793	502.3755	1065.2634	2674.895	
	80%	185.5316	499.5322	1136.9756	2653.57	
/u/	20%	181.5851	466.3882	1213.7435	2614.689	80.58266
	30%	181.9624	476.0003	1206.1687	2613.102	
	40%	182.7677	482.9297	1203.8708	2606.622	
	50%	183.8025	488.2498	1214.8568	2607.204	
	60%	184.91	491.4543	1230.5264	2611.159	
	70%	186.0738	491.4931	1243.5199	2610.377	
	80%	186.9784	487.9431	1249.0159	2604.763	
/i/	20%	177.6551	444.0975	1969.3726	2642.143	79.24623
	30%	177.5232	455.1825	1968.075	2644.352	
	40%	177.9987	463.8834	1962.2018	2644.157	
	50%	178.8537	469.278	1953.6558	2648.557	
	60%	179.8171	472.4822	1947.1768	2640.659	
	70%	180.8424	471.9488	1915.5386	2627.314	
	80%	181.6487	467.0413	1901.7636	2630.928	
/a/	20%	176.9058	547.0173	1720.7056	2582.246	92.56688
	30%	177.2493	568.8575	1721.9476	2565.35	
	40%	177.8755	581.4128	1723.0621	2559.221	
	50%	179.1941	586.879	1727.0272	2544.414	
	60%	180.9244	586.9024	1725.762	2546.502	
	70%	182.8724	577.9854	1720.5034	2528.505	
	80%	184.5074	563.8549	1731.7834	2517.257	

Table II: 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<
/a:/ 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
	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

/uː/ vs /u/	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

Adank, P., Smits, R., and Van Hout, R. (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, P., Van Hout, R., and Smits, R. (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

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

Almurashi, W., Al-Tamimi, J., and Khattab, G. (2019). Static and dynamic cues in vowel production in Hijazi Arabic. In *Proceedings of the 19th ICPhS*, Melbourne, Australia, 3468-3472.

Almurashi, W., Al-Tamimi, J., and Khattab, G. (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, J. (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, J. (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, J., and Ferragne, E. (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.

Arnaud, V., Sigouin, C., and Roy, J. P. (2011). Acoustic description of Quebec French high vowels: First results. In *Proceedings of the 17th ICPhS*, Hong Kong, China, 244-247.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (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, P., and Weenink, D. (1992–2022). Praat: Doing phonetics by computer. http://www.praat.org

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

Chladkova, K., and Hamann, S. (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, C. G., and Pisoni, D. B. (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, I., and Mora, J. C. (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, J., Williams, D., and Escudero, P. (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

Farrington, C., Kendall, T., and Fridland, V. (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, S. H., and Kewley-Port, D. (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, R. A. (1983). Perceptual structure of monophthongs and diphthongs in English. *Language and Speech*, 26(1), 21-60. https://doi.org/10.1177/002383098302600103

Fox, R. A., and Jacewicz, E. (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, M., Miller, J. D., and Meyer, D. J. (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, J., and Cassidy, S. (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, J. M. (2013). Static and dynamic approaches to vowel perception. In G. S. Morrison, and P. F. Assmann (Ed.), *Vowel inherent spectral change* (9-30). Springer, Berlin.

Hillenbrand, J. M., and Nearey, T. M. (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, J. M., Clark, M. J., and Nearey, T. M. (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, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (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, T., Bretz, F., Westfall, P., Heiberger, R.M., Schuetzenmeister, A., and Scheibe, S. (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, C. B. (1992). Modelling human vowel identification using aspects of formant trajectory and context. In Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ed.), *Speech perception, production and linguistic structure* (43-61). IOS Press, Amsterdam.

Jacewicz, E., and Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in American English vowels. In G. S. Morrison, and P. F. Assmann (Ed.), *Vowel inherent spectral change* (177-198). Springer, Berlin.

Jin, S. H., and Liu, C. (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, G. (2007). A phonetic study of gemination in Lebanese Arabic. In *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, 153-158.

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

Manuel, S. Y. (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, K. (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, K. (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, K., and Nolan, F. (2007). Discrimination of speakers using the formant dynamics of /uː/ in British English. In *Proceedings of the 16th ICPhS*, Saarbrücken, Germany, 1825-1828.

Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Le Besnerais, M. (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.

Miller, J. (1989). Auditory-perceptual interpretation of the vowel. *The Journal of the Acoustical Society of America*, 85(5), 2114-2134. https://doi.org/10.1121/1.397862

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

Morrison, G. S., and Nearey, T. M. (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, G. S., and Assmann, P. (2013). *Vowel inherent spectral change*. Springer Science and Business Media. https://doi.org/10.1007/978-3-642-14209-3

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088-2113. https://doi.org/10.1121/1.397861

Nearey, T. M. (2013). Vowel inherent spectral change in vowels in North American English. In G. S. Morrison, and P. F. Assmann (Ed.), *Vowel inherent spectral change* (49-85). Springer, Berlin.

Nearey, T. M., and Assmann, P. F. (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

Oh, E. (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, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., and Van Willigen, B. (2017). Package 'nlme'. Linear and nonlinear mixed effects models. R package version 3.1-152. https://CRAN.R-project.org/package=nlme

Peterson, G. E., and Barney, H. L. (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, B. S., and Pickering, J. B. (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, G. (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, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M., Højsgaard, S., Fox, J., Lawrence, M., Mertens, U., Love, J., Lenth, R., and Christensen, R. (2018). afex: Analysis of factorial experiments. R Package Version 0.28-1. https://CRAN.R-project.org/package=afex.

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

Stevens, K. N., and House, A. S. (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

Strange, W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2081-2087. https://doi.org/10.1121/1.397860

Tiffany, W. R. (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, R. J., and Pols, L. C. (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, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. R package version 7.3-53.1https://CRAN.R-project.org/package=MASS

Watson, J. (2002). *The phonology and morphology of Arabic*. Oxford University Press.

Watson, C. I., and Harrington, J. (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, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York. R package version 3.3.3. https://ggplot2.tidyverse.org

Wickham, H. (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, H., François, R., Henry, L., and Müller, K. (2019). Dplyr: A grammar of data manipulation. R package version 1.0.4. https://CRAN.R-project.org/package=dplyr

Williams, D., and Escudero, P. (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, S. (2015). Package 'mgcv'. R package version 1.8-34 http://cran.r-project.org/web/packages/mgcv/mgcv.pdf

Yang, B. (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, J. (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, S. A., and Jagharghi, A. J. (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