International Research Journal of Mathematics, Engineering and IT
Vol. 2, Issue 10, Oct 2015 IF- 2.868 ISSN: (2349-0322)
© Associated Asia Research Foundation (AARF)
Website: www.aarf.asia Email : editor@aarf.asia , editoraarf@gmail.com

PERFORMANCE OF VOICE ACTIVITY DETECTION METHOD BASED ON ZERO CROSSING RATE AND ENERGY LEVEL IN ARABIC LANGUAGE

Abdelmajid Farchi[1,2], Soufyane Mounir[3], Badia Mounir[4], Jamal Elabbadi[1]

[1]Laboratory of electronics and communication, School Mohammadia of Engineers of Rabat / University Mohamed V, Morocco

[2]Laboratory of mechanical engineering, industrial management and innovation, science and technical faculty of Settat / University Hassan 1er, Morocco

[3]Laboratory of mechanical engineering, industrial management and innovation, National School of Applied Sciences of Khouribga / University Hassan 1er, Morocco

[4]Laboratory of mechanical engineering, industrial management and innovation, graduate school of technology of Safi / University Cadi Ayyad, Morocco

ABSTRACT

This work investigates the detection of voice activity of /CVCVCV/ word for /b, d, k/ introducing vowel /a, u, i/ in Modern Standard Arabic (MSA) using the Zero Crossing Rate (ZCR) and Energy Level algorithm. This algorithm has allowed us to identify with good accuracy the beginning and end of words studied.

KEYWORDS - Modern Arabic Standard, Voice Activity Detection, Zero Crossing Rate, Energy Level, Performance Rate.

1. INTRODUCTION

The Voice Activity Detection allows us to distinguish between segments of an audio signal that include the human voice (period of activity) and non-voice signals (period of non-activity) in the environment of noise, then determine the start and end points of the operation. [1], [2]. Generally, the feature parameters used for endpoint detection are highly sensitive to the environment. Figure.1 represents an example of activity and no activity.

Research has shown the existence of more than half of the errors in the speech recognition caused by inaccurate detection of the end point even in the ideal conditions.[3]. For this, researchers have devoted their work on the determination of the beginning and the end of the word with exactly offering different algorithms.

J. Li & al., adopted a method based on TEO in a noisy environment. It uses three-state transitions, and a judgment mechanism based on double thresholds and the results obtained by performing a comparison with two other endpoint detection algorithms showed the robustness of this algorithm. [2]. J. Wu & X. Zhang presented an algorithm based on statistical models and empirical rules based on an energy detection algorithm through two steps: detection of the parameters characterizing the speech by using the algorithm Detection energy, and offering a Gaussian mixture model to align the endings of their optimal positions. The results obtained show better performance in various noisy scenarios [4]. As the formant structure occurs on the spectrogram, this is called the voice print, Wu & al, used the band spectral entropy (BSE) to trace these characteristics [5]. Another method of adaptive band selection (RABS) is combined with BSE to generate a new parameter called (ABSE). The results show that this parameter is very reliable in various noisy conditions [5]. In this article, we proposed a voice activity detection algorithm based on the Zero Crossing Rate (ZCR) and Energy Level to distinguish the active part of the non-active part of the speech signal in the case of standard Arabic for three places of articulation: bilabial, alveolar and velar. Our study is to make a comparison between these three places by calculating the reliability rate.

