

Exploring Arabic non-sibilant fricative consonants: a comprehensive classification

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

¹IMII Laboratory, Department of Electrical and Computer Engineering, Faculty of Sciences and Techniques, Hassan First University, Settat, Morocco

²LAPSSII Laboratory, High School of Technology, Cadi Ayyad University, Safi, Morocco

Article Info

Article history:

Received Oct 16, 2025

Revised Aug 1, 2026

Accepted Aug 12, 2026

Keywords:

Arabic language

Artificial neural networks

Automatic speech recognition

Energy in frequency bands

Fricative consonants

Non-sibilant

Support vector machines

ABSTRACT

Voice interaction serves as an efficient tool for improving communication between humans and machines, enabling quicker exchanges without the need for physical contact with devices. This study introduces a methodology aimed at identifying Arabic non-sibilant fricative consonants and recognizing their significance within this context. Our approach involves analyzing the normalized energy distribution within consonant-vowel phonetic syllables, which serves as the acoustic input signal for our speech recognition system. Notably, the relatively limited attention given to non-sibilant consonants in existing literature underscores the importance of this research endeavor. The evaluation of our classification method applied to our dataset highlights the pivotal role of vocal signal energy in characterizing Arabic non-sibilant fricative consonants. Achieving a classification accuracy of 90% for distinguishing voiced and voiceless non-sibilant consonants, with a rate exceeding 85% for other categories, underscores the effectiveness of our approach. Noteworthy is the outperformance of our classification technique compared to current algorithms based on support vector machines (SVM) and artificial neural networks (ANNs), indicating its potential to enhance Arabic automatic speech recognition (ASR) systems. Overall, this research contributes to advancing speech recognition technology, particularly in the realm of Arabic linguistics, offering promising implications for future developments in this field.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Youssef Elfahm

IMII Laboratory, Department of Electrical and Computer Engineering Faculty of Sciences and Techniques

Hassan First University

Road of Casablanca B.P: 577, Settat, Morocco

Email: y.elfahm@uhp.ac.ma

1. INTRODUCTION

In the era of technological progress, robots have become increasingly integrated into both industrial and domestic aspects of our daily lives. According to the International Federation of Robotics (IFR), in 2022 alone, the 27 member states of the European Union installed nearly 72,000 robotic units [1], showcasing the pervasive presence of machines. This widespread adoption has spurred the advancement of voice interaction as an efficient mode of communication between humans and machines. To comprehend human speech, machines must be equipped with an automatic speech recognition (ASR) system that translates spoken words into text, facilitating effective communication.

This paper introduces an Arabic speech recognition system employing an analytical approach based on the identification of phonemes, the fundamental units of sound. Phonemes can manifest as consonants or vowels [2], with consonant production involving the obstruction of airflow through the vocal tract by precise positioning and manipulation of articulators such as the tongue, lips, and teeth. This obstruction results in distinct consonant sounds, such as the closure of lips producing the plosive /p/ or the narrowing of space between teeth and lower lip producing the fricative /f/. Contrastingly, vowels are formed by allowing unrestricted airflow through the vocal tract [3]. Our objective is to characterize and identify fricative consonants, a significant component of the Arabic sound system, constituting about half of the consonants in the alphabet. These fricatives are distinguished by the hissing sound generated through the narrowing of the vocal tract, forcing air through a restricted channel [4]. The classification of fricative consonants, particularly sibilant and non-sibilant [5], is explored in the scientific literature. Sibilant fricatives involve directing airflow with the tongue towards closed teeth, producing a whistling noise. In Arabic, sibilant fricatives include /s, sʕ, z, ʒ, and ʃ/, while non-sibilant fricatives include /f, θ, ð, ðʕ, χ, ɣ, ʕ, ħ, and h/. Voicing further categorizes fricatives into voiced and unvoiced, where voiced fricatives involve vocal fold vibration, and unvoiced ones do not.

In this work proposes an analytical approach for accurately identifying non-sibilant Arabic fricatives /f, χ, ɣ, ʕ, ħ, and h/. The method extracts a subset of acoustic indices containing relevant information about these fricatives. Our central research question is: can the energy distribution within specific frequency bands of the vocal signal provide a basis for the accurate classification of Arabic non-sibilant fricatives? This work hypothesizes that analyzing the energy levels in carefully selected frequency bands will yield distinct patterns that allow for the reliable differentiation of these fricatives, potentially offering a computationally efficient alternative to more complex machine learning approaches. The article compares the predictive power of these parameters with the common approach based on neural networks and support vector machines (SVM).

The structure of the article includes a literature review on acoustic cues for fricatives, followed by the presentation of methods and tools used in the study. The results of the analysis and classification algorithms are then discussed, comparing them with two distinct classifiers: neural networks and SVM machines. The conclusion addresses the methods' advantages and limitations, discussing the implications for understanding non-sibilant fricatives in the Arabic language.

The research literature on fricative consonants reveals a wealth of studies exploring these sounds, which have garnered extensive attention. Despite this, our understanding of the relationships between articulatory organs, acoustic changes, and perceptual mechanisms associated with fricatives remains incomplete. The production of fricative consonants involves two parts of the vocal tract approaching each other, creating a narrow passage where turbulent air flow produces the fricative sound [5].

Within the theory of invariants [6], studies emphasize the crucial role of temporal, spectral, and amplitude characteristics in acoustic signals as essential perceptual cues. Stable acoustic features, such as voice onset time (VOT) for plosive consonants and formants for vowels, have been identified. However, the existence of similar invariant acoustic properties for fricatives has sparked controversy, with some studies challenging the notion of universal features and advocating for differentiation between sibilant and non-sibilant varieties. Most studies aimed at characterizing fricative consonants have primarily centered around the voiceless fricatives of English [7]–[9]. Nevertheless, there is a notable deficiency in research concerning fricatives in other languages, with a particular dearth of investigations into the overall consonant system of Arabic, especially focusing on fricatives.

Concerning the effects of various vocal tract configurations during fricative production on various acoustic measures, it is generally accepted that the size and shape of the vocal tract determine the spectrum of a fricative [10]. This is considered well-described by acoustic features such as the location of the spectral peak and the four spectral moments (spectral mean, spread, skewness, and kurtosis) [11]–[14]. Additionally, fricatives are not immune to coarticulation, and articulator movements have significant acoustic consequences for spectral energy distribution. Several studies emphasize the importance of considering language-specific and speaker-specific characteristics when studying fricatives. Consequently, the representation of spectro-temporal distribution has been successfully utilized to examine subtle differences among voiceless fricatives [15]. Furthermore, it has been observed that the vowel following the fricative has a more significant impact on the acoustic characteristics of /f/ than /s/, and even less on /ʃ/ [10]. These findings suggest the necessity of considering various variables that can influence the acoustic properties of fricatives to better understand the underlying mechanisms of their production.

The spectral peak, representing the most intense frequency in the vocal signal, is an acoustical parameter extensively studied in the context of fricatives. Numerous research efforts have focused on understanding the relationship between tongue position and the frequency of the spectral peak during fricative production at different locations in the mouth. It is assumed that this frequency decreases from high to low frequencies as the tongue moves from the front to the back of the mouth [16], [17]. However, this

relationship has not been confirmed for Greek fricatives, as demonstrated by Nirgianaki *et al.* [18]. Moreover, the spectral peak has proven valuable in distinguishing between sibilant and non-sibilant fricatives, specifically differentiating between alveolar and post-alveolar sibilants [8], [19]–[22].

The frequencies and intensity of the friction noise in voiced fricatives are acoustical features that differentiate between consonants based on their articulation point. Labiodental fricatives produce low-intensity noise with diffuse peaks around 3.5 kHz and 8 kHz. Alveolar fricatives, on the other hand, have more intense noise ranging between 4 kHz and 8 kHz, with peaks around 5 kHz and 8 kHz. Finally, for palatal fricatives, turbulence noise is between 2 kHz and 7 kHz, with a diffuse peak averaging around 4 kHz [23]–[25].

