

# Denoising and Enhancement Speech Signal Using Wavelet

Meriane Brahim\*

University of Batna 2 (UB2), Advanced Electronics Laboratory  
Higher School of Technological Education Skikda ENSET, Algeria  
tlcom\_brahim@yahoo.fr

Received: 28/Oct/2020

Revised: 09/Feb/2021

Accepted: 31/Mar/2021

## Abstract

Speech enhancement aims to improve the quality and intelligibility of speech using various techniques and algorithms. The speech signal is always accompanied by background noise. The speech and communication processing systems must apply effective noise reduction techniques in order to extract the desired speech signal from its corrupted speech signal. In this project we study wavelet and wavelet transform, and the possibility of its employment in the processing and analysis of the speech signal in order to enhance the signal and remove noise of it. We will present different algorithms that depend on the wavelet transform and the mechanism to apply them in order to get rid of noise in the speech, and compare the results of the application of these algorithms with some traditional algorithms that are used to enhance the speech. The basic principles of the wavelike transform are presented as an alternative to the Fourier transform. Or immediate switching of the window The practical results obtained are based on processing a large database dedicated to speech bookmarks polluted with various noises in many SNRs. This article tends to be an extension of practical research to improve speech signal for hearing aid purposes. Also learn about the main frequency of letters and their uses in intelligent systems, such as voice control systems.

**Keywords:** Wavelet Transform; Speech Enhancement; Denoising, Discrete Wavelet Ttransforms (DWT); Noise Reduction in Speech Signals.

## 1- Introduction

With the advent of wavelet analysis, it became popular to address unstable physical quantities, such as speech analysis, voice signature detection, and speech recognition. Wavelets have proven successful in front-end speech recognition processors that are an alternative to instant switching using time-wave resolution. For speech recognition, or voice enhancement, wave shapes are based on Henning's window [1-3] The performance of the recognition depends on the frequency domain coverage. The goal of good speech recognition is to increase the wavelength bandwidth without significantly affecting time accuracy. This can be done by collecting the hard-to-detect white noise of the wave and removing it by conventional methods.

Speech components will have large values compared to noise because it is considered an extraneous signal. Transactions are calculated using a multi-precision wave filter bank.[4-13] The filter selection depends on the noise level and other parameters. To obtain a good noise reduction result, a good threshold level must be estimated [2-12].

The wavelet function and the level of decay also play an important and pioneering role in de-noise and noise-canceling signal quality. Recently, various wavelet-based methods have been proposed for the purpose of reducing noise in speech. The wavelength division modulus method is a noise reduction procedure to remove noise by reducing the wavelet coefficients in the wavelet field. The method depends on the threshold in which the signal is each wavelet.

## 2- Speech Signals

Speech is a natural and basic way for humans to convey message and thoughts. Speech frequency normally ranges between 3 Hz to 4 KHz depending upon the character. However the human beings have an audible frequency range of 20 Hz to 20 KHz. The most common problem in speech processing is the effect of meddling of noise in the speech signals. The noise masks the speech signal reduces the quality and the speech is greatly affected by presence of backdrop noise [1-11], Noise shrinking or speech enrichment algorithm is to improve the performance of communication systems when their input or output signals are corrupted by noise signal[14].

Non-stationary signal analysis methods are focused to model the inherent time-varying characteristics of the analyzed signals recorded in several areas namely, communications, speech analysis and synthesis, radar, biomedical, and mechanical engineering [15], The conventional methods based on the Fourier transform are not well suited for spectroscopy of these signals. Moreover, complex biological signals such as a brain signal or a speech signal [15-17].

### 3- Speech Acquisition With Cool Edit

The Cool Edit program for recording, auditory and visual analysis of sounds and their spectra

They will be recorded in monophony at the sampling frequency of 10 kHz with a converter 16 so called the sampling period.

$$T_e = \frac{1}{F_e} = \frac{1}{10 \times 10^3} = 0.1 \times 10^{-3} s \tag{1}$$

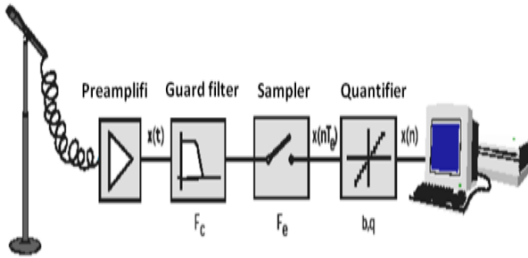


Fig.1 Block of Recording a speech signal

### 4- Wavelet Transform

The wavelet transform (WT) uses wavelet function and varied scales to decompose signals in the T-F domain, and it can guarantee the temporal and spectral resolutions in the entire frequency range. Since introduced in the 1970s, the WT has been used in varied applications, such as signal detection, imaging processing, de-noising of signal, speed improvement, audio classification, etc [2-8]. Zhu and Kim applied the Morlet wavelet transform to analyze the impact noise.

