# Efficiency of the energy contained in modulators in the Arabic vowels recognition 

Nesrine Abajaddi ${ }^{1}$, Youssef Elfahm ${ }^{2}$, Badia Mounir ${ }^{3}$, Laila Elmaazouzi ${ }^{4}$, Ilham Mounir ${ }^{5}$, Abdelmajid Farchi ${ }^{6}$<br>${ }^{1,2,6}$ IMMII Laboratory, Faculty of Sciences and Technics, University Hassan First, Settat, Morocco<br>${ }^{3,4,5}$ LAPSSII Laboratory, Graduate School of Technology, University Cadi Ayyad, Safi, Morocco

## Article Info

## Article history:

Received Jul 8, 2020
Revised Jan 1, 2021
Accepted Jan 13, 2021

## Keywords:

Arabic vowels
Center-of-gravity
Energy bands
Modulators


#### Abstract

The speech signal is described as many acoustic properties that may contribute differently to spoken word recognition. Vowel characterization is an important process of studying the acoustic characteristics or behaviors of speech within different contexts. This current study focuses on the modulators characteristics of three Arabic vowels, we proposed a new approach to characterize the three Arabic vowels /a/, /i/ and / $\mathrm{u} /$. The proposed method is based on the energy contained in the speech modulators. The coherent subband demodulation method related to the spectral center of gravity (COG) was used to calculate the energy of the speech modulators. The obtained results showed that the modulators energy help characterize the Arabic vowels /a/, /i/ and $/ \mathrm{u} /$ with an interesting recognition rate ranging from $86 \%$ to $100 \%$.


This is an open access article under the CC BY-SA license.


## Corresponding Author:

Nesrine Abajaddi
Department of Electrical and Computer Engineering
IMII Laboratory, Faculty of Sciences and Technics
University Hassan First
FST of Settat, Km 3, B.P: 577 Road of Casablanca, Settat, Morocco
Email: n.abajaddi@uhp.ac.ma

## 1. INTRODUCTION

The vowel is a sound of the language characterized, in its production, by a free passage of the air in the cavities located above the glottis, namely the nasal cavity and/or the oral cavity. These cavities serve as filters whose shape and the relative contribution to the airflow affect the quality of the sound obtained [1]. The majority of the vowels are sound, that is, they are pronounced with a vibration of the vocal cords [2]. The phonetic system of the Arabic language has six vowels which are divided into two types, the short vowels (/a/, /i/ and /u/) and the long vowels (/a:/, /i:/ and /u:/). Quality and quantity are two phonetic variables used in the description of vowels. The quantity describes the duration of the vowel while the quality relates to the description of the position of the tongue in the vocal tract, the place of articulation of the vowel, the shape of the lips, the size of the narrowing and the status of the vowel [3].

Several searches have been done on the vowels in order to recognize them or to understand the physiological behavior of the vocal tract during their production. The acoustic indices most used by researchers for the characterization of vowels are the formants and the spectral moments as well as the energy in the frequency bands. Thyer et al. [4] worked on the identification of English vowels of Australia; they found out that the most important cues in the characterization of vowels are the first two formants F1 and F2. In another research related to the classification of vowels. Weber et al. [5] evaluated formant characteristics for vowel classification, they disclosed that MFCCs and formants have almost the same effect
on noisy and clean speech signals. To differentiate between racial populations, other researches have used formants. Xue et al. [6] recorded that, relative to white Africa, Chinese male speakers have a lower F1 form, compared to Afro-Americans, a larger F2 form for Chinese and white Americans.They concluded, on the one hand, that the configuration of the vocal tract differentiates the Chinese from the people of White Africa and the Americans of African descent. On the other hand, the difference is related to the context of the vowels which is influenced by the dialect and the specificities of the language. These findings were also mentioned by Linville et al. [7] who have reported that vocal tract elongation is related to the speaker and articulatory changes in vowels. Savela et al. [8] presented a study on the identification of different vowels of the Udmurt language using spectral moments and formants (F1 and F2). They reported that the spectral moments help get an exhaustive description of the category of the vowel.

