STATIC AND DYNAMIC CUES IN VOWEL PRODUCTION IN HIJAZI ARABIC

Wael Almurashi, Jalal Al-Tamimi and Ghada Khattab

Newcastle University, United Kingdom
W.A.O.Almurashi2@ncl.ac.uk; jalal.al-tamimi@ncl.ac.uk; ghada.khattab@ncl.ac.uk

ABSTRACT

Static cues such as formant measurements obtained at vowel midpoint are regularly taken as the main correlates for the identification of monophthong vowels. However, dynamic cues have been shown to yield better separation of vowels in some languages. This study aims to evaluate the role of static vs dynamic cues in Hijazi Arabic (HA) vowel classification, with vowel duration and F3 as additional cues. Data from 12 male HA speakers producing eight HA vowels in /hVd/ syllables were obtained and evaluated using discriminant analysis. Results show that dynamic cues, particularly the three-point model, had higher classification rates (+98%) than the remaining models. Vowel duration had a significant role in classification accuracy (+11%). Our results are in line with dynamic approaches to vowel classification but also highlight the relative importance of cues across languages; here, the primacy of vowel duration was stark, potentially reflecting the role of length in Arabic phonology.

Keywords: Static cues, dynamic cues, discriminant analysis.

1. INTRODUCTION

Formant frequencies are crucial acoustic correlates for the identification of vowels. For many years, however, the main approach to describing vowels has focused on measuring the first two formants (F1 and F2) at mid-point (e.g. [1], [21], [24], among others). This static approach was followed because it was believed that a vowel's midpoint is the target position which a speaker tries to reach when he/she produces vowels and where a minimal shift in formant value is seen [24]. Nevertheless, subsequent studies have reported that other cues such as dynamic cues—in particular Vowel-Inherent Spectral Changes (VISC) e.g. [3], [19], [22] and the three-point model e.g. [9], [11], [13], —contain essential information, not only for diphthongs but also for monophthong vowels. For instance, discriminant analysis yields better separation of monophthong vowels based on their acoustic measurement when the acoustic parameters are taken from more than one location.

VISC is defined by [22] as the "relatively slowly varying changes in formant frequencies associated with vowels themselves". It is taken from two locations: one around the vowel onset (at 20%) and the other near the vowel offset (at 70-80%) over the full duration of the vowel to eliminate the effects of surrounding consonants [11], [22]. VISC has three primary accounts, namely a) onset + offset (offset model, henceforth), in which the values of the final formant are prioritised, b) onset + slope (slope model, henceforth), which is based on the premise that the rate of change over time is the significant cue, and c) onset + direction (direction model, henceforth), which focusses on the direction of formant frequency changes [8], [18], [22].

Many studies have compared static spectral features with either one of the VISC models (particularly offset) e.g. [10], [11], [12] or all of VISC approaches e.g. [3], and concluded that using VISC models leads to higher correct classification rates than using one point. Moreover, other studies such as [25] and [26] found VISC models to be helpful in improving the separation between lax and tense vowels in English. Regarding the offset, [15] found that Chinese speakers, who have a sparse vowel system, exhibited significantly greater spectral shifts in their productions of vowels than Korean speakers, who have a dense vowel system [16], [17]. Another line of studies e.g. [7], [9], [11], [13], [27], [28], has found a more accurate vowel separation of monophthong vowels when using the three-point model (where formant measures are taken from three locations, namely, at 20% onset, 50% midpoint, and 80% offset during vowel duration) than the midpoint model.

Beyond the first two formants, whose major acoustic correlates of vowel identification all of the aforementioned research has emphasised, the role of third formant (F3) and vowel duration as additional cues have been reported to play a role in vowel discrimination e.g. [11], [12], [26]. For example, [11], who collected their data from /hVd/ syllables, noted that the inclusion of vowel duration increased the separation accuracy of the vowel by 12% in some cases; F3 appeared to have an influence, but not more than the inclusion of vowel duration.

Within research on Arabic, only one dynamic study of vowels has been carried out, but its emphasis

was not on intrinsic dynamic cues; rather, it was focussed on looking at extrinsic dynamic vowel variation (see [2]). Hence, this study is the first step into the field of intrinsic dynamic cues in the Arabic language. With respect to HA vowels, [4], [14], and [20] classified HA vowel production as the following: /i:, a:, u:, i, a, u, e: and o:/. As can be gleaned from the phonemic symbols, Arabic is a quantitative language that relies on vowel duration to form phonemic contrasts [1]. However, there is a debate regarding the tense/lax aspect of Arabic vowels, and a few studies have indicated a difference in both quantity and quality between vowels e.g. [1], [2].