The wavelets have two important properties: first, the scaling factor, and secondly, the transformation and the relationship between them roughly correspond to the measurement process. Compressed waves are used. When the high-bandwidth waves span, they correspond to the low-frequency signals [14-17], at lower bands, it corresponds to rapidly changing signals that consist of

high frequencies. Unlike other transmission tools (Fourier transforms, etc.) used in signal processing, waves allow analysis of signals in both frequency and time domains. There are two types of wavelet transfers: continuous and discrete wavelet transfers. Both transformations are continuous in time (analog), and with their help analog signals can be represented. [5-6-16].

#### 4-1- General theory of CWT

In this work, we stated only some keys equations and concepts of wavelet transform, more rigorous mathematical treatment of this subject can be found in [3-6]. A continuous-time wavelet transform of f(t) is defined as:

$$CWT_{\psi} f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi^* \left( \frac{t-b}{a} \right) dt \tag{2}$$

Here  $a, b \in R, a \neq 0$  and they are dilating and translating coefficients, respectively. This multiplication of  $|a|^{-1/2}$  is for energy normalization purposes so that the transformed signal will have the same energy at every scale. The analysis function  $\Psi(t)$ , the so-called mother wavelet has to satisfy that it has a zero net area, which suggest that the transformation kernel of the wavelet transform is a compactly support function.[9]

A disadvantage of CWT is that the signal representation is often redundant, because a and b are continuous over R (the real number). As the original signal can be completely reconstructed by a model copy of  $W_f(b, a)$ . Usually, we try  $W_f(b, a)$  in a binary network i.e.,  $a = 2^{-m}$  and  $b = n2^m, m, n \in Z^+$ . Substituting the last one into

$$DWT_{\psi} f(a,b) = \int_{-\infty}^{+\infty} f(t) \psi^*(t) dt \tag{3}$$

where  $\Psi_{m,n}(t) = 2^{-m} \Psi(2^m t - n)$  is the dilated and translated version of the mother wavelet  $\Psi(t)$ . [9,14]

SNR is the power of the useful signal to the power of the noise ratio in which meaningful information characterizing the ratio of these power. SNR depended on an additive noise where the undistorted unquantized signal  $s[i]$  and an additional quantization error  $e[i]$  are superposition generated the quantized signal  $sq[i]$ . SNR is usually specified in the logarithmic measure in decibels (dB) in order to cover a wide range of possible SNR values: [10]

$$SNR_{dB} = 10 \log \left( \frac{P_x}{P_e} \right) = 20 \log \left( \frac{A_x}{A_e} \right) \tag{4}$$

Where  $P_x, P_e$  are the average powers of the corresponding signals, and  $A_x, A_e$  is the average value of the amplitude. SNR is often called SQNR

### 4-2- Temporal and Spectral Resolutions in the CWT

Resolutions in the time and frequency domains are critical for evaluation of performance of different wavelets. The temporal resolution in the time domain  $\sigma_t$  and the spectral resolution in the frequency domain  $\sigma_w$  of the CWT can be defined as [5,6]:

$$\sigma_t^2(\gamma) = \int_{-\infty}^{+\infty} (t - u\gamma)^2 |\phi_\gamma(t)|^2 dt \quad (5)$$

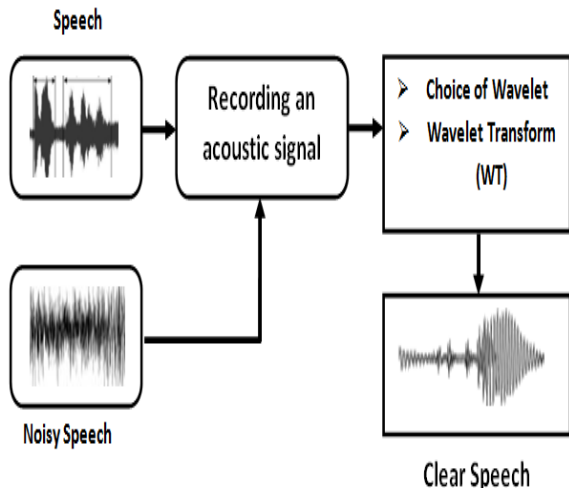
DWT is mathematical tool for decomposing data in a top down fashion. DWT represent a function in terms of a rough overall form, and a wide range of details. Despite of the requirement and type of function i.e., signals images etc. DWT offers sublime methodology and technique for representing the amount of detail present.

