

Classification of Arabic fricative consonants according to their places of articulation

Youssef Elfahm¹, Nesrine Abajaddi¹, Badia Mounir², Laila Elmaazouzi²,
Ilham Mounir², Abdelmajid Farchi¹

¹IMII Laboratory, Hassan First University of Settat, Settat, Morocco

²LAPSSII Laboratory, Superior School of Technology, Safi, Morocco

Article Info

Article history:

Received Jan 4, 2021

Revised Aug 13, 2021

Accepted Aug 29, 2021

Keywords:

Classification

Energy in the bands

Nonsibilant fricatives

Place of articulation

Sibilant fricatives

ABSTRACT

Many technology systems have used voice recognition applications to transcribe a speaker's speech into text that can be used by these systems. One of the most complex tasks in speech identification is to know, which acoustic cues will be used to classify sounds. This study presents an approach for characterizing Arabic fricative consonants in two groups (sibilant and non-sibilant). From an acoustic point of view, our approach is based on the analysis of the energy distribution, in frequency bands, in a syllable of the consonant-vowel type. From a practical point of view, our technique has been implemented, in the MATLAB software, and tested on a corpus built in our laboratory. The results obtained show that the percentage energy distribution in a speech signal is a very powerful parameter in the classification of Arabic fricatives. We obtained an accuracy of 92% for non-sibilant consonants /f, χ, γ, ζ, h, and h/, 84% for sibilants /s, sç, z, 3 and /, and 89% for the whole classification rate. In comparison to other algorithms based on neural networks and support vector machines (SVM), our classification system was able to provide a higher classification rate.

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



Corresponding Author:

Youssef Elfahm

IMII Laboratory, Hassan First University

FST of Settat, Km 3, B.P: 577 Road of Casablanca, Settat, Morocco

Email: y.elfahm@gmail.com

1. INTRODUCTION

The phonetic system of the Arabic language contains 34 phonemes composed of 6 vowels (three short vowels: /a/ /i/ /u/ and three long vowels /a:/ /i:/ /u:/) and 28 consonants (/ʔ/, /b/, /t/, /θ/, /ʒ/, /ħ/, /χ/, /d/, /ð/, /r/, /z/, /s/, /ʃ/, /sʃ/, /dʃ/, /ðʃ/, /tʃ/, /ʒ/, /q/, /f/, /q/, /k/, /l/, /m/, /n/, /h/, /w/ and /j/) [1]. The vowels are produced by allowing air to circulate freely in the vocal tracts, while the consonants are made by compressing air from the lungs. They are defined by the point of articulation (where is the sounds made?) and the mode of production (how is the sounds made?). Regarding the manner of production, consonants can be classified into three groups: Plosives, Fricatives and Sonorants [2]. The plosives are formed when the air within the mouth is compressed by closures and then suddenly bursts with an explosion after those closures are opened. The Arabic plosives include /b/, /t/, /d/, /tʃ/, /q/, /k/, /dʃ/ and /ʔ/ [3]. The sonorant consonants are described as those produced with a relatively free airflow and a position of the vocal cords in such a way that a spontaneous voice is possible. Nasals, laterals, and liquids are among them. /m/, /n/, /l/, /r/, /j/ and /w/ [3]. The fricative consonants are formed by a narrow constriction in the vocal cavity that causes the airflow to be continuously turbulent [4]. The Arabic fricatives are /θ/, /ʒ/, /ħ/, /χ/, /ð/, /z/, /s/, /ʃ/, /sʃ/, /dʃ/, /ðʃ/, /tʃ/, /ʒ/, /f/ and /h/ [3]. It constitutes half of the Arabic consonants (14 among 28 consonants) and are distributed over five places of articulation (either at the level of the larynx, pharynx, tongue, lips, or nose) as shown in Table 1.

Table 1. Five places of articulation of fricative consonants

Place of articulation	Labiodental	Interdental	Alveolar	Velar	Glottal
Consonants	/f/	/θ/ /ð/ /ðˤ/	/z/ /s/ /ʃ/ /ʒ/ /sˤ/	/ɣ/ /χ/ /ʕ/ /ħ/	/h/

Fricative consonants can be classified into voiced consonants, i.e., consonants produced with the vibration of the vocal cords, and unvoiced consonants (without vibration of the vocal cords). Unvoiced Arabic fricatives are /θ/, /ħ/, /s/, /ʃ/, /sˤ/, /χ/, /f/ and /h/, while voiced fricatives are: /z/, /ʒ/, /ʕ/, /ð/ and /ðˤ/ [3]. The majority of studies have classified fricatives into two groups: sibilant and non-sibilant. In phonetics, a sibilant fricative consonant is a consonant whose mode of articulation consists in directing an airflow with the tongue towards the edge of the teeth held closed, producing a hissing sound [5]. The sibilant Arabic fricatives are /s, sˤ, z, ʒ and ʃ/ while non-sibilant fricatives are /f, θ, ð, ðˤ, χ, ɣ, ʕ, ħ, and h/.

Many studies have been carried out on the speech signal characterization which is a critical task in speech recognition. Indeed, when the acoustic signal is converted into linguistic information, the most difficulty that arises is to know the speech perceptual primitives: on which acoustic cues contained in the speech signal does the auditor rely on to recognize phonemes, syllables, or words? Researchers have conducted several works in order to answer this question. Concerning English fricative consonants characterization, Stevens [4] worked on unvoiced fricatives. He claims that non-sibilants /f, h/ have a lower amplitude than sibilants /S, s/. The increased amplitude of sibilants, according to Shadle [6], is due to increased turbulence in the airflow caused by the lower teeth serving as an obstruction to the source of noise from the constricting whistle. Tomiak [7] used spectral moments to classify a set of English fricative consonants composed of /f, θ, s, ʃ and h/, he obtained a classification rate of 74%. He indicated that /h/ has a standard deviation, asymmetry, and kurtosis greater than /f/. Nittrouer [8] and Tjaden [9] have also used spectral moments to differentiate between /S/ and /s/. They concluded that mean, variance, skewness and kurtosis of the friction frequency spectrum help distinguish between /S/ and /s/. Jongman *et al.* [10] found that it is possible to distinguish the articulation place of fricatives by the four spectral moments and spectral changes of the second forming in the following vowel.