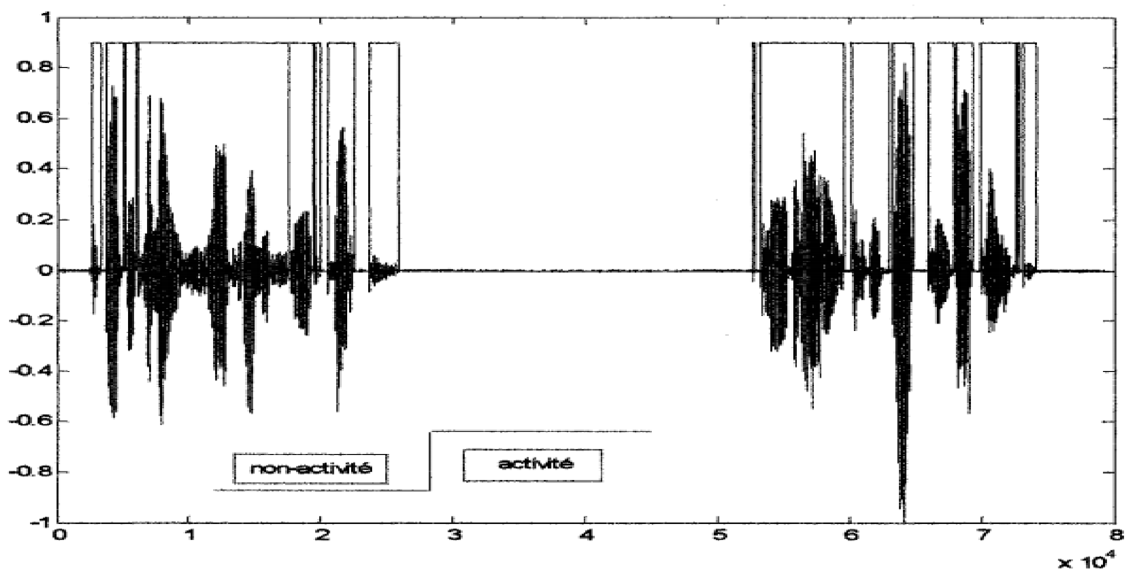

Fig. 1: Example of activity and non-activity [1]

2. ZERO CROSSING RATE AND ENERGY LEVEL

2.1. ZERO CROSSING RATE

For a sampled signal, there is zero crossing when two successive samples have opposite signs [6]. The short-term zero-crossing rate is estimated by the formula:

$$Z_n = \sum_{i=1}^{N} sgn[x(i)].sgn[x(i+1)]$$

With:

$$sgn[x(m)] = \begin{cases} 1, & if\ x(m) \geq 0 \\ 0, & if\ x(m) < 0 \end{cases}$$

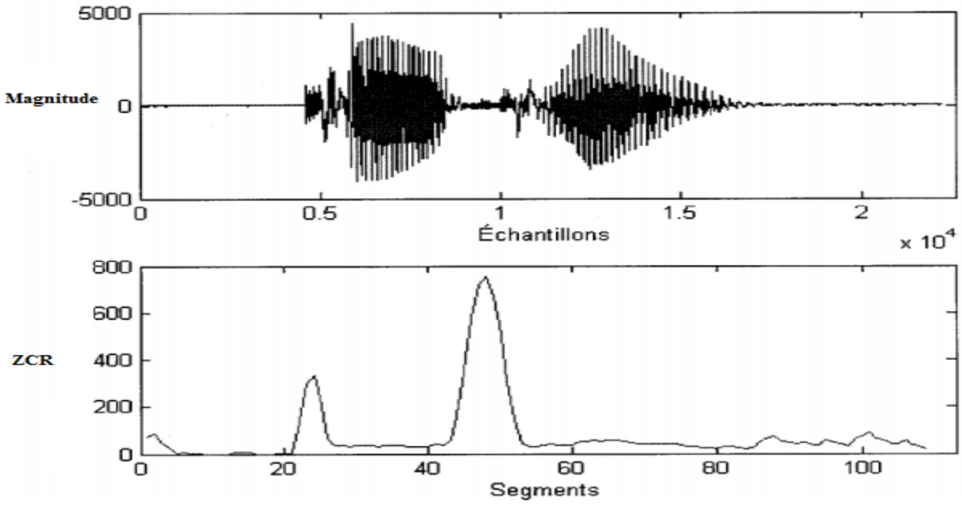

Fig. 2: Zero Crossing Rate of a speech signal

A characteristic for zero crossing rate is that it is high for the unvoiced sound and low for the voiced sound. The zero crossing rate is an important tool to classify voiced / unvoiced and to detect the beginning and end of the word in a speech signal (figure.3).

