Arabic vowels characterization and classification using the normalized energy relating to formants and pitch

Mohamed Farchi, Karim Tahiry, Ahmed Mouhsen
IMII Laboratory, Faculty of Sciences and Technics, Hassan First University of Settat, Settat, Morocco

Article Info

Article history:

Received Jul 23, 2022
Revised Sep 30, 2022
Accepted Nov 27, 2022

Keywords:

Arabic vowels
Classification algorithm
Normalized energy bands
Vowel nucleus

ABSTRACT

Vowels are the primary units of a sound system of a language. The classification of these vowels is therefore very important for the recognition and synthesis of speech. In this paper, we propose a normalized energy-based approach in formants and pitch to characterize Arabic vowels (short vowels: / a /, / i /, / u /; long vowels: / a: /, / i: /, / u: /). The classification was performed using a developed algorithm on records extracted from an Arabic corpus after the extraction of the pitch and the first three formants and the computation of the normalized energy in these bands. The results showed that the algorithm distinguishes Arabic vowels by analyzing the normalized energy in the nucleus of F1, F2, and F3 formants and pitch F0 with a rate of 88.7% for long vowels and a rate of 90% for short vowels.

Corresponding Author:

Mohamed Farchi
IMII Laboratory, Faculty of Sciences and Technics, University Hassan First
Settat, Morocco
Email: simo.farchi@gmail.com

1. INTRODUCTION

The production of vowels is dependent on the flow of air in the cavities above the glottis (oral cavity and nasal cavity). Vowels are usually voiced sounds, created by the vibration of the vocal cords which results in a higher spectrum amplitude in the low and medium frequencies compared to the consonants [1]–[3]. This produces a maximum of spectrum amplitude in the low and medium frequencies during the production of the vowel. These frequencies are called formants. It has been shown that the frequency of the formants is very important in the determination and identification of a vowel. The listening experiments of Peterson and Barney made it possible to map the first two formants (F1 and F2). Higher formants also play a role in the identity of vowels [4]–[9].

The first formant is associated with changes in the opening of the mouth. F1 frequencies of sounds requiring small mouth openings are located at low frequencies and those requiring wide mouth opening at high frequencies while the second formant is associated with changes in the oral cavity such as the position of the tongue and the activity of the lips. On the other hand, the third formant is associated with a front-to-back constriction in the oral cavity [10].

Based on these observations, many studies have been realized on formants to identify vowels. Lulich *et al*. [11] analyzed the second formant in the case of coupling between the vocal tract and the lower respiratory tract (subglottic). He showed that the amplitude discontinuity due to the passage of the second formant (F2) through the second subglottal resonance plays an important role in the perception of vowels. This is consistent with [12], who was able to demonstrate the role of the second subglottic resonance in the distinction of vowels. Natour *et al*. [13] studied the acoustic characteristics of the normal Arabic voice by analyzing the formants F1, F2, and F3. He showed that the frequencies of formants among Jordanian Arabic speakers show a significant decline in men, women and children compared to other racial backgrounds and dialects. Other studies proposed another perspective of analysis of formants. Alotaibi and Hussain [14] analyzed Arabic vowels by studying the values of the first and second formants in a context of consonance-vowel-consonance (CVC) enunciation. This study allowed to classify Arabic vowels using a hidden Markov model (HMM) recognizer.

Vowels have been examined in the past using duration cues. The duration has been shown by Alotaibi and Hussain [15] to be crucial in differentiating between short and long vowels. A correlation between speech rate and vowel duration has been demonstrated by other researchers. Therefore, if speech tempo quickens, vowel duration shortens [16]. According to Mok [17], vowel duration has lately emerged as a crucial acoustic cue for vowel identification and speech understanding. Additionally, Khattab and Al-Tamimi [18] observed no discernible differences in durational outcomes between males and females.