Wavelets perform and offer scale based analysis for a given data. A wide range of applications and usage has been found for these wavelets including signal processing, mathematics and numerical analysis and, for its better performance in signals and image processing it is considered an alternative to Fast Fourier Transform as DWT provide time frequency representation When there is a need for processing and analyzing non stationary tool, DWT can be used .Study shows that discrete wavelets transform have high performance in speech signal processing so far. [18-19].

### 5- Speech enhancement methods

There are various speech enhancement methods proposed for noise reduction and to improve the speech quality and clearness. Only one algorithm. is not enough for all the types of noise present in the surrounding. Hence speech enhancement algorithms are created based on the application she block diagram of speech enhancement is show in figure (2)

Fig. 2 Speech Enhancement Method



The functional diagram of the proposed method has been shown in Fig.3

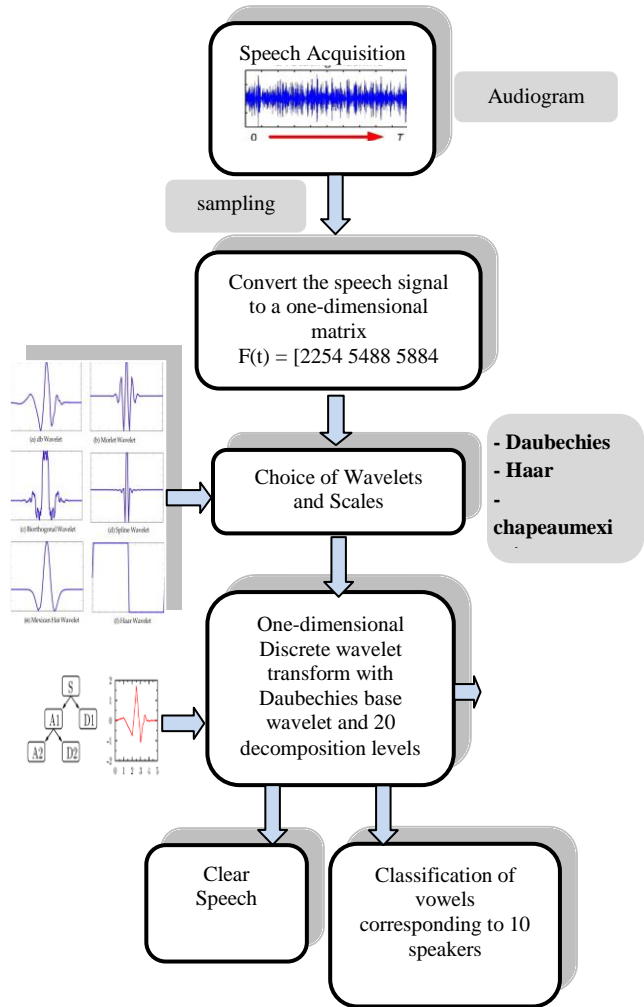


Fig. 3 Diagram of the proposed method

In this method, we rely on the processing of the audio signal stored in the database, where noise can be removed and the main frequencies of each letter can be identified. This method can be used in real time after developing the algorithm in embedded system that manages the voice signal and identifies certain commands that control an automatic system.

### 5-1- Detection of Singularity of the impulse noise signal in CWT

For noise evaluation, the oscillation of the acoustic signals is regarded as a considerable important metric. The CWT is often applied to detect the singularities of a transient signal.

On every stage of numerical simulations standard procedures for calculations of wavelet coefficients were used (in the case of the continuous as well as discrete wavelet transforms), which are integral parts of MATLAB software. For these calculations we employed MATLAB cwt function.

Fig. 4 vowel A time domain

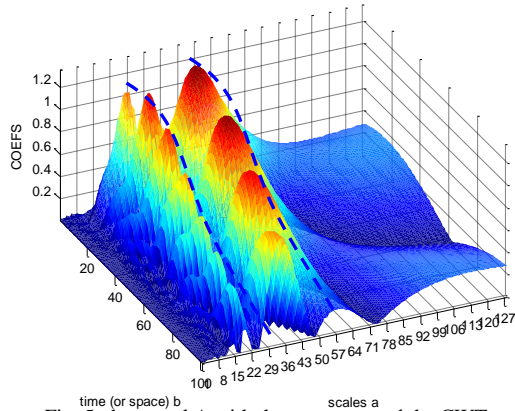
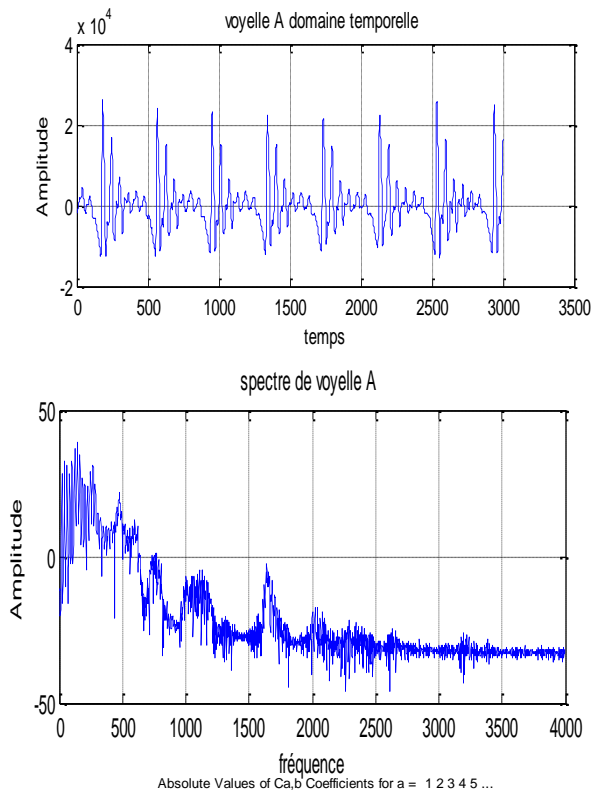