Vaissière [9] studied six focal vowels of the French language. He concluded that these vowels are characterized by a complete fusion of two adjacent formants. He noticed that for $/ \mathrm{i} /$, F3 and F4 (sometimes F4 and F5) are very close. For /u/, F1 and F2 are so close that they appear as a single formant whose values are less than 1000 Hertz. /a/ presents a rapprochement of F1 and F2, F1 has a maximum value around 1000 Hz . Korkmaz and Aytug [10] exploited the formant frequency values (F1, F2 and F3) of eight Turkish vowels within CVC syllables to identify the most efficient classification method. They deduced that support vector machines (SVM) distinguishes Turkish vowels from formant frequencies with a classification success of $90 \%$. Some studies were conducted on Arabic vowels. Alghamdi [11] performed a frequency comparison between short and long vowels for several Arabic dialects (Sudanese, Egyptian and Saudi). He stated that the first two formants of long vowels differ from those of short vowels. Iqbal et al. [12] carried out a new study on the segmentation and identification of vowels using formant transitions in the continuous recitation of Quranic Arabic. they identified the Arabic vowels based on the analysis of the extracted formants. The average accuracy of the vowel identification system is $90 \%$. From another viewpoint, Alotaibi and Hussain [13] analyzed standard Arabic vowels formants. They showed that Formants F1 and F2 contribute to the classification of vowels; The vowel /i/ or /i:/ have a high value of F2 (F2> 1500 Hz ) and the vowel /a/ or /a:/ is characterized by a high value of $\mathrm{F} 1(\mathrm{~F} 1>500 \mathrm{~Hz})$.

Tsukada [14] studied the long and short vowels in Standard Arabic, Japanese and Thai. He informed that the length of long vowels is twice the length of short vowels. He also noticed that in the three languages, the ratio between the length of short and long vowels differed significantly. Natour et al. [15] worked on the examination of Arabic vowels created by men, women, and children in different languages, namely French, German and American English. They indicated that, in general, the Arabic language has a low value of F1 and F2 and a high value of F3. Moreover, the dispersion of the formants of Arabic vowels is identical to that of other studies in other languages. Abuoudeh and Crouzet [16] examined the impact of the length of the Jordanian Arabic vowel on the parameters of the locus equation. They observed that the length of the vowel systematically influences the data of the locus equation and that variations in the length of the vowel are associated with changes in the spectral configuration. Another research on vowels in Palestinian Arabic was carried out by Adam [17] who studied the variation in vowel duration in normal speakers and speakers with Broca's aphasia. The study concluded that, relative to regular speakers, vowel duration was longer for speakers with Broca's aphasia. Tahiry et al. [18] used another approach to characterize Arabic vowels. This approach is based on three methods: formants frequencies extraction, calculation of spectral moments and calculation of energy in six frequency bands. They concluded that for all three methods, the behavior of vowels depends on their place of articulation rather than their duration of production.

Aloqayli and Alotaibi [19] were analysis the values of the formants frequencies and their derivatives to evaluate the vowels in the modern standard Arabic (MSA) dialect. To develop an automatic recognition system of the six MSA vowels, they used artificial neural networks (ANN). The networks were examined with one and two hidden layers, each with a different number of neurons. In the case of one hidden layer, they detected that the system using 16 neurons had an overall performance of $87.96 \%$, while the system using 30 neurons had an overall performance of $86.11 \%$. In the case of two hidden layers, the system with 5 neurons in the first hidden layer and 7 neurons in the second hidden layer had an overall system performance of $83.33 \%$, and the system with 8 neurons in the two hidden layer had an overall system performance of $87.04 \%$. Farchi et al. [20] analyzed long and short Arabic vowels using production duration and energy distribution located in the first and second formant. They spotted that long vowels are twice longer than short vowels and also when the production duration of long vowels increases, the F2 band energy remains constant for all vowels, and the F1 band energy increases or decreases depending on the vowel.

This paper seeks to characterize and recognize Arabic vowels using energy in the modulators to integrate it into many applications of speech recognition and speech intelligibility. Indeed, recognition and speech enhancement systems have gained popularity in recent years. Parallel to that, the modulation domain has been extensively studied in speech applications as it offers a more compact representation. So, this work can be useful in many applications. Its main goal is to determine if and how modulators can contribute to the
identification and recognition of each Arabic vowel. This document is structured as follows. The first part describes the methods and tools used as well as the experiments carried out. The second section is devoted to the analysis and the discussion of the results. Finally, the last part presents a conclusion of this work.

