



Comparison Study of Short Vowels' Formant: MSA, Arabic Dialects and Non-native Arabic

Ghania Droua-Hamdani^(✉)

Centre for Scientific and Technical Research on Arabic Language Development (CRSTDLA),
Algiers, Algeria
gh.droua@post.com, g.droua@crstdla.dz

Abstract. The paper deals with the formants of three short vowels in the Modern Standard Arabic (MSA) language produced by 29 native and non-natives speakers. The studied vowels are /a/, /u/ and /i/. F1, F2 and F3 formants were computed from 145 sentences produced by both groups of speakers. Two experiments were conducted. The first investigation compared formants values of natives speakers with those of non-native speakers. When comparing the MSA L1 vowels against their MSA L2 counterparts, results showed a variation in vowel quality especially for the vowel /a/. A comparative analysis of formants values was also carried out with results of some eight Arabic dialect studies. The outcomes depicted variation especially in the pronunciation of the vowel /u/.

Keywords: Formants · Modern standard Arabic · Native speakers · Non-native speakers · Short vowels · Arabic dialects

1 Introduction

The speech of non-native speakers may exhibit pronunciation characteristics that result from their imperfectly learning the L2 sound system, either by transferring the phonological rules from the first language to the second language or through implementing strategies similar to those used in primary language acquisition [1]. In speech recognizing tasks, distinguishing between native speakers from non-native speakers remains a big challenge to solve. To improve speech recognizers' accuracies, a huge amount of speech data is required in the training and testing phases. Features used for this purpose are usually: MFCC, formants, etc. [2–6]. Formants, whether they are defined as acoustic resonances of the vocal tract, or as local maxima in the speech spectrum, are determined by their frequency and by their spectral width. They were widely studied in speech processing as well as in speech perception, sound production comparison within languages, second language (L2) acquisition, speech pathology studies, recognition tasks, etc. [4, 8–14

Like many other languages, the Arabic language has known several research studies on formants. Regarding the auditory level, we can cite works of [15–17] conducted either on Modern Standard Arabic or on Arabic dialects. Abou Haidar investigated the

MSA vowels system using a set of monosyllabic words to show the vowel cross-dialectal structural stability at the perception level [15]. The comparative study was done on the vowels of eight informants from different linguistic backgrounds (Qatar, Lebanon, Saudi Arabia, Tunisia, Syria, Sudan, United Arab Emirates, and Jordan). Alghamdi based his work on six isolated Arabic syllables that were uttered by 15 native informants representing three Arabic dialects (Egyptian, Saudi and Sudanese) [16]. Newman realized an experimental study on MSA vowels in connected speech. The speech variety that was studied was the Holy Quran recitation [17].

MSA phonetic system is endowed by six vowels: three short vowels (/a/, /u/ and /i/) vs. three long vowels (/a:/, /u:/ and /i:/). However, the English language has fifteen vowel sounds. The investigation examines only short vowels extracted from the speech material. The present study examines vowel variation quality (the first, second and third formants, hereafter F1, F2 and F3) within L1 and L2 Arabic language. The objective is to put forward formant variation in vowel production in Modern Standard Arabic (MSA) spoken by native vs. non-natives speakers. Thus, we examined the foreign accent of speakers within connected sentences produced by Arabic and American participants.

Afterward, we compared our results with those found in some Arabic dialect studies.

The paper is organized as follows. Section 2 exposes speech material and participants used in the study. Section 3 describes formants extraction. Section 4 shows experiments and findings. Section 5 gives the concluding remarks based on the analysis.

2 Materials

In this section, we outline materials used in the experimental setting. Information was given in detail about texts read, speakers, recordings and technical conditions. There are six vowels (three short vowels vs. three long vowels) and two diphthongs in Modern Standard Arabic (MSA) language. In our study, we focused only on formant frequencies as a part of acoustic features computed from short vowels (/a/, /u/ and /i/). Because vowels' formant frequencies can be impacted by phoneme co-articulation, recordings used were speech continuous sentences.