The purpose of the current study is to investigate to what extent the static and dynamic cues, including all VISC models and the three-point model, improve the classification of HA vowels. A second aim it to explore to what extent vowel duration and F3 act as additional cues to classification accuracy. Using these results, we also explore whether HA vowels pairs which differ in phonological length exhibit a difference in term of quality as well as quantity.

2. METHODOLOGY

The participants were 12 male native HA speakers, aged 18 to 30. Recordings were made on a Zoom digital H1 Handy Recorder with a sampling rate of 44,100 Hz and 16-bit amplitude resolution. The HA speakers were asked to produce all vowels in a monosyllabic /hVd/ context within the phrase /ktoːb marteːn/, which means “Write twice” (see Table 1). Together, the HA stimuli comprised 5 repetitions × 8 vowels × 12 HA male participants = 480 items.

Table 1: The set of target words presented to the participants (column 2) alongside nearest real word where the target word use was a non-word.

HA vowel	Target word	Accompanying target word	HA Arabic presentation	English gloss
/u:/	/hu:d/	/hu:d/	هود	Male name
/i:/	/hi:d/	/hi:d/	هيد	Calm down
/e:/	/he:d/	/ze:d/	زيد	Male name
/o:/	/ho:d/	/xo:d/	خود	Take
/a:/	/ha:d/	/ha:d/	هاد	Relaxed
/i/	/hid/	/hidd/	هِد	Destroy!
/a/	/had/	/hadd/	هَد	Drive slowly
/u/	/hud/	/hudd/	هُد	To hit someone’s head

It was difficult to put all of the target vowels into real /hVd/ words in HA, therefore, the nearest real HA words which have the same target vowels in the nonsense /hVd/ syllables were used such as /xoːd/ and /zeːd/. Acoustic analysis was undertaken using PRAAT [5]. The vowel duration and the first three formant values were automatically extracted with the aid of a PRAAT script. The onset and offset of the vocalic segment were manually labelled for each /hVd/ syllable by following the formants homogeneity method. The vowel duration between the start and end boundaries was measured (in ms). F1, F2, and F3 were extracted from one location (50% for the static model), two locations (20% and 80% for VISC models), and three locations (20%, 50%, and 80% for the three-point approach) across the vowel duration. For the offset model, the first three formants were computed as

(1) $\sqrt{\text{Offset}_{80\%} - \text{Onset}_{20\%}}$

whereas for the direction model, the first three formants were computed as

(2) (Offset80% - Onset20%),

and for the slope model, the first three formants were computed as

(3) (Offset80% - Onset20%)/duration

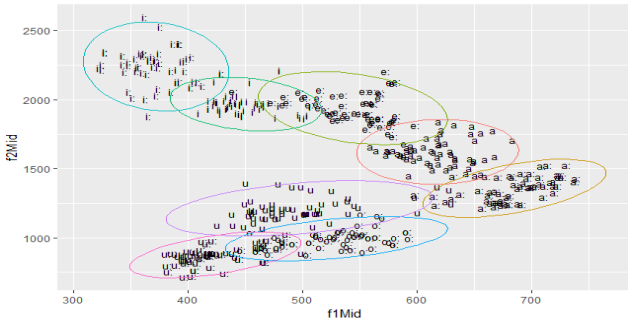
All formant values were checked manually to ensure the accuracy of the results, and any errors in formant estimation were corrected by hand. Discriminant analysis was conducted to evaluate the extent to which the static model, VISC models, the three-point approach, and other acoustic feature sets (F1, F2, F3, and vowel duration) improved vowel classification as reported by [3], [11], [12], among others. A post hoc t-test was used to determine the statistical significance of the study results.