## 2. METHODS

### 2.1. Corpus

The vowels corpus of the Arabic language was constructed by asking nine people (male) in their twenties to pronounce the vowels $/ \mathrm{a} / \mathrm{l} / \mathrm{i} /$ and $/ \mathrm{u} /$ repeatedly. The speakers have repeated each of the vowels 50 times. We have recorded 1350 vowel samples. These recordings were done as a mono sound in the laboratory using the vocal sounds process tool "Praat" with a sampling frequency of 22050 Hz , in a noise-isolated room.

### 2.2. Proposed method

The study existed is treated in the acoustic or spectral domain. In this paper, we focused on a demodulation method to discover the characteristics of the Arabic language in the modulation domain. Such a study constitutes the first step of a larger project which aims to add new characteristics of the Arabic language in this domain. Indeed, to identify the Arabic vowels, we adopted the schematic approach in Figure 1. First of all, the speech signal passed through a filter bank (called the analysis filter bank) in order to decompose it into subband signals. Then, each signal coming from a subband was demodulated so as to extract the modulator on which we calculated the energy. This energy makes it possible to characterize and identify each of the three Arabic vowels.


Figure 1. Block diagram used for processing the speech signal

### 2.3. Analysis filter bank

A filter bank is a set of low-pass, band-pass and high-pass filters. Each filter extracts a single frequency subband from the signal. These filters are used for the spectral decomposition of signals [21]. The input signal is decomposed into $M$ subband signals by applying $M$ analysis filters with different bandwidths. Thus, each of these subband signals contains the input signal information relating to a particular frequency band. The blocks with arrows pointing downwards indicate downsampling by factor N .

### 2.4. Demodulation

Naturally, speech can be described as a modulated signal, expressed as a sum of amplitudemodulated (AM) signals in a set of narrow frequency subbands covering the signal bandwidth [22]. Each subband can be considered as a pair (modulator "envelope"/carrier "fine structure"). Each modulator can be studied, modified and analyzed independently of the carrier or recombined with its corresponding carrier separately. So, in discrete-time, we can describe the full-band speech by:

$$
\begin{equation*}
\mathrm{x}(\mathrm{n})=\sum_{\mathrm{k}=1}^{\mathrm{M}} \mathrm{~s}_{\mathrm{k}}(\mathrm{n})=\sum_{\mathrm{k}=1}^{\mathrm{M}} \mathrm{~m}_{\mathrm{k}}(\mathrm{n}) \circ \mathrm{c}_{\mathrm{k}}(\mathrm{n}) \tag{1}
\end{equation*}
$$

where $\mathrm{x}(\mathrm{n})$ is the observed speech signal, $\mathrm{m}_{\mathrm{k}}(\mathrm{n})$ and $\mathrm{c}_{\mathrm{k}}(\mathrm{n})$ represent the k -th modulator and the carrier waveforms of each subband $s_{k}(n)$, respectively. $M$ is the number of subbands. (o) the operator designates the sample-by-sample multiplication (Hadamard product).

Two classes of methods (coherent and incoherent) are used in the modulation decomposition of the speech signal [23]. Using the non-coherent subband demodulation technique, $\mathrm{m}_{\mathrm{k}}(\mathrm{n})$ and $\mathrm{c}_{\mathrm{k}}(\mathrm{n})$ are calculated from the analytic bandpass subband signal $s_{k}(n)$ obtained from the filter bank. To extract the modulator from each subband, the non-coherent demodulation seems simple and easy to implement. The modulator (also known as the "Hilbert envelope") and the carrier are written, respectively, in the form:

$$
\begin{align*}
& m_{k}(n)=\left|s_{k}(n)\right|  \tag{2}\\
& c_{k}(n)=\exp \left(j \angle s_{k}(n)\right) \tag{3}
\end{align*}
$$

