Contribution to Time and Frequency Analysis of Irregular Sleep Snoring

Mohamed Rezki, and Abdennour Alimohad

Abstract—**The purpose of this paper is to give a summary analysis of human snoring and its episodes. In particular, we consider an acute snoring. In order to extract some frequency information of snoring signal, we apply the Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT) algorithms, Discrete Wavelet Technique, and Power Spectral Density (PSD). Once irregular snoring characterized, we use a Voice Activity Detection (VAD) for snoring episode detection. Furthermore, we give comparative study of three types of thresholds that can control the VAD approach, a fixed threshold, a soft threshold, and a Gaussian threshold. Next, we use a Perceptual Evaluation of Speech Quality (PESQ) method to evaluate the efficiency of the VAD. We find that VAD based on Gaussian threshold is better.**

Index Terms—**Sleep Snoring; Frequency Analysis; Time Analysis; Thresholds; Snore quality measures.**

Original Research Paper
DOI: 10.7251/ELS1923067R

I. Introduction

Snoring is an inspiratory noise caused by vibration of the soft supper-airway tissues, mainly soft palate and posterior faucial pillars [1]. Mostly, snoring can provoke obstructive sleep apnea with the possibilities of higher risks of cardiovascular disease such as heart attacks, strokes, sleep disorders, etc. [2]. With the development of the signal processing tools, many fields of study have been opened and thus non-stationary and complicated signals such as irregular snoring can be analyzed. In addition, one of the most famous processing techniques is the use of wavelet because it can express signals with different frequency components [3].

Also, it offers statistical analysis tools such as histograms. The wavelet-based histograms provide us the ability to naturally extend to multidimensional case, a good optimization [4] and to improve the accuracy substantially over random sampling [5].

Several studies have been done on detection and analysis of sleep snoring (acoustical characterization) based on the classification methods using some experimental databases through the choice of selected features [6]—[8]. Furthermore, other authors treated the sleep snoring by the application of personalized software. Statistical parameters such as the number of the snoring episodes , duration , etc. were calculated in [9]-[10].

However, all these works require a complete characterization with integration of aspects of time and frequency at the same time. In the present work we establish a global model based on time and frequency analysis with integration of the VAD (Voice Activity Detection). In addition, we show the effect of changing the VAD threshold on the efficiency of detection of snoring episodes.

The remainder of the paper is organized as follows: In section 2, we present the different processing frequency techniques used to analyze the sleep snoring, Section 3 is devoted to describe the theoretical and practical approach for time analysis of sleeps snoring. Finally, we give a summary conclusion in section 4.

Manuscript received 28July 2019. Received in revised form 9 October 2019 and 2 November 2019.

Mohamed Rezki, is with Department of Electrical Engineering, Faculty of Applied Technology, Bouira University, Algeria. E-mail: mohamedrezki197@yahoo.fr

Abdennour Alimohad is with Department of Electrical Engineering, Faculty of Applied Technology, Bouira University, Algeria, and Research Laboratory in Electrical Engineering and Automatic LREA, University of MEDEA, Algeria. E-mail: alimohad@msn.com

II. Theoretical Approach for Frequency Analysis

The signal recorded in [11] is considered the strongest snoring signal characterized also by its high irregularity. So we can consider it as an obstructive sleep apnea (OSA) signal, which contains an apnea part (see Fig. 1 [8]).

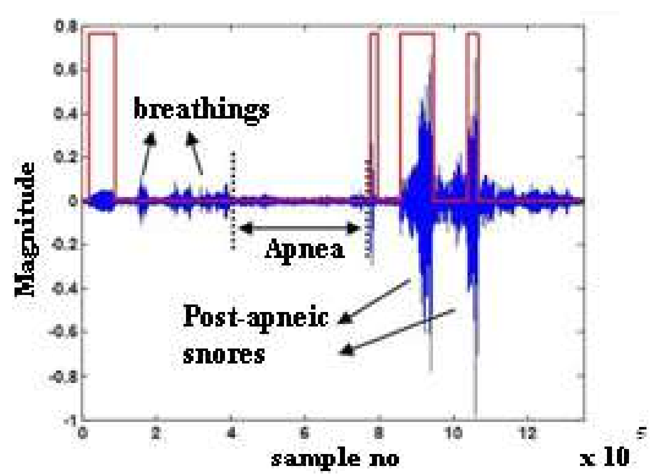


Fig. 1. Detection of snoring episodes for the OSA patient having both regular and post-apneic snoring sounds (rectangular pulse in red represents sound segments). [8].

Fig.1 shows four breathing cycles separated by four snoring periods. And we observe that when the breathings become much attenuated during a long time it becomes an episode of apnea. After the apnea episode the snoring becomes quite intense and very irregular what is called post- apneic episodes.

A. Processing techniques

A.1. Fast Fourier Transform

The Fast Fourier Transform (FFT) was derived from the Discrete Fourier Transform (DFT), which is given by [12]:

$$y_k = \sum_{n=0}^{N-1} \omega_N^{nk} x_n \quad (1)$$

where K = 0,…, N-1, ω_N is the primitive root of unity

$\exp\left(\frac{-2\pi i}{N}\right)$, and N is the size of the input x.

