Classification of the Fricative and Occlusive Consonants According to the Place and the Mode of Articulation

Soufyane Mounir*, Karim Tahiry and Abdelmajid Farchi
*Corresponding author email id: soufyanemounir@gmail.com

I. INTRODUCTION

Several methods can be adopted to improve the speech recognition rate. Among these methods are the extraction of characteristics that are characterized by observation vectors determined by time methods such as linear predictive coding (LPC) or Mel Frequency Cepstral Coding (MFCC). The feature extraction phase is a very important factor in the development of a recognition system [1, 2, 3].

Nishinuma studied the French language where he tried in his research to define protocols for detecting consonant clusters based on temporal size. He used bisyllabic and trisyllabic words where he inserted target syllables CCV, VCC, CV and VC associated with the vowels /i, a, â/ [4].

Using statistical analysis, he came to retrieve five parameters which we quote relate voicing, manner of articulation of the fricatives namely: spectral moments, F2onset frequency, locus equation, slope of the spectrum, location of spectral peaks, measurement of static and dynamic amplitudes and the duration of the noise were based on discrete Fourier transforms [22 23 24]. They concluded that there is no invariance in the acoustic signal and, therefore, the categorization of speech by the listeners requires a massive integration of signals as well as mechanisms of compensation able to manage the contextual influences.

Spinus and Lilley were interested in the classification of fricatives. For this, they examined two methods. From a corpus of Romanian fricatives and for the coding of speech, the first method is based on the comparison of two acoustic measurements: the spectral moments and the cepstral coefficients. For the second method, they aimed at extracting measurements in segment areas after comparing two techniques of their determination [25]. For the first method, Spinu and Lilley divided the phonetic segments into three zones of almost equal duration, while in the second method, they used hidden Markov models (HMM) to break each segment into three regions. For the 2nd method, they aimed at extracting measurements in segments areas after comparing two techniques for their determination [25]. About the 1st method, Spinu and Lilley divided the phonetic segments into three zones of almost equal duration, whereas in the second method, they used models of Markov hidden (HMM) to decompose every segment into three regions of such kind to minimize the variances of the measures in every region. Having classified fricatives according to the place of articulation, the harmonization, the state of palatalization and the sex by using the logistic regression, they found relevant results at the level of the use of the cepstral coefficients which are more reliable than the spectral moments at the level of the classification. On the other hand, they ended that the use of zones identified by HMM possesses a rate of classification higher than the use of regions of equal duration.

In our study we try to classify the occlusive and fricatives consonants in standard modern Arabic language for three places of articulation: (bilabial, alveolar or dental, velar) respectively by means of spectral moments: spectral mean (m1), standard deviation (2), skewness (3) and kurtosis (4). We also try to make a comparison between these two modes of articulation.

II. CORPUS

Articulatory data were collected for fifteen Moroccan men, by pronouncing the CV syllable on four occasions,
where \( C \) is the consonant and \( V \) the vowel is. The concerned consonants are (bilabial: /b/ = /v/, alveolar or dental consonant: /d/ = /l/, velar (k / /j/ / for occlusives) and (bilabial: /f/ = /v/, interdental consonant: /θ/ = /s/, velar: /γ/ = /k/ for fricatives). For the vowels, we used the short vowels /a, i, u/.

In an isolated room and using "Praat" software, we used for the recording a microphone (Labtec AM-232, sensitivity: 35 dB, Impedance: 2.2 kOhm, bandwidth: 20a 8500 Hz) at 20 cm on a PC. With a frequency of sampling of 22050 Hz, the sound is directly scanned on a PC. We used the same software to segment the syllables CV.

### III. SPEECH SIGNAL PROCESSING

#### A. Pretreatment

The pretreatment of the voice signal for the automatic speech recognition of the word is a compression of the data to facilitate a real time estimation. The estimation itself can be made in the temporal domain or on the result of an analysis court-term made by the pretreatment. That will be useless to deal all of the signal (word / not word), for it we need to isolate the vocal activity by using a combination of two techniques: energy level and the passage by zero.

#### B. Preemphasis

We meet a problem of decrease of amplitude in the spectrogram, for it we have to accentuate the speech x (n) by calculating the magnitude x’(n) = x(n) - αx(n - 1). It is the filter which serves to amplify high frequencies. More \( α \) est grand, more the magnitude is raised in high frequency.

In our experience, we chose \( α = 0, 95 \) obtained from the following formula \( α = e^{(-2\pi 100 \pi f)} \).

#### C. Windowing and FFT

Before extracting the parameters of the speech signal, it is essential to break it down into segments because it is of a non-stationary nature. By multiplying each segment by a Hamming window, we succeed in weakening the discontinuities at the ends. This window is given by the following equation:

\[
x_1(n) = x_2(n). (0.54 + 0.46 \cos \left( \frac{2\pi n}{N - 1} \right))
\]

Where N is the number of samples.

The FFT step transforms the speech signal into a frequency domain [2]:

\[
X_n = \sum_{k=0}^{N-1} x_k e^{-\frac{2\pi ink}{N}}, n = \{0, 1, ..., N - 1\}
\]

#### D. Calculating Spectral Moments

In our work, we are interested in the method of spectral moments after transforming the signal into a frequency domain. The four spectral moments concerned are: spectral mean \( m_1 \), standard deviation \( m_2 \), skewness \( m_3 \) and kurtosis \( m_4 \).