Another acoustical parameter frequently employed in the study of fricatives is the first spectral moment, which can be defined either as the average of the spectral energy distribution or as the center of gravity (CoG) of the fricative signal [26]. Research indicates that the CoG can distinguish between sibilant and non-sibilant consonants, as well as different types of sibilants. Studies have shown that sibilants have higher CoG values than non-sibilants [27], [28], and the consonant /s/ has a higher CoG than the consonant /ʃ/ [17], [29]–[32].

As for the second spectral moment, which measures the amount of spread or variance of spectral energy around the mean, previous research has demonstrated that sibilant fricatives exhibit lower spectral variance than non-sibilants [27], with the post-alveolar fricative /ʃ/ having the lowest spectral variance [11]. Other studies have reported findings regarding the third and fourth moments of the spectrum, namely skewness and kurtosis. Skewness characterizes the slope of the spectrum and provides a measure of the overall asymmetry of the energy distribution. A skewness value of zero indicates a symmetric distribution centered around the mean. Positive skewness implies a negative slope with energy concentration in low frequencies, while negative skewness indicates a positive slope and energy prevalence in higher frequencies [33], [34]. Kurtosis, on the other hand, refers to the “peakedness” or flatness of the distribution. A kurtosis of 3 indicates a normal distribution, while a value lower than 3 suggests a flat distribution, and a higher value means a sharper distribution [33], [34]. Researchers have suggested that skewness and kurtosis can potentially differentiate between the consonants /s/ and /ʃ/, finding that /s/ exhibits negative skewness while the fricative /ʃ/ has positive skewness [27], [31], [35]. However, other studies have also reported that /s/ had a higher positive skewness than /ʃ/ [27]. In terms of kurtosis, research has revealed that /s/ has a large positive value while /ʃ/ has a small positive or negative value [17], [27], [31], [35].

Regarding the characterization of Arabic fricatives, the work of Al-Khairy [36] discovered that non-voiced Arabic fricatives have a longer average duration of friction noise than voiced fricatives. Sibilant fricatives have a longer absolute average duration of friction noise compared to non-sibilant fricatives. He found that amplitude measurements can distinguish between sibilant and non-sibilant fricatives but cannot identify individual consonants within each group. By using the location of the spectral peak of fricatives, Al-Khairy [36] found that it is possible to distinguish between sibilant and non-sibilant fricatives but not between the post-alveolar non-sibilants /ʃ/ and the alveolar ones /s/ and /z/. In the same context, the results of the study conducted by Benamrane [23] classified Arabic fricatives into three groups based on their place of articulation: anterior fricatives /f, θ, ð, ðʃ/, sibilant fricatives /s, sʃ, z, ʒ, ʃ/, and posterior fricatives /x, ɣ, ʕ, h, and h/. Anterior fricatives exhibit a flat spectrum with few energy peaks and relatively low intensity. Their frequency CoG is moderately high but is the highest among all fricatives for /ð and ðʃ/. Sibilant fricatives have a compact spectrum, with a main region of higher energy than the other groups, as well as high intensity and frequency CoG. Fricatives /s/ and /sʃ/ have a higher CoG than /z, ʒ, and ʃ/. Posterior fricatives have intermediate or high intensity and a moderate or low frequency CoG.

Building upon our previous work characterizing Arabic fricatives using energy distribution in frequency bands [37], this study addresses the more challenging task of distinguishing between individual non-sibilant fricatives. While prior research, including our own, has successfully differentiated between sibilant and non-sibilant fricatives, a finer-grained classification of non-sibilants remains largely unexplored, particularly within the Arabic language. The rarity of research focusing specifically on individual non-sibilant fricatives is largely due to the inherent acoustic complexities and the significant difficulty in accurately classifying and identifying these phonemes compared to their sibilant counterparts. This research is important because accurate identification of these phonemes is crucial for advancing Arabic ASR systems, improving their robustness to variations in speech and accents. Furthermore, our novel approach leverages a computationally efficient method, analyzing energy distribution in carefully selected frequency bands, offering a potential advantage over complex machine learning models in resource-constrained environments. These findings not only contribute to a more detailed understanding of the acoustic properties of Arabic non-sibilant fricatives but also pave the way for more efficient and accurate Arabic speech recognition technologies.

2. METHOD

In this section, the overall principle of this proposed method is firstly described. After that, this work describes the details of the key steps in the following subsections. The method used to characterize non-sibilant Arabic fricative consonants is described as follows: in the case of a consonant–vowel (CV) type signal, this work has developed a method to analyze the energy distribution in the speech signal, as illustrated in Figure 1. Our approach relies on the identification of two specific intervals: the middle of the consonant and the middle of the vowel.

The key acoustic parameters of this method include the acoustic energy in the middle of the consonant, denoted by the acronym “MCi”, and the energy in the middle of the vowel, denoted by “MVi”. Here the subscript (i) signifies the relevant frequency band, because this work intends to examine the energy distribution in specific frequency bands. The precision of determining these two intervals is closely linked to the identification of the moment delimiting the transition between the consonant and the vowel. Once this time point is identified, this work selects the second third of the signal to characterize the consonant interval and similarly determine the corresponding interval for the vowel.

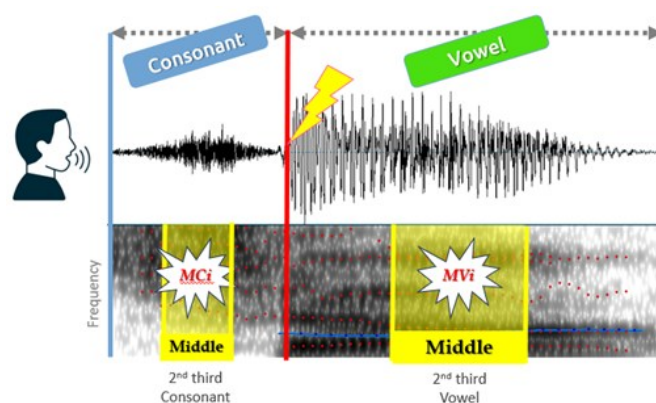


Figure 1. Energy distribution for a CV-type vocal signal

To do this, this work used the following methodology (see pipeline Figure 2): initially, this work built a database, i.e. a corpus of CV-type audio files. The phonemes in this database were then subjected to preprocessing to extract acoustic features, including energy in the bands. Then, we proceeded to the selection of the vector of indices, which constitutes the input of the classification system responsible for recognizing the non-sibilant fricative consonant: voiced, unvoiced, and pharyngeal.

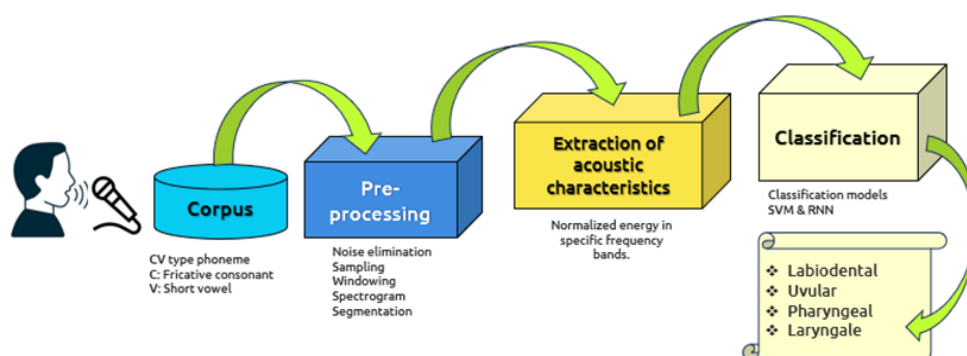


Figure 2. Speech signal classification processing pipeline

2.1. Corpus

To build our own corpus, this work invited 20 Moroccan speakers to repeat a CVCVCV-type sequence four times. This sequence is illustrated graphically in Figure 3. These sequences were recorded

using Praat software, with a sampling frequency of 22.05 kHz. Then, proceeded to segment these sequences in order to isolate the CV unit, ultimately resulting in a corpus comprising 4,320 CV units. To guarantee absolute precision and avoid coarticulation noise, the initial segmentation of the transitions between the consonant and the vowel was performed manually on the first half of the dataset using Praat.