Researchers have also considered spectral moments to characterize vowels based on the fact that the vowel spectrum reflects both the glottal source's characteristics and the vocal tract filter's functions, and that the spectral moments combine the glottal source's spectral amplitude model and the vocal tract filter's resonances [19]. Tahiry *et al*. [20] showed that center of gravity (CoG) and STD reveals two phases of production of vowel. The start of production is characterized by a transient regime followed by a stationary state as the duration increases. Savela *et al*. [21] found that the identification of vowels can be based on spectral moments. Indeed, several spectral attributes in the space of the vowels allow the evaluation of the quality and the classification of the vowels. Pentti [22] analyzed the spectral moments to compare the production of /S/ of young children to adults. The results of this study showed significant effects of vowel coarticulation on spectral characteristics especially in symmetrical vowel contexts.

Another aspect of the study was able to demonstrate that the energy distribution in the frequency bands can characterize Arabic vowels according to the duration of production. Furthermore, the percentage distribution of energy in these frequency bands appears to be unaffected by the time of production [23]. Due to the distribution of formants F1 and F2, they demonstrated that the majority of energy for the vowel /a/ is located in the first five bands B1, B2, B3, B4, and B5. The distance between these formants (F1>600 Hz and F2>1,000 Hz) leads to an energy distribution that covers the five frequency bands. The vowel /i/'s energy is concentrated in the first and fifth bands (B1 and B5). This is explained by the fact that the first formant F1, in addition to the energy produced by vocal sounds glottic vibration (400 Hz), is positioned in the low frequencies (F1 300 Hz). Because the first two formants F1 and F2 are concentrated in the low frequencies, the main energy for the vowel /u/ is concentrated in the first band B1. Another feature extraction approach based on a frequency response model of the vocal tract has been presented. Paulraj *et al*. [24] examined the mean and maximum energy amplitudes as characteristics in fixed frequency frames between 20 and 2,500 Hz. The energy characteristics obtained made it possible, using multinomial logistic regression (MLR), to detect the vowels /a/, /e/, /i/, /o/, and /u/.

Inspired by previous works related to Arabic vowels recognition, the purpose of this work is to define another aspect of study of the Arabic vowels. The findings presented in this article are based on an acoustic study of Arabic vowels. The primary goal is to identify Arabic vowels (short and long ones) according to the normalized energy bands characterizing the first three formants and pitch.

This paper is arranged in the following manner. Section 2 defines the baseline methods and means employed and describes the experiments achieved. Section 3 discusses the results. Section 4, describes the summary and conclusions of this work.

2. METHOD

2.1. General processing

This part defines the methodology used to describe the behavior of Arabic vowels through a series of experiments. It also provides a detailed description of the data collected as well as the tools used in this study. Ten Moroccan were asked to pronounce isolated syllables CV (C is a consonant and V is one of the vowels /a/, /i/ or /u/) with short and long vowels. We constructed the corpus using the consonant /ʔ/: /ء/ given its minimal effect on the vocal tract (Table 1).

The recordings were digitized with a sampling frequency of 22,050 Hz. By isolating each vowel via "Praat," useful signals were extracted from the recordings. These signals are then segmented into 11.6 ms segments and windowed using the "hamming window," followed by a 512-point fast fourier transform (FFT) and zero-padding. For this study, the first three formant frequencies were calculated using the linear predictive coding (LPC) method. Then, the energy in each of these three formants was deduced.

2.2. Details formants extraction method

The basic problem of the LPC system is to determine the formants from the speech signal in the form of a linear combination of previous samples. The LPC method conducts a spectral analysis of the input signal by cutting it into frames, based on the premise that the speech signal is generated by a buzzer placed at the end of a tube. A linear combination of filter parameters and previous samples predicts output samples. The buzz is produced by the gap between the vocal cords named glottis and is distinguished by its strength and frequency (pitch). The vocal tract constitutes the tube which is characterized by its spectral envelope peaks. These peaks are called formants. Figure 1 shows the pre-processing steps performed on the speech signal to extract the formants.