However, the modulator and the carrier are not limited in the predefined frequency band then it's not possible to modify them individually. For this reason, we adopted a coherent subband demodulation. We used the spectral center-of-gravity (COG) method [24] to evenly track the evolution of spectral concentration within subband over time, as an estimate of the instantaneous frequency $f_{k}(n)$. The carrier phase is defined in (4). Coherent demodulation defines the modulator according to the estimated carrier in (5) and (6).

$$
\begin{align*}
& \phi_{\mathrm{k}}(\mathrm{n})=\sum_{\mathrm{p}=0}^{\mathrm{n}} \mathrm{f}_{\mathrm{k}}(\mathrm{p})  \tag{4}\\
& \mathrm{c}_{\mathrm{k}}(\mathrm{n})=\exp \left(\mathrm{j} \angle \mathrm{~s} \phi_{\mathrm{k}}(\mathrm{n})\right)  \tag{5}\\
& \mathrm{m}_{\mathrm{k}}(\mathrm{n})=\mathrm{s}_{\mathrm{k}}(\mathrm{n}) \cdot \mathrm{c}_{\mathrm{k}}^{*}(\mathrm{n}) \tag{6}
\end{align*}
$$

### 2.5. Energy band

The analysis of the temporal variations and the average energy of the modulators reveals that five bands allow identifying each of the vowels. The energy distribution of the speech signal in these frequency bands makes it possible to determine the modulators which characterize each vowel $/ \mathrm{a} / \mathrm{/} / \mathrm{u} /$, or $\mathrm{i} /$. The five frequency bands used in this study are as follows: Band $1(100-400 \mathrm{~Hz})$; Band $2(400-800 \mathrm{~Hz})$; Band 3 $(800-2000 \mathrm{~Hz})$; Band $4(2000-3500 \mathrm{~Hz})$ and Band $5(3500-5000 \mathrm{~Hz})$. The modulator of each subband was hamming windowed ( 256 points with an overlap of $50 \%$ ). Then, the energy in each window of the modulator was calculated by (7):

$$
\begin{equation*}
\mathrm{E}(\mathrm{~b}, \mathrm{j})=\sum_{\mathrm{k}=0}^{\mathrm{i}-1}|\mathrm{~m}(\mathrm{~b}, \mathrm{k})|^{2} \tag{7}
\end{equation*}
$$

where $b$ is the number of modulators, $j$ is the number of windows and $i$ is the length of the window. Next, we calculated the percentage energy of each window for each modulator as (8):

$$
\begin{equation*}
\mathrm{P}_{\mathrm{ET}}=\frac{\mathrm{E}(\mathrm{~b}, \mathrm{j})}{\mathrm{E}_{\mathrm{T}}(\mathrm{j})} * 100 \tag{8}
\end{equation*}
$$

With $E_{T}(j)$ is the total energy of all modulators in the same window, calculated by (9):

$$
\begin{equation*}
E_{T}(j)=\sum_{n=0}^{b} E(j, n) \tag{9}
\end{equation*}
$$

## 3. RESULTS AND DISCUSSION

### 3.1. The energy in the modulators and recognition system

The aim of this part is to analyze the modulators of the three Arabic vowels (/a/, /i/ and $/ \mathrm{u} /$ ). We calculated the percentage of total energy in each modulator by (8) for all samples in our vocal corpus as shown in Figures 2, 3 and 4. Then, we calculated the average percentage of each modulator relative to the others in order to extract the dominant modulators for each vowel. Based on the average energy percentage value for each modulator for the three vowels, we can see that:

The vowel /a/: the majority of energy is located in the second modulator (49\%), the third modulator $(35 \%)$, and the first modulator which contains less important energy of the order of $14 \%$ as shown in Figure 2. These results can be justified by the fact that for the vowel/a/, the first formant F1 is located around 600 Hz which corresponds to the second modulator $(400-800 \mathrm{~Hz})$. On the other hand, the second formant F2 is located around 1000 Hz , hence the energy contained in the third modulator ( $800-2000 \mathrm{~Hz}$ ) [18].

