International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 2, April 2023, pp. 1997~2008
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i2.pp1997-2008

Characterization of Arabic sibilant consonants

Youssef Elfahm[1], Nesrine Abajaddi[1], Badia Mounir[2], Laila Elmaazouzi[2], Ilham Mounir[2], Abdelmajid Farchi[1]
[1]IMII Laboratory, Faculty of Sciences and Technics, University Hassan First, Settat, Morocco
[2]LAPSSII Laboratory, High School of Technology, University Cadi Ayyad, Safi, Morocco

Article Info

Article history:

Received Jan 25, 2022
Revised Sep 15, 2022
Accepted Oct 12, 2022

Keywords:

Alveolar
Classification
Energy bands
Post-alveolar
Sibilant fricatives

ABSTRACT

The aim of this study is to develop an automatic speech recognition system in order to classify sibilant Arabic consonants into two groups: alveolar consonants and post-alveolar consonants. The proposed method is based on the use of the energy distribution, in a consonant-vowel type syllable, as an acoustic cue. The application of this method on our own corpus reveals that the amount of energy included in a vocal signal is a very important parameter in the characterization of Arabic sibilant consonants. For consonants classifications, the accuracy achieved to identify consonants as alveolar or post-alveolar is 100%. For post-alveolar consonants, the rate is 96% and for alveolar consonants, the rate is over 94%. Our classification technique outperformed existing algorithms based on support vector machines and neural networks in terms of classification rate.

Corresponding Author:

Youssef Elfahm
Laboratory IMII, Electrical and Computer Engineering Department, Faculty of Sciences and Technologies, Hassan First University
Road of Casablanca B.P: 577, Settat, Morocco
Email: y.elfahm@uhp.ac.ma

1. INTRODUCTION

The field of automatic speech processing has undergone considerable development in recent years. This development allowed humans to communicate with machines. As a result, speech recognition systems are used in a wide range of activities, both professional and public, such as newspaper writing, controlling industrial machinery, and so on. Knowledge of various fields, such as signal processing, linguistics, phonology, computer science, and statistics, is required in the subject of automatic speech recognition. from a phonetic perspective, vowels and consonants are the two basic kinds of vocal sounds. The creation of vowels necessitates open air circulation in the vocal tract, whereas the generation of consonants necessitates an interruption or disturbance in the flow of air at one point [1], [2]. Occlusive and fricative consonants are the two basic modalities of consonantal articulation in articulatory phonetics. Occlusives are noisy sounds of short duration marked by quiet caused by the complete closure of the vocal tract at a specific point (as in the consonant /k/). Fricatives, on the other hand, are noisy sounds created by turbulent airflow. There is a frictional noise (as in the consonant: /s/) when this flow hits a constriction [3], [4]. Fricatives can be grouped into the sibilant and non-sibilant categories. In the opposite of the non-sibilant consonants, the sibilant ones are produced by directing a flow of air with the tongue towards the edge of the teeth kept closed, resulting in a distinctive hissing sound [5]. Figure 1 reports the classification of Arabic consonants.

The researchers conducted many studies in order to design a voice recognition system and/or improve its performance. They used a variety of acoustic indices in their research, including the duration and amplitude of the frication, the center of gravity (CoG) value, spectral moments (skewness, mean, kurtosis, and standard deviation), gammatone filter outputs, Mel-frequency cepstral coefficients (MFCCs), and so on.

For sibilant consonants, several studies were conducted. Indeed, Behrens and Blumstein [6] undertook an examination of the temporal changes of the spectral features of English sibilants (/s/ and /ʃ/) as part of their work on the characterization of sibilant consonants. According to their findings, monitoring the frequency of the peak at the start, middle, and end of the consonant allows for highly accurate identification of these sounds. The consonant /s/ had a greater peak frequency than the consonant /ʃ/. The spectrum and intensity of fricative consonants can be used to determine the place of articulation, according to Borden [7]. When compared to non-sibilants, sibilants (/s/, /z/, /sh/, and /zh/) have unusually steep high frequency spectral peaks and comparatively high intensity levels. The alveolar cells' spectral peak (/s/ and /z/) is around 4 kHz. For a typical male speaker, the post-alveolar (/sh/ and /zh/) frequency is around 2.5 kHz. The duration and amplitude of the frication are also related to the articulation point, allowing to discriminate between sibilants and non-sibilants [8].

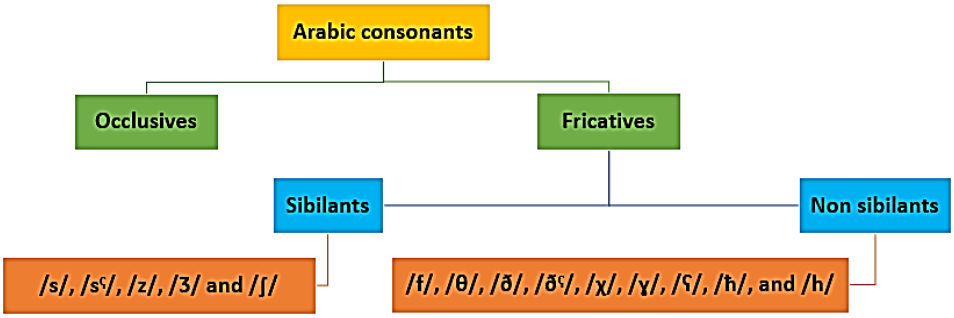

Figure 1. Diagram classifying the Arabic fricative consonants

To identify English fricative sounds, Ali *et al.* [9] used the maximum normalized spectral slope (MNSS) and the spectral CoG (SCoG). They stated in their paper that the detection of sibilants is done in two stages: the first is determining the voicing, and the second is determining the articulation location. They were 87 percent accurate on average. Regarding the recognition of English sibilants in terms of alveolar and post-alveolar, Ali *et al.* [10] found that alveolar peaks around 5 kHz while the post-alveolar peaks around 3 kHz. The CoGs, which have been identified between 2 and 4 kHz for the post-alveolar and between 4 and 8 kHz for the alveolar, can be used to distinguish the two classes of sibilants [11], [12]. The front cavity of the post-alveolar consonant /ʃ/ is larger than that of the alveolar consonant /s/ from an articulatory standpoint. This difference is accompanied by a qualitative difference in the shape of the front cavity: for /ʃ/, the tongue is positioned so that a sublingual cavity would be formed behind the lower incisors, whereas for /s/, the tongue is positioned in a way in which the underside of the tongue tip comes into contact with the lower incisors, obviating the need for a sublingual cavity [13]–[15]. Non-sibilant English fricatives have bigger standard deviations, lower overall amplitudes, and shorter durations than sibilant ones, according to spectral moments. The palato-alveolar junction place /ʃ/ (4.7 kHz) had a lower spectral mean than the alveolar one /s/ (7.1 kHz). The asymmetry average for the consonant /s/ was negative in all female productions and considerably positive in all male outputs [16], [17]. Kong *et al.* [18] focused on classifying English fricatives as alveolar, post-alveolar, or non-sibilant. The data was analyzed using spectral characteristics, gammatone filter outputs, and MFCCs. They achieved an accuracy of 88 percent with gammatone filter outputs and 87 percent with non-gammatone filter outputs. Kochetov [19] looked at the CoG, formants F1, F2 and F3 during the next vowel, and length of the four unvoiced Russian sibilants (/s, sʲ, ʂ, and ʃʲ /). During the frication area, he discovered that the CoG aids detect anterior versus posterior contrast. F1 and especially F2 at the beginning and middle of the next vowel distinguished the palatalized versus non-palatalized difference. Only /ʃʲ/ is distinguished by the fricative duration.