3. RESULTS

3.1. Static and dynamic cues

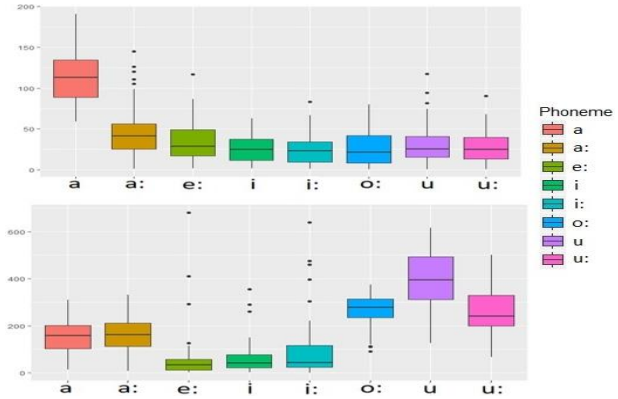
Beginning with the static model, Figure 1 shows a clear and significant separation in the vowel space between the HA vowels, in particular short and long pairs ($p<.001$). The results also showed lax vowels to be more centralised than their long counterparts.

Figure 1: Scatterplot of the midpoints of the first two formant values of Hijazi Arabic vowels.



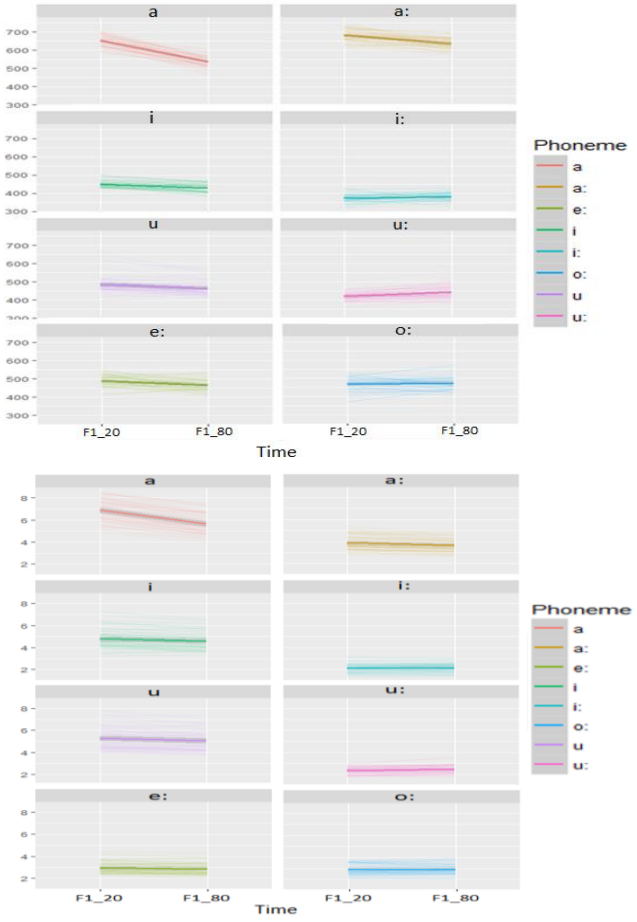
Regarding dynamic cues, particularly the offset model, the amount of overall spectral shifts for HA vowels was significant (see Figure 2; in particular between the first quartile, median, and third quartile).

Figure 2: Boxplot of the offset model for the eight HA vowels (F1- above and F2 - below).



The results of the direction and the slope model in Figure 3 varied among the HA vowels. Most importantly, the direction and slope of the F1 spectral change of short vowels displayed a significantly decreasing spectral shift compared to their long counterparts.

Figure 3: F1 results of the direction model (above) and the slope model (below).



3.2. Discriminant analysis

The results of the three proposed approaches were evaluated via Discriminant Function Analyses which were run in three stages: the classification accuracy of all eight HA vowels, followed by the correct classification rates of the lax and tense HA vowels (/i,/ /a/, /u/ vs /i:/, /a:/, /u:/) as a group and finally the HA vowel pairs (/i:/ vs /i/, /a:/ vs /a/ and /u:/ vs /u/).

In general, the discriminant analysis results showed that taking three measurements from the vowel resulted in the highest classification accuracy (from 93% to 97%) for all eight HA vowels, followed by the offset model (from 91% to 97%), the static cues (from 90% to 96%), then other VISC models, namely, the slope model (from 61% to 74%), and the direction model (from 57% to 74%). The correct classification rate for the duration alone was 24% (see Table 2).

Table 2: Discriminant analysis results showing the classification accuracy of vowels trained on various combinations of parameters ("No Dur" indicates that the duration was not included, whereas "Dur" means the duration was included).