Fig. 5 the vowel A with the spectrum and the CWT

Table 1: shows the different frequencies of the vowel " A, E, I, O, U " respectively

| Frequency Hz \ Vowel | Frequency F1 | Frequency F2 | Frequency F3 |
|----------------------|--------------|--------------|--------------|
| A                    | 625          | 1491         | 2356         |
| E                    | 387          | 1985         | 2875         |
| I                    | 246          | 2018         | 3196         |
| O                    | 313          | 756          | 2271         |
| U                    | 312          | 750          | 2079         |

We usually focus on three main frequencies for letter and word identification, after removing noise and applying wavelet transform.

In practice, the third frequency can be neglected because it may be close in several letters, and we are satisfied with the first and second frequency only, especially if the noise is removed at a satisfactory rate.

The figure opposite shows the letter E after noise cancellation

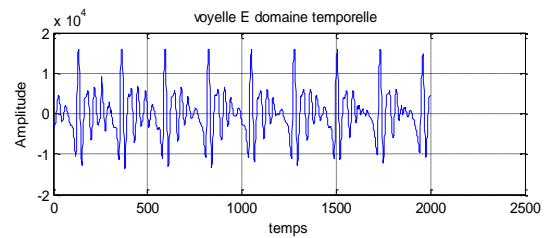


Fig. 6 vowel E time domain

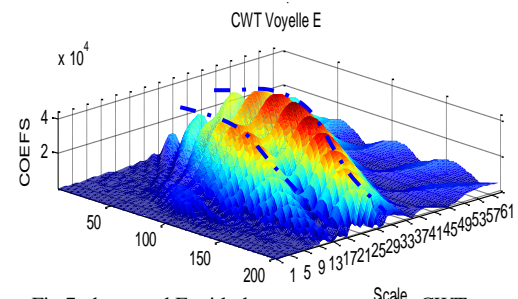
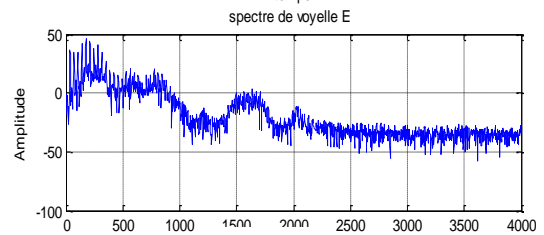


Fig.7 the vowel E with the spectrum and the CWT

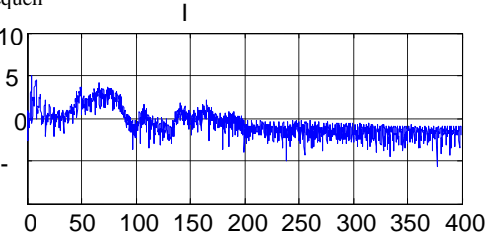
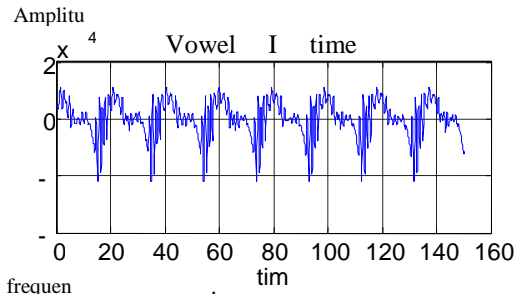