Cooper *et al.* [20] studied unvoiced fricatives using spectral moments, median power, and fricative duration as acoustic indicators when working on Arabic fricatives. This study demonstrated that spectral asymmetry may be used to determine consonant articulation points. The asymmetry value increases when the point of articulation is moved from the front to the back of the vocal tract. The value of /s/ was bigger than that of /ʃ/ in terms of spectral mean. The greatest values were found in alveolar fricatives, followed by post-alveolar fricatives, while the lowest values were found in non-sibilant fricatives. The spectral standard deviation values of sibilant fricatives were lower than those of non-sibilant fricatives. In his research on Arabic fricative consonants, Al-Khair [21] found that the spectral position of the peak is an acoustic index that permits to distinguish between alveolar sibilant consonants /s and z/ and post-alveolar /ʃ/. Arabic

sibilants have a compact spectrum with a greater intensity and frequency CoG than non-sibilants, according to Benamrane [22]. The consonants /s and s/ have a high CoG in comparison to the consonants /z, ʒ and ʃ/. Mokari and Mahdinezhad [23] have conducted a comparison of tow Azerbaijani fricative classifiers. The first system uses spectral moments, spectral peak, amplitude, and duration, whereas the second one employs cepstral coefficients. This comparison shows that the cepstral coefficients were more trustworthy predictors in the categorization of the nine Azerbaijani fricatives. Based on the energy in the bands as an acoustic indication, Elfahm *et al.* [24] developed a technique for categorizing Arabic fricative consonants into two main groups: sibilant and non-sibilant. They discovered that sibilant consonants had zero energy in the band (800 to 2,000 Hz), while non-sibilant had the lowest energy in the region (5,000 Hz to 8,000 Hz).

As can be seen from this overview, the major works were limited to classify the tow sibilant consonants /s and ʃ/ using spectral moments and CoG values as acoustic cues. In this study, our contribution is to extend the classification to the other Arabic sibilant consonants /s, sˤ, z, ʒ, ʃ/. Our algorithm uses the energy distribution in syllable to classify these consonants into two groups: alveolar /s, sˤ, and z/ and post-alveolar /ʒ and ʃ/. Then, it recognizes the consonants of each group. This paper is organized as follows: The methodology and instruments employed, as well as the experiments conducted, are presented in the first section. The results are presented and discussed in the second section. A summary of the findings and a presentation of the conclusions are included in the final section.

2. METHOD

This study took place in two phases: a phase of construction and segmentation of the corpus, a second phase concerning the processing and the acoustic analysis of the signal. The purpose of the first phase is to record vocal sequences and segment these sequences into syllabic units of the consonant-vowel CV type. In the signal processing part, we calculated the landmarks and the energy in frequency bands in order to use it in the acoustic analysis of the voice signal.

2.1. Corpus and signal processing

The data used for the acoustic analyzes presented in this study include audio recordings from our own corpus. We asked nine male Moroccan speakers to repeat a CVCVCV sequence four times, see Figure 2. All footage is recorded in an isolated chamber via the Labtech AM232 microphone which was placed 20 cm from the corner of the mouth and at a 45° angle to increase recording quality. The audio files were recorded in rural areas, using Praat software, at a sampling frequency of 22.05 KHz. From this dataset, we performed a segmentation operation, exploiting landmarks, to extract a CV sequence.

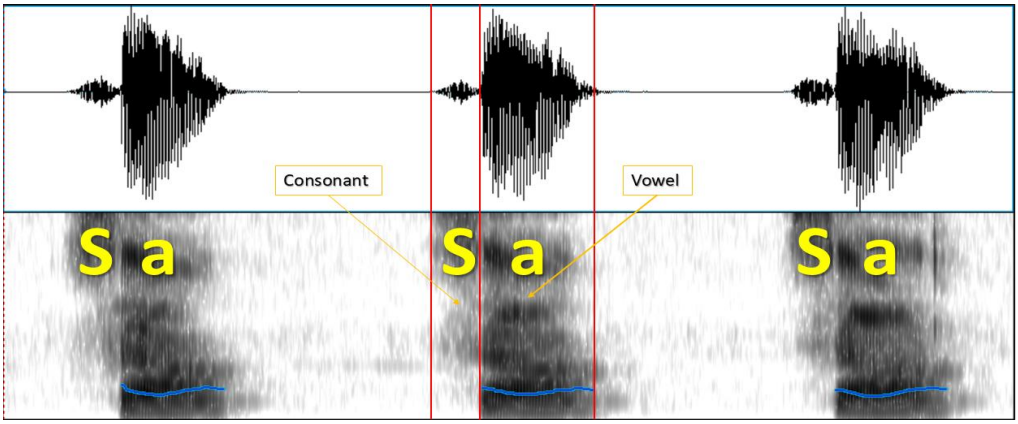

Figure 2. Using the Praat software, recording the sibilant consonant /s/ followed by the vowel /a/

The speech signal's spectrogram was computed using MATLAB software as follows: The signal is initially divided into 11.6 ms segments, with adjacent segments overlapping by 9.6 ms. To obtain appropriate frequency resolution, these segments underwent Hamming windowing, which was preceded and followed by zero padding. After that, the fast Fourier transform (FFT) is computed. To get the normalized energy *EB(n)* of vowels and fricative consonants, use (1).

$$E_B(n) = \sum_i 10.\log(|X(n,i|^2) \quad (1)$$

X(n,i) is the amplitude of the spectrum smoothed by a moving average of 20 points along the time index (*n*). The frequency band is represented by *B*. The frequency index, (*i*) is calculated using the DFT indices that reflect the bottom and upper boundaries of each band. Then, using (2), we computed energy percentage E_{Bn} (*n*) of band *B* for each window *n*.

$$E_{Bn}(n) = \frac{E_B(n)}{E_T(n)} \quad (2)$$

$E_T(n)$ represents global energy in segment *n*.

In order to identify a landmark, it is necessary to measure the rate of change of a number of characteristics that were derived from the speech signal over a brief period of time. The ensuing equation was applied to determine the rate of change (ROC) of energy in band *b*.

$$R_{EB}(n) = E_B(n) - E_B(n - J) \quad (3)$$

The time step is represented by the letter *J*. The difference in energy value between the current window *n* and the one preceding it by *J* windows is shown by this measurement.