This meticulous manual annotation served as a reliable ground truth to study and validate the acoustic behavior of landmarks. Once the manual boundaries were established, this work validated that our automatic detection, based on the rate of change (ROC) of normalized energy, perfectly matched the manual references. Consequently, the second half of the corpus was segmented automatically, providing a highly reliable and robust set of 4,320 samples without requiring multiple human annotators.

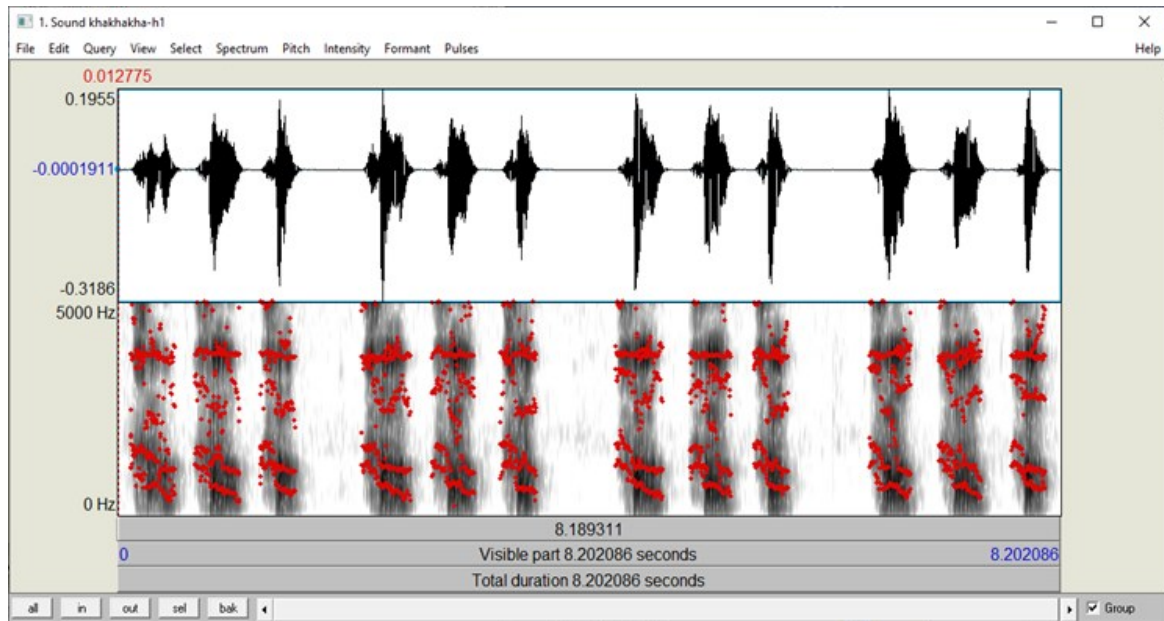


Figure 3. Illustration of a CVCVCV-type sequence for the consonant / χ / followed by the vowel / a /

2.2. Preprocessing

Initial preprocessing of a speech signal sample involves the application of a pre-emphasis filter, aimed at improving speech quality by improving intelligibility and attenuating background noise. Given the stochastic nature of the speech signal, a windowing step follows to decrease variance effects and improve temporal resolution, using a Hamming window which has a Gaussian shape in the time domain. Zero-padding is used to improve frequency resolution, and to address signal discontinuities a window overlaps of 9.6 ms is implemented. The next step consists of calculating the fast Fourier transform (FFT) to move from the time domain to the frequency domain, thus facilitating the generation of a spectrogram. Subsequently, this spectrogram is used to determine the energy in specific frequency bands.

2.3. Energy calculation

The energy per frequency band, denoted “ E_b ”, was calculated by adding the squares of the amplitude of each segment of the spectrogram according to (1).

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

Where $X(n, k)$ represents the amplitude of the spectrum smoothed using a 20-point moving average at time index n . The index b represents the considered frequency band. The frequency index k varies between the indices of the Fourier transform corresponding to the lower and upper bounds of each band.

The calculation of the logarithm of the energy in each frequency band is also justified by the functioning of the human auditory system since we do not perceive sound intensity on a linear scale. To mitigate the influence of speaker-related variations, a normalization step was implemented. This involved dividing the energy in each band, represented by “ b ”, by the total energy, ensuring a normalized

measurement for analysis and comparison. Using the notations “PE_b” to denote the percentage of energy in band (b) at frame (n), and “ET” to represent the total energy across all bands as shown (2).

$$PE_b(n) = 100 \cdot \frac{E_b(n)}{E_T(n)} \quad (2)$$

2.4. Location of consonants

This study approach uses two acoustic parameters: the energy at the middle of the consonant and the energy at the middle of the vowel. The effectiveness of this method depends on the temporal pointy delimiting the transition between the consonant and the vowel. To identify this point, we exploited landmark points, which are moments relative to abrupt changes in the acoustic spectrum due to articulatory movements associated with the production of sounds (consonant/vowel). In this study need the (g) landmark, which characterizes the vibration of the vocal cords, allowing us to locate the beginning and end of a vowel precisely. To determine the g landmark, this study calculated the ROC of the normalized energy in the B1 band [100-400 Hz] using (3).

$$ROC_{Eb}(n) = E_b(n) - E_b(n - t)) \quad (3)$$

In Figure 4, especially in the lower part illustrating the ROC, two types of peaks (one positive and one negative) can be observed. Each landmark presents two peaks, corresponding to the beginning and end of an acoustic phenomenon. The acoustic landmark (g) corresponds to the times when the vocal cords begin to vibrate (g+) and when this vibration ends (g-). These times are identified by the crossing points of the ROC of the normalized energy curve for the band [0-400 Hz] with the thresholds of 9 dB (g+) and -9 dB (g-) respectively.

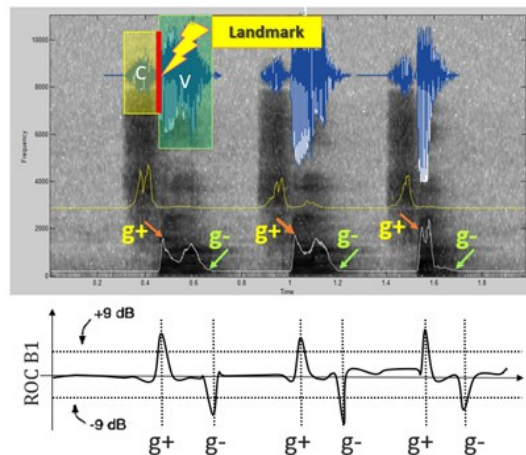


Figure 4. Landmark points detection processing. The bottom curve is the ROC in energy within the [0-400 Hz] band

2.5. Classification algorithms

In the classification phase, this study employed two classification models to assess the performance of our own classification method: artificial neural networks (ANN) and SVM. Neural networks, capable of supervised learning, consist of computational units called neurons. Each neuron undergoes a linear step, summing weighted inputs, followed by a non-linear step with an activation function like the sigmoid. The arrangement of neurons in the layers of the ANN is flexible during design. The process of using an ANN involves choosing the number of layers, neurons per layer, and activation functions, followed by random weight initialization, training to adjust weights, and finally, the prediction phase [38]–[40].

As for the SVM classifier, also known as a large-margin separator, it relies on a supervised machine learning concept to find an optimal hyperplane for separating different classes. The SVM's operation includes selecting a kernel function, a mathematical transformation making data linearly separable in a high-dimensional space. Next, the SVM is trained by identifying the hyperplane that maximizes the margin between classes. Once trained, the SVM can be used to predict new data points [41], [42].