The duration and the amplitude of the friction are linked to the place of articulation and make it possible to distinguish mainly sibilants from non-sibilants. Ladefoged studied the acoustics of English sibilants /s/ and /ʃ/. His work has shown that unvoiced sibilants /s/ and /ʃ/ have a relatively high acoustic intensity compared to non-sibilant fricatives (labial or interdental) [11]. Nissen and Fox achieved a 65% classification rate for adult productions of /f, θ, s, ʃ/ using spectral moments, duration, normalized amplitude, and spectral slope. They also found that the spectral slope, as well as the variance, made it possible to discriminate between sibilants /s/ and /ʃ/, on the one hand, and between sibilants versus non-sibilants (/f/ and /θ/) [12]. Jesus and Shandle [13] studied the fricative consonants /s, ʃ and χ/. He observed that the amplitude peaks are high at the back of the articulation. He mentioned in his study that the consonant /s/, which is located in the front cavity, had a broad peak at 8 kHz, /ʃ/ in the center had a peak at 3.5 kHz while /χ/ with, a longer front cavity, had a series of peaks at 1.3 and 2.4 kHz.

In previous language studies (English, French and Japanese), the acoustic differences between alveolar and post-alveolar sibilants have been shown to be well represented by centre of gravity (CoG) values [14], [15]. The post-alveolar sibilants' CoGs are between 2 and 4 kHz, and, while the alveolar ones were significantly higher, typically between 4 and 8 kHz. According to Tabain [14], sibilant fricatives have very little variability in production, while non-sibilant dental have great variability. Meunier [16] carried out an acoustic phonetic study on French fricatives. She reported that voiced fricatives have a lower noise intensity due to the vocal cords vibration which decreases the supraglottic pressure. The frequency and intensity of friction noise depends on the place of articulation of the consonant. There are three places of articulation for the French language fricatives: the space between the lower lip and the upper incisors (labiodentals /f/ and /v/), the space between the tongue and the alveolis (alveolar /s/ and /z/) and the space between the tongue and the hard palate (palatals /ʃ/ and /ʒ/). Labiodental's noise is of low intensity (it is besides their principal characteristic), it presents diffuse peaks between 3.5 and 8 kHz. That of the alveolar, more intense, is between 4 and 8 kHz with peaks around 5 kHz and 8 kHz. Finally, for the palatals, the turbulence noise is between 2 and 7 kHz with a diffuse peak whose average is around 4 kHz [16].

In 2009, Driaunys *et al.* [17] worked on the hierarchical classification of Lithuanian phonemes. They have developed algorithms for recognizing plosive, sonorant, and fricative consonants. They mentioned in their study that the energy in the high frequencies of the Lithuanian fricatives /S, Š, Ž and Z/ is higher than the energy in the low frequencies compared to the sonorant consonants. Sung and Cho carried out an acoustic study to distinguish Korean sibilant (/s/ and /s*/) from English sibilant fricatives (/s/ and /ʃ/). They found that there is a significant distinction in the duration of friction between the fricatives of the two languages. However, the center of gravity and the average frequencies of the major spectral peak were not major acoustic cues to distinguish between the fricatives of the two languages [18]. In order to investigate the acoustic cues signaling the place of articulation, Alwan *et al.* [19] investigated speech perception for plosives and fricatives in

American English in the presence of noise. They found that the frequency of the formants and the relative spectral amplitude were the acoustic cues which allow a better perception of the place of articulation. Based on the location of articulation, Malde *et al.* [20] grouped English fricative consonants into five categories: glottal, post-alveolar, alveolar, dental, and labiodental. They used characteristics based on the modulation spectrogram. They managed to classify the consonants according to the place of articulation with an accuracy of 89.09%. They also obtained a rate of 87.51% for the phonemes classification [20]. Elina Nirgianak worked on Greek fricatives. The normalized amplitude distinguished fricative non-sibilants from sibilants, according to her findings. She also discovered that normalized duration and normalized amplitude are parameters that differentiate Greek voiced fricatives from voiceless fricatives, with a classification rate of 67.7% for voiced and 83.2% for voiceless fricatives. According to Nirgianaki [21], the parameters that characterized the five points of articulation were the spectral mean and the F2 onset. Spinu and Lilley [22] made a comparison between a method based on spectral moments and a new method based on cepstral coefficients, to classify five pairs /f-f^j, v-v^j, z-z^j, ʃ-ʃ^j and x/h-ç^j/ voiced and unvoiced of simple and palatalized Romanian fricatives. They obtained a classification rate of 95.3%. Prasad and Yegnanarayana [23] have proposed a classification method for English fricatives based on the use of parameters such as: dominant resonance frequencies, spectral moments, and center of gravity. They showed that voiced non-sibilants have a lower identification rate than unvoiced due to their vowel-like spectral characteristics.