A.2. Short Time Fourier Transform (STFT)

Short Time Fourier Transform (STFT) gives spectrograms that are created either by approximation as a filter Bank or from Fourier transform calculation. The continuous –time STFT for a given time signal x(t) is written as:

$$STFT(x(t)) = X(\tau, f) = \int_{-\infty}^{+\infty} x(t).w(t-\tau).e^{-j2\pi ft}dt \quad (2)$$

where w is the windowing function and f is the frequency. In contrast with the Fourier transform techniques which give the frequency part, the STFT allows to operate in frequency and time.

A.3. Wavelet Transform

The Wavelet Transform is a particular technique that describes signals in both time and frequency domains**.** Its physical form is a small window (mother wavelet) which is used to scan a macro signal [3]. The wavelet equation as a function of mother wavelet is given by:

$$\varphi_{a,b}(t) = |a|^{-\frac{1}{2}} \varphi\left(\frac{t-b}{a}\right) \quad (3)$$

where a, b are, respectively, the dilation and translation parameters, and $\varphi(t)$ is called the single mother wavelet.

In literature, there exits two large groups of wavelets: continuous and discrete. The Continuous Wavelet Transform (CWT) for a given function $f(t)$ is defined as [3]:

$$W_f(a,b) = |a|^{-\frac{1}{2}} \int_R f(t) \overline{\varphi\left(\frac{t-b}{a}\right)} dt \quad a,b \in R; a \neq 0 \quad (4)$$

where $\overline{\varphi\left(\frac{t-b}{a}\right)}$ denotes the complex conjugate of the wavelet function $\varphi\left(\frac{t-b}{a}\right)$.

Next, the Discrete Wavelet Transform (DWT) is a sampling wavelet using a bank of filters. It is given by [13]:

$$W(l,s) = 2^{\frac{s}{2}} \sum_n x(n)\Psi(2^s n - l) \quad (5)$$

where n=1, 2,…N and N is the total number of samples, l describes the shifting and s is the scale.

Examples of the discrete mother wavelets are presented in Fig. 2.

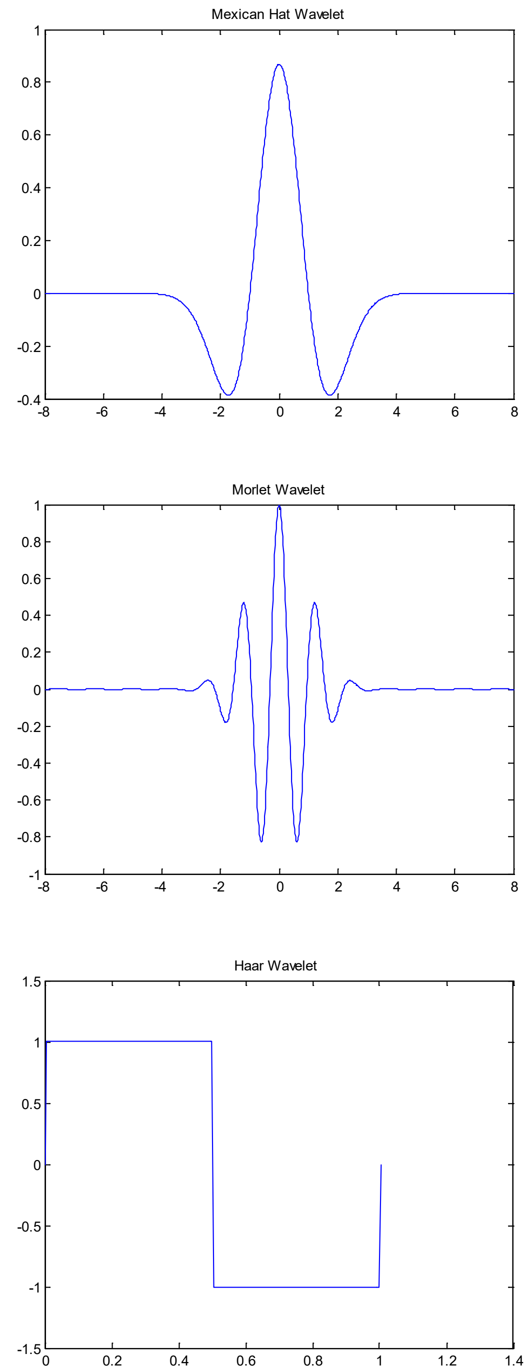


Fig. 2. Discrete mother wavelets examples.

A.4. Spectral analysis

The power spectral density (PSD) estimation has many applications such as the elimination of the wide-band noise mixed to the useful signal. PSD allows also, seeing the real distribution of the signal power depending on its frequency. In other word, the power spectral density represents the power content of a signal in an imperceptible frequency band contrary to the power spectrum that shows the frequencies which contain the signal's power. In order to estimate the power spectral density, we use the Welch's method that is described as follows [14]-[15]:

First, the original data is split up into overlapped data segments. Second, a window signal is applied to each segment.

