

Arabic vowels characterization and classification using the normalized energy in frequency bands

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 Jun 27, 2022

Revised Sep 15, 2022

Accepted Oct 5, 2022

Keywords:

Arabic vowels

Classification algorithm

Normalized energy bands

Vowel nucleus

ABSTRACT

The main objective of this work is to conduct an acoustic study of Arabic vowels (/a/, /a:/, /u/, /u:/, /i/ and /i:/) in order to determine the most relevant characteristics that allow recognizing these vowels. The analysis of vowel spectrograms reveals that the energy distribution as a function of time and frequency clearly differs according to the considered vowel. Thus, we used the normalized energy in frequency bands to classify these vowels. Thereafter, we have exploited the obtained results to develop algorithms that allow the classification of vowels and the distinction of the long vowels from the short ones. The efficiency of these algorithms was evaluated by testing their performances on our Arabic corpus.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mohamed Farchi

IMII Laboratory, Faculty of Sciences and Technics, University Hassan First

Settat, Morocco

Email: simo.farchi@gmail.com

1. INTRODUCTION

Sounds are broadly classified into voiced and unvoiced speech. Vowels are spoken sounds that generate acoustically filtered, quasi-periodic air pulses as they flow through the vocal tract. Vowels and consonants differ primarily in that vowels resonate in the throat [1], [2]. As well, the vowels are sounds produced without constriction in vocal tract unlike other types of sounds which are all produced by narrowing the flow of air through the vocal tract. However, the vowels are articulated by lifting a part of the tongue body, the location of the vowel refers to the part of the tongue that is highest in its production [3].

Vowel identification is crucial in the process of continuous speech recognition. Hence, efforts were developed in analyzing and characterizing vowels. Despite the fact that analyzing speech signals in the frequency domain is extremely important in studying its acoustic properties, formants remain the classical way of classifying vowels. Especially, the two first formants are the most significant acoustic parameters that determine vowels. Some researchers have exploited formant frequencies to develop algorithms for identification and classification of vowels in continuous speech [1]-[4]. Other researchers have developed a recognition system based on the frequencies of the first and second formants (F1 and F2) [5]-[9]. They reported that the formants of the long vowels are peripheral to those of the short ones. Natour *et al.* [10] have found that the front vowel /i/ has a high frequency for F2, while the back vowel /u/ has a very low F2. The F2 values for the central vowel /a/ are between these two extremes. Several additional cues that distinguish vowels have also been found in earlier studies. Alotaibi and Hussain [5] showed that the duration is very important to distinguish between short and long vowels. Others researchers have shown a relationship between speech rate and vowels duration. So, if the speech rate increases the vowel duration becomes short [11]. More recently, researchers have indicated that vowel duration is a hugely important acoustic cue for vowel identification and speech

intelligibility [12], [13]. Moreover, Khattab and Al-Tamimi [14] have found no significant difference between males and females regarding to the durational results. Other investigators have developed a method based on the wavelet transform and spectral analysis for speech consonant and vowel segmentation in Arabic language without linguistic information [15]. They have reported that there are a significant difference among the long and short vowels in both quantity and quality [16].

The literature review shows that the different vowel classification methods (/a/,/u/and/i/) used are effective. However, the distinction between short and long vowels remains difficult to implement. Indeed, researchers report that long vowels have a longer production time than short vowels [17], [18] and long vowel formants are peripheral to those of short ones [19]. However, the variability of the speech signal makes the determination of the production time and the ranges of the formants of a short or long vowel very difficult. In our previous study [20], have found that both the formant frequencies and the normalized energy bands can differentiate between short Arabic vowels. Additionally, they have discovered that the spectral moments of long vowels (CoG and STD) show that their generation occurs in two stages: a transient phase at the start of vowel formation and a steady state phase as duration lengthens. Furthermore the difference between the short vowels and long ones is the fact that the equilibrium position is maintained longer in the production of a long vowel and the rate of change of the formants or the normalized energy (percent of energy) in the frequency bands can be a good indication for distinguishing between the long vowels and those short. According to [21] and [22], the spectral moments enable a more thorough definition of the vowel category. Korkko [23] used spectral moments to examine how young children produced the consonant /s/ in contexts with symmetrical vowels (such as /isi, usu, ysy,/s/). His research shows how vowel co-articulation has a considerable impact on its spectral properties.