Totally 29 speakers participated in the experiment (15 natives/14 non-natives). They uttered five sentences taken from script 1 of the West Point corpus [18]. The English language is the mother tongue of non-native speakers. The first language of non-native speakers was the English language. Technical conditions of recording were as follows: normal speech rate, a sampling frequency of 22.05 kHz. A total of 145 recordings were used in the analysis. Table 1 shows the number and gender of speakers in the sample.

Table 1. Distribution of native and non-native speakers per gender

Native speakers		Non-native speakers	
Male	Female	Male	Female
5	10	6	8
Total	15	14	

Formants are distinctive frequency features that refer to frequency resonances. To compute speech formants, we annotated and segmented manually all speech material i.e. 145 recordings of the dataset onto their different segmental units (vowels and consonants) using Praat software [19]. Then, we extracted all vowels of the speech material, which constituted our speech signal material.

3 Feature Extraction

All speech signals of the vowel's dataset were first pre-emphasized with a pre-emphasis filter that is used to enhance the high-frequency components of the spectrum. This is performed by applying the following formula:

$$x'_n = x_n - ax_{n-1} \quad (1)$$

where $a\psi$ is the pre-emphasis coefficient which should be in the range $0 \leq a < 1$.

Hamming window function was applied on each frame of the signal, to reduce boundary effects. The impulse response of the Hamming window is defined as follows:

$$(n) = 0.54 - 0.46\cos(2\pi n / (N - 1)) \quad (2)$$

where $0 \leq n \leq N - 1$.

The normal frame length was about 30 ms to ensure stationary property

Most formant tracking algorithms are based on Linear Predictive Coding (LPC). LPC coefficients are extracted from pre-processed speech. The estimated future speech samples from a linearly weighted summation of past p -samples using the method of least squares is

$$\hat{x}(n) = - \sum_{k=1}^p a(k)x(n-k) \quad (3)$$

where $x(n)$ and $\hat{x}(n)$ are speech samples and their estimates and

$a(k) = [a(1), a(2), \dots, a(p)]^T$ is the LPC parameters and p is the linear predictive (LP) filter order.

4 Results and Discussions

Formants refer to resonant frequencies of the vocal tract that appear as clear peaks in the speech spectrum. Formants display a large concentration of energy within voiced phonemes as vowels. Most often, the two first formants, F1 and F2, are sufficient to identify the kind of vowel. Nevertheless, in the study, we performed three formants (F1, F2 and F3) values for each vowel to show variation in pronunciation between native and non-native speakers.

The first investigation seeks to provide the assessment of the mean values of all formants for both corpora categories i.e. short vowels produced by native and non-native MSA speakers groups. The calculation was conducted regardless gender of speakers (male/female). Table 2 outlines in detail the average and standard deviation values of F1, F2 and F3 measured for each short vowel /a/, /u/ and /i/.

Table 2. Means and standard deviation of F1, F2 and F3 (Hertz) of native (L1) and non-native (L2) speakers

		/a/	std	/u/	std	/i/	std
F1	L1	756,50	125,39	456,54	130,13	435,19	116,58
	L2	722,95	130,50	544,68	200,89	487,36	217,41
F2	L1	1643,09	244,87	1284,66	441,14	2139,58	357,23
	L2	1514,58	264,37	1420,32	472,96	2089,83	368,68
F3	L1	2735,39	325,26	2712,75	253,23	2929,33	289,58
	L2	2702,39	275,37	2785,36	356,50	2819,84	301,39

Figure 1 displays formants measures computed from L1 and L2 MSA for each short vowel /a/, /u/ and /i/. When comparing the L1 vowels against their L2 counterparts, we can observe that the averages of F1, F2, and F3 calculated in /a/ produced by the non-natives were less compared to their homologs in native speech. In contrast, for /u/ uttered by L2 participants, the values of F1, F2, and F3 increased compared to /u/ produced by Arabic speakers. In /i/, a slight variation was noticed in the second formant between both groups. F1 of native speakers was less compared to those of non-native speakers. In the case of F3, natives gave higher value.