2.2. Segmentation

Vowels and consonants are produced when the vocal tract suddenly constricts. This articulatory action is mirrored in the speech signal's spectrum by a sudden change at the moment in time when the sound is closed or released [25], [26]. These time points serve as markers for determining the beginning and the end of a consonant or vowel. We employed two sorts of landmarks point in our approach. The first is the acoustic cue (g), which indicates when the vocal cords begin to vibrate (g+) and when they stop vibrating (g–). These times correspond to the crossing points of the ROC curve of the first band B1 above and below the threshold values of 9 dB (g+) and -9 dB (g-) respectively. The acoustic cue (b) Burst is the second kind, with (b+) indicating the start of the frication noise for fricative consonants or the commencement of the explosion for plosive consonants, and (b-) indicating the conclusion of the frication or suction noise. Between the points (g-) and (g+), the landmark point (b) is positioned at the most important peak of the ROC curve of the bands B2 to B5. The following intervals correlate to a consonant or vowel's location: A vocal consonant or vowel is expressed by (+g, -g). (+b, +g, -g): A syllable that starts with a frication, with (+b) indicating that the frication is present. (+b, -b, +g, -g): initial plosive syllables, (+b, -b) denoting the start and end of the liberation.

2.3. Support vector machine and artificial neural network methods

An artificial neural network (ANN) is a mathematical model that imitates the functions of the human brain. Today, the multilayer perceptron (MLP) is a form of neural network that is widely used in classification. The input and output nodes are separated by one or more layers in this feed-forward network. With one and two hidden layers, each with a different number of neurons, we put the MLP network to the test (s). The output layer is made up of two neurons, whereas the input layer is made up of four neurons [27]. When determining the number of neurons per hidden layer, there are several guidelines to follow. The size of hidden layer must be either the same as the size of the input layer [28] or 75 percent of its [29].

Support vector machine (SVM) is a classification technique based on supervised machine learning. The objective is to find a decision function that uses the optimal hyperplane margin separation as a starting point. Support vectors are the data points closest to the hyperplane. SVM transforms the representation space of the input data into a higher-dimensional space where a linear separation is more likely when the data to be processed is not linearly separable [30]. This is accomplished through the usage of a kernel function. The polynomial kernel is the most often used kernel in SVMs.

3. RESULTS AND DISCUSSION

Our classification algorithm works in three steps. It all starts with recognizing the vowel that follows the consonant. Then and for the same vowel, our algorithm divides the consonants into two categories, alveolar and post-alveolar. Finally, it separates the consonants that belong to each of the two categories.

3.1. Vowel classifications

We will detail the operation and show the results of the classification method for the three Arabic vowels in this section. We first divided each time domain vowel in three equal segments which are: onset, middle and offset. For each vowel frequency band as shown in Table 1, we then calculated the normalized energy in the middle of each vowel as shown in Figure 3.

We see that the vowel /a/ is characterized by a high energy in the BV1 band (more than 50%), while the BV2 band has only 40%. The energy in the B3 band is about 10%. In the case of the vowel /u/, band BV1 carries the most energy, roughly 80%. The BV2 band has a 20% energy level. The band BV3 has the lowest energy value. The vowel /i/ varies from the other vowels in that it has essentially little energy in the BV2 band and a lot in the BV3 band (around 30 percent). These findings are consistent with those of Abajaddi *et al.* [31].

Based on these remarks, we developed the algorithm depicted in Figure 4 to classify the three Arabic vowels. First, we look at the energy in the middle of the vowel in the second band BV2 (MV2). If it is greater than 5%, one proceeds to the analysis of the energy in the middle of the vowel in band BV1, if not one proceeds to the examination of the distribution of energy in the middle of the vowel in band BV3. The vowel sought is /a/ if the energy in band BV1 is less than 55%; otherwise, the searched vowel is /u/. The vowel is /i/ if the energy of the BV3 band is more incredible than 5%, otherwise, the vowel is /u/. This algorithm's evaluation showed a very high classification rate (over 98 %). Table 2 summarizes the results that our algorithm achieved.

Table 1. Vowel frequency bands

BV1	BV2	BV3	BV4
100 to 600 Hz	600 to1800 Hz	1,800 to 4,600Hz	4,600 to 8,000Hz

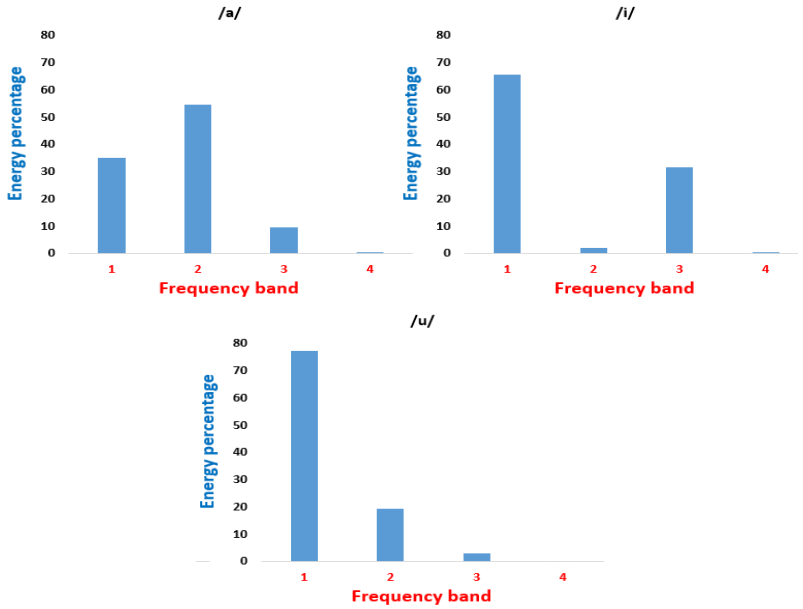

Figure 3. The energy distribution in the middle of the three Arabic vowels in the four frequency bands

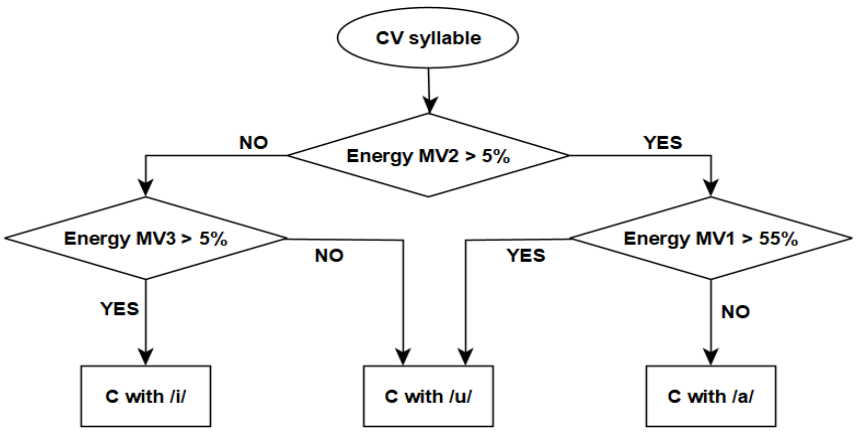

Figure 4. The algorithm for classifying the three Arabic vowels /a/, /i/, and /u/