2.2. ENERGY LEVEL

One of the tools to provide a faithful representation of changes in the amplitude of the voice signal x (n) over time is energy short term [6]. In general, the energy of the frame of a signal is given by:

$$E_n = \frac{1}{N} \sum_{i=1}^{N} [x(i)]^2, 1 \leq i \leq N$$

$w(n)$: Hamming window

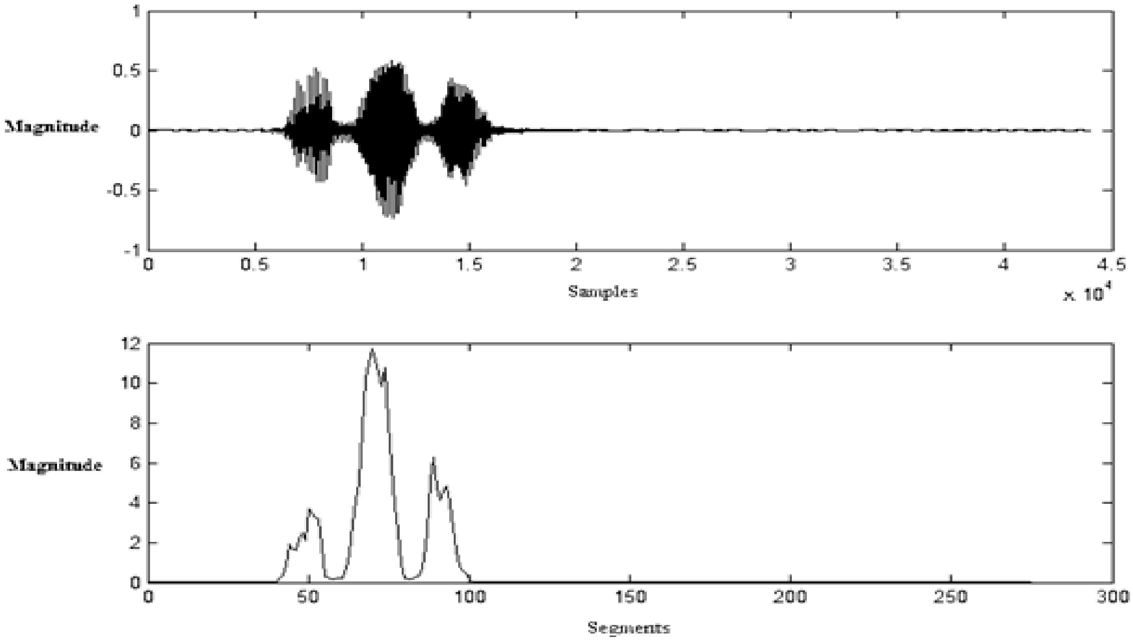

Fig. 3: Short term energy of a speech signal

3. METHODOLOGY

3.1. CORPUS

Four Moroccan adult speakers (men) speaking Modern Standard Arabic has been invited to pronounce a series of words CVCVCV (C: consonant V vowel) four times to three different places of articulation (/ b /: Bilabial; / d /: Alveolar, / k /: velar) with three short vowels (/ a, i, u /). The recording was made using a microphone (AM-232 Labtec; Sensitivity: -35dB, Impedance: 2.2 kOhm, bandwidth: 20-8500 Hz) at a distance of 20 cm in an isolated and quiet room via the software "Praat". The sound is digitized directly to a PC with a sampling frequency of 22050 Hz because the maximum possible frequency is 11025 Hz beyond this frequency, the signal is extremely poorly sampled and the resulting sound is unusable. The quantization used is a 16-bit linear quantization to reduce the quantization error. The recording time is 2 seconds for each syllable. The results are obtained by applying the recordings to a program made by us in matlab.

3.2. ALGORITHM

In this study, we proposed a voice activity detection algorithm based on the zero crossing rate and energy shown in Figure 3 by using high and low thresholds of zero crossing rate and energy level.