Figure 2 illustrates the plot of the first two formants computed of both groups. The chart shows that the short vowels of non-native speakers are positioned slightly more central than their native counterparts: this tendency is more outspoken for /u/ and /i/. As it can be seen, the vowel diagram plotted for native speakers is larger than the triangle obtained by the second group. The diagram shows also that the latter is almost located inside the triangle drawn for native subjects. Statistical analysis indicates a significant effect in L2 pronunciation in both F1 and F2 formant for the vowel /a/. Likewise, results of /u/ and /i/ show also a significant effect on F1 for both vowels.

Comparative analysis was also carried out between the outcomes of this study and some other previous works [14–16]. Indeed, we compared our results with those obtained from researches conducted on Arabic dialects (7 and 1 dialects respectively) by [14] and [15] and another study realized on Modern Standard Arabic (MSA) [16]. The Arabic dialect accents are the Qatari, the Lebanese, the Saudi, the Tunisian, the Syrian, the Sudanese, the Jordanian and the Egyptian. The authors built their experimental studies basing on a different kind of corpus (syllable, words, and connected speech).

Figure 3 depicts the location of MSA native and non-native vowels between all different dialect pronunciations. As it can be seen from Fig. 3, all vowels /u/ of Arabic dialects are positioned outside the MSA vowel diagram. However, almost all dialect values of /a/ and /i/ are within the MSA triangle. The chart shows also, that /u/ of non-native speakers is more central than all other dialects' values. Regarding, /a/ and /i/ of the same group the plot displays that almost localized in the periphery of the graph.

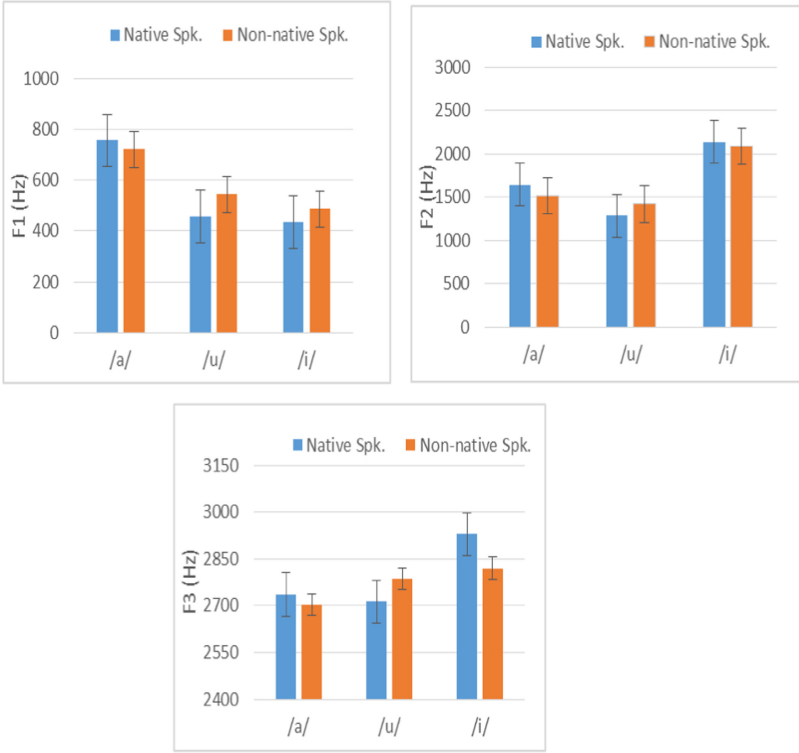


Fig. 1. Comparison of formant averages for both native and non-native MSA short vowels

■ MSA native spk. □ MSA non-native spk.