Concerning Arabic fricatives characterization, Al-Khairy [24] noted that the duration of unvoiced fricatives' frication noise (average 134.21 ms) has a longer average than that voiced fricatives (average 92.05 ms). He discovered that non-sibilant fricatives have a shorter absolute overall frication noise duration (109.34 ms) than sibilants (average of 138.09 ms). Furthermore, Al-Khairy concluded that the amplitude measurements can differentiate sibilant fricatives /s, s^s, z and ʃ/ as a class of non-sibilants /f, θ, ð and ð^s/ without identifying the consonants of two classes. Using the spectral localization of the fricative peak, the results of Al-Khairy showed that it is possible to distinguish sibilant fricatives from non-sibilant ones. On the other hand, the peak spectral localization which allows the distinction between post-alveolar fricatives /ʃ/ and alveolar fricatives /s and z/ does not succeed in distinguishing the non-sibilants. Bendahmane [25] also worked on Arabic fricatives. She noted that they can be divided according to their places of articulation into three groups: the anterior /f, θ, ð and ð^s/, the sibilants /s, s^s, z, ʃ and ʃ/ and the posterior /χ, ʁ, ʕ, h, and h/. The anterior have a flat spectrum with a few energy peaks, are low in relative intensity, and have a relatively high frequency CoG, but it is the highest of all the fricatives for /ð and ð^s/. The sibilants have a compact spectrum, the main energy region has a higher frequency than that of the other groups, their intensity and their frequency CoG are high. The fricatives /s and s^s/ have a higher CoG than /z, ʃ and ʃ/. The posterior ones are described by a moderate to high intensity. Their CoG frequency is medium or low. The literature shows that research works carried out on fricative consonants were interested in their classification into sibilants/non-sibilants or voiced/non-voiced. The most used acoustic parameters in the characterization of these consonants are center of CoG, spectral moments, amplitude measurements, noise duration, F2 onset frequency and spectral peak location. In this work, we develop a classification system of Arabic fricative consonants according to their places of articulation based on the distribution of energy in four frequency bands as shown in Table 2 for a syllable of consonant-vowel (CV) type. The fricatives which are the subject of the study are illustrated in Table 3. Other algorithms such as neural networks and support vector machine (SVM) were used to compare their classification performance to that of our algorithm.

This article begins, first, with a presentation of the methods and tools used as well as the tests carried out. The second section presents the results obtained. The last section discusses the results obtained by the proposed algorithm and compares them with other algorithms like SVM and neural network. At the end of this work, we present the conclusions retained.

Table 2. The four frequency bands in Hz

Band 1	Band 2	Band 3	Band 4
0-400	800-2000	2000-5000	5000-8000

Table 3. The fricatives objects of the study

Non-sibilants	Sibilants
h, h̄, ʕ, χ, ʁ and f	ʃ, ʃ, s, s ^s and z

2. RESEARCH METHOD

2.1. Corpus construction

Nine Moroccan speakers, aged from 21 to 35 years, were invited to repeat four times a type of sequence CVCVCV where C is one of the Arabic fricative consonants mentioned in Table 3, V can be the vowel /i/, /u/ or /a/. In total, we have 1188 sequences (CVCVCV *11 consonants *3 vowels *4 times *9 men). The sequences were recorded in an isolated room with a quality microphone (Labtec AM232). In order to avoid turbulence due to direct airflow, the microphone was placed at an angle of about 45° and 20 cm from the

speaker's mouth's corner. Recordings are made using Praat software. Using the acoustic landmarks, we segmented each CVCVCV sequence into CV syllable which finally gave a corpus of 3564 CV. All recordings were sampled at a frequency of 22050 Hz.

2.2. Signal processing

To calculate the energy in the bands of each CV syllable, it is necessary to do a preprocessing of the signal. Each CV was windowed (11.6 ms with an overlap of 9.6 ms). Then, the fast Fourier transform (FFT) was calculated for 512 points. For each frame, a 20 point moving average taken along the time index i smooths the magnitude spectrum. Peaks in four distinct frequency bands were chosen from the smoothed spectrum $X(i, k)$ as (1):

$$E_b(i) = 10 \cdot \log(\max_k |X(i, k)|^2) \quad (1)$$

the band index b can range from 1 to 4. The frequency index k is obtained from discrete Fourier transform (DFT) indices that reflect each band's lower and upper limits.

2.3. Rate of change calculation

Landmark detection entails detecting regions with substantial variance in the rate of change (ROC) of a series of parameters derived from the speech signal over a short period of time. To obtain the rate of change of the parameters, a calculation of the rate of change depending on the first difference operation with a fixed time step is commonly used. The ROC calculation for the band energy parameter $E_b(i)$ is as (2):

$$rE_b(i) = E_b(i) - E_b(i - J) \quad (2)$$

the time step is denoted by the letter J . This metric depicts the energy value difference between the current frame I and the frame preceding J frames.

2.4. Localization of consonants

Consonants are made when a constriction in the vocal tract is suddenly formed and released. This articulation movement affects the acoustics so that the spectrum of the speech signal suddenly changes at the point in time when the consonant closes or releases [26]. These time points are called landmarks that help locate the start and end of a sound. The first step in identifying a consonant is to locate these points in a speech sequence. For this purpose, we used the landmark method of Liu who proposed three types of landmarks: g (glottis), b (burst) and s (syllabicity). Glottis points mark the times when the vocal cords start ($g+$) or stop ($g-$) the free vibration. These times correspond to the ROC crossing points in the first band above and below threshold values of 9 dB ($g+$) and -9 dB ($g-$) respectively. The onset ($b+$) of the burst of air after an affricate, stop consonant release, or onset of frication noise for fricative consonants, and the offset ($b-$) where aspiration or frication noise abruptly stops due to a stop closure are marked by burst points. The most prominent peak in the ROC of bands 2, 3, and 4, between the ($g-$) and ($g+$) points, is where this burst occurs. The onset ($s+$) or offset ($s-$) of voiced sonorant consonants are marked by syllabicity points. In our process, we exploited, only, the two types of landmarks g and b . Here are examples of the most common types of syllabics [27]: ($+g, -g$): Denotes a voiced consonant or a vowel. ($+b, +g, -g$): A syllable that starts with a fricative, with ($+b$) denoting the presence of frication. And ($+b, -b, +g, -g$): Syllables with an initial plosive, with ($+b, -b$) indicating the release's start and finish.