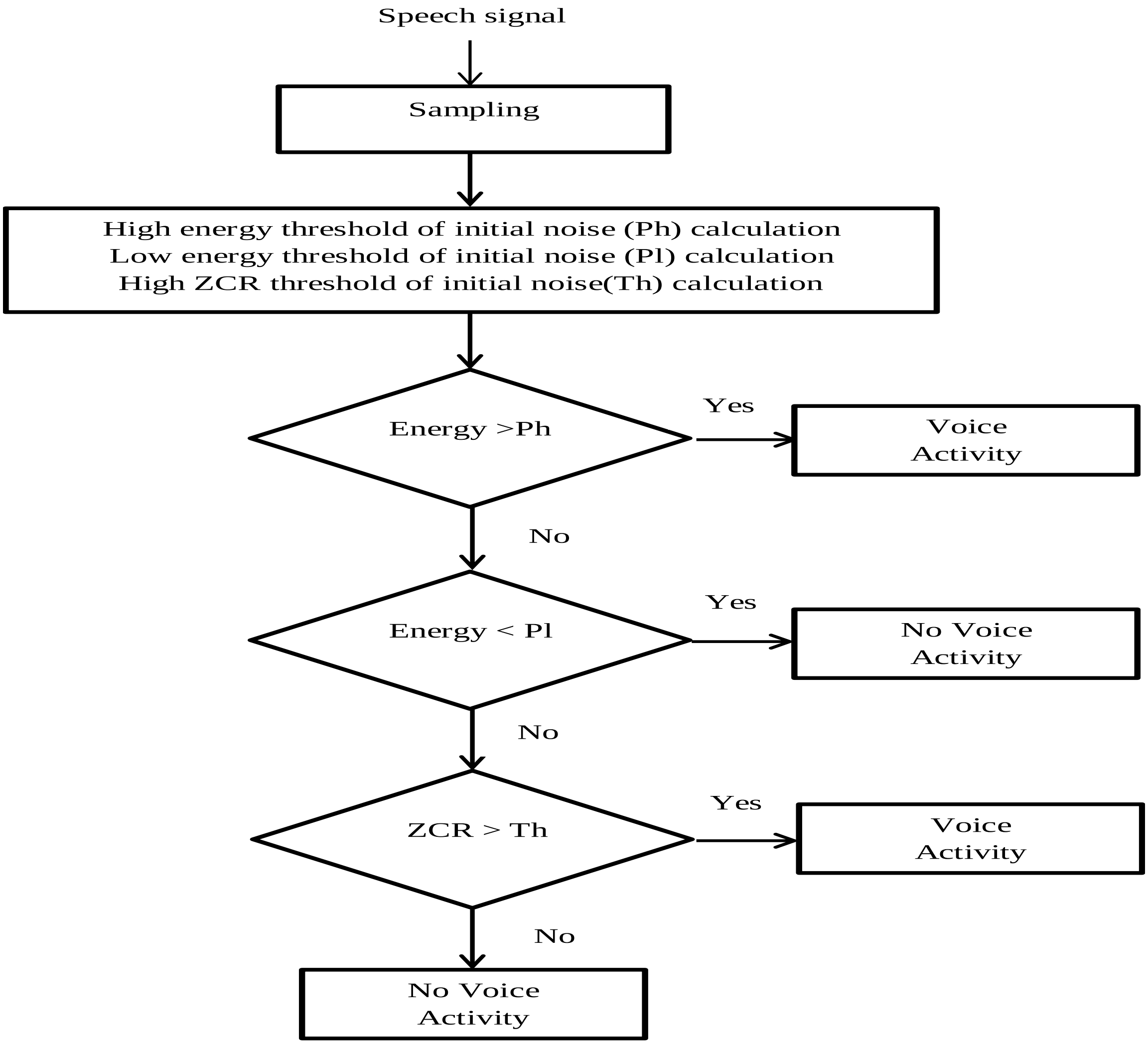
Speech signal
Sampling
High energy threshold of initial noise (Ph) calculation
Low energy threshold of initial noise (Pl) calculation
High ZCR threshold of initial noise(Th) calculation
Energy >Ph
Yes
Voice Activity
No
Energy < Pl
Yes
No Voice Activity
No
ZCR > Th
Yes
Voice Activity
No
No Voice Activity

Figure 4 : VAD algorithm based ZCR

4. RESULTS AND DISCUSSIONS

We realized a program in Matlab code based on this algorithm. The results obtained showed that the performance rate for the alveolar (90%) is larger than bilabial (76%) than velar (59%) (Figure 5).