Noticing that the windowing operation is what makes the Welch method a "modified" periodogram.

B. Practical aspects of snoring signal frequency analysis

Our experiments were performed on data given by [11], which consists of mp3 sounds. The data were recorded in a high quality format with a sampling frequency of 44100 Hz.

B.1. FFT technique

The application of FFT algorithm to a snoring signal is given in the following figure.

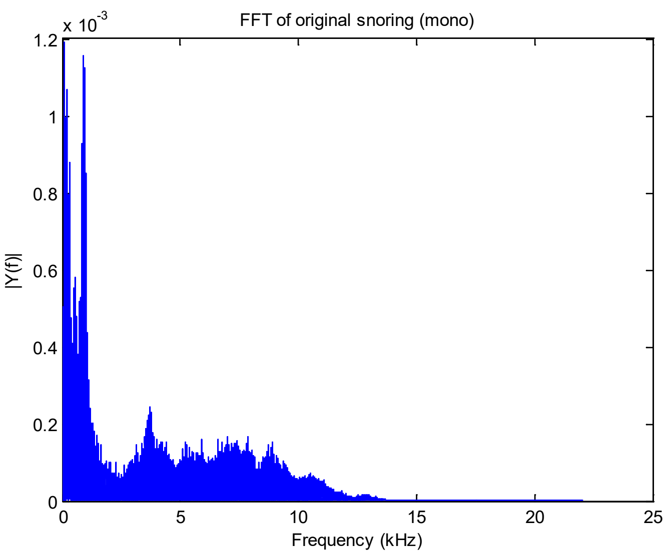


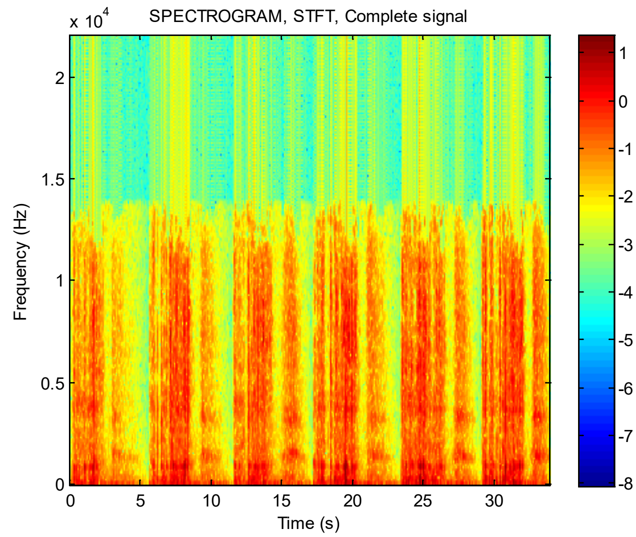
Fig. 3. FFT of original snoring.

As we can see, Fig.3 displays only the positive half of the frequency spectrum and discards the redundant negative half. Therefore, the spectral component extends to half the sampling frequency (around 22 kHz).

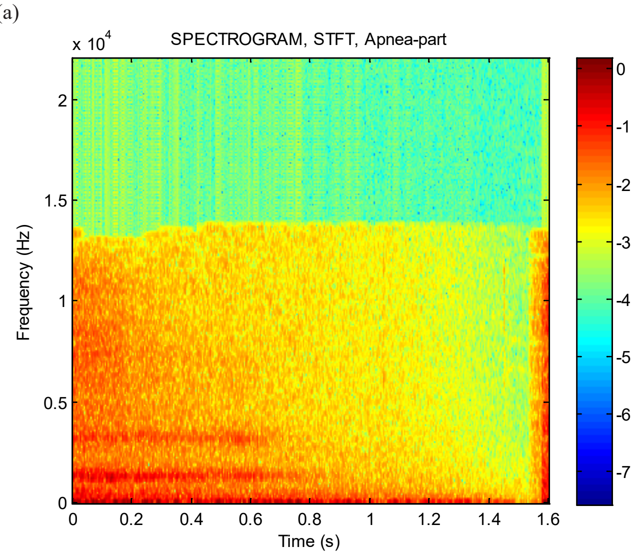
We can deduce from Fig. 3 that the main part of the snoring signal reaches 14 kHz and the FFT signal has two highest picks at 0.06 kHz and 0.88 kHz.

B.2. STFT technique

In this part we compute the STFT for the snoring signal and the apnea part, which are presented, respectively, in Fig.4 (a) and Fig. 4 (b).



(a)



(b)

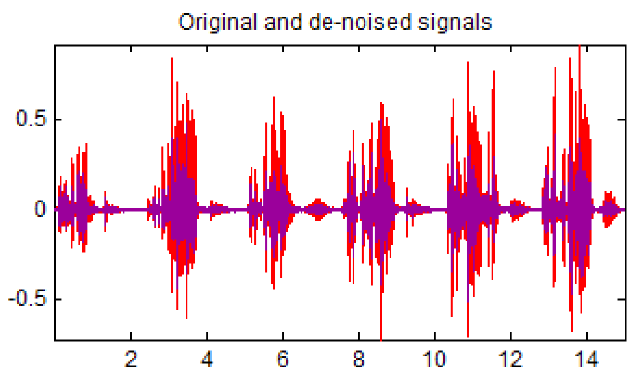
Fig. 4. STFT of original snoring: (a) Complete signal, (b) Apnea-part of snoring signal.