The spectral mean, the standard deviation, the skewness and the kurtosis are respectively given by:

\[
m_1 = \frac{\sum_{i=0}^{n/2} f_i P(f_i)}{\sum_{i=0}^{n/2} P(f_i)}
\]

\[
m_2 = \left( \frac{\sum_{i=0}^{n/2} (f_i - m_1)^2 P(f_i)}{\sum_{i=0}^{n/2} P(f_i)} \right)
\]

\[
m_3 = \left( \frac{\sum_{i=0}^{n/2} (f_i - m_1)^3 P(f_i)}{\sum_{i=0}^{n/2} P(f_i)} \right)
\]

\[
m_4 = \left( \frac{\sum_{i=0}^{n/2} (f_i - m_1)^4 P(f_i)}{\sum_{i=0}^{n/2} P(f_i)} \right)
\]

### IV. RESULTS AND DISCUSSIONS

The value of "m1" of the consonant / k / is greater than / b / and / d /.

The standard deviation \( \sigma / / f / / \) has the largest value followed by / b / then by / γ / (Figure 1, 2). From these results, we find that the spectral mean allows to classify the velar occlusive consonants, on the other hand, it allows to classify all the places of articulation of the fricative consonants studied. We also find that in the case of occlusive consonants, when the consonant is produced at the level of the posterior cavity, the spectral mean is greater than that of the consonant produced at the level of the anterior oral cavity. So here we are talking about the size of this organ. These latter results are in harmony with what is found by Nitrouer and Stevens who stated that the oral cavity and the spectral mean are dependent [27 28]. For the standard deviation, the values of the alveolar occlusive consonants are larger than those of the velar than those of the bilabial ones. As for the fricative consonants, also the interdentals are the most dispersed followed by the bilabial and the velar.

The third spectral moment provides information on the location of the spectrum compared to the normal distribution. All occlusive consonants have positive values, which shows that their spectrum is offset to the left, that the totality of the acoustic energy is contained in the low frequencies. More precisely, the velar consonants are the closest to the axis of symmetry of the normal distribution, unlike the bilabial consonants. On the other hand, at the level of consonants fricatives, we find two opposite (opposed) signs, the spectrum of interdental consonants is moved to the right of the axis of symmetry, but it remains closer, on the other hand, two other places of joints (articulations) possess spectrum where the maximal energy is contained in the low frequencies where the spectrum of the bilabial stays farthest of this axis.

The results obtained from the last moment for the occlusive consonants show that the spectral distribution of the bilabial is the narrowest seen that it corresponds to the biggest value of the kurtosis, on the other hand that of the velar is the most flattened. For fricatives, the spectrum of the velar is the most flattened with compared with the other places of articulation. We can say that according to the place of articulation. The moments "m1", "m2" and "m3" allow to classify occlusive consonants and fricatives consonants.
Fig. 1. The four spectral moments of the occlusive consonants for three places in context CV

Fig. 2. The four spectral moments of the fricative consonants for three places in context CV
We notice that at the level of the spectral mean, the bilabial and interdental fricatives present the biggest values that those of the occlusive consonant bilabial and alveolar (dental consonants) respectively (Figure 3). On the other hand, the velar fricatives present a spectral mean less important than that of the velar occlusive consonants. About the second moment, the spectral distributions of bilabial and dental occlusive consonant are less dispersed than those of the bilabials and interdental consonants fricatives respectively, but no difference is noticed about velar occlusive and velar fricative.

Concerning the third moment, the spectral distributions of bilabial and velar occlusives and fricatives carry the totality of energy in the low frequencies, but the first ones are the most offset compared to the second ones. The big difference is noted about the dentals and interdentals where they have a right-hand asymmetry (negative value), while the total energy of the spectral distributions of the dentals is contained in the low frequencies and this can be due to the little difference in the place of articulation.

Remarkable results at the fourth moment showed that bilabial, interdental, and velar spectrum are narrower than those of bilabial, alveolar and velar occlusive respectively.

We can say that "m1" makes it possible to differentiate between the occlusive consonants and the fricative consonants according to the place of articulation. The first moment is less important in bilabial and alveolar occlusive compared to bilabial and interdental fricative. On the other hand, it is more important in velar occlusive than in fricatives. The moment "m4" makes it possible to classify the occlusive consonants in front of the fricative consonants whatever the place of articulation studied since the values of the two moments of the occlusive for all these places are more important than those of the fricatives.

V. CONCLUSION

This study, based on the standard modern Arabic language, where we treated two different modes of articulation (occlusive and fricative) for three different sites of articulation (bilabial, dental / interdental, velar), revealed several interesting results. "m1", "m3" and "m4" can classify consonants according to the place of articulation for a single mode of articulation. The moment "m4" makes it possible to distinguish between the occlusives and the fricatives irrespective of the place of articulation.

REFERENCES


by integrating cues computed relative to

–

n of cepstral

[27]

[26]

[25]

[24]

[23]

[22]

[21]

[20]

[17]

[16]

[15]

[14]

[13]

[12]

[11]

[10]

[9]

[8]

[7]

[6]

[5]

[4]

[3]

[2]

[1]

Acoust. Soc. Am. Vol. 97, pp. 520

production at di

Nittrouer, S., 1995, "Children learn separate aspects of speech

anticipatory e

Feng, Y., Hao, G, J., Xue, S, A., Max, L, 2011, "Detecting

Romanian fricatives

coefficients and spectral moments in the classification of

Spinu, L., & Lilley, J., 2016, «

Journal of the International Phonetic Association, 42(3), 281

affricates and their motivation: Evidence from Czech and Polish


1

spectral place cues using multitaper techniques


Research, 56, 1175

from adolescents


Tomiaik, G., 1990, « An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents (Doctoral dissertation) », State University of New York, Buffalo