3. RESULTS AND DISCUSSION

In our previous studies, our initial investigation focused on the categorization of fricatives based on their mode of articulation, specifically whether they are sibilant or non-sibilant. We observed that the majority of non-sibilant consonants are characterized by low energy in high frequencies (5,000-8,000 Hz) and significant energy in low frequencies (800-2,000 Hz) [37]. Additionally, we noted a substantial difference in energy distribution for the same consonant followed by three different vowels. To mitigate the vowel's effect on the consonant, we opted to conduct a classification within groups of consonants followed by the same vowel. This decision led us to undertake vowel classification, where we found that the vowel /a/ distinguishes itself from other vowels with a significant amount of energy in the band (600-1,800 Hz). Similarly, the vowel /i/ stands out from the vowel /u/ due to notable energy in the band (1,800-4,600 Hz). In this study algorithms for classifying non-sibilant fricatives operates as follows: within a given vowel context, the algorithm categorizes consonants into two groups: voiced and unvoiced. Subsequently, it proceeds to classify the consonants within each group.

3.1. Energy distribution analysis of non-sibilant fricatives

This section outlines the findings from the analysis of normalized energy distribution within non-sibilant fricative consonants in the Arabic language (/f, χ, γ, ʕ, ħ, and h/) followed by one of the three short vowels (/a, i, and u/). This study analysis is based on the examination of energy distribution within specific frequency bands. Specifically, this study scrutinized the following bands: [100-400 Hz], [400-600 Hz], [600-1000 Hz], [400-1,000 Hz], [1,000-2,000 Hz], [1,000-2,600 Hz], [2,000-3,200 Hz], [3,200-4,200 Hz], [4,200-4,600 Hz], and [4,600-8,000 Hz].

An analysis of Figures 5 to 7 revealed that non-sibilant consonants, whether voiced or voiceless, exhibit significant energy in the first frequency band [100-400 Hz]. This characteristic can be explained by the fact that these consonants are produced in the posterior part of the oral cavity, promoting the emergence of formant structures for fricative consonants, as suggested by Stevens [19]. However, this band does not differentiate between voiced and voiceless non-sibilant consonants. To characterize voicing in these consonants, this study examined the second frequency band [400-600 Hz]. Our observations showed that voiced consonants /γ/ and /ʕ/ exhibited energy exceeding 10%, regardless of the following vowel. In contrast, this frequency band contained no energy for voiceless consonants /f, χ, ħ, and h/. Furthermore, this study observed a difference in energy distribution in frequency bands for the same consonant followed by one of the three vowels. This observation indicates that vowels have a significant effect on non-sibilant consonants, which can be attributed to coarticulation. The following provides a detailed acoustic analysis of the distribution of normalized energy in the middle of all non-sibilant fricative consonants in the mentioned frequency bands.

3.1.1. The /ʕ/ consonant

For the /ʕ/ consonant followed by the /a/ vowel, the analysis of normalized energy distribution (Figure 5(a)) reveals that nearly 30% of the energy is concentrated in the [100-400 Hz] and [600-1000 Hz] bands, with less than 10% in the [400-600 Hz] band and approximately 20% in the [1,000-2,600 Hz] band. However, when the /ʕ/ consonant is followed by the /i/ vowel, the energy distribution changes, with around 30% in the [100-400 Hz] band, over 20% in the [400-600 Hz] band, about 10% in the [600-1,000 Hz] band, over 15% in the [2,000-3,200 Hz] band, and around 10% in the [1,000-2,600 Hz] band. For the /u/ vowel, there is a migration of energy from high to low frequencies, resulting in a concentration of 40% in the [100-400 Hz] band, over 30% in the [400-600 Hz] band, and about 15% in the [600-1,000 Hz] band.

3.1.2. The /γ/ consonant

Based on the analysis of normalized energy distribution graphs for the /γ/ consonant followed by the /a/ or /i/ vowels (Figure 5(b)), it was observed that the [100-400 Hz] frequency band represents approximately 60% of the total energy, while the [400-600 Hz] band contains over 10% of the energy. In contrast, the [600-1,000 Hz] band has almost no energy, less than 5%. The [3,200-4,200 Hz] band contains about 10% of the total energy. When the /γ/ consonant is followed by the /u/ vowel, the energy distribution remains the same for the [100-400 Hz] and [400-600 Hz] bands. However, a difference was observed for the third band [600-1,000 Hz], which contains approximately 10% of the total energy. The other bands have less than 5% of the total energy.

3.1.3. The /ħ/ consonant

For the /ħ/ consonant followed by the /a/ vowel, energy distribution is such that about 10% of energy is in each of the [100-400 Hz], [600-1,000 Hz], and [1,000-2,600 Hz] frequency bands, with a total absence of energy in the [400-600 Hz] band. The [1,000-2,600 Hz] and [3,200-4,200 Hz] bands each contain over 20% of energy (Figure 6(a)). When the /ħ/ consonant is followed by the /i/ vowel, approximately 10% of

the energy is concentrated in the [100-400 Hz] band, while the [2,000-3,200 Hz] and [3,200-4,200 Hz] bands each contain over 30% of energy. The other bands have almost no energy. When the /h/ consonant is followed by the /u/ vowel, the [100-400 Hz] band shows a concentration of over 50% of energy, while the [1,000-2,600 Hz] band contains about 10% of energy, and the [3,200-4,200 Hz] band around 20%. The other bands have low energy.

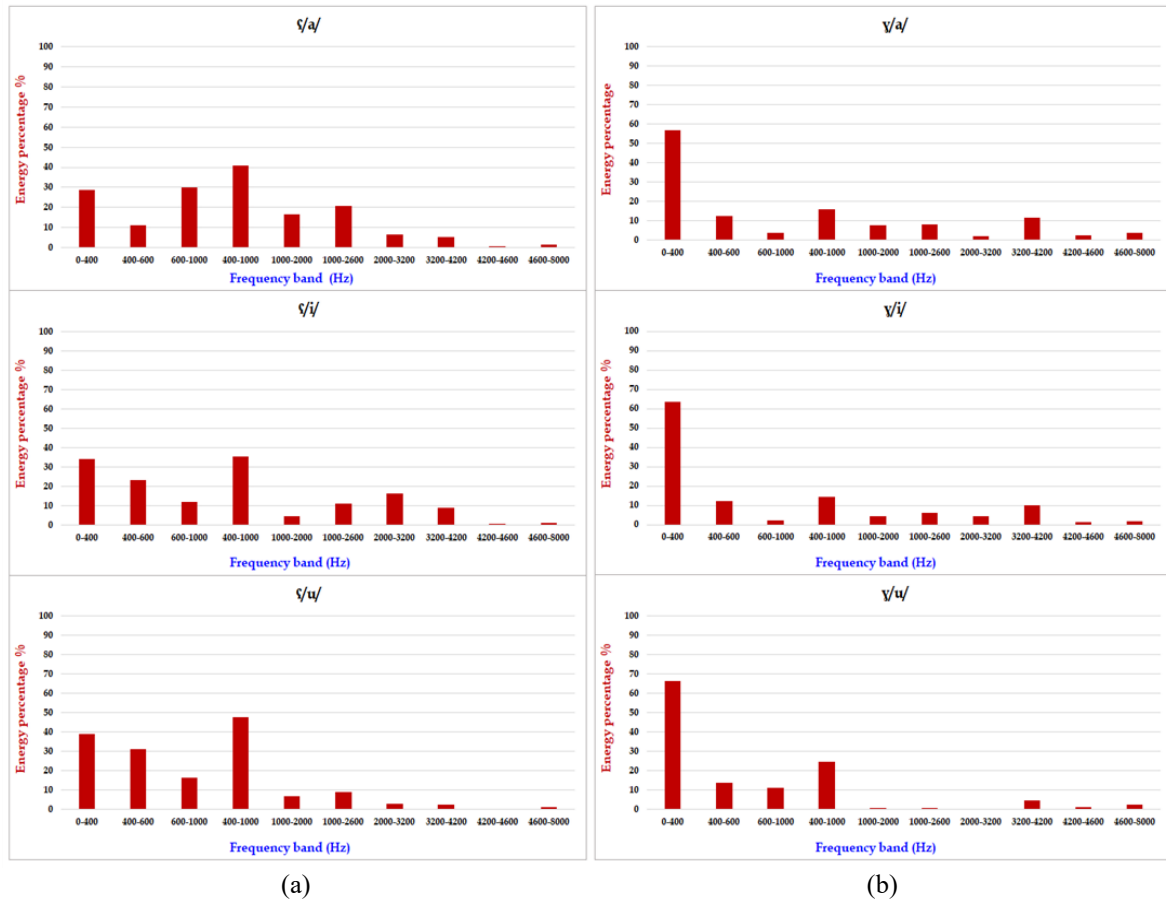


Figure 5. Energy distribution in the middle of voiced consonants: (a) /ʕ/ and (b) /ɣ/