The vowel /i/: the energy is focused in the first modulator ( $77 \%$ ) and the fourth modulator ( $20 \%$ ) as shown in Figure 3. We can explain this result by the fact that the first formant F1 of the vowel /i/ is about

300 Hz which implies the first modulator $(100-400 \mathrm{~Hz})$. On the other hand, the second formant F2 is located around 2000 Hz , hence the energy located in the fourth modulator ( $2000-3500 \mathrm{~Hz}$ ) [18].

The vowel /u/: the energy is concentrated on the first two modulators ( $27 \%$ for M1 and $68 \%$ for M2) as shown in Figure 4. This is justified by the fact that the first two formants F1 and F2 are concentrated on the low frequencies and the distance between them is small (F1> 300 Hz and $\mathrm{F} 2<1000 \mathrm{~Hz}$ ). This shows the existence of the first two modulators (M1: $100-400 \mathrm{~Hz}$ and M2: $400-800 \mathrm{~Hz}$ ) [18].


Figure 2. The calculation results of (a) The modulators energy percentage distribution of the vowel /a/, (b) The modulators energy percentage average of the vowel /a/


Figure 3. The calculation results of (a) The modulators energy percentage distribution of the vowel $/ \mathrm{i} /$, (b) The modulators energy percentage average of the vowel /i/


Figure 4. The calculation results of (a) The modulators energy percentage distribution of the vowel $/ \mathrm{u} /$, (b) The modulators energy percentage average of the vowel /u/

Therefore, we can notice that each of these Arabic vowels can be at most characterized by two main modulators. So, we have developed a simple recognition system that classifies these three Arabic vowels based on the previously proposed modulators. Figure 5 presents the adopted algorithm.


Figure 5. Algorithm of classification of Arabic vowels

### 3.2. Recognition system performance

To test the performance of our approach, it was implemented under MATLAB and tested utilizing the records from our corpus. The number of all vowels in this experiment is 1350 . To evaluate our identification algorithm, we calculated the confusion matrix as shown in Table 1. This confusion matrix represents the recognition percentage of Arabic vowels. The last column (Unclassified) represents the number of cases that could not be identified.

Table 1. Vowels confusion matrix

| cable 1. |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | $/ \mathrm{a} /$ | li/ | /u/ | Unclassified |
| $/ \mathrm{a} /$ | 100 | 0 | 0 | 0 |
| $/ \mathrm{i} /$ | 0 | 95.33 | 4.66 | 0 |
| $/ \mathrm{u} /$ | 13.77 | 0 | 86.22 | 0 |

We can see from the results of the adopted approach that the identification error is null, all the vowels have thus been identified. The values $13.77 \%$ and $4.66 \%$ represent the confusion percentage of a vowel with another ( $/ \mathrm{a} /$ with $/ \mathrm{u} /$ and $/ \mathrm{u} /$ with $/ \mathrm{i} /$ respectively). The confusion between the vowel $/ \mathrm{u} / \mathrm{and} / \mathrm{a} /$ or /i/ respectively is due to the energy distribution of /u/ which is concentrated on the first two modulators M1 and M2. Thus, an interlocutor variation may lead to drop-in energy in the first modulator following a value of the first F1 formant greater than 400 Hz . Hence, a confusion between /u/ and /a/. In the same way, an interlocutor variation can generate a higher value of the second formant F2 (> 3500 Hz ), the energy in the fourth modulator decreases, and consequently, confusion is induced between $/ \mathrm{u} /$ and $/ \mathrm{i} /$.

### 3.3. Discussion

The characterization of Arabic vowels based on modulators reveals that modulators can classify Arabic vowels (/a/, /i/ and /u/). Each of these vowels can be characterized by two main modulators. According to the analysis of the modulators in the subbands, we can notice that all the vowels have a significant energy percentage in the first modulator because they are all voiced sounds. However, this percentage of energy varies from one vowel to another ( $/ \mathrm{a} /: \approx 13 \%$; $/ \mathrm{i} /: \approx 77 \%$ and $/ \mathrm{u} /: \approx 27 \%$ ).