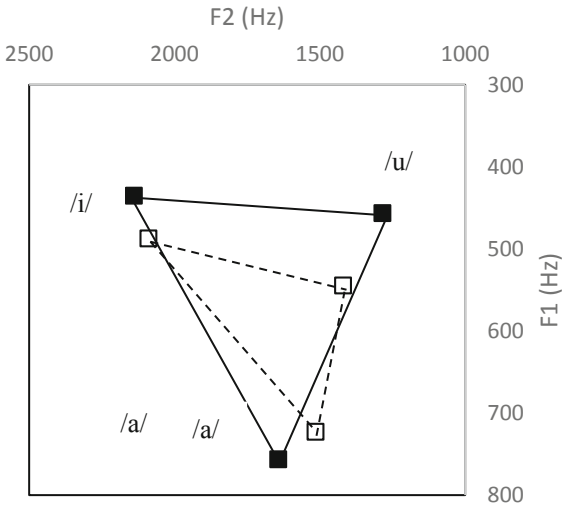


Fig. 2. Vowel diagram of native vs. non-native speakers

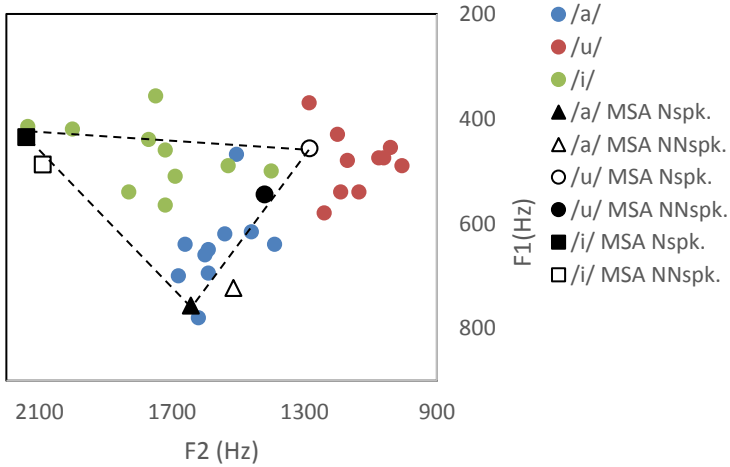


Fig. 3. Comparison of MSA native and non-native vowels with dialect vowels

5 Conclusion

The study deals with the formant analysis of Modern Standard Arabic (MSA) vowels produced by 15 native (L1) and 14 non-natives (L2) speakers. Three formants were computed from all short vowels (/a/, /u/ and /i/) of the speech material. We computed for all vowels the three first formants F1, F2 and F3. When comparing the L1 vowels against their L2 counterparts, results show that formants calculated in /a/ produced by the non-natives were less compared to their homologs in native speech. In contrast, for /u/ uttered by L2 participants, the values of F1, F2, and F3 increased compared to /u/ produced by Arabic speakers. Regarding the vowel /i/, a slight variation was noticed in the second formant between both groups. The vowel diagram plotted from F1 and F2 values for both native and non-native speakers show that the short vowels of L2 speakers are positioned slightly more central than their native counterparts. This tendency is more outspoken for /u/ and /i/. Non-native triangle is almost located inside the triangle drawn for native subjects.

A comparative analysis of formants values was also carried out between native and non-native MSA with results of some eight Arabic dialect studies. The findings show that that /u/ of non-native speakers is more central than all other values computed in Arabic dialects. Concerning /a/ and /i/ uttered by L2 participants, the outcomes show that the latter are localized in the periphery of the vowel diagram.

References

1. MacDonald, M.: The influence of Spanish phonology on the English spoken by United States Hispanics. In Bjarkman, P., Hammond, R. (eds.) *American Spanish pronunciation: Theoretical and applied perspectives*, Washington, D.C.: Georgetown University Press, pp. 215–236, ISBN 9780878404933. (1989)