3.1.4. The /χ/ consonant

After examining the energy distribution graphs of the /χ/ consonant followed by the /a/ or /i/ vowels (Figure 6(b)), it was observed that the high frequencies [4,200-8,000 Hz] contain about 20% of the total energy, while the [100-400 Hz] and [3,200-4,200 Hz] bands contain about 30% of the energy. The other bands have almost no energy. However, when the /u/ vowel follows the /χ/ consonant, a significant increase in energy in the [100-400 Hz] band was noted, representing over 60% of the total energy. The [400-1,000 Hz] and [4,200-8,000 Hz] bands contain about 10% of energy each, while the other bands have no energy.

3.1.5. The /h/ consonant

The energy distribution analysis of the /h/ consonant followed by the /a/, /i/, or /u/ vowels (Figure 7(a)) shows a strong concentration of energy in the [100-400 Hz] band. The other bands have low energy. The dominant low-frequency presence highlights the acoustic traits of /h/, mainly shaped by glottal airflow.

3.1.6. The /f/ consonant

After examining the graphs of normalized energy distribution for the /f/ consonant followed by the /a/, /i/, or /u/ vowels (Figure 7(b)), it was observed that the [100-400 Hz] band contains about 30% of the total energy. The [3,200-4,200 Hz] and [4,200-8,000 Hz] bands each contain approximately 20% of the energy, while the other bands have low energy. However, if the /u/ vowel follows the consonant, an increase in energy in the [100-400 Hz] band was noted, representing over 40% of the total energy. The other bands are similar to those of the /a/ and /i/ vowels.

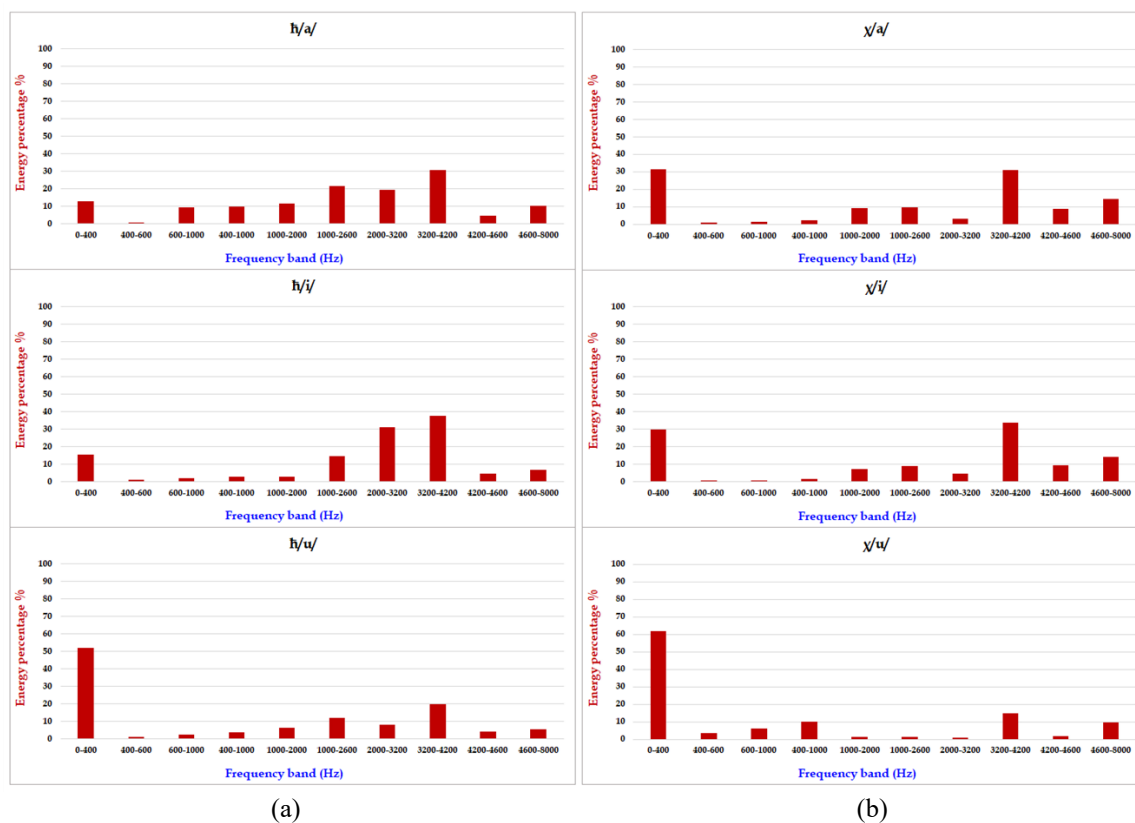


Figure 6. Energy distribution in the middle of unvoiced consonants: (a) /h/ and (b) /χ/

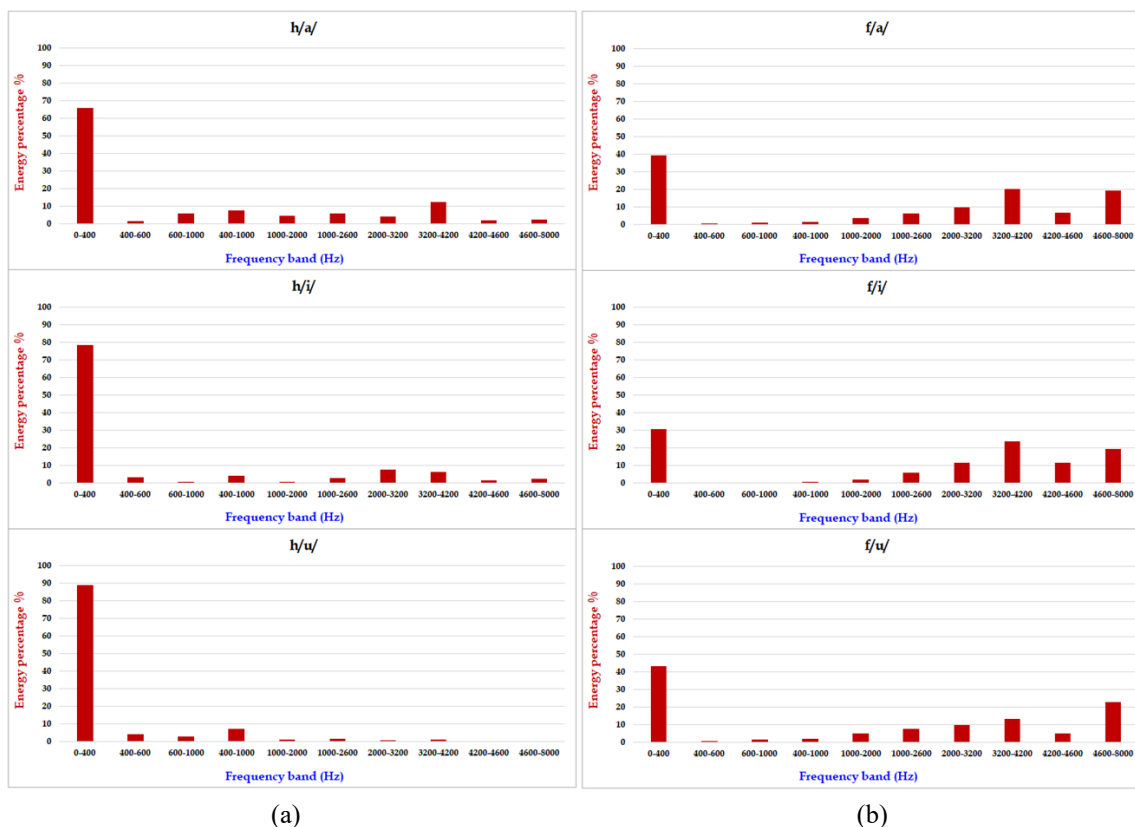


Figure 7. Energy distribution in the middle of unvoiced consonants: (a) /h/ and (b) /f/

3.2. Non-sibilant fricative consonant classification algorithms

3.2.1. Classification of non-sibilant fricative consonants into voiced/unvoiced categories

Based on the results of acoustic analysis concerning the distribution of normalized energy, the algorithm for classifying non-sibilant fricative consonants into voiced and unvoiced categories was developed. Figure 8 illustrates this algorithm, functioning as follows: If the energy in the frequency band B2 (400-600 Hz) exceeds 5%, the consonant is identified as voiced. Otherwise, the consonant is recognized as unvoiced.