It's clear from Fig.4(b) that the STFT of the apnea-part has greatly lower magnitude then the STFT of the complete signal (Fig 4.a). This result confirms the fact that the apnea part follows a quite audible snore. And the presence of apnea is related to the existence of a strongly snore (see Fig. 1).

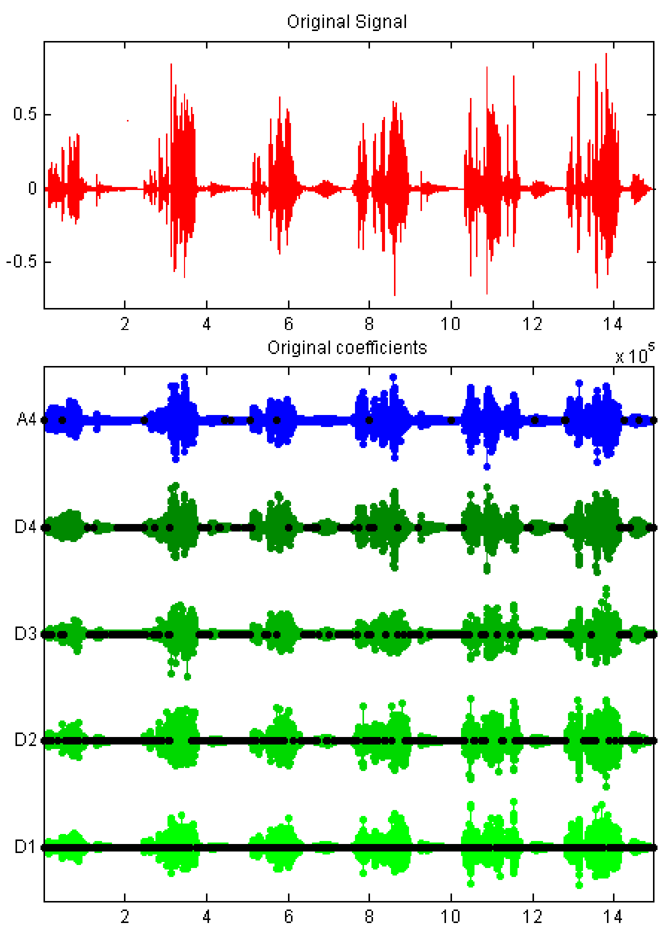
B.3. Discrete Wavelet Transform technique

Here, we analyze the snoring signal using the Discrete Wavelet Transform with four level Haar wavelet. This level presents the advantage of having minimal coefficients of decomposition (A4, D1, D2, D3, D4).

In practice, we find that over 4 levels, the signals will have a different shape compared to the original signal. So, we compute 4 levels of decomposition and the obtained results are shown in Fig. 5.



(a) Loaded and de-noised signals (red & purple colors respectively)



(b) Original details coefficients

Fig. 5. Wavelet decomposition of the snoring signal and its de-noised operation.

The outputs “**A**” and “**D**” are the reconstruction wavelet coefficients [16]:

- The approximation output A: is the low frequency content of the input signal component.
- The multidimensional output D: gives the details or the high frequency components of the input signal at various levels.

From Fig.5 (a), we noticed that the wavelet de-noising of the signal gives a nuanced result with attenuation of amplitude and a slight decimation of the fluctuations.

In order to derive statistical information from the wavelet analysis, a histogram is used. The representation of this histogram is given in Fig. 6.

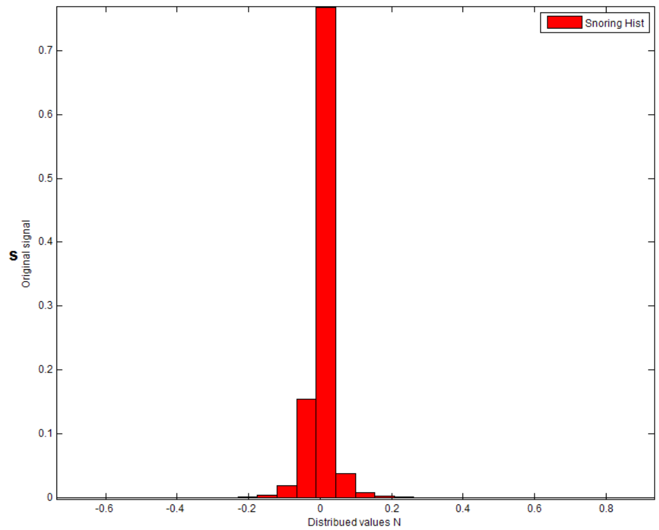


Fig. 6. Histogram representation of the snoring signal.

From Fig. 6, we observe that the distribution of the positive and negative values is not the same (considering the non-stationary character of the signal). Also, the major concentration of the values (almost 85 %) is around zero.