The peak energy variation in the frequency band from 0 to 400 Hz (band1) is used to detect voicing offsets ($g-$) and voicing onsets ($g+$); the peak energy is calculated as (3):

$$E_g(i) = 10 \log_{10}(\max_{k1 \leq k \leq k2} |X(i, k)|^2) \quad (3)$$

where $k1 \leq k \leq k2$, $k1$ and $k2$ being the DFT indices corresponding to 0 and 400 Hz respectively. A rate of rise measure of $E_g(i)$ is computed with a time step of 50 ms ($J = 50$) as (4):

$$rE_g(i) = E_g(i) - E_g(i - J) \quad (4)$$

the voicing offset and onset points are determined by crossing points $rE_g(i)$ below and above threshold values of -9 dB and +9 dB, respectively. The most prominent peak in the ROC, between the ($g-$) and ($g+$) points, has an intervocalic burst onset ($b+$).

2.5. Calculation of the energy in the bands

We define the normalized band energy in each frame of the syllable as (6):

$$E_{bn}(i) = E_b(i) - / E_T(i) \tag{5}$$

the normalized band energy b in frame (i) is represented by $E_{bn}(i)$, the overall energy in frame (i) is represented by $E_T(i)$, and the band energy b in frame (i) is represented by $E_b(i)$.

2.6. Artificial neural network (ANN) algorithm

The artificial neural network is a mathematical model that works much like the human brain. The multilayer perceptron (MLP) is one of the most widely used neural networks today for classification. It is a feed-forward network of one or maybe more layers of nodes between the input and output nodes hidden between them. We tested the MLP network with one and two hidden layers as shown in Figures 1 and 2, each with a different number of neurons in the hidden layer (s). The output layer consists of two neurons while the input layer consists of four neurons [28]. There is no universal rule for calculating the number of neurons per hidden layer (s), but there are some guidelines. The size of the hidden layer must either be equal to that of the input layer [29], or equal to 75% of it [30].

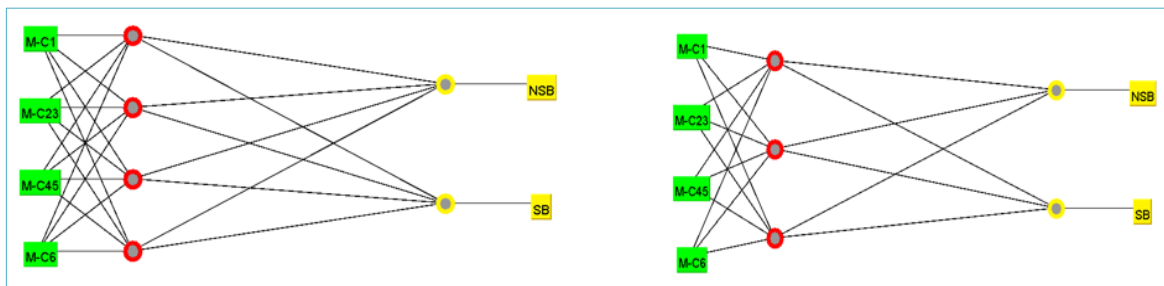


Figure 1. Hidden single layer neural network of four neurons on the left and three neurons on the right

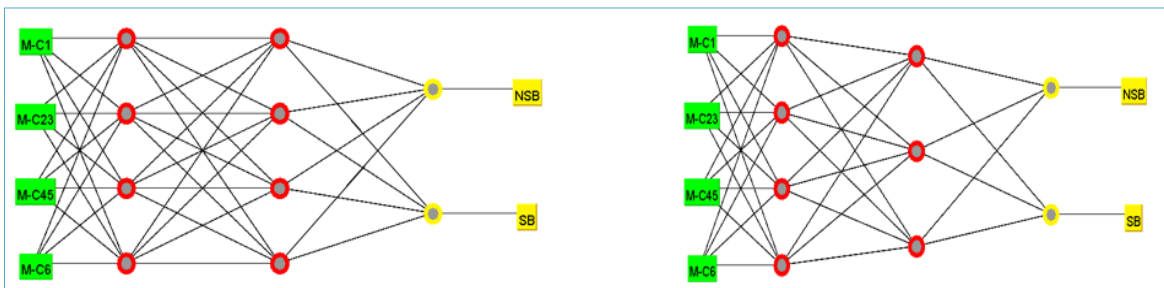


Figure 2. Hidden double-layer neural network of (4, 4) neurons on the left and (4, 3) neurons on the right

2.7. Support vector machine (SVM) algorithm

SVM is a classification algorithm that uses supervised machine learning. The task is to find a decision function based on the best hyperplane margin separation. Support vectors are the data points closest to the hyperplane. SVMs are a generalization of linear classifiers. In the case where the data to be processed is not linearly separable, SVM converts the input data's representation space into a higher-dimensional space in which a linear separation is more probable [31]. This is achieved through a kernel function. The usual kernel used with SVMs is the polynomial kernel.