Table 1. Arabic corpus of long and short vowels

Vowel /a/	Vowel /i/	Vowel /u/
/a/	/i/	/u/
/aa/	/ii/	/uu/
/aaa/	/iii/	/uuu/
/aaaa/	/iiii/	/uuuu/
/aaaaa/	/iiiii/	/uuuuu/

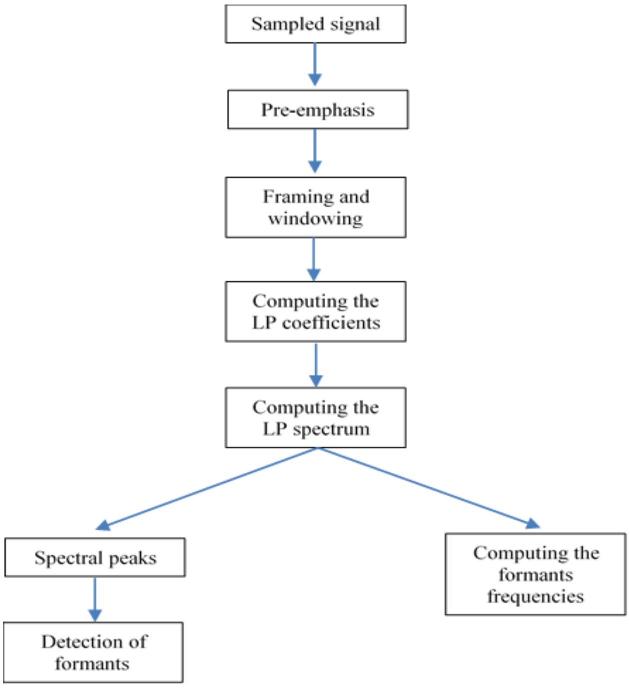

Figure 1. Chart of the detection procedure of formants with LPC [13]

After evaluating the tendency of the first three formants (F1, F2, and F3) for Arabic vowels (/a/, /i/, et /u/), we found that these three frequency bands present significant acoustic changes defined for each vowel. Furthermore, the pitch frequency (F0) is to consider since the vowels are voiced sounds. The three frequency bands considered in this work are:

- /a/ Band F0: 0–400 Hz; F1: 500–800 Hz; Band F2: 1,000–1,500 Hz; Band F3: 2,200–2,800 Hz
- /i/ Band F0: 0–400 Hz; Band F1: 100–400 Hz; Band F2: 2,000–3,000 Hz; Band F3: 2,800–3,400 Hz
- /u/ Band F0: 0–400 Hz; Band F1: 300–600 Hz; Band F2: 600–1,100 Hz; Band F3: 2,400–3,000 Hz

2.3. Normalized energy computation in formants

The amplitude of the spectrum was smoothed in each segment by taking 20 average points along the time index n. The calculations focused on the three bands that represent each vowel's three formants, F1, F2, and F3. The frequency index k varies between each band's lower and upper limits. The energy in each band was measured using the following formula:

$$E_b(n) = \sum_k 10\ \log_{10}(|X(n,k)|^2) \tag{1}$$

where |X(n, k)| denotes the spectrum amplitude and $E_b(n)$ denotes the band energy b in segment n. The normalized energy band for each segment was then determined as (2):

$$E_{bn}(n) = \frac{E_b(n)}{E_T(n)} \tag{2}$$

Given a segment n, $E_{bn}(n)$ is the normalized band energy b in this segment and $E_T(n)$ is the total energy.

3. RESULTS AND DISCUSSION

3.1. Energy distribution in formants

The objective of this section is to analyze the normalized energy distribution of Arabic vowels (three short: /a/, /i/, /u/ and three long vowels: /a:/, /i:/, /u:/) in the frequency bands relating to the first three formants (F1, F2, and F3) and pitch F0. The first three formants are more significant for the characterization of Arabic vowels. The results obtained are summarized in Table 2. The analysis of these data allowed us to observe that a minimum of 70% of the energy is focused in the frequency bands F0 (pitch), F1, F2, and F3 (formants) for each of the vowels /a/, /i/ and /u/.