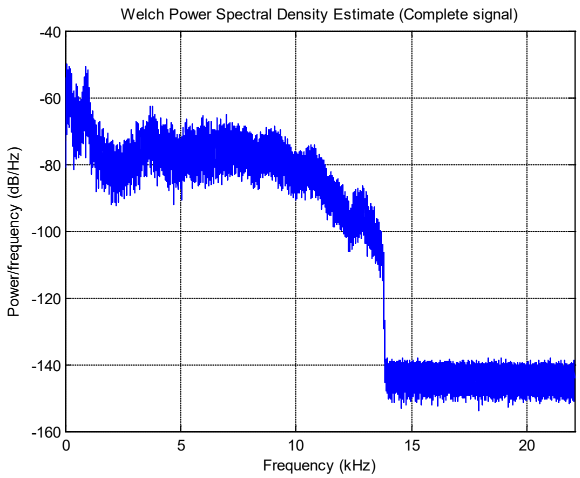
Finally, we noticed that the shape of the histogram reflects the Gaussian nature of the signal distribution which will be exploited in section III.

B.4. Power Spectral Density technique and Spectral Envelope

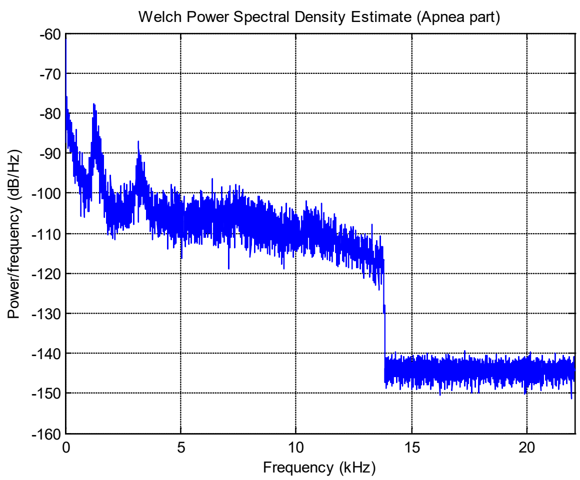
The computation of the Power Spectral Density (PSD) is realized using the Welch’s method. Therefore, the PSD of the snore signal is shown in Fig. 7.

The Fig. 7 shows that the smooth character of the complete snoring signal is better than the apnea part, and the most energy of the signal (complete or partial) is concentrated in the frequency interval [0 14] (kHz). We can also see the existence of some peaks such as at 1.25 kHz and at 3.2 kHz. These peaks are important because they help to locate formants of an acoustic signal. These formants are used, as parameters, in many applications such as acoustic synthesis and speech recognition. In addition, we can see the existence of an abrupt decrease in energy around the frequency 14 kHz (exactly 13.8 kHz). This is due to the non-regular nature of the snoring signal.

Next, to confirm the results of the power spectral density technique, we apply the spectral envelope method that represents the spectral magnitude versus the frequency. And gives an envelope to the spectrum by linking the peaks. In Fig. 8, we show the spectral envelope of the snoring signal.

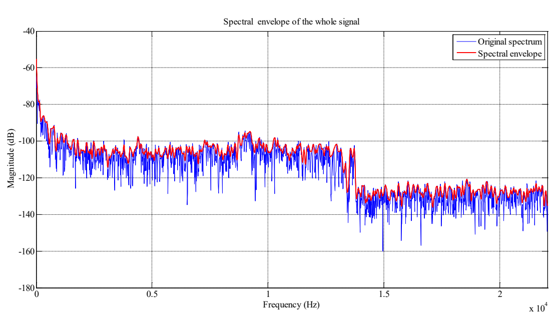


(a)

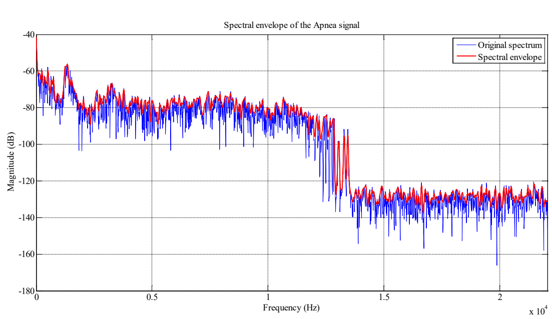


(b)

Fig. 7. Welch power spectral density estimation of: (a) Complete signal, (b) Apnea-part of snoring signal



(a)



(b)

Fig. 8. The spectral envelope of: (a) Complete signal, (b) Apnea-part of snoring signal

The results shown in Fig.8 are similar to those of Fig. 7, and we have the same collapse of power around the frequency 14 kHz.

III. Theoretical Approach of Time Analysis

The International Telecommunication Union (ITU) had imposed and coordinates some standards for telecommunications to better exploit the frequency band of speech. Among these standards, the silence elimination in speech keeps only the voice. This latter guarantees the effectiveness of transmission, because only speech activity is detected and treated.

The most used technique for silence elimination is VAD (Voice Activity Detection), it consists of detecting the presence or absence of human speech.