Table 2. Accuracy of vowel classification

Vowel	/a/	/i/	/u/
Classification rate	98.5%	98.9%	97.4%

3.2. Classification of alveolar/post-alveolar consonants followed by vowels /a/, /i/ and /u/

3.2.1. Alveolar/post-alveolar consonants followed by vowels /a/ or /i/

To characterize the consonants, we took the following steps: we divided the time domain of each consonant into three equal segments. Then, the normalized energy in the middle of each consonant for each consonant frequency band in Table 3 was calculated as shown in Figure 5. By analyzing the energy distribution graphs of alveolar and post-alveolar consonants as shown in Figure 5, we discovered that the energy follows the same evolution in the band B1, B2 and B3. On the other hand, the energy distribution is different in bands B4 and B5. In the B4 band, alveolar consonants have an energy proportion of less than 30%, whereas in the B5 band, it is larger than 50%. The energy distribution for post-alveolar consonants is flipped, with more than 60% of the energy in the B4 band and less than 10% in the B5.

Table 3. Consonant frequency bands

B1	B2	B3	B4	B5
100 to 400 Hz	400 to 1,600 Hz	1,600 to 3,000 Hz	3,000 to 5,000 Hz	5,000 to 8,000 Hz

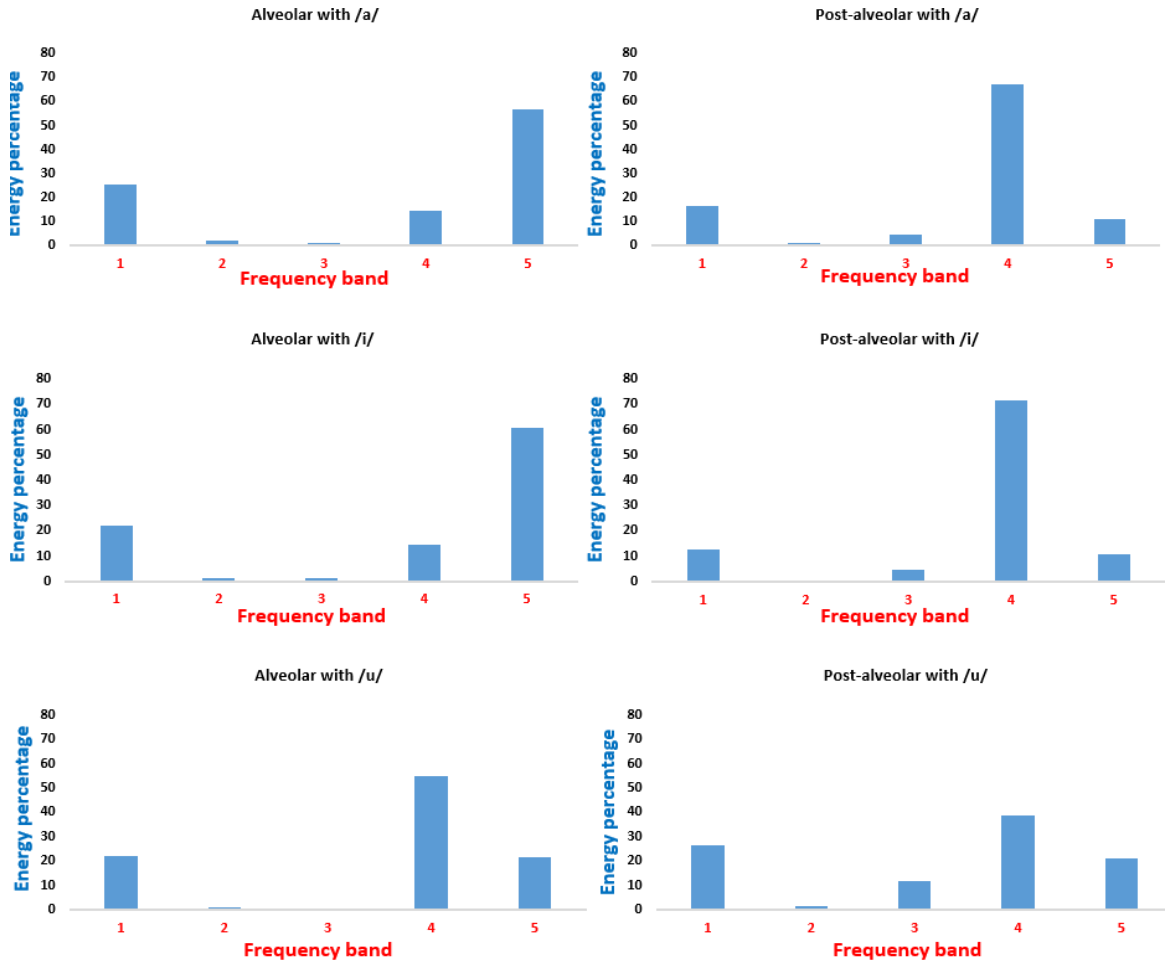

Figure 5. The energy distribution of alveolar and post-alveolar consonants followed by the vowel /a/, /i/ and /u/

Based on the findings, we devised the algorithm shown in Figure 6, which classifies sibilant consonants into alveolar and post-alveolar consonants when they are followed by one of the two vowels /a/ or /i/. The following is how the algorithm works: The consonant is classed as alveolar if the energy percentage

in the middle of the consonant in the B4 band: (3,000 to 5,000 Hz) is less than 35 percent. The consonant is post-alveolar otherwise. This method has a perfect classification rate of 100%.

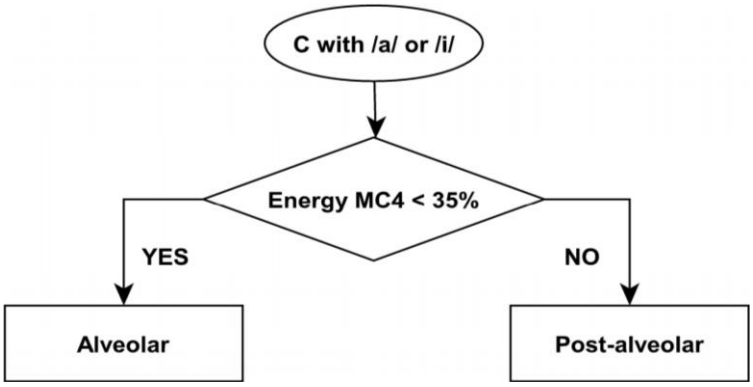

Figure 6. Alveolar and post-alveolar consonants classification algorithm followed by the vowel /a/ or /i/

3.2.2. Alveolar/post-alveolar consonants followed by the vowel /u/

The energy distribution of alveolar and post-alveolar consonants followed by the vowel /u/ is shown in Figure 5. The difference in energy distribution between consonants followed by the vowels /a/ and /I/ and those followed by the vowel /u/ is the first observation. The vowel /u/, on the other hand, shifted energy from higher to lower frequencies. The second discovery is that alveolar consonants have essentially little energy in the B3 band, but post-alveolar consonants have energy in this band. The energy in the other bands evolves in the same way as the two consonant classes.