We notice an important energy in the band F0 for the three vowels. This is explained by the vibration of the vocal cords at the moment of the production of the vowel. For the vowel /a/, the bands F0 and F1 are isolated due to the position of the tongue (Figure 2). Indeed, this case can be explained by the fact that the closer the tongue is to the roof of the mouth, the lower is the frequency of the first formant.

Table 2. Average energies in bands F0, F1, F2 and F3

	Em_F0	Em_F1	Em_F2	Em_F3	Total
/a/	21.42	26.20	18.73	1.84	68.19
/a:/	40.44	17.91	16.39	2.12	76.85
/i/	51.75		9.55	6.91	68.22
/i:/	60.10		9.90	7.52	77.53
/u/	38.93	27.87	11.19	0.48	78.46
/u:/	59.34	15.07	15.42	0.14	89.97

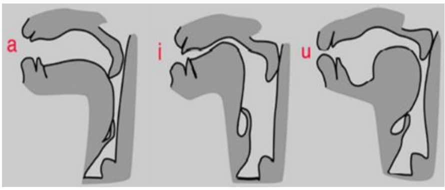

Figure 2. Vowel articulatory movements (/a/, /i/ and /u/) from [10]

The high energy density in the band F0 for the vowel /i/ is explained by the fact that the pitch and first formant frequency bands are combined. The fact that the frequency bands corresponding to the pitch frequency and the first formant overlap for the vowel /u/ explains the significant energy in the bands F0 and F1. We notice for the long vowels an increase of the energy of the pitch due to the important energy demand for the vibration of the vocal cords in order produce a long vowel.

We observed that the vowel /i/ has the highest percentage of energy in the band F1 compared to the vowels /a/ and /u/ after examining the formants bands energy distribution. In the frequency band corresponding to the first formant, the vowel /i/ (short and long) has an energy rate greater than 50%. The vowel /i:/ can be easily distinguished from the vowel /i/. In comparison to the vowel /i/, the vowel /i:/ has a higher energy rate than the vowel /i/. Moreover, the scrutiny of the energy distribution in the band F2 reveals that unlike the vowels /a/ and /i/, this band is important in distinguishing between short and long vowels for the vowel /u/. Regarding the distribution of energy in the band F3, we noticed that this data is not significant for the characterization of Arabic vowels.

3.2. Algorithm

The results outlined in the previous section demonstrate that it is possible to identify Arabic vowels based on the percentage of energy distribution in the frequency bands characterizing the first three formants and pitch. These results allowed us to develop an algorithm, which help to distinguish between short and long Arabic vowels. The speech input undergoes several stages of processing. The spectrogram is first calculated. Then, an energy waveform is constructed in the first band, the energy derivative is calculated and peaks in the derivative are detected. The results of this step feed the g-landmark processing step to determine vowel from the syllables CV (Figure 3). The frequency bands characterizing the first three formants and pitch are then determined to calculate the normalized energy equivalent to these bands.

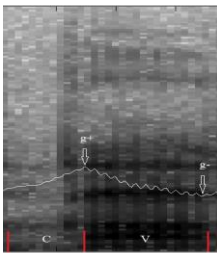

Figure 3. Vowel (v) localized between g+ and g−landmarks (white represent energy in first band)

The result of this treatment is injected at the classification stage to identify the concerned vowel. The results of the classification stage are implemented in the algorithm given in Figure 4, which allows perceiving short and long vowels. The efficiency of this algorithm was determined using our corpus. The obtained results are summarized in Tables 3 to 5. It can be seen that, for vowels /a/ and /u/, the certainty of this algorithm is above 90% (96.55% for /a/, 91.67% for /a:/, 94.29% for /u/ and 94.44% for /u:/). For the vowel /i/, the algorithm is less efficient with an accuracy of around 80% (80.56% for /i/ and 80% for /i:/).