3. RESULTS AND DISCUSSION

3.1. Results obtained by our algorithm

By analyzing the band energy percentage distribution in fricative consonants (/ʒ/, /h/, /χ/, /z/, /s/, /ʃ/, /sʃ/, /ʒ/, /χ/, /f/ and /h/) we observed that this distribution varies according to the consonant's articulation point

as shown in Figures 3 and 4. We have noticed that when the fricative consonant is sibilant, the band which contains the least amount of energy is the second band B2 (800-2000 Hz). When the fricative consonant is non-sibilant, the least amount of energy is in the fourth band B4 (5000-8000 Hz). This behavior can be explained by the effect of the shape and position of the constriction of the vocal tract.

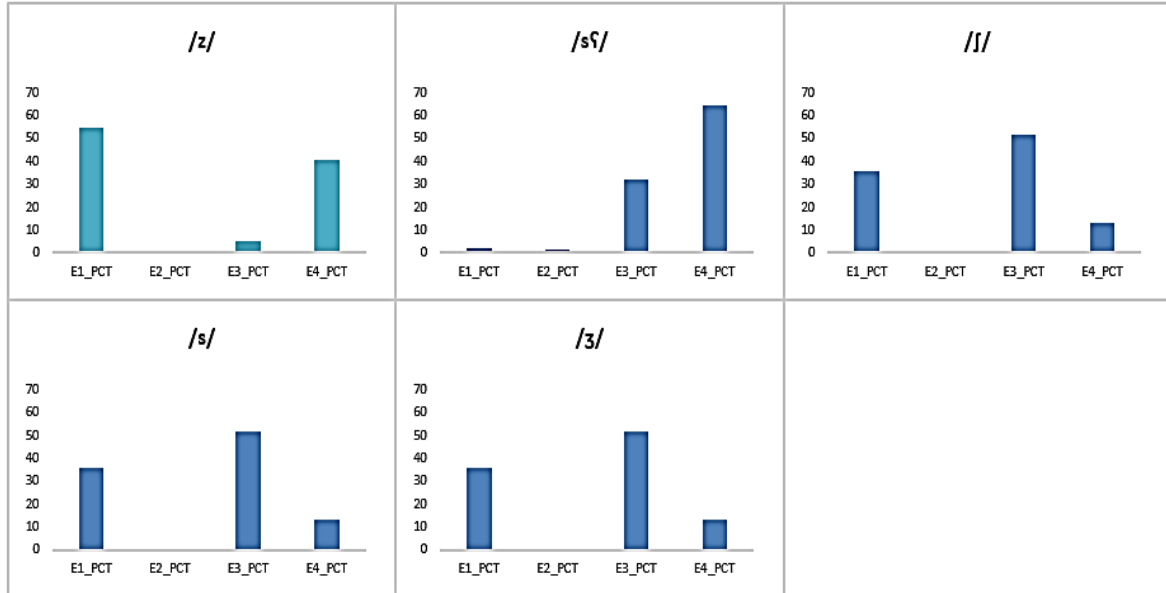


Figure 3. The energy percentage distribution of sibilant fricatives

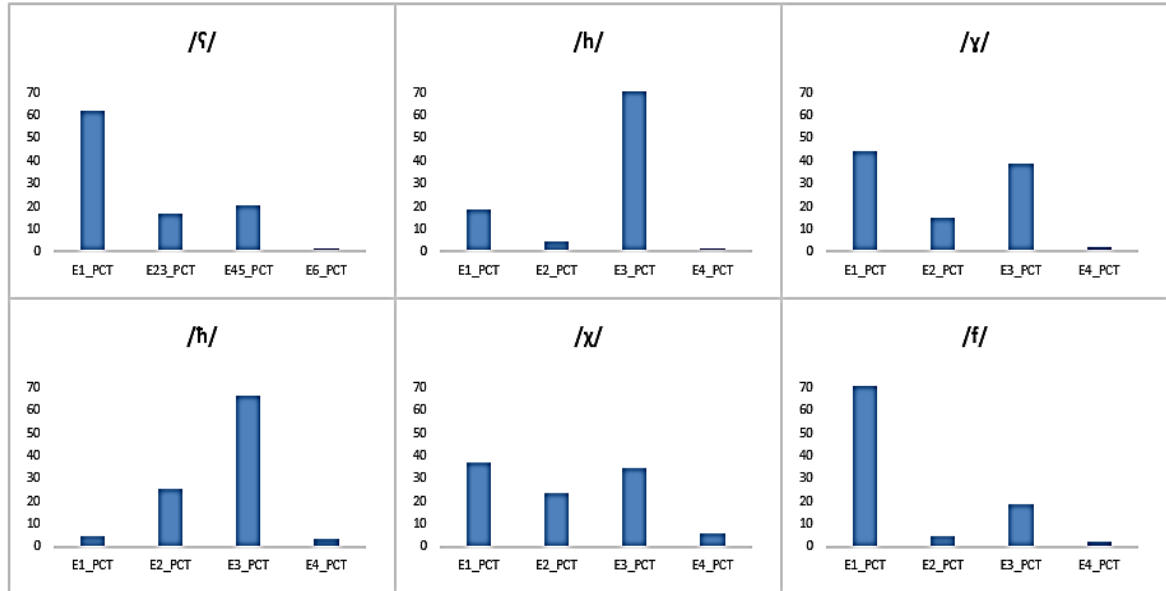







Figure 4. The energy percentage distribution of non-sibilant fricatives

The energy in the four bands as shown in Figure 3 in the middle of the sibilant consonants indicates that the consonant /ʒ/ has energy in band B1, while the consonant /s/ has low energy in band B1, low energy in band B2, near to zero energy in band B2, low energy in band B3, and high energy in band B4. The consonant /ʃ/ has a low energy in B1 and B2, a high energy in B3 and an energy of 20% in the band B4. For the consonant /sʃ/, the energy is close to zero in B1, low in B2 and more than 35% in B3 and B4. The consonant /z/ is distinguished by a high energy in band B1, a near-zero energy in band B2, a medium energy in band B3, and a 40% energy in band B4. This analysis shows that sibilant fricatives are characterized by zero energy in the

band B2 with the exception of the consonant /s/ which has a low energy in this band. They are also characterized by high energy in the high frequencies (B3 and B4). Low energy in band B2 (close to zero), high energy in band B3 and energy greater than 20% in band B4. Therefore, the common feature among hissing fricative consonants is the low energy (close to zero) of the second band B2. This is due to the fact that the production of sibilant consonants involves a strong lingual tension: along the entire length of the tongue, a canal is hollowed out, with the air passing through a small circular opening at the point of articulation. The Table 4 shows the articulation of sibilant consonants.