In our case, we consider the snoring signal as the voice and the long delay between two episodes of snore as an apnea sequence.

Various VAD algorithms have been developed. These algorithms used an energy threshold to separate the presence or absence of voice. In this work, we propose to introduce the VAD technique in our system to detect apnea episodes in the snoring signal. This is shown in the following figure (Fig. 9).

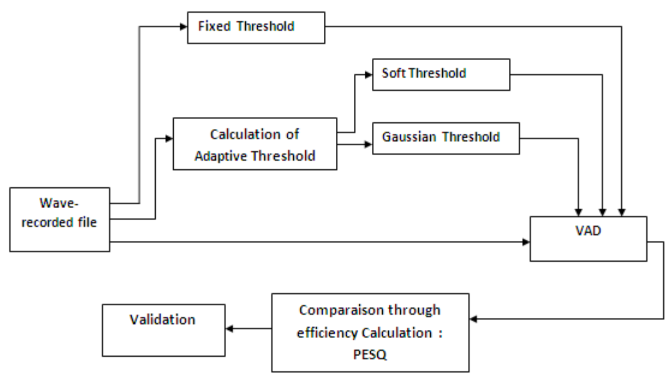


Fig. 9. Flow chart of the proposed system

We have applied three (03) different thresholds. First, we consider a fixed threshold with an experimentally value of 0.05. Then, we use an adaptive threshold which varies according to each frame of the snoring signal. For this latter threshold, we have selected two (02) techniques, a soft threshold and a Gaussian threshold.

The soft threshold was given by D.L. Donoho in [17] as:

$$tw2 = k.w_{mean}.(\sigma_w) \tag{6}$$

w_{mean} is a parameter depending on the frames number, the energy of each frame and the zero crossing number. k is a constant which can be determined empirically. $\sigma_w{}^2$ is the variance. Finally, the Gaussian threshold is expressed as follows:

$$tw3 = \frac{FWHM}{2.\sqrt{2}\,\log(2)} \tag{7}$$

where $FWHM$ is the Full Width at Half Maximum, which expressed the difference between the two extreme values of the Gaussian distribution. Since we have shown the Gaussian nature of our snore signal (see Fig. 6), we apply this last threshold to each frame of the signal. Thresholds mean is then taken (Fig. 10).

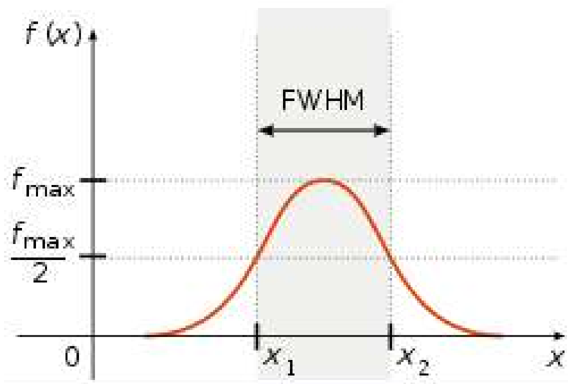


Fig. 10. Shape of a Full width at half maximum

As an example, Fig.11 presents the Gaussian distribution corresponding to the 10th frame.

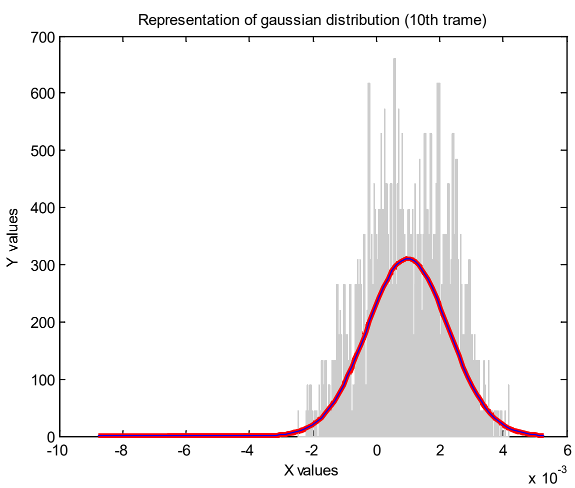
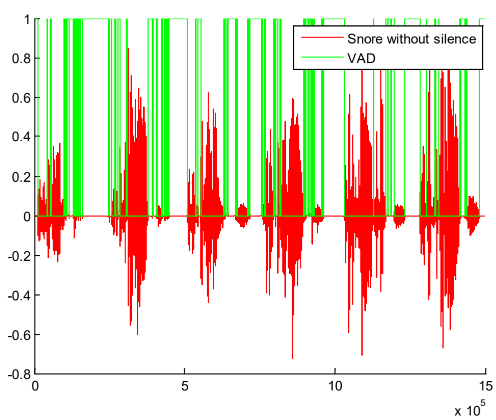
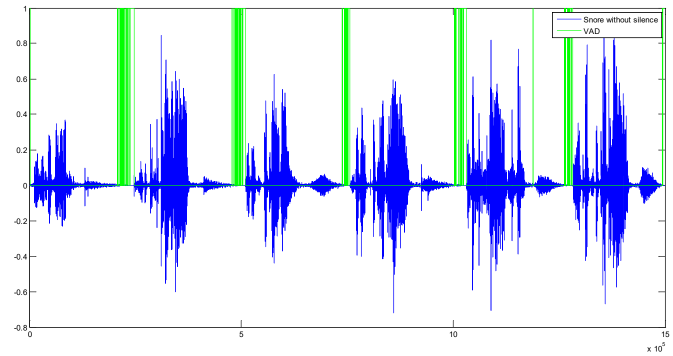


Fig. 11. FWHM of the 10th frame.