The main goal of this research is to add to the body of knowledge regarding Arabic vowels in experimental literature. The conclusions reported here are based on an acoustic analysis of Arabic vowels and the creation of an algorithm for classifying long and short vowels. The main objective is to detect the Arabic vowels using the bands' normalized energy.

The structure of this paper is as follows. The procedures used, the equipment used, and the experiments conducted are described in the first section. The results are presented and discussed in the second section. A summary of the results and a discussion of the conclusions are presented in the final section.

2. METHOD

2.1. General processing

A set of experiments were conducted with the aim to describe how Arabic vowels behave. The methodology for these studies is outlined in this section, together with information on how the data was gathered and the tools that were employed. A corpus of Arabic language with short and long vowels (/a/, /a:/, /u/, /u:/, /i/, and /i:/) was created. Twenty Moroccan speakers between the ages of 20 and 40 were asked to repeat the isolated syllables CV, which contain both short and long vowels. Isolated syllables rather than words were selected to reduce the influence of other phonemes on the vowels under study. Additionally, the vowel's length can be freely increased. Since producing it involves little strain on the vocal tract, the consonant /ʔ/: /ʕ/ was chosen for the entire corpus (Table 1). The speech was separated into 11.6 ms time segments with a 9.6 ms overlap and sampled at 22,050 Hz. After Hamming windowing and zero-padding each segment, a 512-point fast fourier transform was computed.

Table 1. Arabic corpus of long and short vowels

Vowel/a, -/	Vowel /i, -/	Vowel /u, -/
/أ/ : /ʔa/	/إ/ : /ʔi/	/ؤ/ : /ʔu/
/ئا/ : /ʔaa/	/ئي/ : /ʔii/	/ؤو/ : /ʔuu/
/ئاا/ : /ʔaaa/	/ئيي/ : /ʔiii/	/ؤوو/ : /ʔuuu/
/ئااا/ : /ʔaaaa/	/ئييي/ : /ʔiiii/	/ؤووو/ : /ʔuuuu/
/ئاااا/ : /ʔaaaaa/	/ئيييي/ : /ʔiiiiii/	/ؤوووو/ : /ʔuuuuu/

2.2. Computation of energy band

The magnitude spectrum was smoothed out per frame along the time index n using a 20-point moving average. Six distinct frequency bands (band 1: 0-400 Hz; band 2: 400-800 Hz; band 3: 800-1200 Hz; band 4: 1200-2000 Hz; band 5: 2000-3500 Hz and band 6: 3500-5000 Hz) were chosen from the smoothed spectrum X(n,k). The energy in each band was calculated as (1):

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

where the band index b ranges from 1 to 6. The frequency index k ranges from the DFT indices representing the lower and upper boundaries for each band. Then, the normalized energy band for each frame was determined by:

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

where $E_{bn}(n)$ denotes the normalized band energy b in the frame n , $E_T(n)$ is the frame's overall energy, and $E_b(n)$ denotes the band energy b in the frame n .

The vowel is composed of three segments, the onset is the first segment, the closing segment is the coda and the nucleus is the central segment of the vowel. The energy in vowel nucleus was calculated as follows:

$$E_{nucleus} = \sum_{\frac{1}{3}d_v}^{\frac{2}{3}d_v} E_{bn}(n) \quad (3)$$

Where $E_{nucleus}$ is the normalized band energy in vowel nucleus, d_v is the vowel overall duration.

3. RESULTS

3.1. Band energy

The first purpose of this part is to study the energy distribution of vowels (/a/, /i/ and /u/) in the predefined frequency bands (B1, B2, B3, B4, B5, and B6) according to production duration. The Tables 2-4 summarize the obtained results. We can notice that all vowels have a significant energy in the first band since they are voiced sounds. Additionally, the percentage distribution of energy bands is unaffected by the production duration.