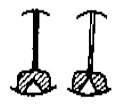

Table 4. The sibilant fricatives articulation [32]

/s/	/z/	/s ^h /	/ʃ/	/ʒ/
				
Unvoiced consonant produced by the approach of the tip of the tongue towards the alveolar region.	Voiced consonant. Same articulation as /s/, but with vibration of the vocal cords.	Unvoiced consonant. The back of the tongue is hollowed out in a channel and approaches the anterior or central part of the hard palate.	The tongue is pressed against the alveoli. Lips are often rounded or protruding forward when hissing.	Voiced consonant. Same articulation as /ʃ/, but with vibration of the vocal cords.

The energy distribution of the bands of the non-sibilants as shown in Figure 4 reveals that the consonant /ʃ/ is characterized by a low energy in B1, a high energy in B2, an energy close to 20% in B3 and too low energy in B4. The consonant /h/ has an energy of 20% in the band B1, a low energy in B2, an energy of 80% in B3 and an energy close to zero in B4. For the consonant /ç/, it has a high energy in B1 and B4, an energy of 15% in the band B2 and a low energy in B4. The consonant /x/ is described by an energy greater than 30% in B1 and B3, an energy greater than 15% in B2 and an energy of 7% in B4. Regarding the consonant /ħ/, the energy is close to zero in B1, greater than 20% in B2, strong in B3 and too weak in B4. Finally, the consonant /f/ has too low energy in B1, high energy in B2 and B3 and low energy in B4. This examination shows that non-sibilant fricatives have a low energy in band B4 except for the fricative /ç/ and a high energy in band B2 for all consonants except the fricative /h/.

Hence, the non-sibilant fricative consonants are all characterized by the low energy (close to zero) in the fourth band B4. This behavior can be justified by the fact that the production of non-sibilant consonants takes place in the posterior part of the vocal tract with the exception of the consonant /f/. Table 5 summarizes the articulation of non-sibilant consonants.

Table 5. The non-sibilant fricatives articulation [32]

/x/	/ç/	/ħ/	/ʃ/	/h/	/f/
					
Unvoiced consonant. The posterior part of the back of the tongue retracts very strongly towards the soft palate, near the uvula.	Same articulation as /x/, but with vibration of the vocal cords.	Unvoiced consonant. The root of the tongue is pushed back strongly and approaches the posterior wall of the pharynx. There is a strong friction. The articulatory tension is very strong.	Same articulation as /h/, but with vibration of the vocal cords.	Unvoiced consonant. The glottis is almost entirely closed except for a narrow opening in its upper part at the level of the arytenoid cartilages	The lower lip is close to the upper teeth and can sometimes brush against them with its upper outer part or, sometimes, with its inner part.

The findings from the preceding part indicate that the percentage distribution of the band's energy in fricative consonants is dependent on the consonant's articulation point. These consonants can, however, be distinguished. Indeed, the percentage of energy present in bands 2 and 4 allows two classes of consonants (/s, s^h, z, ʒ and /ʃ and /f, ç, x, ħ and h/) to be distinguished. In the first group, called sibilant consonants, the energy in B2 tends towards zero and it is greater than 20% in B4. In the second group, named non-sibilant consonants, the energy in B4 tends towards zero and it exists in B2 with significant percentages.

Based on these findings, we developed an algorithm to classify fricative consonants into two categories: sibilant/non-sibilant. The operating principle of our algorithm is described as follows: first, the voice signal goes through the general processing step. The spectrogram is then computed and split into four frequency bands. In each of the four bands, an energy waveform is constructed, the energy derivative is calculated, and peaks in the derivative are identified. These peaks help us in the processing of the g-b landmarks. The landmarks of each consonant are then computed, as well as their normalized energy and rate of change in the four frequency bands. These outputs are used during the classification phase as shown in Figure 5 in order to recognize the group of the consonant concerned (sibilants/non-sibilants).

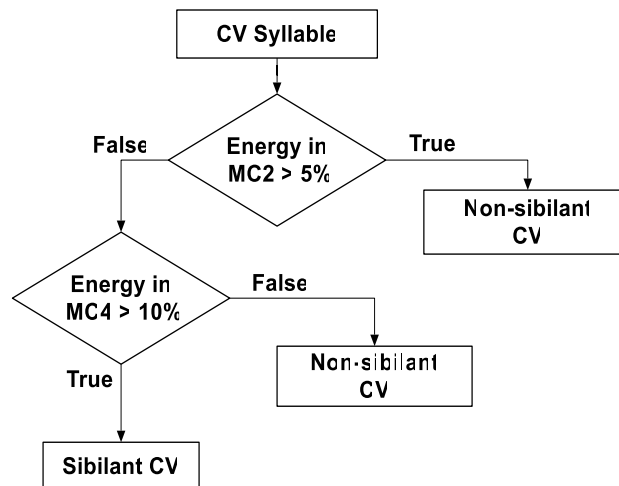


Figure 5. Classification algorithm for sibilant and non-sibilant fricatives. (MC2: energy in the middle of the consonant in band 2. MC4: energy in the middle of the consonant in band 4)

In order to evaluate our algorithm's performance, it was tested using the records from our corpus. The total number of fricative consonants in these experiments was 3564 CV of which 1944 CV are non-sibilant consonants and 1620 CV are sibilant fricatives. The results obtained showed that 1804 non-sibilant fricatives were correctly classified, which gives an accuracy of 92% and 1373 sibilant fricatives were correctly identified with an accuracy of 84%. We can see that, for all fricative consonants, the overall classification rate of our algorithm is greater than 89%. We then compared the performance of our algorithm with that of the algorithms ANN and SVM.