Fig.8 the vowel E with the spectrum and the CWT

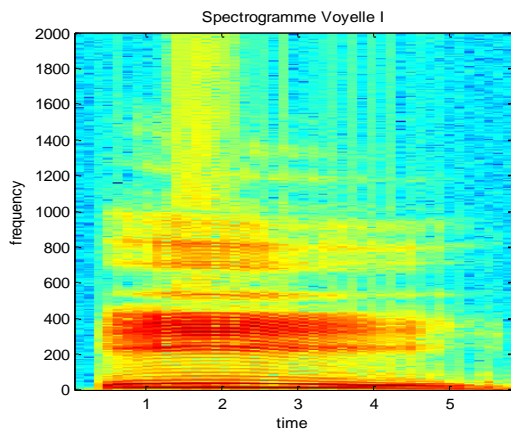


Fig.9 Spectrogram vowel i

Table 2: Presents the results obtained for various simulation cases following the order of Daubechies wavelets, applied to the vowel 'a'

| Vowel a |    | Baub.2   | Baub.4   | Baub.8   | Baub.16  | Baub.20  |
|---------|----|----------|----------|----------|----------|----------|
| Appro   | A1 | 7268.9   | 7269.8   | 7268.4   | 7265.6   | 7264.3   |
| Decom   | D1 | 209.2332 | 146.5979 | 144.4224 | 144.1895 | 137.3471 |
| Appro   | A2 | 5139.5   | 5139.6   | 5137.7   | 5133.7   | 5131.8   |
| Decom   | D2 | 147.9397 | 103.6432 | 102.0855 | 101.8817 | 97.0284  |
| Appro   | A3 | 5139.0   | 5138.1   | 5134.2   | 5126.3   | 5122.5   |
| Decom   | D3 | 0.2686   | 0.0442   | 0.0555   | 0.0760   | 0.0567   |
| Appro   | A4 | 5138.5   | 5136.7   | 5130.8   | 5119.0   | 5113.2   |
| Decom   | D4 | 0.0529   | 0.0104   | 0.0174   | 0.0269   | 0.0285   |
| Appro   | A5 | 5138.5   | 5135.2   | 5127.3   | 5111.7   | 5104.0   |
| Decom   | D5 | 0.0529   | 0.0092   | 0.0068   | 0.0208   | 0.0192   |

### 5-2- Enhancement vowel "A", by the Wavelets

The following figure shows the resulting multi-resolution vowel "A", by the Wavelets

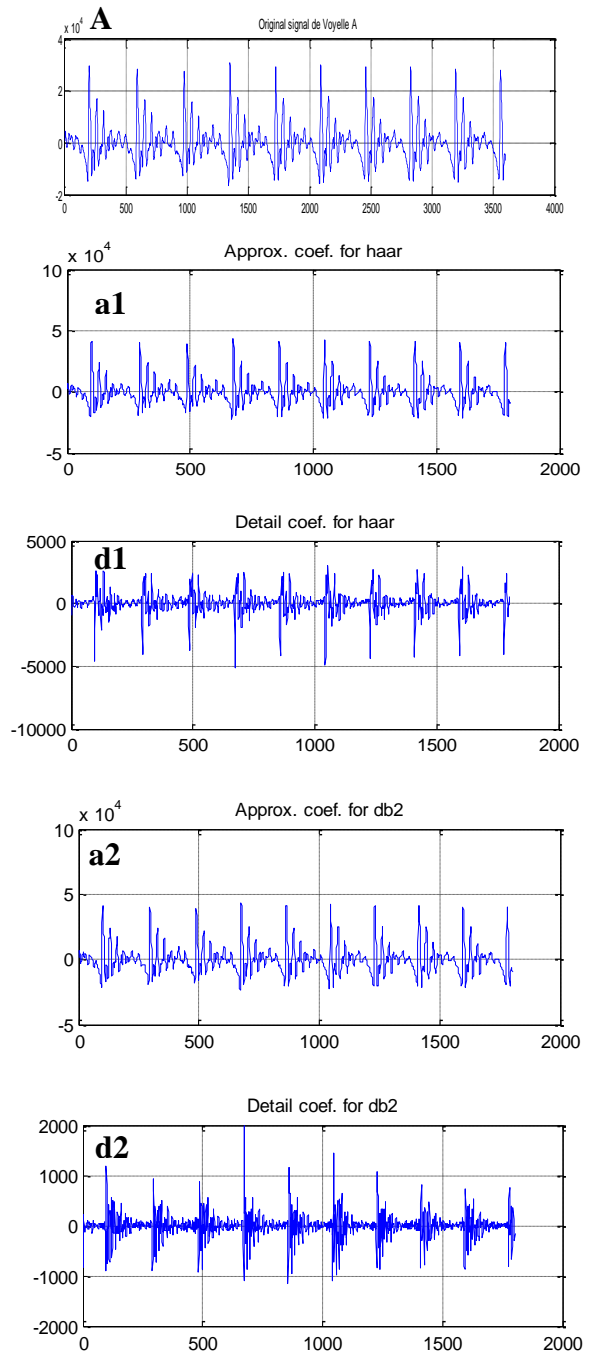


Fig.10 vowel "A", Levels of Decomposition by the Wavelets A: original speech signal, a1: appro 1, d1: level 1, a2: appro 2, d2:level 2, A=a1 + d1 + a2 + d2



The method is based on thresholding in the signal that each wavelet coefficient of the signal is compared to a given threshold. Using wavelets to remove noise from a signal requires identifying which components contain the noise, and then reconstructing the signal without those components.