We can also see that for /a/, more than 70% of the total energy of the vowel is concentrated in bands B1 and B2 in equal parts ($\approx 35\%$). On the other hand, for /u/, more than 70% of its energy is located in band B1 and ($\approx 20\%$) in band B2. For the vowel /i/, more than 70% of its energy is concentrated in the band B1 against ($\approx 20\%$) in the band B5.

Table 2. Distribution of the energy percentage in the bands according to the production time of the vowel /a/

Production duration (s)	B1 (%)	B2 (%)	B3 (%)	B4 (%)	B5 (%)	B6 (%)
0,316	43,16	41,81	8,19	4,89	1,54	0,37
0,636	37,91	40,06	9,91	6,80	4,26	1,02
0,867	40,97	42,02	8,25	5,71	2,62	0,39
0,983	36,43	44,81	10,62	4,64	2,09	0,37

Table 3. Distribution of the energy percentage in the bands according to the production time of the vowel /i/

Production duration (s)	B1 (%)	B2 (%)	B3 (%)	B4 (%)	B5 (%)	B6 (%)
0,370	76,07	2,71	0,009	0,006	19,21	1,95
0,651	76,46	3,32	0,008	0,007	18,4	1,72
0,720	76,36	2,99	0,008	0,007	18,9	1,66
0,967	71,32	2,89	0,010	0,008	23,6	2,08

Table 4. Distribution of the energy percentage in the bands according to the production time of the vowel /u/

Production duration (s)	B1 (%)	B2 (%)	B3 (%)	B4 (%)	B5 (%)	B6 (%)
0,431	87,73	11,64	0,52	0,01	0,05	0,03
0,558	86,98	12,46	0,47	0,00	0,03	0,02
0,639	80,80	17,95	1,13	0,03	0,05	0,00
0,948	82,57	15,38	1,90	0,03	0,09	0,00

The examination of the bands energy distribution of the vowels /a/, /i/, /u/ (see Figures 1-3) reveals two phases:

- A transient phase which represents the beginning of vowel production and characterized by large changes in values of normalized energy bands.
- A steady phase when increasing vowel production time, where normalized energy in the bands (B1, B2, B3 and B5) represents significant variations. These results are consistent with those of [20].