3.2. Results obtained by ANN and SVM algorithms

The two algorithms (ANN and SVM) are executed in the Weka software, we used, for both algorithms, cross validation. This is a standard evaluation technique; it allows you to perform repeated percentage splits. Divide the data set into 10 pieces ("folds"), then hold each piece in turn to test it and practice the algorithm on the other 9 sets. Tables 6 and 7 summarize the results obtained by the ANN algorithm while the Table 8 presents the results obtained by the SVM classifier.

We can see that the best recognition rate of the ANN algorithm is that obtained by the perceptron of two hidden layers of four neurons per layer (88.60%). The recognition rate given by the SVM algorithm is 84.62%. By comparing the results obtained by the ANN and SVM algorithms with those of our algorithm, we can see that our algorithm remains the most efficient with an overall rate of 89%.

Table 6. Results obtained by the ANN algorithm for a single hidden layer perceptron

Number of neurons per layer	Overall recognition rate	Recognition rate of non-sibilants	Recognition rate of sibilants
4	88.02%	85.0%	91.7%
3	87.20%	85.2%	89.6%

Table 7. Results obtained by the ANN algorithm for a double hidden layer's perceptron

Number of neurons per layer	Overall recognition rate	Recognition rate of non-sibilants	Recognition rate of sibilants
4, 4	88.60%	85.9%	91.9%
4, 3	87.73%	84.6%	91.5%

Table 8. Results obtained by the SVM algorithm

Kernel	Overall recognition rate	Recognition rate of non-sibilants	Recognition rate of sibilants
Polynomial	84.62 %	84.3 %	85.1 %

3.3. Discussion

The main aim of this study is to characterize and classify the sibilant /s, s^ʕ, z, ʒ and ʃ/ and non-sibilant /f, χ, γ, ʕ, ħ and h/ Arabic fricative consonants using the energy normalized in the four different frequency bands (Band 1: 0-400 Hz; Band 2: 800-2000 Hz; Band 3: 2000-5000 Hz and Band 4: 5000-8000 Hz). Looking at the energy distribution of each CV syllable in the four bands, we can see that sibilant fricatives have higher frequency sounds than non-sibilant fricatives. From the articulation point of view, speech generally appears as a regular alternation of more or less large openings and more or less complete closings of the vocal tract. Fricative consonants are articulated with a critical constriction of the vocal tract: between the lower lip and the upper incisors for the labiodental /f/, between the tip of the tongue and the alveoli for the alveolar /s, s^ʕ and z/, between the blade and the back of the alveoli for the post-alveolar /χ and γ/, between the root of the tongue and the posterior walls of the pharynx for the pharyngeal /ʕ and ħ/ and in the glottis for the laryngeal /h/. This narrowing, putting the exhaled air into turbulence, produces a noise whose amplitude and frequency structure depend on the air pressure in the constriction area, its diameter, and the location of obstacles in the column of the air leaving the fricative channel [33]. These factors make it possible to separate sibilant fricatives from non-sibilant ones. The sibilants show a greater intensity of friction noise because the jet of expelled air strikes the obstacle at a right angle proposed by the upper incisors for /s, s^ʕ, z, ʒ and ʃ/, which has the effect of increase turbulence [34]. On the other hand, the absence of a real obstacle or the existence of a less frontal obstacle to the air flow for non-sibilant consonants /f, χ, γ, ʕ, ħ, and h/, hardly amplifies the noise. The spectral distribution of energy is also different depending on the place of articulation. The sibilants have more polarized energy on a specific frequency band than the non-sibilants. Thus, maximum energy around 2000-7000 Hz (band B3 and band 4) typically emerges for /s, s^ʕ, z, ʒ and ʃ/. The voiced fricatives differ from the unvoiced by a higher overall energy in the low frequencies (the band B1) because of the vibration of the vocal cords. Non-sibilants have an energy distribution in the medium or low frequencies (bands B2 and B3).

4. CONCLUSION

The classification of fricative consonants is a more difficult task of speech recognition. In this work, we investigated the characterization and classification of Arabic fricative consonants based on the percentage of energy in four frequency bands. The results obtained show that the energy in B2 and B4 helps to classify the sibilants consonants /s, s^ʕ, z, ʒ and ʃ/ from the non-sibilants ones /f, χ, γ, ʕ, ħ, and h/. Therefore, the energy distribution of frequency bands presents indices which are rich in information content and useful in the classification of fricative consonants. Classification experiments were carried out on the fricative consonants extracted from our Arabic corpus. The results gave an overall rating of 89%. Future work will be directed towards the identification and classification of sibilant consonants.