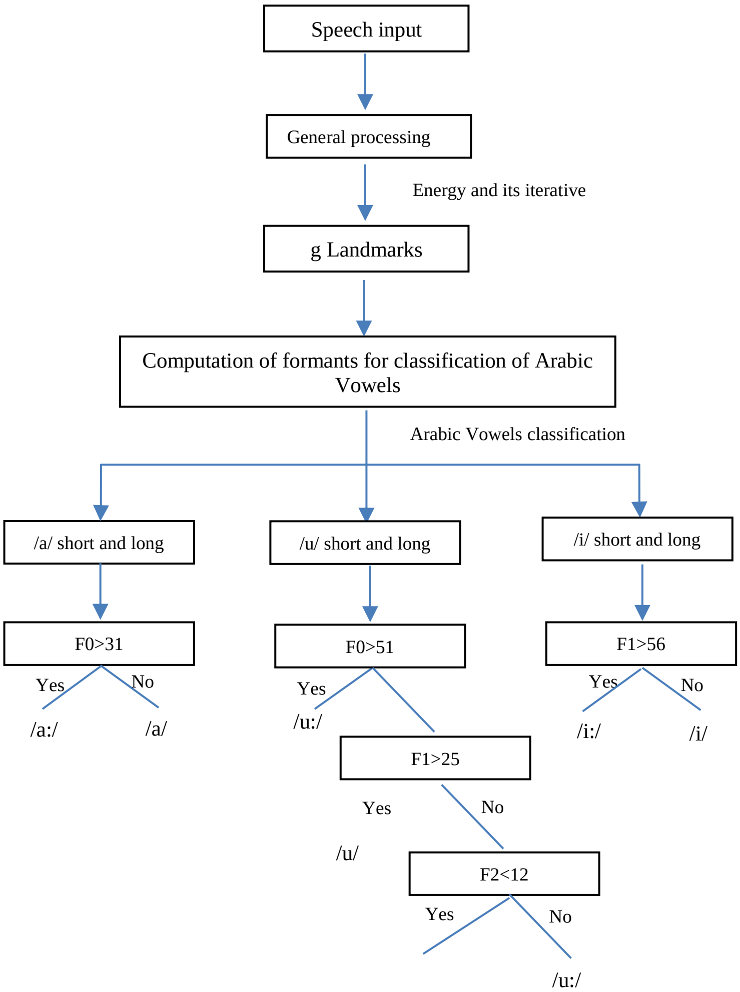

Figure 4. Algorithm of classification of Arabic vowels

Table 3. Confusion matrix of vowels /a/ and /a:/

	/a/	/a:/	Overall percent-tage of errors (%)
/a/	28	1	3.45
/a:/	4	44	8.33

Table 4. Confusion matrix of vowels /u/ and /u:/

	/u/	/u:/	Overall percent-tage of errors (%)
/u/	33	2	5.71
/u:/	2	34	5.56

Table 5. Confusion matrix of vowels /i/ and /i:/

	/i/	/i:/	Overall percent-tage of errors (%)
/i/	29	7	19.44
/i:/	8	32	20.00

The energy distribution of Arabic vowels shows that their behavior is dependent on where they are articulated. The energy in the frequency bands of formants F0 (pitch), F1, F2, and F3 has been found to characterize Arabic vowels in this research. It was possible to differentiate between short and long vowels. These results are very competitive compared to those reported in the literature [23], [25]–[27].

4. CONCLUSION

This work has shown a new approach for the characterization and classification of Arabic vowels according to acoustic cues. This study was based on the energy percentage in the formants and pitch frequency bands (/a/: Band F0: 0–400 Hz; Band F1: 500–800 Hz; Band F2: 1,000–1,500 Hz; Band F3: 2,200–2,800 Hz, /i/: Band F0: 0–400 Hz Band F1: 100–400 Hz; Band F2: 2,000–3,000 Hz; Band F3: 2,800–3,400 Hz, /u/: Band F0: 0–400 Hz Band F1: 300–600 Hz; Band F2: 600–1,100 Hz; Band F3: 2,400–3,000 Hz). The obtained results show that the energy in bands related to the formants F1, F2 and F3 as well as the pitch allow first to identify the nature of the Arabic vowel then to distinguish between the long and short ones. Classification experiments were performed on Arabic vowels extracted from our Arabic corpus. The findings yielded to an overall classification of 89.29%.