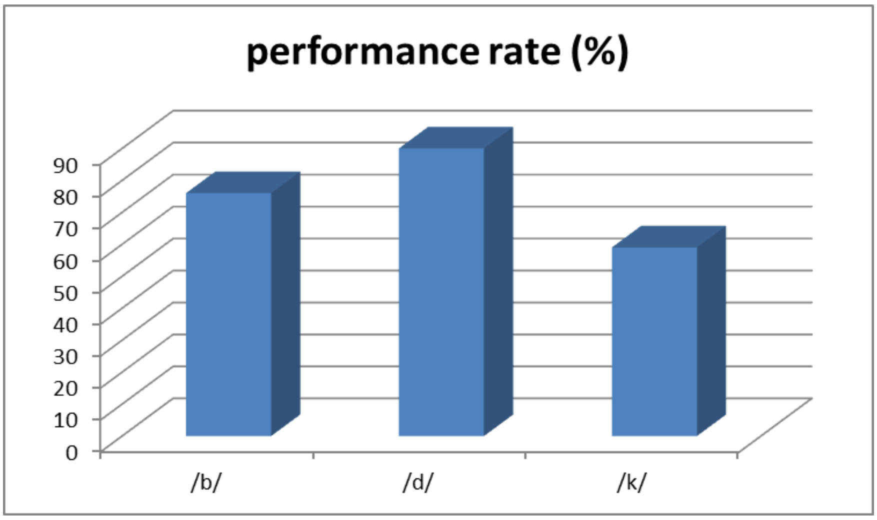

Figure 5: Performance rate of the algorithm ZCR/energy level for / b /, / d / and / k /

Figure.6 indicates the result of our algorithm, it shows accurately the beginning and the end of the speech signal in the case of /bababa/.

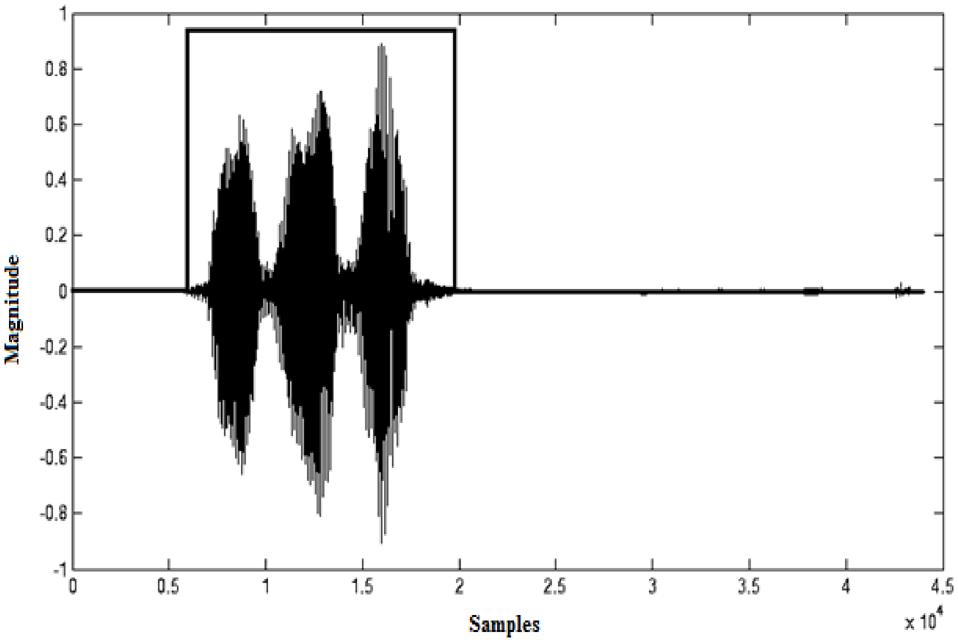

Figure 6: Voice Activity for /bababa/

Figure 7 shows the results of VAD / kakaka /. It is clear that there is difficulty in determining the vocal part of speech and this is due to the energy of the consonant / k / which is very low and confused noise.