REFERENCES

- [1] J. Cantineau, "Study of Arabic linguistics," (in French), Universitas Michigan, 1960.
- [2] A. Juneja and C. E. Wilson, "Segmentation of continuous speech using acoustic-phonetic parameters and statistical learning," *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP '02)*, vol. 2, 2002, pp. 726-730, doi: 10.1109/ICONIP.2002.1198153.
- [3] A. Al-Nassir, "Sibawayh the phonologist: A critical study of the phonetic and phonological theory of sibawayh as presented in his treatise Al-Kitab," *London and New York: Kegan Paul International*, 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, pp. 1180-1192, 1971, doi: 10.1121/1.1912751.
- [5] P. Ladefoged and I. Maddieson, "The sounds of the world's languages," Wiley-Blackwell, 1996.
- [6] C. Shadle, "The acoustics of fricative consonants," in *Research Laboratory of Electronics*, Tech. Rep., Cambridge, MA, 1985.
- [7] G. R. Tomiak, "An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents," Ph.D. Dissertation, State University of New York at Buffalo, 1990.
- [8] S. Nittrouer, "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *The Journal of the Acoustical Society of America*, vol. 97, pp. 520-530, 1995.
- [9] K. Tjaden and G. S. Turner, "Spectral properties of fricatives in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 40, pp. 1358-1372, 1997.
- [10] 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.
- [11] P. Ladefoged and K. Johnson, "A course in phonetics," 4th Ed, TX: Harcourt College Publishers, 2001.
- [12] S. Nissen and R. A. Fox, "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2570-2578, 2005, doi: 10.1121/1.2010407.

- [13] L. M. T. Jesus and C. H. Shadle, "Acoustic analysis of European Portuguese uvular [χ, ʁ] and voiceless tapped alveolar [t̪] fricatives," *Journal of the International Phonetic Association*, vol. 35, no. 1, pp. 27-44, 2005.
- [14] M. Tabain, "Variability in fricative production and spectra: implications for the hyper- and hypo- and quantal theories of speech production," *Language and Speech*, vol. 44, pp. 57-94, 2001, doi: 10.1177/00238309010440010301.
- [15] M. Toda, "Speaker normalization of fricative noise: considerations on language-specific contrast," in *Proceedings of the 16th ICPhS*, pp. 825-828, 2007.
- [16] C. Meunier, "Acoustic phonetics," Auzou P., Ed., *Les dysarthries*, (in French), 2007, pp. 164-173.
- [17] K. Driaunys, V. Rudzionis, and P. Žvinys, "Implementation of hierarchical phoneme classification approach on LTDIGITS Corpora," *Information Technology and Control*, vol. 38, no. 4, pp. 303-310, 2009.
- [18] E. K. Sung and Y. J. Cho, "An acoustic study of korean and english voiceless sibilant fricatives," *Phonetics and Speech Sciences*, vol. 2, no. 3, pp. 37-46, 2010.
- [19] A. Alwan, J. Jiang, and W. Chen, "Perception of place of articulation for plosives and fricatives in noise," *Speech Communication*, vol. 53, no. 2, pp. 195-209, 2011, doi: 10.1016/j.specom.2010.09.001.
- [20] K. D. Malde, A. Chittora, and H. A. Patil, "Classification of fricatives using novel modulation spectrogram based features," *International Conference on Pattern Recognition and Machine Intelligence*, 2013, pp. 134-139, doi: 10.1007/978-3-642-45062-4_18.
- [21] E. Nirgianaki, "Acoustic characteristics of Greek fricatives," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 2964-2976, 2014, doi: 10.1121/1.4870487.
- [22] L. Spinu and J. Lilley, "A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives," *Journal of Phonetics*, vol. 57, pp. 40-58, 2016, doi: 10.1016/j.wocn.2016.05.002.
- [23] R. S. Prasad and B. Yegnanarayana, "Identification and classification of fricatives n speech using zero-time windowing method," *Interspeech*, Hyderabad, pp. 187-191, 2018.
- [24] M. A. Al-Khairy, "Acoustic characteristics of Arabic fricatives," Ph.D. Dissertation, University of Florida, 2005.
- [25] A. Bendahmane, "Acoustic study of standard Arabic fricatives (Algerian speakers)," (in French) M.S. Thesis, École Doctorale Des Humanités, University of Strasbourg, 2013.
- [26] A. L. Sharlene, "Landmark detection for distinctive feature-based speech recognition," *The Journal of the Acoustical Society of America*, vol. 100, no. 5, pp. 3417-3430, 1996, doi: 10.1121/1.416983.
- [27] S. Boyce, H. Fell, and J. Macausian, "SpeechMark: landmark detection tool for speech analysis," *Interspeech*, pp. 1849-1897, 2012.
- [28] R. P. Lippmann, "Review of neural networks for speech recognition," in *Neural Computation*, vol. 1, no. 1, pp. 1-38, Mar. 1989, doi: 10.1162/neco.1989.1.1.1.
- [29] B. Wierenga and J. Kluytmans, "Neural nets versus marketing models in time series analysis: A simulation study," *Marketing: Its Dynamics and Challenges*, 1994.
- [30] V. Venugopal and W. Baets, "Neural networks and statistical techniques in marketing research: a conceptual comparison," *Marketing Intelligence and Planning*, vol. 12, no. 7, pp. 30-38, 1994, doi: 10.1108/02634509410065555.
- [31] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998, doi: 10.1023/A:1009715923555.
- [32] J. M. C. Thomas, L. Bouquiaux, and F. Cloarec-Heiss, "Initiation to phonetics: Articulatory and distinctive phonetics," (in French), *Presses universitaires de France*, 1976.
- [33] G. Fant, "Acoustic theory of speech production," La Haye: Mouton, 1960.
- [34] C. H. Shadle, "Articulatory-acoustic relationships in fricative consonants," *Speech Production and Speech Modelling*, vol. 55, pp. 187-209, 1990, doi: 10.1007/978-94-009-2037-8_8.