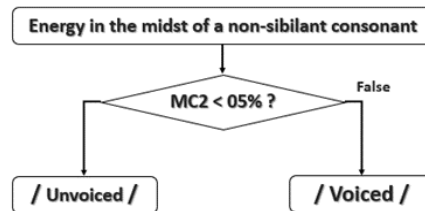


Figure 8. Algorithm for classifying non-sibilant fricatives

3.2.2. Classification of voiced non-sibilant fricative consonants

Acoustic analysis of voiced non-sibilant fricative consonants, followed by the vowels /a/ or /i/, led to the development of algorithms presented in Figure 9. For non-sibilant fricatives followed by the vowels /a/ or /i/ (Figure 9 (a)), if the energy in band B3 (600-1,000 Hz) exceeds 5%, the consonant is classified as /ʕ/. Otherwise, it is identified as /ɣ/. For voiced non-sibilant fricatives accompanied by the vowel /u/ (Figure 9 (b)), the algorithm operates as follows: if the energy in band B2 (400-600 Hz) exceeds 20%, the consonant is categorized as /ʕ/. Otherwise, the consonant is identified as /ɣ/.

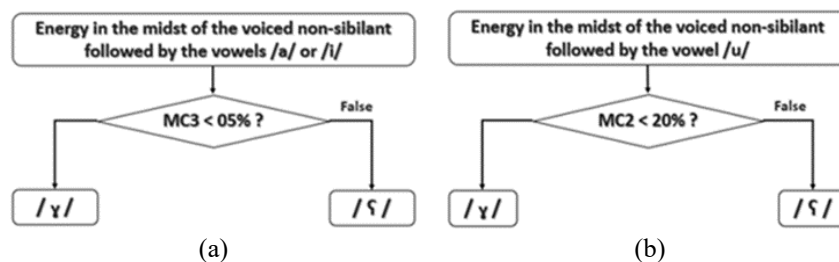


Figure 9. Algorithm for non-sibilant fricative consonants followed by (a) vowels /a/ and /i/ and (b) followed by the vowel /u/

3.2.3. Classification of unvoiced non-sibilant fricatives followed by the vowel /a/

The classification of unvoiced non-sibilant fricatives followed by the vowel /a/ was carried out based on acoustic analysis of the consonants /f/, /χ/, /h/, and /ħ/, leading to the algorithm presented in Figure 10: if the energy in band B6 (2,000-3,200 Hz) is greater than 10%, the consonant is identified as /ħ/. Otherwise, if the energy in band B4 (1,000-2,000 Hz) is greater than 5%, the consonant is classified as /χ/. If the sum of energy in bands B2 and B3 (400-1,000 Hz) is less than 5%, the consonant is identified as /f/. Otherwise, the consonant is recognized as /h/.

3.2.4. Classification of unvoiced non-sibilant fricatives followed by the vowels /i/ and /u/

For the consonants /f/, /χ/, /h/, and /ħ/ followed by the vowels /i/ and /u/, classification algorithms were established (in Figure 11) based on observed acoustic characteristics. The algorithm for the vowel /u/ is presented in Figure 11(a) and operates according to the following criteria: if the energy in band B9 (4,600-8,000 Hz) is less than 5%, the consonant is identified as /ħ/. Otherwise, if the energy in band B5 (1,000-2,600 Hz) is greater than 5%, the consonant is classified as /χ/. If the energy in band B9 is less than 10%, the consonant is identified as /h/; otherwise, it is recognized as /f/. Conversely, Figure 11(b) presents the algorithm for the vowel /i/: if the energy in band B7 (3,200-4,200 Hz) is less than 5%, the consonant is identified as /ħ/. Otherwise, if the energy in band B6 (2,000-3,200 Hz) is greater than 15%, the consonant

is classified as /h/. If the energy in band B4 (1,200-2,000 Hz) is less than 5%, the consonant is identified as /f/; otherwise, the consonant is recognized as /x/.

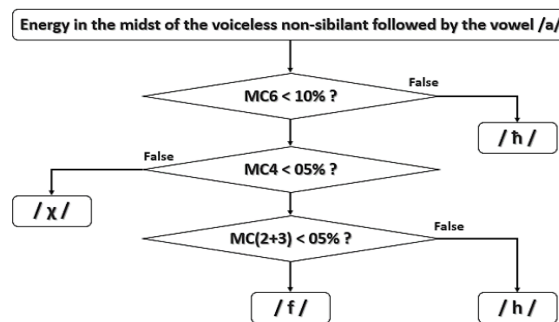


Figure 10. Algorithm for classifying unvoiced non-sibilant fricatives followed by the vowel /a/

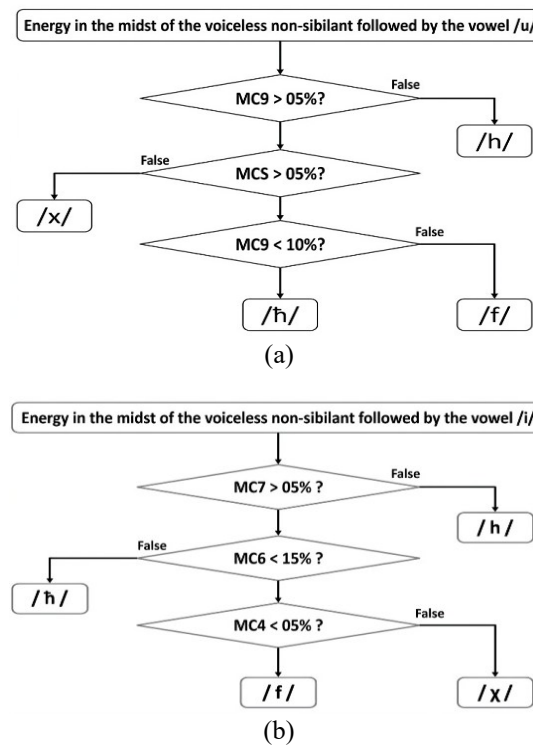


Figure 11. Algorithm for classifying unvoiced non-sibilant fricatives of (a) followed by vowel /u/ and (b) followed by vowel /i/

3.3. Comparative analysis of algorithm performance for classifying non-sibilant fricative consonants in Arabic

In this comparative study of the performance of the algorithm for classifying non-sibilant fricative consonants in Arabic examined three main approaches: our own method, a model based on ANN, and finally a model based on an SVM. To ensure the robustness of our models and to rigorously prevent overfitting, the evaluation of all classification algorithms (our proposed method, ANN, and SVM) was conducted using a 10-fold cross-validation approach. The total corpus of 4,320 samples was systematically partitioned so that, in each iteration, 90% of the data was utilized for training the models while 10% was reserved for testing. This guarantees that our reported accuracies reliably reflect the models' ability to generalize.