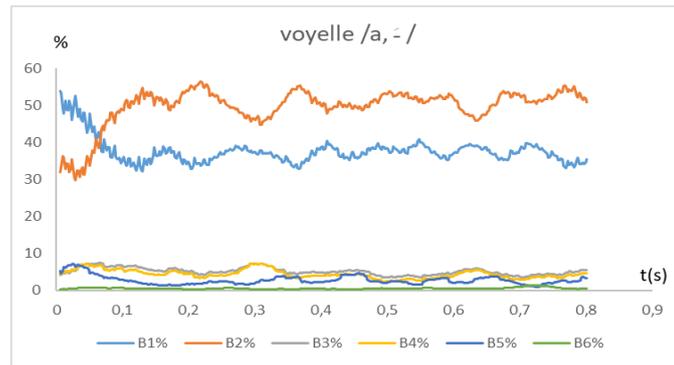


Figure 1. The bands energy distribution of the vowels /a/

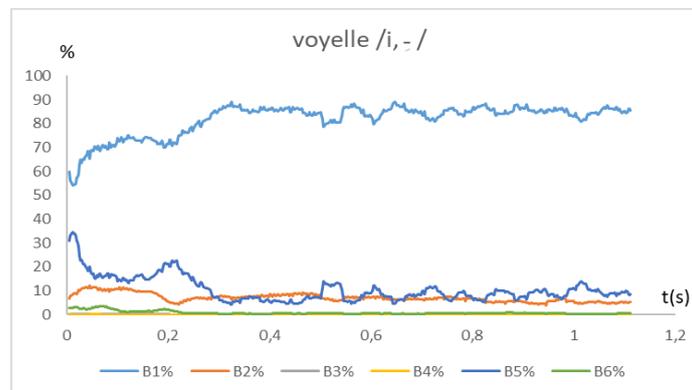


Figure 2. The bands energy distribution of the vowels /i/

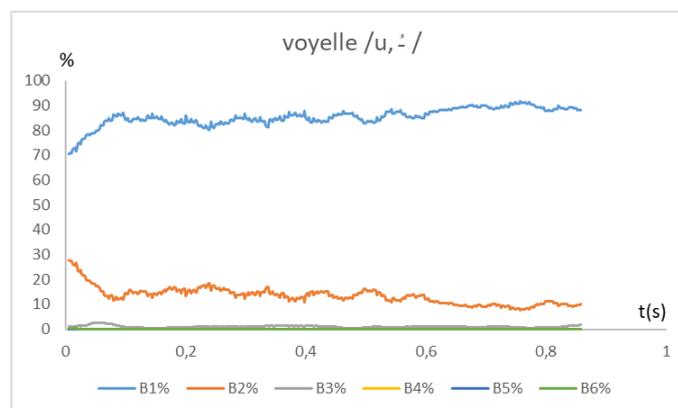


Figure 3. The bands energy distribution of the vowels /i/

3.2. Algorithm

Based on the results obtained for the energy distribution in the predefined frequency bands of vowels (/a/, /i/ and /u/) according to production duration, we have developed an algorithm which allows the classification of vowels. This algorithm consists of two parts: the first part is used to recognize the vowel /a, a:/, /i, i:/ or /u, u:/ and the second part is used to decide if this vowel is short or long.

Vowel recognition: by analyzing the results of the energy distribution in the six frequency bands [20], it can be seen that the normalized energy in the B1 band distinguishes the vowel /a, a:/. The normalized energies of the B2 and B5 bands make it possible to distinguish /u, u:/ from /i, i:/. Figure 4 shows the algorithm that makes this classification possible.

To classify the long vowels from the short ones, we have calculated the average rate of change of normalized energy in the bands (B1, B2, B3, and B5) in the vowel nucleus. If this average is less than three dB, it is a long vowel otherwise it is a short vowel. To determine the vowel nucleus, the vowel is divided into three parts of the same length (each part represents 1/3 of the total length of the vowel): the first third is the beginning of the vowel (onset), the second third is the nucleus of the vowel (nucleus) and the last third is the end of the vowel (offset). The algorithm that distinguishes the short vowel from the long one is given in Figure 5.