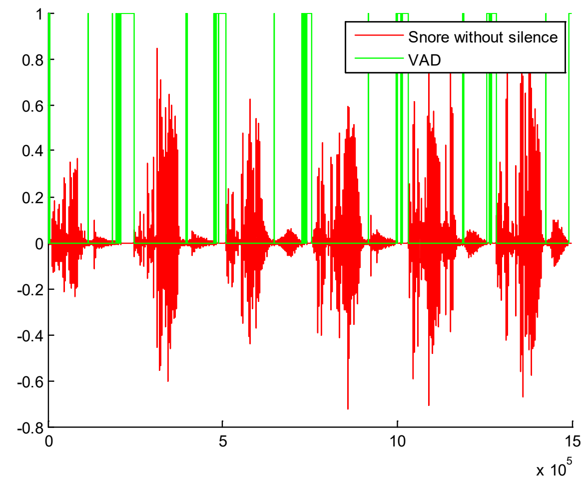
The application of the VAD technique based on the different thresholds, cited before, gives the following results:



(a)



(b)



(c)

Fig. 12. Detection of snoring episodes by applying the VAD with: (a) fixed threshold, (b) soft threshold, and (c) Gaussian threshold.

We remark from Fig.12 that the duration of the snoring episode depends on the VAD, and more precisely on the threshold level. Notice also that the decision of considering a silence as a sleep apnea episode depends on its duration, its repetition per hour, and a subject examination by a doctor.

A. Performance evaluation measure of the proposed system:

Measuring the quality of a sound signal is an important and a very difficult task. There are objective and subjective measures. These latter are based on the tests of direct listening (qualified normative auditors who hear). These measures are expensive and difficult to apply on a snoring signal. Therefore, they have been replaced by objective methods [18].

Several objective speech quality measures were evaluated such as: segmental Signal to Noise Ratio (segSNR), distortion (signal distortion, background distortion), PESQ (Perceptual Evaluation of Speech Quality) and others [19].

Perceptual Evaluation of Speech Quality, is a standardized objective method of measuring speech quality. A detailed description of PESQ can be found in ITU-T Recommendation P.862 (02/01) (standards for telecommunications and Information Communication Technology of the International Telecommunication Union 'ITU') [20]. The original and degraded signals are aligned in time to correct for delays, and then processed through an auditory transform to obtain the loudness spectra. The absolute difference between the degraded and original loudness spectra is used as a measure of audible error in the next stage of PESQ computation. The objective evaluation allows transformation of the reference and degraded signal to an internal representation resembling a psychophysical representation of audio signals in the human auditory system [21].

The final PESQ score is computed as a linear combination of the average symmetrical disturbance value d_{sym} and the average asymmetrical disturbance value d_{asym} as follows [22]:

$$PESQ = a_0 + a_1.d_{sym} + a_2.d_{asym} \tag{8}$$

where $a_0 = 4.5$, $a_1 = -0.1$ and $a_2 = -0.0309$.

We have chosen PESQ in order to show that the quality of a signal containing the information (speech, snoring, etc.) still remains acceptable after truncation following the use of the VAD. By applying the PESQ_MOS (wideband measurements) to our system with two sampling frequencies: 44100Hz (original recording frequency) and 16000 Hz (standard frequency for PESQ), we get the results shown in Table 1. Notice that the PESQ calculation was done using the software given in [19].

Table I

Performance Evaluation measure by PESQ

Sampling frequency (Hz)	PESQ (fixed threshold)	PESQ (soft threshold)	PESQ (Gaussian threshold)
44100	2.9706	4.0261	4.5000
16000	2.5648	3.9832	4.4095

According to Table I, the best PESQ score is given for the Gaussian threshold, this confirms the superiority of adaptive threshold compared to the fixed threshold.

IV. Conclusion

In this work, we analyzed an irregular sleep snoring signal using several methods in time and frequency. The results in frequency domain were obtained by applying the Fast Fourier Transform (FFT), the Short Time Fourier Transform (STFT), the Discrete Wavelet Transform (DWT), and Power Spectral Density (PSD). Which show the existence of some peaks that can be used to locate formants of the snoring signal and its frequency band. These results were also confirmed by using the spectral envelope method. Concerning the time domain, Three variants of VAD technique were used to detect the snoring/ apnea episodes. Where, the VAD uses respectively, a fixed threshold, an adaptive soft threshold, and a Gaussian threshold. In order to choose the best threshold, a performance study of the method was done through the use of two (02) criterias. Firstly by subjective criteria via a simple listening of the sound signal issued from the VAD. Secondly by calculating the PESQ (Perceptual Evaluation of Speech Quality) score which is an objective quality measure. Both methods have confirmed that the Gaussian threshold is the best.