Unlike STFT which has constant resolution at all times and at all frequencies, WT has good temporal resolution and low frequency resolution at high frequencies, and good frequency resolution and low temporal resolution at low frequencies.

In The figure 7, the red represent the maximum spectral intensity, while the blue represents the minimum frequency.

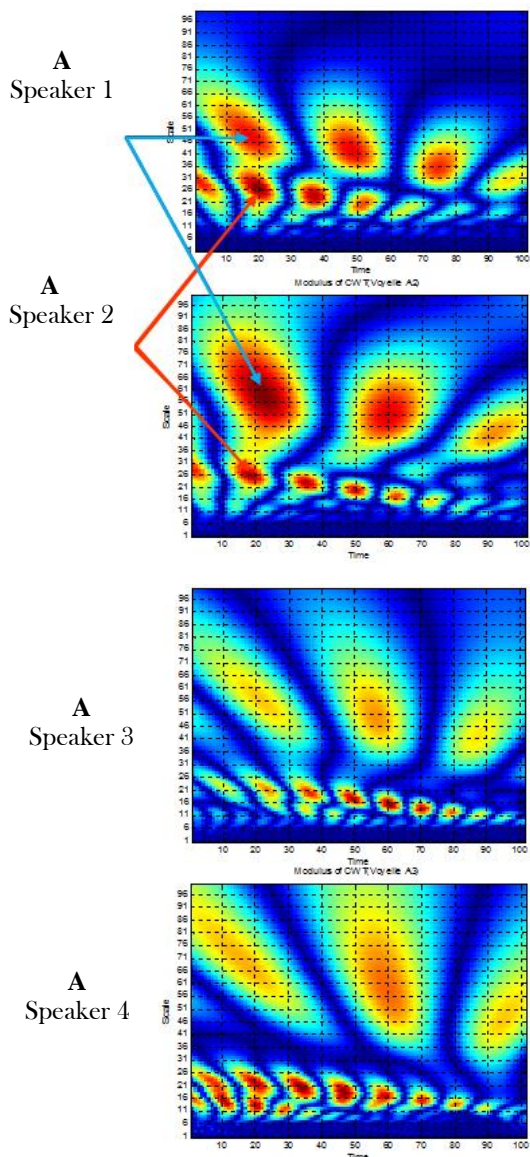


Fig.11 the CWT, Daubechies db12

Table 3 : shows the wavelet transform on the letter a after the noise is removed

|                 | Speaker 1<br>/A/ | Speaker 2<br>/A1/ | Speaker 3<br>/A2/ | Speaker 4<br>/A3/ |
|-----------------|------------------|-------------------|-------------------|-------------------|
| The first Scale | 26               | 25                | 26                | 24                |
| Second Scale    | 51               | 49                | 50                | 53                |

Table 4 : In the same method we find the following results for the Vowel U

|                 | Speaker 1<br>/U/ | Speaker 2<br>/U1/ | Speaker 3<br>/U2/ | Speaker 4<br>/U3/ |
|-----------------|------------------|-------------------|-------------------|-------------------|
| The first Scale | 41               | 42                | 44                | 41                |
| Second Scale    | 89               | 82                | 80                | 85                |

Does the system recognize single words or continuous speech? Obviously, it is easier to recognize isolated words well separated by periods of silence than to recognize the sequence of words constituting a sentence. Indeed, in the latter case, not only is the border between words no longer known but, moreover, the words become strongly articulated (i.e. the pronunciation of each word is affected by the word preceding as well as by the one that follows - a simple and well-known example being the links of French).

The following figure shows the Slice of a vowel A zone of the word /slap/ with CWT from Daubechies.

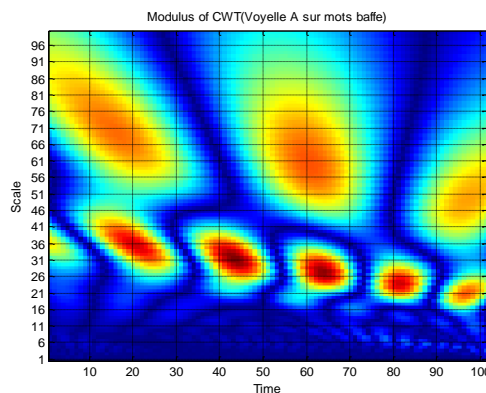


Fig.12 .vowel A the word /slap/ with CWT from Daubechies

Note that the formants presented on the same scale, the following.