The results show that vowel/a/ is identified by the two main modulator M2 (second modulator) and M3 (third modulator). These results may be justified by the fact that the vowel/a/ is produced by lifting the central part of the tongue to the junction between hard and soft palates [25]. Thus, the distribution of Formants F1 and F2 is in agreement with our results: F1 is greater than 600 Hz , this energy is thus found in the second modulator M2 $(400-800 \mathrm{~Hz})$. F2 is greater than 1000 Hz , then the third modulator M3 ( $800-2000 \mathrm{~Hz}$ ) contains an important part of energy [18].

The vowel /i/ is distinguished by its energetic distribution in the fourth modulator because it is produced by raising the tip of the tongue towards the hard palate [25] so F2 moves towards the high frequencies ( $\mathrm{F} 2>2000 \mathrm{~Hz}$ ) and therefore the energy is concentrated in M4. On the other hand, /i/ also
contains a significant percentage of energy in the first modulator. This concentration of energy in the first modulator M1 can be explained by the value of the first formant F1 ( $\mathrm{F} 1<300 \mathrm{~Hz}$ ) which is located in the low frequencies in addition to the energy due to the glottal vibration ( $<400 \mathrm{~Hz}$ ) generated by the voiced sounds [18].

For the vowel $/ \mathrm{u} /$, most of the energy is focused on the first modulator (M1: 100-400 Hz) and the second modulator (M2: $400-800 \mathrm{~Hz}$ ). This behavior is due to the two first formants F1 and F2 which are so close and concentrated in low frequencies ( $\mathrm{F} 1>300$ and $\mathrm{F} 2<1000 \mathrm{~Hz}$ ) [18]. These results may be justified by the fact that the vowel $/ \mathrm{u} /$ is a back vowel and is produced by raising the back of the tongue towards the soft palate [25].

Therefore the obtained results in this work are consistent with those reported by other research studies on the classification of vowels using formants in the spectral domain (Alghamdi [11]; Alotaibi [13]; Vaissière [9]; Tahiry [18]; Aloqayli and Alotaibi [19]). On the other hand, the recognition system of Arabic vowels proposed in this work presents interesting advantages. Indeed, it is a simple system in its implementation whose recognition performance is very interesting ( $100 \%$ for $/ \mathrm{a} /$, $95.33 \%$ for $/ \mathrm{i} /$ and $86.22 \%$ for $/ \mathbf{u}$ ) comparing with the other methods [19] as shown in Table 2.

Table 2. Recognition rate comparison

|  | The proposed <br> method $(\%)$ | ANN (One hidden <br> layers with 16 <br> neurons) $(\%)$ | ANN (One hidden <br> layers with 30 <br> neurons) $(\%)$ | ANN (Two hidden <br> layers with [5-7] <br> neurons) $(\%)$ | ANN (Two hidden <br> layers with [8-8] <br> neurons) (\%) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| /a/ | 100 | 100 | 88.89 | 83.33 | 83.33 |
| /i/ | 95 | 83.33 | 83.33 | 72.22 | 77.78 |
| /u/ | 86.22 | 72.22 | 83.33 | 83.33 | 100 |

## 4. CONCLUSION

In this article, a new approach is presented to recognize Arabic vowels based on modulators energy. This method is based on the calculation of the percentage of energy contained in modulators in predefined frequency bands. The results showed that each vowel can be identified by two main modulators whose energy percentage is quite important. The recognition rate of the vowels is largely satisfactory to ensure the robustness of this algorithm. These results encourage us to characterize the remaining consonants in future work in order to build a simple and efficient speech recognition system that can be used in denoising applications, sound separations and segmentation.