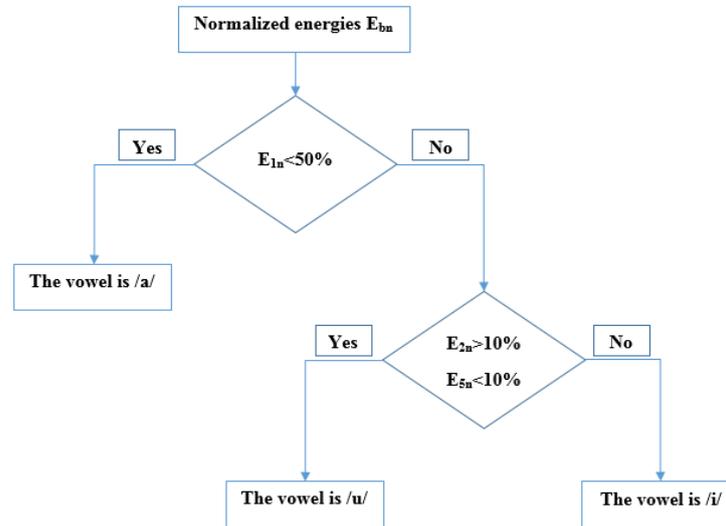


Figure 4. Flowchart of Arabic vowels classification /a/, /u/ and /i/

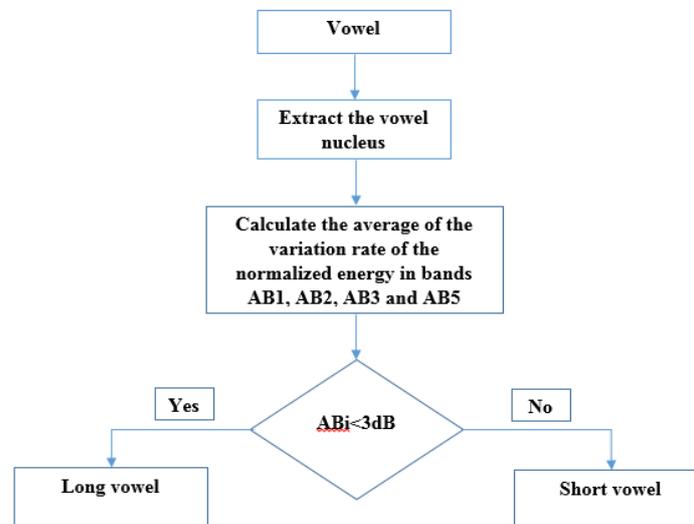


Figure 5. Flowchart of distinction between short and long vowel

3.3. Algorithm performance evaluation

Our algorithm was implemented in MATLAB and tested using the data from our corpus to determine its performance. The number of all short and long vowels in this experiment is 1200. 1167 vowels are accurately classified according to the data, giving the classification process a total accuracy of 97.25%. Table 5 offers more thorough outcomes. As we can see, there were a relatively high number of errors made when identifying the letters "u" and "i".

For the distinction between the long and short vowel, we conducted classification tests on our corpus: 300 short vowels and 900 long vowels (400 records for each vowel: 100 short and 300 long). The classification results are summarized in the Tables 6-8. We can see that this algorithm allows a correct classification of 1136

vowels, hence a recognition rate of 94%. The overall recognition rate of the six short and long vowels is 92%. These results are very competitive compared to those reported in the literature [17], [19], [24], [25].

Table 5. Confusion matrix of vowels

	/a/	/u/	/i/	Overallpercentage of errors (%)
/a/	396	3	1	0.33
/u/	0	387	13	1.08
/i/	2	14	384	1.33

Table 6. Confusion matrix of short and long vowels of /a/

	/a/	/a:/
/a/	91	9
/a:/	10	290

Table 7. Confusion matrix of short and long vowels of /u/

	/u/	/u:/
/u/	89	11
/u:/	7	293

Table 8. Confusion matrix of short and long vowels of /i/

	/i/	/i:/
/i/	82	12
/i:/	15	285

4. CONCLUSION

The main contribution of this paper is the development of an algorithm that allow to recognize each Arabic vowel (/a/, /a:/, /u/, /u:/, /i/ and /i:/). Based on the fact that the energy distribution over time and frequency of each sound depends on the articulator used and the place and manner of production, we conducted an acoustic study based the energy percentage in the bands (band 1: 0–400 Hz; band 2: 400–800 Hz; band 3: 800–1200 Hz; band 4: 1200–2000 Hz; band 5: 2000–3500 Hz and band 6: 3500–5000 Hz). The results demonstrate that each vowel can be classified using the normalized energy in the frequency bands. The algorithms proposed in this work use these indices to recognize each vowel. The performance tests of these algorithms on our Arabic corpus show recognition rates of 92% for the vowels.

As perspectives of this research work, several axes can be explored. We can explore the deportation of those algorithms on platforms and test their robustness in a noisy environment. Searches can also be oriented towards the characterization of other phonemes.