Based on this analysis, we developed the classification algorithm for sibilant consonants accompanied by the vowel /u/, as shown in Figure 7. This method works as follows: in the band B3 (1600 to 3000 Hz), the consonant is classed as alveolar if the energy percentage in the middle of the consonant is less than 5%. Otherwise, the consonant is considered post-alveolar. This algorithm has an accuracy of 90.5%.

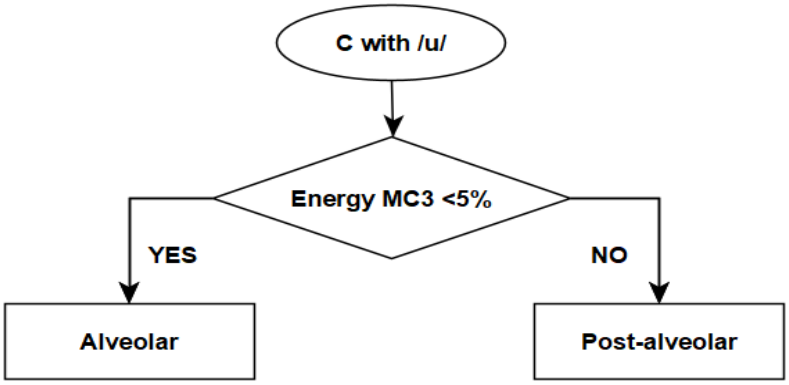

Figure 7. Classification algorithm for alveolar and post-alveolar consonants followed by the vowel /u/

3.3. Classification of post-alveolar consonants /ʒ/ and /ʃ/ accompanied by the three vowels

The distribution of the energy percentage in the middle of the two post-alveolar consonants (/ʒ/ and /ʃ/) in the five frequency bands is shown in Figure 8. It can be observed that the energy in the bands B2, B3, B4, and B5 follows the same distribution. The energy in the B1 band, on the other hand, allows for differentiation between the two post-alveolar cells. The consonant /ʒ/ is distinguished by the existence of energy in the first band, whereas the consonant /ʃ/ has no energy in this band. On the basis of this finding, we proposed the algorithm shown in Figure 9, which permits the categorization of post-alveolar consonants (/ʒ/ and /ʃ/) as follows: if the energy in band B1 is zero, the consonant is classified as /ʃ/; otherwise, the consonant is classified as /ʒ/. This method has an accurate classification rate of 96.76 percent.

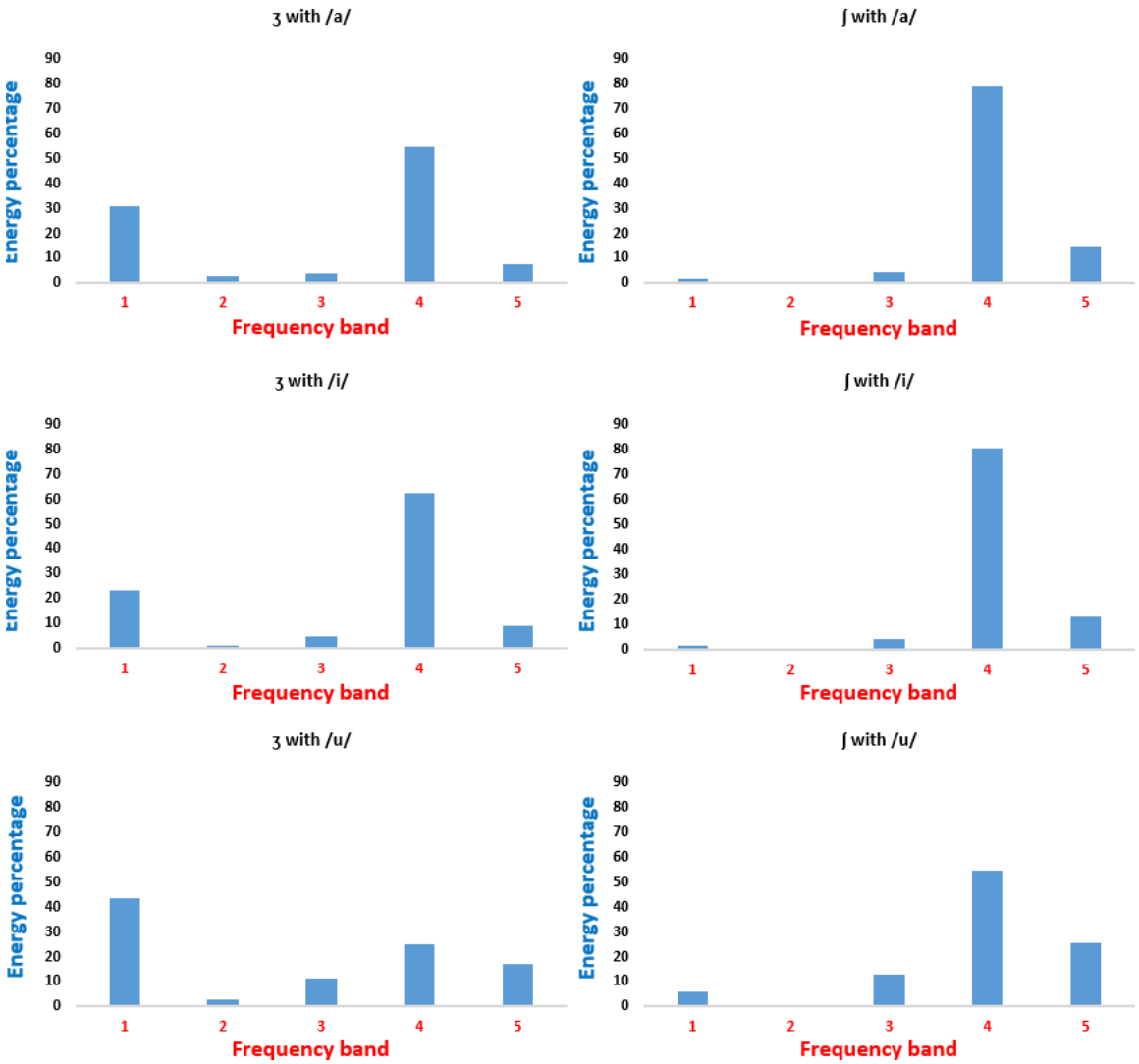

Figure 8. The energy distribution of post-alveolar consonants followed by the three vowels

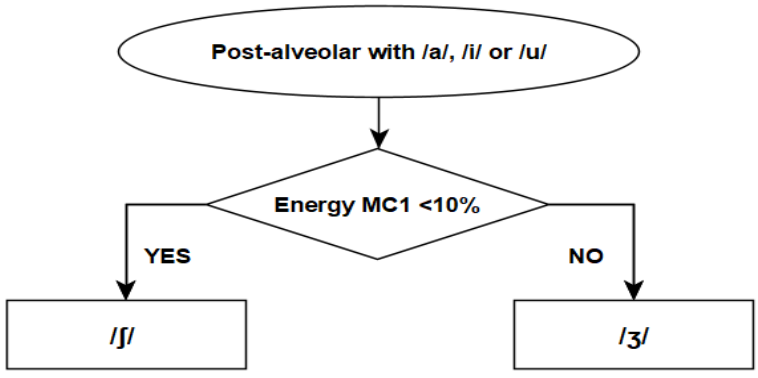

Figure 9. Classification algorithm for post-alveolar consonants followed by the three vowels.