2. Nicolao, M., Beeston, A.V., Hain, T.: Automatic assessment of English learner pronunciation using discriminative classifiers. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, QLD, (2015). pp. 5351–5355, doi: <https://doi.org/10.1109/ICASSP.2015.7178993>.
3. Marzieh Razavi, M., Magimai Doss M.: On recognition of non-native speech using probabilistic lexical model INTERSPEECH 201. In: 15th Annual Conference of the International Speech Communication Association, 14–18 September, Singapore (2014)
4. Alotaibi, Y.A., Hussain, A.: Speech recognition system and formant based analysis of spoken Arabic Vowels. In: Lee, Y.-H., Kim, T.-H., Fang, W.-C., Ślęzak, D. (eds.) *FGIT 2009*. LNCS, vol. 5899, pp. 50–60. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10509-8_7
5. Droua-Hamdani, G., Selouani, S.A., Boudraa, M.: Speaker-independent ASR for modern standard Arabic: effect of regional accents. *Int. J. Speech Technol.* **15**(4), 487–493 (2012)
6. Droua-Hamdani, G., Sellouani, S.A., Boudraa, M.: Effect of characteristics of speakers on MSA ASR performance. In: *IEEE Proceedings of the First International Conference on Communications, Signal Processing, and their Applications (ICCSPA 2013)*, pp. 1–5. (2013).
7. Droua-Hamdani, G.: Classification of regional accent using speech rhythm metrics. In: Salah, A., Karpov, A., Potapova, R. (eds) *Speech and Computer. SPECOM 2019*. Lecture Notes in Computer Science, vol 11658, pp. 75–81. Springer, Cham (2019)
8. Droua-Hamdani, G.: Formant frequency analysis of msa vowels in six algerian regions. In: Karpov, A., Potapova, R. (eds.) *SPECOM 2020*. LNCS (LNAI), vol. 12335, pp. 128–135. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60276-5_13
9. Farchi, M., Tahiry, K., Soufyane, M., Badia, M., Mouhsen, A.: Energy distribution in formant bands for Arabic vowels. *Int. J. Elect. Comput. Eng.* **9**(2), 1163–1167 (2019)
10. Mannepalli, K., Sastry, P.N., Suman, M.: analysis of emotion recognition system for Telugu using prosodic and formant features. In: Agrawal, S.S., Dev, A., Wason, R., Bansal, P. (eds.) *Speech and Language Processing for Human-Machine Communications. AISC*, vol. 664, pp. 137–144. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6626-9_15
11. Korkmaz, Y., Boyacı, A.: Classification of Turkish Vowels Based on Formant Frequencies. In: *International Conference on Artificial Intelligence and Data Processing (IDAP)*, pp. 1–4, Malatya (2018). <https://doi.org/10.1109/IDAP.2018.8620877>
12. Natour, Y.S., Marie, B.S., Saleem, M.A., Tadros, Y.K.: Formant frequency characteristics in normal Arabic-speaking Jordanians. *J. Voice* **25**(2), e75–e84 (2011)
13. Abd Almisreb, A., Tahir, N., Abidin, A.F., Md Din, N.: Acoustical Comparison between /u/ and /u:/ Arabic Vowels for Non-Native Speakers. *Indonesian J. Elect. Eng. Comput. Sci.* **11**(1), 1–8 (2018). <https://doi.org/10.11591/ijeecs.v11.i1>.
14. Rusza, J., Cmejla, R.: Quantitative acoustic measurements for characterization of speech and voice disorders in early-untreated Parkinson's disease. *J. Acoust. Soc. Am.* **129**, 350 (2011). <https://doi.org/10.1121/1.3514381>
15. Abou Haidar, L.: Variabilité et invariance du système vocalique de l'arabe standard, Unpubl. PhD thesis, Université de Franche-Comté. (1991).
16. Alghamdi, M.: A spectrographic analysis of Arabic vowels: a cross-dialect study. *J. King Saud Univ.* **10**(1), 3–24 (1998)
17. Newman, D.L., Verhoeven, J.: Frequency Analysis of Arabic Vowels in Connected Speech Antwerp papers in linguistics, Vol. 100, pp. 77–87 (2002)
18. Linguistic Data Consortium LDC. <http://www ldc.upenn.edu>.
19. Boersma, P., Weenink, D.: Praat: doing phonetics by computer. <http://www.praat.org>. Accessed Mar 2010