	Static 50%		Direction 20–80%		Slope 20–80%		Offset 20–80%		Three-Point Model 20% & 50% & 80%		
	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	Dur
F1, F2	90	96	57	73	61	74	91	97	93	97	
F1, F2, F3	92	96	59	74	61	74	92	97	95	97	
Duration											24

In terms of the classification of lax and tense HA vowels as a group, which showed a higher improvement in the classification accuracy in comparison to Table 2, the three-point approach and the static model obtained the best rates (97–99% and 96–99%, respectively), followed by the offset model (between 94% and 98%), then the slope model (between 77% and 91%) and direction model (between 73% and 90%). Additionally, the tense/lax vowel group was classified with 31% accuracy based on duration (without formant values) (see Table 3).

Table 3: The correct classification rates of HA vowels (lax and tense).

	Static 50%		Direction 20–80%		Slope 20–80%		Offset 20–80%		Three-Point Model 20% & 50% & 80%		
	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	Dur
F1, F2	96	98	73	90	77	91	94	98	97	99	
F1, F2, F3	97	99	73	90	77	91	95	98	98	99	
Duration											31

Running a discriminant analysis on vowel pairs naturally presents a noticeable improvement in the classification accuracy compared to Table 2 and 3. The three-point model had a better rate (99%),

followed by the static model (between 96% and 99%), then by the offset model (between 95% and 99%) whereas it was between 78% and 99% for the slope model, and between 74% and 99% for the direction model. In addition, the classification rate of HA vowel pairs was between 96% and 99% for the duration alone for each of these pairs (see Table 4).

Table 4: The correct classification rates of HA vowel pairs.

		Static 50%		Direction 20–80%		Slope 20–80%		Offset 20–80%		Three-Point Model 20% & 50% & 80%		
		No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	Dur
/u/ vs. /u:/	F1, F2	99	99	86	99	86	99	98	99	99	99	
	F1, F2, F3	99	99	90	99	89	99	99	99	99	99	
	Duration											99
/i/ vs. /i:/	F1, F2	96	99	74	97	78	98	96	99	99	99	
	F1, F2, F3	96	99	75	97	78	98	96	99	99	99	
	Duration											96
/ɑ/ vs. /ɑ:/	F1, F2	97	99	82	97	95	98	95	99	99	99	
	F1, F2, F3	98	99	84	97	95	98	99	99	99	99	
	Duration											97

The inclusion of vowel duration with the formant frequencies in any model led to a substantial improvement in vowel separation up to 25% (average +11%) while F3 improved the discrimination rate by between 1% and 4% overall (average +1%).

4. DISCUSSION AND CONCLUSION

The data demonstrate that the three-point model is the best model and is the most accurate for classifying HA vowels in all three stages (with average of 98.1%) in comparison to the other proposed models. Such a finding is in line with previous studies e.g. [7], [9], [11], [13], [27], [28]. The offset model, on the other hand, comes in second (with average of 97.5%), which also supports other research e.g. [11], [10], [12]. Interestingly though, the data reveal that the static approach was sufficient, obtaining higher accuracies (with average of 97.1%), and was superior to the other proposed VISC models based on direction and slope. Such a result is contrary to expectations of other studies e.g. [3], [22]. The interpretation of this result could be illustrated as follows: those studies which found that direction and slope models outperformed single-point models in classification accuracy examined both models in different phonetic environments than /hVd/, and according to [6], [9], [23], [26], the /hVd/ context is acoustically least comparable to other consonantal contexts. [6] found that by using the discriminant analysis, the recognition scores are least accurate from tokens taken from /hVd/ compared to other contexts. This could be due to the phonological voicing status of the following coda, which might significantly alter spectral characteristics and vowel duration. Putting such findings together, it seems to be the case that there are experimental results in which vowels with other consonantal context transitions, which provide additional information regarding the vowel's phonetic identity, are identified more accurately by all VISC models than vowels in isolation or /hVd/ [23], which do not contain as many transitions. Hence, it is likely the differences in findings between this paper and those of other studies e.g. [3], [22] are due to contextual and language differences.