3.4. Classification of alveolar consonants /z/, /s/ and /sˤ/ accompanied by the three vowels

The categorization of the three alveolar consonants /z/, /s/, and /sˤ/ according to the energy in the middle of the consonant is a bit tricky, as seen in Figure 10. In particular, the two consonants /s/ and /sˤ/ have the identical energy distribution in all bands, independent of the vowel that follows the consonant. However,

owing to the band B1, we can distinguish the sound /z/ from the two consonants /s/ and /sˤ/. The consonant /z/ has an energy percentage of more than 45 percent in band B1, but the two consonants /s/ and /sˤ/ have an energy percentage of zero in this band.

We may extract the algorithm of Figure 11 from this result, which allows us to classify the alveolar consonants /z/, /s/, and /sˤ/. This algorithm operates as follows: if the energy in band B1 is more than 30%, the consonant is designated as /z/; otherwise, the consonant is designated as /sˤ/ or /s/. The accuracy of this algorithm is 94.75 percent. The results of the three methods are summarized in the Table 4. We can see that our approach consistently outperforms or equals the results provided by other algorithms (ANN and SVM).

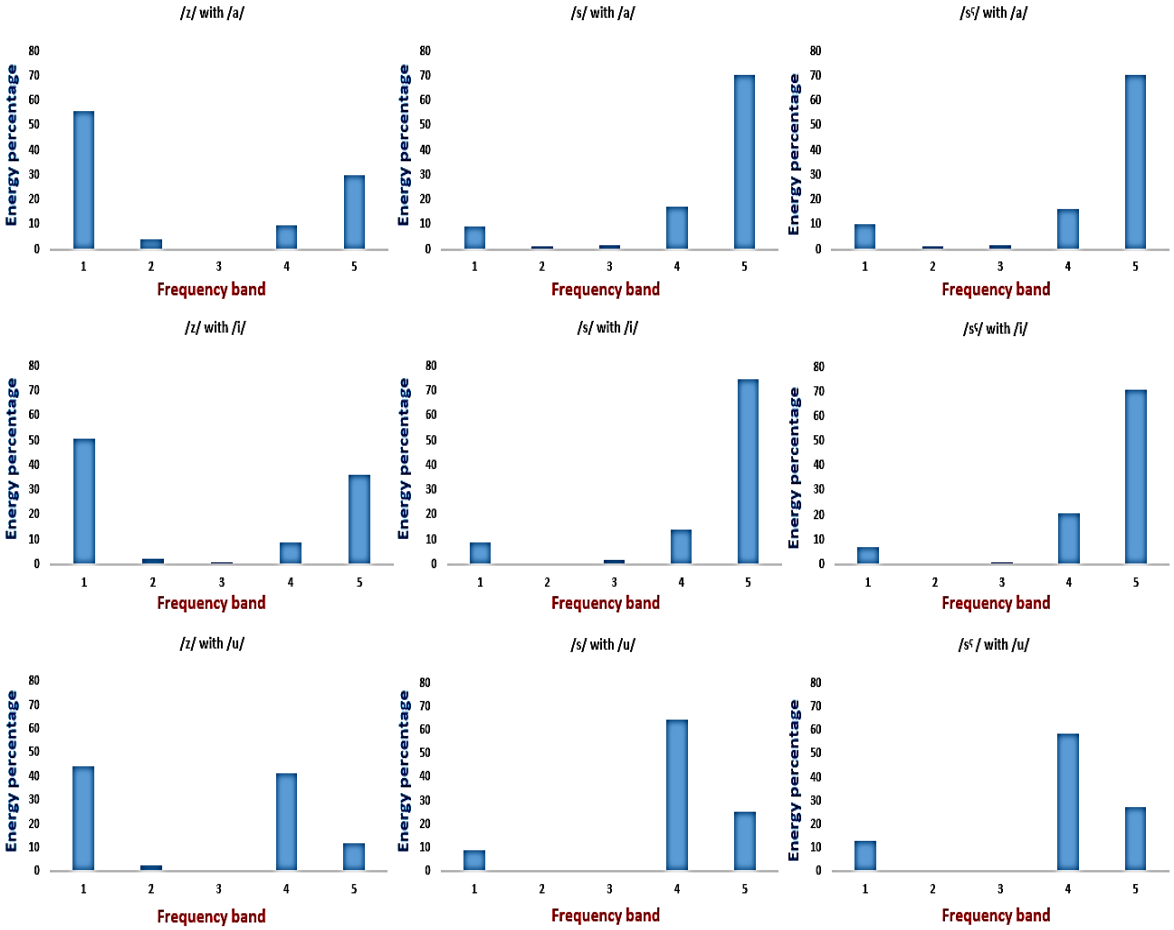

Figure 10. The energy distribution of alveolar consonants followed by the vowel /a/

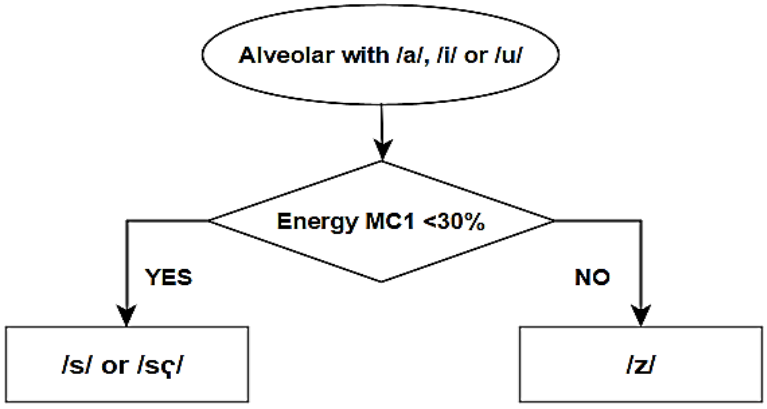

Figure 11. Alveolar consonant classification algorithm followed by the vowel /a/, /i/ or /u/

Table 4. A comparison of our algorithm's performance with that of the ANN and SVM algorithms

	Our algorithm	ANN algorithm	SVM algorithm
Vowel classifications	93 %	83.64 %	83.33 %
Classification of alveolar/post-alveolar consonants followed by vowels /a/ and /i/	100 %	100 %	99.63 %
Classification of alveolar/post-alveolar consonants followed by the vowel /u/	90.50 %	90.22 %	86.11 %
Classification of post-alveolar consonants /ʒ/ and /ʃ/ accompanied by the three vowels	96.76 %	94.77 %	93.80 %
Classification of alveolar consonants /z/, /s/ and /sˤ/ accompanied by the three vowels	94.75 %	93.35 %	93.15 %