REFERENCES

- [1] B. Prica and S. Ilic, "Recognition of Vowels in Continuous Speech by Using Formants," *Electronics and Energetics*, vol. 23, no. 3, pp. 379-393, December 2010, doi: 10.2298/FUEE1003379P.
- [2] A. W. Howitt, "Vowel Landmark Detection," *The Journal of the Acoustical Society of America*, vol. 112, p. 2279, 2002, doi: 10.1121/1.4779139.
- [3] G. Fant, "Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations," *Gravenhage, Mouton: English*, 1960.
- [4] 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," *New Challenges in Applied Intelligence Technologies*, SCI 134, pp. 155–162, 2008, doi: 10.1007/978-3-540-79355-7_15.
- [5] 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, June 2010.
- [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*, pp. 174–177, 2003, 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," *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, 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," *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, 2008, pp. 1-5, doi: 10.1109/ICTTA.2008.4530026.
- [9] D. Gargouri, M. A. Kammoun, and A. B. Hamida "A Comparative Study of an All Pole Speech Analysis for Formant Extraction," *Third International Conference on Systems, Signals & Devices, Sousse, Tunisia*, 2005.
- [10] 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. e75–84, 2011, doi: 10.1016/j.jvoice.2010.10.018.
- [11] H. K. D. Souza and J. C. Mora, "Effects of speech rate on the intelligibility of non-native vowels," In *Proceedings of the International Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics*, vol. 5, pp. 337-352, 2014.
- [12] S. H. Ferguson and D. K.-port, "Talker differences in clear and conversational speech: acoustic characteristics of vowels," *Journal of Speech Language and Hearing Research*, vol. 50, pp. 1241–1255, 2007, doi: 10.1044/1092-4388(2007)087).
- [13] P. P. K. Mok, "Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation," *Language and Speech*, vol. 54, pp. 527–545, 2011, doi: 10.1177/0023830911404961.
- [14] G. Khattab and J. Al-Tamimi, "Durational cues for gemination in Lebanese Arabic," *Language and Linguistics*, vol. 22, pp. 39–55, 2008.
- [15] M. F. Tolba, T. Nazmy, A. A. Abdelhamid, and M. E. Gadallah, "A Novel Method for Arabic Consonant/Vowel Segmentation using Wavelet Transform," *International Journal on Intelligent Cooperative Information Systems IIJICIS*, vol 5, no. 1, pp. 353-364, July 2005.
- [16] A. A. M. Ahmed, "Production and Perception of Libyan Arabic Vowels," PhD Thesis, Newcastle University, pp. 153–167, 2008.
- [17] K. Tsukada, "An Acoustic Comparison of Vowel Length Contrasts in Standard Arabic, Japanese and Thai," *2009 International*

- Conference on Asian Language Processing*, 2009, pp. 76-79, doi: 10.1109/IALP.2009.25.
- [18] M. Abuoudeh and O. Crouzet, "Vowel Length Impact on Locus Equation Parameters: An Investigation on Jordanian Arabic," *Interspeech*, pp. 184-188, 2014, doi: 10.13140/2.1.5092.0966.
- [19] Y. A. Alotaibi and A. Hussain, "Speech recognition system and formant based analysis of spoken Arabic vowels," In *Proceedings of First International Conference, December, FGIT*, pp. 50-60, 2009, doi: 10.1007/978-3-642-10509-8_7.
- [20] K. Tahiry, B. Mounir, I. Mounir and A. Farch, "Energy bands and spectral cues for Arabic vowels recognition," *International Journal of Speech Technology*, vol. 19, no. 4, pp. 707-716, 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)*, pp. 384-388, 2007.
- [22] P. Milenkovic and K. Forrest, "Classification of vowels using spectrum moments," *Journal of the Acoustical Society of America*, vol. 83, no. S67, 1988, doi: 10.1121/1.2025467.
- [23] P. Korkko, "Spectral moments analysis of /s/ coarticulation development in Finnish-speaking children," In *ICPhS Proceedings of the 18th International Congress of Phonetic Sciences*, ISBN 978-0-85261-941-4, 2015.
- [24] 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, pp. 1025-1033, 2017, doi: 10.1121/1.4998706.
- [25] M. Stanek and L. Polak, "Algorithms for vowel recognition in fluent speech based on formant positions," *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, 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 & 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. Her research interests include speech recognition, 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.