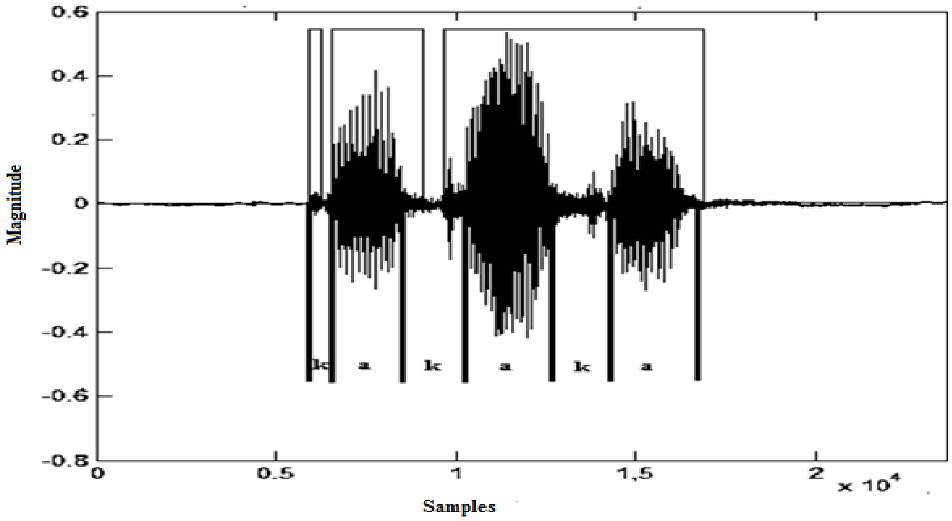

Figure 7: Voice Activity for /kakaka/

According spectrograms words / dadada / and / kakaka /, we note that the energy level of the consonant / d / is higher than / k /: the gray level of the consonant / d / is dense than / k / (Figure 8 and Figure 9).