References

[1] M. Partinen, "Chapter 18 - Epidemiology of sleep disorders", *Handbook of Clinical Neurology*, vol. 98, p.p. 275-314, 2011.

[2] (How to Stop Snoring: Causes, Aids, Remedies) Siamak N. Nabili. Available: https://www.medicinenet.com/snoring/article.htm, (Date accessed: 04 July 2019).

[3] Z. Yinhong, L.Quanlu, "The Study of the Snoring Signal Analysis Wavelet Transform", in *Recent Advances in Computer Science and Information*, Lecture Notes in Electrical Engineering, vol. 129. Springer-Verlag, Berlin Heidelberg New York, p.p.67-75, 2012.

[4] Y. Matias, J. S.Vitter, and M. Wang, "Dynamic Maintenance of Wavelt-Based Histograms", *Proceedings of the 26th International Conference on Very Large Data Bases, VLDB'00*, pp. 101-110, 2000.

[5] T. Li, Q. Li, S. Zhu, and M. Ogihara, "A Survey on Wavelet Applications in Data Mining", *SIGKDD Explorations*. vol.4, no 2, pp.49- 68, 2002.

[6] C. Wang , and J. Peng, "The Methods of Acoustical Analysis of Snoring for the Diagnosis of OSAHS", *Journal of Sleep Medicine & Disorders,* vol, no2, pp.1-7, 2017.

[7] M.M.D. De Meyer, W. Jacquet, O.M. Vanderveken, L.A.M. Marks, "Systematic review of the different aspects of primary snoring ", *Sleep Medicine Reviews,* p.p. 1-32, 2019.

[8] M. Cavusoglu, C.F. Poets, and M.S.Urschitz., "Acoustics of snoring and automatic snore sound detection in children", *Physiological Measurement*, vol.38, no.11, pp.1919-1938, 2017.

[9] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, T. Akcam, Y. Serinagaoglu, "SASA: A Software System for Sleep and Snoring Analsis", *HIBIT'07- international symposium on health informatics and bioinformatics,* pp.1-8, 2007.

[10] B. Calabrese, F. Pucci, M. Sturniolo, P.H. Guzzi, P. Veltri, A. Gambardella, M. Cannataro, " *A* System for the Analysis of Snore Signals", *International Conference on Computational Science, ICCS 2011, Procedia Computer Science,* vol.4, pp. 1101–1108, 2011.

[11] Salamisound: Sound Effect. Available: http://www.salamisound.com/2045645-heavy-snoring-very-irregular (Date accessed: 22 December 2017).

[12] S. G. Johnson, M. Frigo, "A Modified Split-Radix FFT With Fewer Arithmetic Operations", *IEEE Transactions on Signal Processing*, vol. 55, no.1, p.p.111-119, 2007.

[13] M. Aqil, A. Jbari, A. Bourouhou, , "ECG Signal Denoising by Discrete Wavelet Transform", *International Journal of Online and Biomedical Engineering (iJOE)*, vol.13, no.9,pp. 51-68, 2017.

[14] Spectral density – Using pwelch to a set of signals: some questions (Matlab). Available: https://stackoverflow.com/questions/35528353/using-pwelch-to-a-set-of-signals-some-questions-matlab, (Date accessed: 04 July 2019)

[15] P. Sysel, Z. Smekal, "Enhanced estimation of power spectral density of noise using the wavelet transform" in 2007, *IFIP International Federation for Information Processing*, vol. 245.Personal Wireless Communications, Boston: Springer, pp. 521-532, 2007.

[16] Discrete Wavelet Transform (DWT). Available: https://hvdc.ca/webhelp/Master_Library_Models/CSMF/Discrete_Wavelet_Transform_(DWT)/Discrete_Wavelet_Transform_(DWT).htm, (Date accessed: 07 July 2017).

[17] D. L. Donoho, "De-noising by soft-thresholding", *IEEE Transactions on Information Theory*,vol.41. pp. 613–627, 1995.

[18] M. Guéguin, "Evaluation objective de la qualité vocale en contexte de conversation (Doctoral thesis),", Lab. Traitement du signal et de l'image, Université Rennes 1, 2006.

[19] H. Yi , and P.C. Loizou , "Evaluation of objective measures for speech enhancement", *IEEE Transactions on Audio, Speech, and Language Processing*,vol.16, no.1, pp. 229-238, 2006.

[20] ITU-T Recommendation P.862. "Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-To-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs," February 2001.

[21] W. Zhang, Y. Chang, Y. Liu, and L. Xiao, “A New Method of Objective Speech Quality Assessment in Communication System”, *Journal of multimedia*, vol. 8, no. 3, pp.291-298, 2013.
[22] A.W. Rix, J. G. Beerends, M.P. Hollier and A.P. Hekstra, “Perceptual Evaluation of Speech Quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs”, *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP ‘01),* volume: 2, pp. 749-752, 2001