REFERENCES

[1] B. Prica and S. Ilic, "Recognition of vowels in continuous speech by using formants," *Facta universitatis - series: Electronics and Energetics*, vol. 23, no. 3, pp. 379–393, 2010, doi: 10.2298/FUEE1003379P.
[2] A. W. Howitt, "Vowel landmark detection," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2279–2279, Nov. 2002, doi: 10.1121/1.4779139.
[3] K. N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1872–1891, Apr. 2002, doi: 10.1121/1.1458026.
[4] J. M. Pickett, *The sounds of speech communication: a primer of acoustic phonetics and speech perception*, 1st ed. Austin, TX: University Park Press, 1980.
[5] H. R. Iqbal, M. M. Awais, S. Masud, and S. Shamail, "On vowels segmentation and identification using formant transitions in continuous recitation of Quranic Arabic," in *New Challenges in Applied Intelligence Technologies*, vol. 134, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 155–162, doi: 10.1007/978-3-540-79355-7_15.
[6] Y. Kanamori and K. Kido, "Recognition of vowels in connected speech by use of the characteristics on perception of vowel," in *ICASSP '76. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1976, pp. 174–177, doi: 10.1109/ICASSP.1976.1170083.
[7] T. M. Thasleema, V. Kabeer, and N. K. Narayanan, "Malayalam vowel recognition based on linear predictive coding parameters and k-NN algorithm," in *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, Dec. 2007, pp. 361–365, doi: 10.1109/ICCIMA.2007.372.
[8] M. Nazari, A. Sayadiyan, and S. M. Valiollahzadeh, "Speaker-independent vowel recognition in Persian speech," in *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, Apr. 2008, pp. 1–5, doi: 10.1109/ICTTA.2008.4530026.
[9] D. Gargouri, M. Frikha, and A. B. Hamida, "A comparative study of an all pole speech analysis for formant extraction," 2005.
[10] F. W. Householder and G. Fant, "Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations," *The Slavic and East European Journal*, vol. 5, no. 3, pp. 285–286, 1961, doi: 10.2307/304731.
[11] S. M. Lulich, A. Bachrach, and N. Malyska, "A role for the second subglottal resonance in lexical access," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2320–2327, Oct. 2007, doi: 10.1121/1.2772227.
[12] K. N. Stevens, *Acoustic phonetics*. Cambridge, MA: MIT, 1998.
[13] Y. S. Natour, B. S. Marie, M. A. Saleem, and Y. K. Tadros, "Formant frequency characteristics in normal Arabic-speaking Jordanians," *Journal of Voice*, vol. 25, no. 2, pp. 75–84, Mar. 2011, doi: 10.1016/j.jvoice.2010.10.018.
[14] Y. A. Alotaibi and A. Hussain, "Speech recognition system and formant based analysis of spoken Arabic vowels," in *Future Generation Information Technology: First International Conference, FGIT 2009*, 2009, pp. 50–60, doi: 10.1007/978-3-642-10509-8_7.
[15] Y. A. Alotaibi and A. Hussain, "Comparative analysis of arabic vowels using formants and an automatic speech recognition system," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 3, no. 2, pp. 11–22, 2010.
[16] S. H. Ferguson and D. Kewley-Port, "Talker differences in clear and conversational speech: acoustic characteristics of vowels," *Journal of Speech, Language, and Hearing Research*, vol. 50, no. 5, pp. 1241–1255, Oct. 2007, doi: 10.1044/1092-4388(2007/087).
[17] P. P. K. Mok, "Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation," *Language and Speech*, vol. 54, no. 4, pp. 527–545, Dec. 2011, doi: 10.1177/0023830911404961.
[18] G. Khattab and J. Al-Tamimi, "Durational cues for gemination in Lebanese Arabic," *Language and Linguistics*, vol. 11, no. 22, pp. 39–56, 2008.
[19] P. Milenkovic and K. Forrest, "Classification of vowels using spectrum moments," *The Journal of the Acoustical Society of America*, vol. 83, no. 1, May 1988, doi: 10.1121/1.2025467.
[20] K. Tahiry, B. Mounir, I. Mounir, and A. Farchi, "Energy bands and spectral cues for Arabic vowels recognition," *International Journal of Speech Technology*, vol. 19, no. 4, pp. 707–716, Dec. 2016, doi: 10.1007/s10772-016-9363-3.
[21] J. Savela, S. Ojala, O. Aaltonen, and T. Salakoski, "Role of different spectral attributes in vowel categorization: the case of Udmurt,"