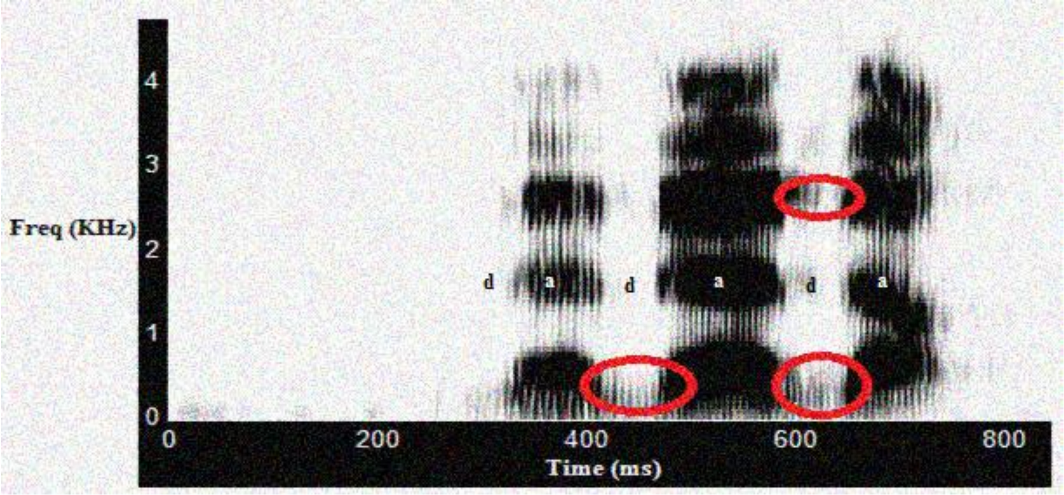

Figure 8: Spectrogram of /dadada/

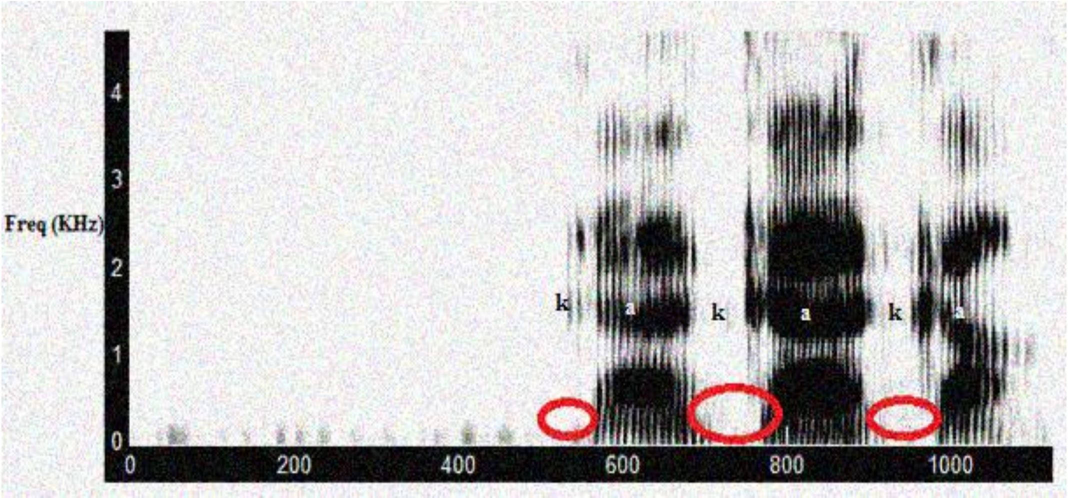

Figure 9: Spectrogram of /kakaka/

These results are consistent with those obtained in the work of Mounir & al [7], where this is explained by the phenomenon of coarticulation in the work of Munir et al, where they showed that the application of the equation of the locus CV context allowed on the one hand, to identify the place of articulation of the consonants according to their virtual locus (bilabials ≈ 1200 ≈ 1800 and velar alveolar ≈ 2600). On the other hand, the locus equation indicates that the slopes of velar have the largest degree of coarticulation and alveolar have the smaller. This result indicates that the intervention of the language in the production of the consonant is inversely proportional to the degree of coarticulation, and therefore the energy of the consonant / k / is less important compared to other consonants / b, d /.This means the energy level is nearer to that of the noise, and it is considered by the algorithm as no voice activity.

5. CONCLUSION

In this work, the results obtained showed a good performance of the algorithm ZCR / energy level for the alveolar, but less performance for velar where we noticed that this algorithm considers that there is no voice activity during the pronunciation of the consonant / k / for most recordings. This is explained by the fact that the energy level of the velar close to that of the noise. This low power is shown in the work of Mounir & al [7].

REFERENCES

[1] S. Robidas, *Comparaison de methods pour la détection d'activité vocale à bande large sous différents bruits*, doctoral diss., School of Engineering and Information Technology, Ottawa, Canada, 2006.

[2] L. Jie, Z. Ping, J. Xinxing, D. Zhiran, *Speech Endpoint Detection Method Based on TEO in Noisy Environment*, Procedia Engineering, *29*, 2012, 2655-2660.

[3] J. Zhang, F. Jiang, H. LIU, *Study on endpoint detection based on multi-characteristic jointed in noisy environment*, Computer Engineering and Application, *45*, 2009, 114-116.

[4] J. Wu, X. Zhang, *An efficient voice activity detection algorithm by combining statistical model and energy detection*, EURASIP Journal on Advances in Signal Processing, 2011.

[5] B.F. Wu, K.C. Wang, *Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments*, IEEE Trans. Speech Audio Process, *13*, 2005, 762-775.

[6] L.R. Rabiner, M.R. Sambur, *An algorithm for determining the endpoints of isolated utterances*, Bell System Technical Journal, *54*, 1975, 297-315.

[7] S. Mounir, I.Mounir, B. Mounir, A. Farchi, Z. Hachkar, *Anticipatory and carry-over effects of vowel to vowel coarticulation in Arabic language*, International Journal of Research in Computer Engineering and Electronics, *1(* 3), 1975, 1-4.