Table 5: shows the comparison between Vowel / A / from word slap and vowel / A / separated

|                 | Vowel /A/ separated | Vowel /A/ from word slap |
|-----------------|---------------------|--------------------------|
| The first Scale | 26                  | 28                       |
| Second Scale    | 51                  | 56                       |

The following figure shows the Slice of a vowel / U / of the word / Une / with CWT from Daubechies

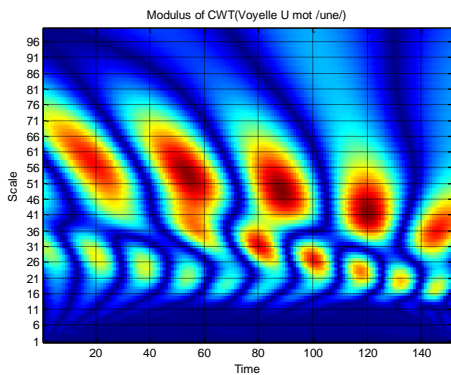


Fig.13 vowel U the word / Une / with CWT from Daubechies

Table 6 : shows the comparison between the Vowel /U / of word Une and vowel / U / separated,

|                 | Vowel /U/ separated | Vowel /U/ from word Une |
|-----------------|---------------------|-------------------------|
| The first Scale | 42                  | 39                      |
| Second Scale    | 82                  | 79                      |

From the previous results, it can be said that it can be difficult to identify the frequencies of words or letters without removing noise.

The results also indicate that the choice of wavelet type affects the degree of filtering and noise cancellation.

The figure follows presented the word Une with CWT from Daubechies.

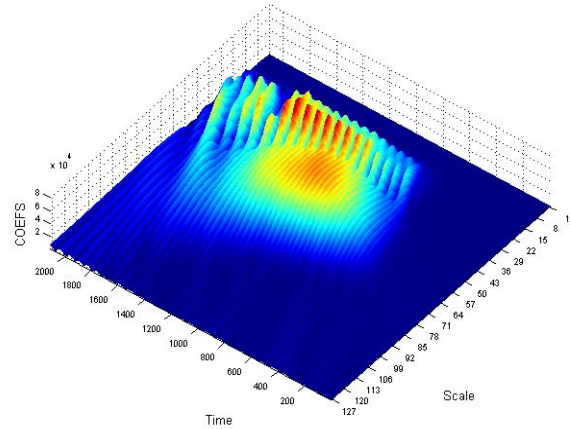


Fig.14 the word one with CWT from Daubechies

The figure (15), presented the classification of the vowels forming according to the scales corresponding with 10 speakers.

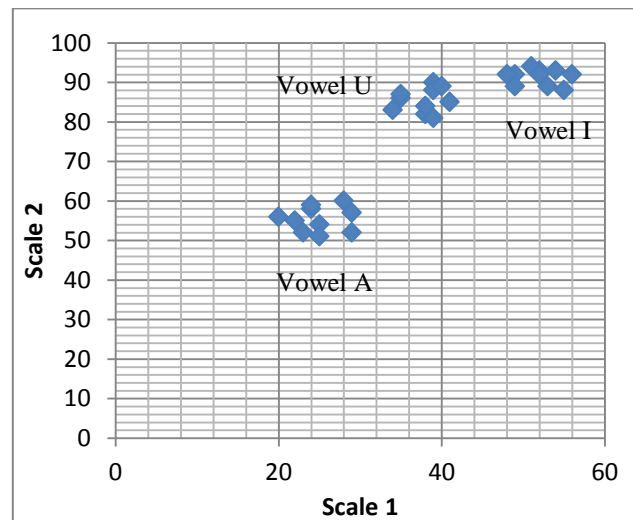


Fig.15 the classification of the vowels

After recording the speech signal for several people, and when applying the proposed method, the vowels were divided into three groups, so we can say that there is a convergence of the first and second scale in some values for each group.

## 6- Conclusion

The analog acquisition of the speech signal is the first thing to consider for spectral analysis of vowels or isolated words. Without it, it would be impossible to decompose this signal correctly and accurately in order to study it. This representation is not always the best for most signal processing applications. In many cases, the most relevant information is hidden in the frequency component of the signal. The frequency SPECTRUM of a signal is constituted by the frequency components of this signal. The frequency spectrum of a signal indicates which frequencies exist in the signal. The wavelet decomposition is similar to the Gabor decomposition: a speech signal is written in the form of a superposition of such offset and dilated wavelets.

If we get rid of the noise in the audio signal, the main frequencies are recognized in a short time and with great precision, especially in word recognition programs, and this is what we have discussed in this article where the wavelet conversion can be used to reduce noise and gain an understanding of sound.