in *Proceedings of the 16th Nordic Conference of Computational Linguistics, NODALIDA 2007*, 2007, pp. 384–388.
[22] P. Körkkö, "Spectral moments analysis of/s/coarticulation development in Finnish-speaking children," in *ICPhS Proceedings of the 18th International Congress of Phonetic Science*, 2015, pp. 1–4.
[23] M. Farchi, K. Tahiry, and A. Mouhsen, "Arabic vowels characterization and classification using the normalized energy in frequency bands," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 268–274, Feb. 2023, doi: 10.11591/eei.v12i1.4395.
[24] M. P. Paulraj, S. Yaacob, and M. Y. S. Azmi, "Vowel classification based on frequency response of vocal tract," in *2008 International Conference on Computer and Communication Engineering*, May 2008, pp. 1125–1130, doi: 10.1109/ICCCE.2008.4580782.
[25] K. Tsukada, "An acoustic comparison of vowel length contrasts in standard Arabic, Japanese and Thai," in *2009 International Conference on Asian Language Processing*, Dec. 2009, pp. 76–79, doi: 10.1109/IALP.2009.25.
[26] D. Friedrichs, D. Maurer, S. Rosen, and V. Dellwo, "Vowel recognition at fundamental frequencies up to 1 kHz reveals point vowels as acoustic landmarks," *The Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 1025–1033, Aug. 2017, doi: 10.1121/1.4998706.
[27] M. Stanek and L. Polak, "Algorithms for vowel recognition in fluent speech based on formant positions," in *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2013, pp. 521–525, doi: 10.1109/TSP.2013.6613987.

BIOGRAPHIES OF AUTHORS

Mohamed Farchi was born in 1993. He received the engineering degree in Electrical Engineering and Power Systems from Mohammadia School of Engineers, Rabat, Morocco, in 2015. He is currently a Ph.D. student in Engineering, mechanical, Industrial Management and Innovation (IMII) Laboratory research Laboratory, Faculty of Sciences and Technics, Hassan First University with a thesis on automatic speech recognition. He can be contacted at email: simo.farchi@gmail.com.

Karim Tahiry was born in Settat, Morocco, in 1988. He received the Ph.D. degree from the Faculty of Sciences and Technics, Hassan first University, Morocco in 2018 in Electronics and Telecommunications. In 2011, he received the Master degree in Automatic, Signal processing, Industrial Computing from University Hassan first, Settat, Morocco. Member of Engineering, mechanical, Industrial Management and Innovation (IMII) Laboratory. His research interests include speech recognition and signal processing. He can be contacted at email: karim.tahiry@gmail.com.

Ahmed Mouhsen was born in May 1960. He received the Ph.D. degree in electronics from the University of Bordeaux I, France. He is currently a Professor of Electronics in FSTS university Hassan 1st Settat Morocco. He is involved in the design of hybrid, active, and passive microwave electronic circuits, digital systems, and IoT. He can be contacted at email: mouhsen_ahmed@yahoo.fr.