3.5. Discussions

The goal of this study is to develop an algorithm that can identify the Arabic sibilant consonants (/s/, /sˤ/, /z/, /ʒ/ and /ʃ/) followed by the three Arabic vowels (/a/, /i/, and /u/) using the normalized energy distribution of a speech signal in the previously described frequency bands. Our algorithm begins by identifying the vowel that follows the consonant. The energy in the middle of the vowel as determined by an acoustic analysis revealed that the three vowels contain a large amount of energy in the B1 band. The fact that vowels are voiced sounds, that is, sounds generated by the vibrating of the vocal cords, justifies this behavior. The energy for the vowel /i/ is focused in two bands: the first (100 to 600 Hz) and the third band (1,800 to 4,600 Hz), with the first band containing the majority of the energy. The value of the first formant (F1<300 Hz) which is found in the low frequencies, in addition to the energy owing to the vibration of the vocal cords created by spoken sounds, explains the high energy concentration in the B1 band. The value of the second formant (F2>2,000Hz), which is found in the high frequencies, is responsible for the quantity of energy present in the third band. The most anterior and closed vowel in terms of articulation is /i/. As a result, it has the smallest front cavity and, as a result, a very big rear cavity, resulting in a very high F2 and a very low F1. Due to the distribution of the formants F1 and F2 (F1> 600 Hz and F2>1000 Hz), the majority of the energy for the vowel /a/ is placed in the B1 and B2 bands. In articulatory phonetics, /a/ being the least anterior and most open front vowel, has a rear cavity affiliated with a very high F1, exhibiting intermediate lowering and advancement of the tongue. Because the first two formants F1 and F2 are concentrated in the low frequencies (F1>100 Hz and F2 1,000 Hz), the energy of the vowel /u/, which is the most closed, posterior, and rounded, is concentrated in the first band B1 [32], [33].

Once the vowel has been identified, the second phase of our algorithm consists of recognizing the consonant. Figure 5 depicts the energy distribution of the two types of sibilant consonants: alveolar and post-alveolar consonants. We discovered that when post-alveolar consonants are followed by the vowels /a/ and /i/, the majority of their energy is concentrated in the fourth band (3,000 to 5,000 Hz), whereas alveolar consonants have a substantial energy share in the fifth band (5,000 to 8,000 Hz). The point of constriction of the vocal tract justifies this energy distribution from an articulatory standpoint. Between the tip of the tongue and the alveoli, the alveolar sibilants (/s/, /sˤ/, and /z/) are articulated. The alveolar consonants offer a maximum of energy in the high frequencies due to the pressure of expelled air at the level of this constriction. Between the lamina and the rear of the alveoli, the post-alveolar sibilants (/ʒ/ and /ʃ/) are articulated. The energy has been decreased towards the band (3,000-5,000 Hz) at this point of articulation. We discovered that when consonants are followed by the vowel /u/, the consonant's energy migrates to the lower frequency ranges (1,600 to 3,000 Hz). This is due to the influence of the vowel /u/ coarticulation on the consonant it precedes. The sound consonants (/ʒ/ and /z/) differ from the deaf consonants (/s/, /sˤ and /ʃ/) in that they have more overall energy in the low frequencies. As previously stated, the sound consonants are created by a vibration of the vocal cords, which explains this behavior [34], [35].

4. CONCLUSION

The Arabic sibilant fricative consonants (/s, sˤ, z, ʒ and ʃ/), followed by the three vowels (/a/, /i/, and /u/), were classified in this work. The suggested method's key characteristic is the use of normalized energy as an acoustic index in frequency ranges. Our findings indicate that the energy contained in the speech signal is a critical element in sound characterization. The rate of proper categorization surpasses 90%. The characterization of non-sibilant consonants will be the focus of future research.

REFERENCES

[1] J. Cantineau, "Arabic phonetics lessons," (in French), Paris: Klincksieck, 1960.
[2] A. Juneja and C. Espy-Wilson, "Segmentation of continuous speech using acoustic-phonetic parameters and statistical learning,"