Furthermore, results of this study are highly competitive when compared to recent baseline architectures in the literature. Recent studies have demonstrated that classical baselines, such as Mel-frequency cepstral coefficients (MFCC) combined with multi-layer perceptron (MLP) ANN, achieve

recognition accuracies around 86% to 90% for Arabic letters. Similarly, convolutional neural networks (CNN) combined with rectified linear unit (ReLU) activation reach approximately 84% accuracy for spoken Arabic digits [43] and [44]. This proposed energy-band approach not only surpasses these benchmarks for certain non-sibilant categories but does so with a lighter computational footprint. This study results indicate that, in the majority of cases, algorithm demonstrated a higher classification rate than those based on neural networks and SVM machines. The different classification rates obtained are presented in Tables 1 to 3, demonstrating the effectiveness of approach in the classification of non-hissing fricative consonants in Arabic.

Table 1. Classification rate of the three algorithms for non-sibilant consonants

Method	Overall accuracy	Accuracy/voiced (%)	Accuracy/unvoiced (%)	Average error
Our method	89.57	90.50	88.65	0.20
ANN	89.60	87.20	90.80	0.20
SVM	86.41	66.80	96.20	0.12

Table 2. Classification rate of the three algorithms for voiced non-sibilant consonants

Following vowel	Method	Overall accuracy (%)	Accuracy /ɣ/ (%)	Accuracy /ʕ/ (%)	Average error
Consonants followed by /a/	Our method	95.73	97.20	93.50	0.06
	ANN	93.05	95.40	90.70	0.07
	SVM	93.98	99.10	88.90	0.05
Consonants followed by /i/	Our method	90.20	97.20	83.03	0.18
	ANN	89.80	94.40	85.20	0.17
	SVM	80.55	99.10	62.00	0.19
Consonants followed by /u/	Our method	85.00	99.07	71.30	0.21
	ANN	79.60	87.00	72.20	0.29
	SVM	77.31	78.70	75.90	0.19

Table 3. Classification rate of the three algorithms for unvoiced non-sibilant consonants

Following vowel	Method	Overall accuracy	Accuracy				Average error
			/f/ (%)	/x/ (%)	/h/ (%)	/h/ (%)	
Consonants followed by /a/	Our method	87.03	83.30	94.40	87.90	82.40	0.10
	ANN	87.03	83.30	94.40	91.70	78.70	0.09
	SVM	77.54	64.80	78.70	83.30	83.30	0.27
Consonants followed by /i/	Our method	84.50	76.00	89.80	82.40	89.80	0.09
	ANN	87.73	80.60	89.80	89.80	90.70	0.09
	SVM	83.33	71.30	85.20	84.30	92.60	0.27
Consonants followed by /u/	Our method	86.80	82.40	87.94	81.40	95.30	0.10
	ANN	86.11	88.90	86.10	82.40	87.00	0.10
	SVM	76.62	71.10	66.70	67.60	98.10	0.27

4. CONCLUSION

This article classifies Arabic non-sibilant fricative consonants (/f, x, ɣ, ʕ, h, and h/), followed by the three vowels (/a/, /i/, and /u/). The primary feature of the proposed method is the use of normalized energy, within the middle of the consonant, as an acoustic index in specific frequency bands. This study findings suggest that the energy contained within the vocal signal is a crucial element in sound characterization. The correct categorization rate exceeds 85%. This study anticipates that in future research, can leverage this algorithm in a comprehensive digital process for real-time detection and recognition of all Arabic language consonants. Finally, it is important to acknowledge that the dataset used in this research was recorded by Moroccan speakers. While dialectal variations pose a significant challenge in Arabic ASR, the core methodology of this study focuses on the fundamental CV syllable. Because our approach aims to characterize the isolated physical sound and energy distribution rather than semantic meaning, it captures the intrinsic articulatory properties of the Arabic non-sibilant fricatives regardless of the specific dialect. Testing this generalized acoustic baseline through cross-dialect validation will be a primary focus of our future work to build a fully speaker-independent Arabic ASR system.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to all speakers who participated in the recording of the speech corpus. We also acknowledge the support of the IMII Laboratory (Hassan First University of Settat) and LAPSSI Laboratory (Cadi Ayyad University of Safi) for providing the technical facilities required to conduct this research.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Youssef Elfahm	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Nesrine Abajaddi	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Badia Mounir	✓	✓		✓	✓	✓	✓			✓		✓	✓	
Laila Elmaazouzi		✓		✓	✓	✓				✓	✓	✓		
Ilham Mounir		✓		✓	✓	✓				✓		✓		
Abdelmajid Farchi	✓	✓		✓	✓		✓			✓		✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing-Original Draft

E : Writing-Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [YE], upon reasonable request.

REFERENCES

- [1] International Federation of Robotics, "Robot sales surge in Europe, Asia and the Americas," ifr.org. Accessed: Feb. 3, 2024. [Online]. Available: <https://ifr.org/ifr-press-releases/news/robot-sales-surge-in-europe-asia-and-the-americas>
- [2] J. Cantineau, *Cours de phonétique arabe*. Paris, France: Klincksieck, 1960.
- [3] A. A. Al-Nassir, *Sibawayh the phonologist: a critical study of the phonetic and phonological theory of Sibawayh*. London: United Kingdom: Kegan Paul International, 1985.
- [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*. Oxford, United Kingdom: Blackwell Publishers, 1996.
- [6] K. N. Stevens, "On the quantal nature of speech," *Journal of Phonetics*, vol. 17, no. 1–2, pp. 3–45, 1989, doi: 10.1016/s0095-4470(19)31520-7.
- [7] W. Jassem, "Peter Ladefoged, phonetic data analysis: an introduction to fieldwork and instrumental techniques," *Journal of the International Phonetic Association*, vol. 35, no. 1, pp. 113–115, 2005, doi: 10.1017/S0025100305001908.
- [8] 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, 1988, doi: 10.1016/s0095-4470(19)30504-2.
- [9] K. D. Malde, A. Chittora, and H. A. Patil, "Classification of fricatives using novel modulation spectrogram based features," in *Pattern Recognition and Machine Intelligence*, Berlin, Heidelberg: Springer, 2013, pp. 124–139, doi: 10.1007/978-3-642-45062-4_18.
- [10] K. N. Stevens and G. Weismer, "Acoustic phonetics," *The Journal of the Acoustical Society of America*, vol. 109, no. 1, pp. 17–18, 2001, doi: 10.1121/1.1327577.
- [11] C. H. Shadle and S. J. Mair, "Quantifying spectral characteristics of fricatives," in *International Conference on Spoken Language Processing, ICSLP, Proceedings*, 1996, vol. 3, pp. 1521–1524, doi: 10.1109/ICSLP.1996.607906.
- [12] B. McMurray and A. Jongman, "What information is necessary for speech categorization? harnessing variability in the speech signal by integrating cues computed relative to expectations," *Psychological Review*, vol. 118, no. 2, pp. 219–246, 2011, doi: 10.1037/a0022325.
- [13] L. M. T. Jesus and C. H. Shadle, "A parametric study of the spectral characteristics of European Portuguese fricatives," *Journal of Phonetics*, vol. 30, no. 3, pp. 437–464, 2002, doi: 10.1006/jpho.2002.0169.
- [14] L. M. T. Jesus and P. J. B. Jackson, "Frication and voicing classification," in *Computational Processing of the Portuguese Language*, Berlin, Germany: Springer, 2008, pp. 11–20, doi: 10.1007/978-3-540-85980-2_2.
- [15] P. F. Reidy, "Spectral dynamics of sibilant fricatives are contrastive and language specific," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2518–2529, 2016, doi: 10.1121/1.4964510.
- [16] G. W. Hughes and M. Halle, "Spectral properties of fricative consonants," *The Journal of the Acoustical Society of America*, vol. 28, no. 2, pp. 303–310, 1956, doi: 10.1121/1.1908271.
- [17] 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.