Denoising of speech signals has been achieved successfully using wavelets. This paper provides a practical approach on how noisy audio (in wavelet form) incorporated with white gaussian noise can be denoised by using the coiflet wavelet.

## References

- [1] Rupali V. Mane, and Dr.M.T.Kolte, " Implementation of Adaptive Filtering Algorithm for Speech Signal on FPGA " , International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 2, Issue3, March 2014.
- [2] J. Kim, " Time-frequency characterization of hand-transmitted, impulsive vibrations using analytic wavelet transform", Journal of Sound and Vibration, vol.308, pp. 98-111, Nov 2007.
- [3] M. HemaLatha and Dr.S, " Resolution enhancement of low resolution satellite images using Dual tree complex wavelet transform", International Journal of Scientific & Engineering Research, Volume 8, Issue 5, May-2017.
- [4] I. Daubechies, Ten Lectures on Wavelets. Philadelphia: SIAM, 1992.
- [5] U. Jayakrishnan, G. Dhavale and P. Khandelwal, Wavelet Denoising Of Discrete-Time Signals, EE678 Wavelets Application Assignment, 2005.
- [6] Y. Hoshino, Wavelet Transform Analysis the Recognizing Brain Activities for Development the Palm-Size and Simplification Near-Infrared Spectroscopy Prototype System by Using Arduino,2018.
- [7] K. Borna, and S. Palizdar, "Short Time Price Forecasting for Electricity Market Based on Hybrid Fuzzy Wavelet Transform and Bacteria Foraging Algorithm" Journal of Information Systems and Telecommunication, Vol. 4, No. 4, October-December 2016.
- [8] S. Rani, and R. Kaur, "review: audio noise reduction using filters and discrete wavelet transformation", Journal of The International Association of Advanced Technology and Science (JIAATS), ISSN-5563 1682, Vol. 16 , June 2015.
- [9] C. Gargour, M. Gabrea, V. Ramachandran, and J. Lina, "A Short Introduction to Wavelets and Their Applications", IEEE Circuits and Systems Magazine, ISSN: 1531-636X, vol. 2, pp. 57-67, 2009.
- [10] R. Torkamani, and Sadeghzadeh, "Wavelet-based Bayesian Algorithm for Distributed Compressed Sensing", Journal of Information Systems and Telecommunication, Vol. 7, No. 2, April-June 2019.
- [11] N. Kaladharan, "Speech Enhancement by Spectral Subtraction Method", International Journal of Computer Applications (0975 – 8887) Volume 96– No.13, June 2014.
- [12] P. Goli, "Speech Intelligibility Improvement in Noisy Environments for Near-End Listening Enhancement", Journal of Information Systems and Telecommunication, Vol. 4, No. 1, January-March 2016.
- [13] M. R. Kahrizi, "Long-Term Spectral Pseudo-Entropy (LTSPE): A New Robust Feature For Speech Activity Detection", Journal of Information Systems and Telecommunication, Vol. 6, No. 4, October-December 2018.
- [14] K. Wang, "Wavelet packet analysis for speaker-independent emotion recognition", Neurocomputing 398 (2020) 257–264.
- [15] A. Bhattacharyya, " Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals ", m5G; v1.232; Prn:7/03/2018; 12:30] P.1 (1-12).
- [16] A. Upadhyay, and R.B. Pachori, "Speech enhancement based on mEMD–VMD method, Electron", ELECTRONICS LETTERS 30th March 2017 Vol. 53 No. 7 pp. 502–504.
- [17] P. Singh, S.D. Joshi, R.K. Patney, and K. Saha, "The Fourier decomposition method for nonlinear and non stationary time series analysis", Proc. R. Soc. A 473 (2017).
- [18] Yaseen, G. Young Son, and Soonil Kwon, " Classification of Heart Sound Signal Using Multiple Features", Appl. Sci. 2018, 8, 2344; doi:10.3390/app8122344.
- [19] M. Pourseidrezaei, "Prediction of Psychoacoustic Metrics Using Combination of Wavelet Packet Transform and an Optimized Artificial Neural Network ", Archives of Acoustics – Volume 44, Number 3, 2019.

**Meriane brahim** was born in Skikda, Algeria, in April 1982. He received the Dipl.-Ing. degree in electronics in 2006, the M.Sc. degree in electronics in 2009, It is currently working toward the Ph.D. degree in the faculty of Electrical from Batna University, Algeria. Currently, he is an Assistant Professor in the Department of Technology at Higher School of Technological Education Skikda ENSET , Algeria. His research interests include data compression, and Astronomical Image and speech noise cancelation and signal processing in the field of digital audio, member of the advanced electronics laboratory team, He is supervised in the doctoral thesis Pr. Benatia Djamel, Mr. Meriane was an IEEE member in 2015.