in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02.*, 2002, vol. 2, pp. 726–730, doi: 10.1109/ICONIP.2002.1198153.
[3] A. A. al Nassir, "Sibawayh the phonologist : A critical study of the phonetic and phonological theory of Sibawayh as presented in his treatise ?Al Kitab?," University of York, 1993.
[4] K. N. Stevens, "Airflow and turbulence noise for fricative and stop consonants: Static considerations," *The Journal of the Acoustical Society of America*, vol. 50, no. 4B, pp. 1180–1192, Oct. 1971, doi: 10.1121/1.1912751.
[5] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Wiley, 1996.
[6] S. J. Behrens and S. E. Blumstein, "Acoustic characteristics of English voiceless fricatives: a descriptive analysis," *Journal of Phonetics*, vol. 16, no. 3, pp. 295–298, Jul. 1988, doi: 10.1016/S0095-4470(19)30504-2.
[7] L. J. Raphael, G. J. Borden, and K. S. Harris, *Speech science primer: Physiology, acoustics, and perception of speech*. Williams & Wilkins, 1984.
[8] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1252–1263, 2000, doi: 10.1121/1.1288413.
[9] A. M. Abdelatty Ali, J. Van Der Speigel, and P. Mueller, "Auditory-based speech processing based on the average localized synchrony detection," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, 2000, vol. 3, pp. 1623–1626, doi: 10.1109/ICASSP.2000.862016.
[10] A. M. Abdelatty Ali, J. Van der Spiegel, and P. Mueller, "Acoustic-phonetic features for the automatic classification of fricatives," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2217–2235, May 2001, doi: 10.1121/1.1357814.
[11] J. Goodacre and Y. Nakajima, "The perception of fricative peaks and noise bands," *Journal of Physiological Anthropology and Applied Human Science*, vol. 24, no. 1, pp. 151–154, 2005, doi: 10.2114/jpa.24.151.
[12] M. Toda, "Speaker normalization of fricative noise: Considerations on language-specific contrast," in *Proceedings of the16th International Congress on Phonetic Sciences*, 2007, pp. 825–828.
[13] M. Toda and K. Honda, "An MRI-based cross-linguistic study of sibilant fricatives," in *Proceedings of the 6th International Seminar on Speech Production*, 2003, pp. 1–6.
[14] J. S. Perkell *et al.*, "The distinctness of speakers' /s/—/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 6, pp. 1259–1269, Dec. 2004, doi: 10.1044/1092-4388(2004/095).
[15] S. McLeod, A. Roberts, and J. Sita, "Tongue/palate contact for the production of /s/ and /z/," *Clinical Linguistics & Phonetics*, vol. 20, no. 1, pp. 51–66, Jan. 2006, doi: 10.1080/02699200400021331.
[16] K. Maniwa, A. Jongman, and T. Wade, "Acoustic characteristics of clearly spoken English fricatives," *The Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 3962–3973, Jun. 2009, doi: 10.1121/1.2990715.
[17] K. L. Haley, E. Seelinger, K. C. Mandulak, and D. J. Zajac, "Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach," *Journal of Phonetics*, vol. 38, no. 4, pp. 548–554, Oct. 2010, doi: 10.1016/j.wocn.2010.07.006.
[18] Y.-Y. Kong, A. Mullangi, and K. Kokkinakis, "Classification of fricative consonants for speech enhancement in hearing devices," *PLoS ONE*, vol. 9, no. 4, Apr. 2014, doi: 10.1371/journal.pone.0095001.
[19] A. Kochetov, "Acoustics of Russian voiceless sibilant fricatives," *Journal of the International Phonetic Association*, vol. 47, no. 3, pp. 321–348, Dec. 2017, doi: 10.1017/S0025100317000019.
[20] D. S. Cooper, C. Scholl, L. Petrosino, R. C. Scherer, and L. H. Small, "The acoustics of fricative consonants in gulf spoken Arabic," ProQuest Dissertations Publishing, Bowling Green State University, Ohio, 2005.
[21] M. A. Al-Khair, "Acoustic characteristics of Arabic fricatives," University of Florida, 2005.
[22] A. Benamrane, "Acoustic study of standard Arabic fricatives (Algerian speakers)," (in French), Université de Strasbourg, Strasbourg, 2013.
[23] P. Ghaffarvand Mokari and N. Mahdinezhad Sardhaei, "Predictive power of cepstral coefficients and spectral moments in the classification of Azerbaijani fricatives," *The Journal of the Acoustical Society of America*, vol. 147, no. 3, pp. EL228–EL234, Mar. 2020, doi: 10.1121/10.0000830.
[24] Y. Elfahm, N. Abajaddi, B. Mounir, L. Elmaazouzi, I. Mounir, and A. Farchi, "Classification of Arabic fricative consonants according to their places of articulation," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 1, pp. 936–945, Feb. 2022, doi: 10.11591/ijece.v12i1.pp936-945.
[25] S. A. Liu, "Landmark detection for distinctive feature-based speech recognition," *The Journal of the Acoustical Society of America*, vol. 96, no. 5, pp. 3227–3227, Nov. 1994, doi: 10.1121/1.411152.
[26] S. Boyce, H. Fell, and J. MacAuslan, "SpeechMark: Landmark detection tool for speech analysis," *13th Annual Conference of the International Speech Communication Association*, pp. 1894–1897, 2012.
[27] R. P. Lippmann, "Review of neural networks for speech recognition," *Neural Computation*, vol. 1, no. 1, pp. 1–38, Mar. 1989, doi: 10.1162/neco.1989.1.1.1.
[28] J. Bloemer, J. Lemmink, and H. Kasper, "Neural nets versus marketing models in time series analysis: A simulation study," in *Proceedings of the 23rd Annual Conference of the European Marketing Academy*, 1994, pp. 1139–1153.
[29] V. Venugopal and W. Baets, "Neural networks and statistical techniques in marketing research," *Marketing Intelligence & Planning*, vol. 12, no. 7, pp. 30–38, Aug. 1994, doi: 10.1108/02634509410065555.
[30] C. J. C. Burges, "Tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.
[31] N. Abajaddi, Y. Elfahm, B. Mounir, L. Elmaazouzi, I. Mounir, and A. Farchi, "Efficiency of the energy contained in modulators in the Arabic vowels recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3601–3608, Aug. 2021, doi: 10.11591/ijece.v11i4.pp3601-3608.
[32] Y. A. Alotaibi and A. Hussain, "Speech recognition system and formant based analysis of spoken Arabic vowels," in *Future Generation Information Technology*, 2009, pp. 50–60.
[33] Y. Korkmaz and A. Boyaci, "Classification of Turkish vowels based on formant frequencies," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, Sep. 2018, pp. 1–4, doi: 10.1109/IDAP.2018.8620877.
[34] M. Toda, "Articulatory and acoustic study of sibilant fricatives," (in French), Ph.D. thesis, Université Paris III, 2009.
[35] Y. Meynadier, "Elements of acoustic phonetics," (in French), in *Méthodes et outils pour l'analyse phonétique des grands corpus oraux, Hermes Science Publications*, 2010, pp. 25–83.

BIOGRAPHIES OF AUTHORS

Youssef Elfahm received a master's degree specializing in automatic control, signal processing and industrial computing from the University of Hassan First in Settat, Morocco, in 2017. Currently, he is a professor in the Department of Electrical Engineering at Alkhaouarizmi Technical High School. His research interests include speech recognition systems, speech production, and artificial intelligence. He can be contacted at y.elfahm@uhp.ac.ma.

Nesrine Abajaddi was born in Casablanca, Morocco, in 1994. received a master's degree specializing in automatic control, signal processing and industrial computing from the University of Hassan First in Settat, Morocco, in 2017. She is currently a Ph.D. student in Engineering, mechanical, Industrial Management, and Innovation Laboratory research Laboratory, Faculty of Sciences and Technics, Hassan First University. She can be contacted at n.abajaddi@uhp.ac.ma.

Badia Mounir was born in Casablanca, Morocco, in 1968. He received an engineer degree in 1992 in automatic and industrial computing, The Mohammadia School of Engineering, Rabat, Morocco. She is an assistant professor at Graduate School of Technology, University Cadi Ayyad since 1992. Habilitaded to supervise research (HDR) since 2007 and professor of higher education (PES) since 2017. Member of Laboratory of Process, Signals, Industrial Systems, Informatic (LAPSSII). Her research interests include speech recognition, signal processing, energy optimization and modeling. She can be contacted at Mounirbadia2014@gmail.com.

Laila Elmazouzi Ing Ph. D. in Telecommunication and Networks. Habilitaded to supervise research (HDR) at High School of Technology- Cadi Ayyad University. Member of the LAPSSII Laboratory (Laboratory of Process, Signals, Industrial Systems, Informatic). Her research interests include telecommunication, signal processing, emotion recognition, machine learning. She can be contacted at Elmazouzi2001@yahoo.fr.

Ilham Mounir is a Ph.D. in applied mathematics. She habilitaded to supervise research (HDR) at High School of Technology, Cadi Ayyad University. Member of LAPSSII (Laboratory of Process, Signals, Industrial Systems, Informatic). Her research interests include applied mathematics, signal processing, emotion recognition, speech recognition, and energy: optimization and modeling. She can be contacted at ilhamounir@gmail.com.

Abdelmajid Farchi Ing received a Ph.D. in electric engineering and telecommunications and is now a chief of research team "Signals and Systems" in Laboratory of Engineering, Industrial Management and Innovation. He is an educational person responsible for the cycle engineer electrical systems and embedded systems of the faculty of the sciences and technology of Settat, Morocco. He can be contacted at abdelmajid.farchi1@gmail.com.