## REFERENCES

[1] "Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet," Cambridge University Press, 1999.
[2] K. N. Stevens, "Acoustic Phonetics," MIT Press, Cambridge, 1998.
[3] E. Saadah, "The Production of Arabic Vowels by English L2 Learners and Heritage Speakers of Arabic," Dissertation, University of Illinois, 2011.
[4] Thyer, N., Hickson, L., and Dodd, B., "The perceptual magnet effect in Australian English vowels," Perception \& Psychophysics, vol. 62, no. 1, pp. 1-20, 2000, doi: 10.3758/BF03212057.
[5] De Wet, F., Weber, K., Boves, L., Cranen, B., Bengio, S., and Bourlard, H., "Evaluation of formant-like features on an automatic vowel classification task," J Acoust Soc Am., vol. 116, no. 3, pp. 1781-1792, 2004, doi: 10.1121/1.1781620.
[6] Xue, S. A., Hao, G. J. P., and Mayo, R., "Volumetric measurements of vocal tracts for male speakers from different races," Clinical Linguistics \& Phonetics, vol. 20, no. 9, pp. 691-702, 2006.
[7] Linville, S. E., and Rens, J., "Vocal tract resonance analysis of aging voice using long-term average spectra," Journal of Voice, vol. 15, no. 3, pp. 323-330, 2001.
[8] J. Savela, "Role of selected spectral attributes in the perception of synthetic vowels," Dissertation, University of Turku, 2009.
[9] J. Vaissière, "On the acoustic and perceptual characterization of reference vowels in a cross-language perspective," in Proceedings of the ICPhS, 2011, pp. 52-59.
[10] Y. Korkmaz and B. Aytuğ, "Classification of Turkish Vowels Based on Formant Frequencies," International Artificial Intelligence and Data Processing, 2018, doi: 10.1109/IDAP.2018.8620877.
[11] M. M. Alghamdi, "A spectrographic analysis of Arabic vowels: A cross-dialectal study," Journal of King Saud University, vol. 10, pp. 3-24, 1998.
[12] Iqbal, H. R., Awais, M. M., Masud, S., and Shamail, S., "On Vowels Segmentation and Identification Using Formant Transitions in Continuous Recitation of Quranic Arabic," New Challenges in Applied Intelligence Technologies Springer, SCI 134, pp. 155-162, 2008.
[13] Y. Alotaibi and A. Hussain, "Speech recognition system and formant based analysis of spoken Arabic vowels," in Proceedings of First International Conference, FGIT, 2009, pp. 10-12.
[14] K. Tsukada, "An Acoustic Comparison of Vowel Length Contrasts in Standard Arabic, Japanese and Thai," International Conference on Asian Languages Processing, 2009, doi: 10.1109/IALP.2009.25.
[15] Natour, Y. S., Marie, B. S., Saleem, M. A., and Tadros, Y. K., "Formant frequency characteristics in normal Arabic speaking Jordanians," Journal of Voice, vol. 25, no. 2, pp. 75-84, 2010.
[16] M. Abuoudeh and O. Crouzet, "Vowel Length Impact on Locus Equation Parameters: An Investigation on Jordanian Arabic," 15th Annual Conference of the International Speech Communication Association, 2014.
[17] H. Adam, "Acoustical analysis of vowel duration in Palestinian Arabic speaking aphasics," Am. J. Psychiatry Neurosci., vol 2, pp. 13-17, Jan. 2014.
[18] Tahiry, K., Mounir, B., Mounir, I., and Farchi, A., "Energy bands and spectral cues for Arabic vowels recognition," International Journal of Speech Technology, vol. 19, pp. 707-716, 2016, doi: 10.1007/s10772-016-9363-3.
[19] F. M. Aloqayli and Y. A. Alotaibi, "Spoken Arabic Vowel Recognition Using ANN," European Modelling Symposium (EMS) IEEE, 2017.
[20] Farchi, M., Tahiry, K., Mounir, S., Mounir, B., and Mouhsen, A., "Energy distribution in formant bands for Arabic vowels," International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 2, pp. 1163-1167, 2019.
[21] A. Mertins, "Signal Analysis: Wavelets, Filter Banks," Time-Frequency Transforms and Applications, Wiley, 1999.
[22] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," IEEE Trans. Inf. Theory, vol. 4, no. 1, pp. 53-57, Mar. 1958.
[23] P. Clark and L. E. Atlas, "Time-frequency coherent modulation filtering of non-stationary signals," Trans. Signal Processing, vol. 57, no. 11, pp. 4323-4332, Nov. 2009.
[24] P. J. Loughlin and B. Tacer, "On the amplitude- and frequency-modulation decomposition of signals," J Acoust. Soc. Am., vol. 100, no. 3, pp. 1594-1601, 1996.
[25] G. Fant, "Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations," Gravenhage, Mouton Co, 1960.