- [18] E. Nirgianaki, A. Chaida, and M. Fourakis, "Acoustic structure of fricative consonants in Greek," in *Proceeding of the 3rd Tutorial and Research Workshop on Experimental Linguistics / Proceedings Papers*, 2010, pp. 125–128, doi: 10.36505/ExLing-2010/03/32/000152.
- [19] P. Strevens, "Spectra of fricative noise in human speech," *Language and Speech*, vol. 3, no. 1, pp. 32–49, 1960, doi: 10.1177/002383096000300105.
- [20] W. Jassem, "The formants of fricative consonants," *Language and Speech*, vol. 8, no. 1, pp. 1–16, 1965, doi: 10.1177/002383096500800101.
- [21] J. M. Heinz and K. N. Stevens, "On the properties of voiceless fricative consonants," *The Journal of the Acoustical Society of America*, vol. 33, no. 5, pp. 589–596, 1961, doi: 10.1121/1.1908734.
- [22] S. L. 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.
- [23] A. Benamrane, "Étude acoustiques des fricatives de l'arabe standard (Locuteurs Algériens)," Ph.D. dissertation, Université de Strasbourg, Strasbourg, France, 2013.
- [24] P. Ladefoged, "Speech acoustics and phonetics: selected writings. By Gunnar Fant. (text, speech and language technology 24.) Dordrecht: Kluwer, 2006. pp. xiv, 322. ISBN 1402027893. \$64.95.," *Language*, vol. 83, no. 1, pp. 189–191, 2007, doi: 10.1353/lan.2007.0027.
- [25] O. Niebuhr and C. Meunier, "The phonetic manifestation of French /s#/ and /ʃs/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation," *Phonetica*, vol. 68, no. 3, pp. 133–160, 2011, doi: 10.1159/000331235.
- [26] K. Forrest, G. Weismer, P. Milenkovic, and R. N. Dougall, "Statistical analysis of word-initial voiceless obstruents: preliminary data," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 115–123, 1988, doi: 10.1121/1.396977.
- [27] G. Tomiak, "An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents," Ph.D. thesis, SUNY at Buffalo, University at Buffalo, Buffalo, United States, 1991.
- [28] 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, no. 1, pp. 57–94, 2001, doi: 10.1177/00238309010440010301.
- [29] J. Padgett and M. Zygis, "Evolution of sibilants in Polish and Russian," *ZAS Papers in Linguistics*, vol. 32, pp. 155–174, 2003, doi: 10.21248/zaspil.32.2003.190.
- [30] S. Funatsu and S. Kiritani, "Perceptual properties of Russians with Japanese fricatives," in *5th International Conference on Spoken Language Processing (ICSLP 1998)*, 1998, doi: 10.21437/icslp.1998-719.
- [31] S. Nittrouer, M. S. -Kennedy, and R. S. McGowan, "The emergence of phonetic segments," *Journal of Speech, Language, and Hearing Research*, vol. 32, no. 1, pp. 120–132, 1989, doi: 10.1044/jshr.3201.120.
- [32] K. Tjaden and G. S. Turner, "Spectral properties of fricatives in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 40, no. 6, pp. 1358–1372, 1997, doi: 10.1044/jslhr.4006.1358.
- [33] K. M. Newell and P. A. Hancock, "Forgotten moments: a note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions," *Journal of Motor Behavior*, vol. 16, no. 3, pp. 320–335, 1984, doi: 10.1080/00222895.1984.10735324.
- [34] G. Peeters, "A large set of audio features for sound description," *CUIDADO IST Project Report*, vol. 54, pp. 1–25, 2004.
- [35] D. H. McFarland, S. R. Baum, and C. Chabot, "Speech compensation to structural modifications of the oral cavity," *The Journal of the Acoustical Society of America*, vol. 100, no. 2, pp. 1093–1104, 1996, doi: 10.1121/1.416286.
- [36] M. A. Al-Khairy, "Acoustic characteristics of Arabic fricatives," Ph.D. thesis, Department of Linguistic, Graduate School of the University of Florida, Gainesville, United States, 2005.
- [37] 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*, vol. 12, no. 1, pp. 936–945, 2022, doi: 10.11591/ijece.v12i1.pp936-945.
- [38] P. Besse, "Apprentissage statistique," *Département Génie Mathématique et Modélisation Institut National des Sciences Appliquées de Toulouse, INSA*, 2006.
- [39] Z. Zhang, "A gentle introduction to artificial neural networks," *Annals of Translational Medicine*, vol. 4, no. 19, 2016, doi: 10.21037/atm.2016.06.20.
- [40] S. Dikaleh, D. Xiao, C. Felix, D. Mistry, and M. Andrea, "Introduction to neural networks," in *27th Annual International Conference on Computer Science and Software Engineering*, 2017, doi: 10.5555/3172795.317283.
- [41] H. Guo and W. Wang, "Granular support vector machine: a review," *Artificial Intelligence Review*, vol. 51, no. 1, pp. 19–32, 2019, doi: 10.1007/s10462-017-9555-5.
- [42] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.
- [43] S. Bhatia, A. Devi, R. I. Alsuwailam, and A. Mashat, "Convolutional neural network based real time Arabic speech recognition to Arabic braille for hearing and visually impaired," *Frontiers in Public Health*, vol. 10, 2022, doi: 10.3389/fpubh.2022.898355.
- [44] S. Haboussi, N. Oukas, T. Zerrouki, and H. Djettou, "Arabic speech recognition using neural networks: concepts, literature review and challenges," *Journal of Umm Al-Qura University for Applied Sciences*, pp. 462–484, 2025, doi: 10.1007/s43994-025-00213-w.

BIOGRAPHIES OF AUTHORS



Youssef Elfahm    received his M.Sc. degree in Automatic Control, Signal Processing, and Industrial Computing in 2017, and his Ph.D. degree in Electronics, Electrotechnics, and Automatics from Hassan First University, Settat, Morocco, in 2023. His doctoral dissertation focused on the development of a classification system for Arabic fricative consonants. Currently, he is a professor of Engineering Sciences at Lycée Mohammed VI d'Excellence (LM6E), Benguerir, Morocco. His research interests include speech recognition systems, speech production, signal processing, and artificial intelligence. He can be contacted at email: 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 Hassan First University 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 Techniques, Hassan First University. She can be contacted at email: n.abajaddi@uhp.ac.ma.



Badia Mounir    was born in Casablanca, Morocco in 1968. Engineer degree (1992) in "Automatic and Industrial Computing". The Mohammadia School of Engineering, Rabat, Morocco. Assistant professor at Graduate School of Technology, Cadi Ayyad University since 1992. Habilitated to supervise research (HDR) since 2007 and professor of higher education (PES) since 2017. Member of the Laboratory of Process Signals, Industrial Systems, Informatic (LAPSSII) Laboratory. Her research interests include speech recognition, signal processing, energy optimization and modeling. She can be contacted at email: mounirbadia2014@gmail.com.



Laila Elmazouzi    Ing Ph.D. in Telecommunication and Networks. Habilitated to supervise research (HDR) at High School of Technology-Cadi Ayyad University. Member of the LAPSSII Laboratory (Laboratory of Process Signals, Industrial Systems, Informatics). Research interests: telecommunication, signal processing, emotion recognition, machine learning. She can be contacted at email: elmazouzi2001@yahoo.fr.



Ilham Mounir    Ph.D. in Applied Mathematics, habilitated to supervise research (HDR) at the High School of Technology-Cadi Ayyad University. Member of the LAPSSII Laboratory (Laboratory of Process, Signals, Industrial Systems, Informatics). Research interests: applied mathematics, signal processing, emotion recognition, speech recognition and energy: optimization and modeling. She can be contacted at email: ilhamounir@gmail.com.



Abdelmajid Farchi    Ing Ph.D. in Electrical Engineering and Telecommunications, chief of the research team (Signals and Systems) in the Laboratory of Engineering, Industrial Management and Innovation. Educational person responsible for the Electrical Systems and Embedded Systems Engineering Cycle of the Faculty of Sciences and Techniques, Hassan First University, Settat, Morocco. He can be contacted at email: abdelmajid.farchi@gmail.com.