The slope and direction models provide a better overview of the characterisation of dynamic cues of the HA vowels, particularly the tense/lax pairs. Such a result is consistent with [25] and [26]. In addition, such results support other studies e.g. [1], [2], which argue that Arabic tense and lax vowels are different in terms of their quantity as well as their quality. This study found that HA vowels displayed great spectral movement and that due to that the low-density languages would have more space and freedom to produce their vowels compared to high-density languages; such a consequence is in agreement with other studies e.g. [15], [16], [17]. Our results demonstrate that the effectiveness of the first two formant frequencies is indisputable for the overall classification of HA vowels. Although duration alone was not sufficient for the distinction of all HA vowels combined (Table 2) or for the distinction between tense-lax as a group (Table 3), when combined with formants, it adds to the overall classification of HA vowels. However, when looking at the tense-lax pairs (Table 4), the results show that the role of formant patterns and vowel duration is almost comparable, which is expected as Arabic vowel pairs are extensively distinguished by duration [1]. Therefore, this study highlights the importance of duration in HA vowels due to the prominent role of phonological length in Arabic phonology. This conclusion is in line with many previous studies e.g. [11], [12], [26]. Vowel duration in this study has more influence on overall vowel classification than has been found elsewhere, with a substantial improvement in vowel separation (up to 25%) while in [11] study was only up to 12%. F3 appears to have little influence on the classification accuracy of HA vowels, which is in agreement with other studies e.g. [11].

To sum up, our results are found to be more consistent with dynamic theories of vowels, as they provide evidence that monophthong vowels are dynamic and that vowel duration is the most useful additional feature to differentiate between phonemes. These results could be extended to look at contexts beyond hVd, as suggested by many researchers e.g. [11], [26], in order to dig deeper into dynamic properties in various consonantal contexts and provide further comparative research, which will be our next step.

5. REFERENCES

[1] Almbark, R., Hellmuth, S. 2015. Acoustic analysis of the Syrian vowel system. In: *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. University of Glasgow. ISBN 9780852619414.

[2] Al-Tamimi, J. 2007. 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.

[3] Arnaud, V., Sigouin, C., Roy, J. P. 2011. Acoustic description of Quebec French high vowels: First results. *Proceedings of the 17th ICPhS*. Hong Kong, 244-247.

[4] Bakalla, M. 1981. The contribution of the Arabs and Muslims to the study of vowel length. A manuscript. Arabic Language Institute. University of Riyadh, Saudi Arabia.

[5] Boersma, P., Weenink, D. 1992–2014. Praat: Doing phonetics by computer.

[6] Elvin, J., Williams, D., 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.

[7] Ferguson, S. H., 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.

[8] Gottfried, M., Miller, J. D., Meyer, D. J. 1993. Three approaches to the classification of American English vowels. *Journal of Phonetics*. 21: 205-229.

[9] Harrington, J., 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.

[10] Hillenbrand, J. M., Nearey, T. M. 1999. Identification of resynthesized /hvd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*. 105, 3509-3523.

[11] Hillenbrand, J. M., Getty, L. A., Clark, M. J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*. 97: 3099-3111.

[12] Hillenbrand, J. M., Clark, M. J., Nearey, T. M. 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, 109(2), 748-763.

[13] Huang, C. B. 1992. Modelling human vowel identification using aspects of formant trajectory and context. *Speech perception, production and linguistic structure*, 43-61

[14] Jarrah, M. A. 1993. The Phonology of Madina Hijazi Arabic: A Non-linear Analysis. PhD., University of Essex.

[15] Jin, S. H., 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), EL363-EL369.

[16] 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.

[17] Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., 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, 723-726.

[18] Morrison, G. S., Nearey, T. M. 2007. Testing theories of vowel inherent spectral change. *The Journal of the Acoustical Society of America*. 122 (1): EL 15-22.

[19] Morrison, S., Assmann, P. 2012. *Vowel inherent spectral change*. Springer Science & Business Media.

[20] Mousa, A. 1994. The Interphonolgy of Saudi Learners of English. Ph.D. University of Essex.

[21] Munro, M. J. 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36(1), 39-66.

[22] Nearey, T. M., 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.

[23] 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.

[24] Peterson, G. E., 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.

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

[26] Watson, C. I., Harrington, J. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America* 106: 458-468.

[27] Yuan, J. 2013. The spectral dynamics of vowels in Mandarin Chinese. In *INTERSPEECH*. pp. 1193-1197.

[28] Zahorian, S. A., 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.