International Journal of Electrical and Computer Engineering (IJECE)
Vol. 6, No. 5, October 2016, pp. 2150~2157
ISSN: 2088-8708, DOI: 10.11591/ijece.v6i5.10826

Adaptive Speech Compression Based on Discrete Wave Atoms Transform

Bousselmi Souha[1], Aloui Nouredine[2], Cherif Adnane[1]
[1]Department of Physics, Faculty of Sciences of Tunis, Farhat Hached University, Tunisia
[2]Centre for Research on Microelectronics & Nanotechnology, Sousse Technology Park, Tunisia

Article Info

Article history:

Received Mar 21, 2016
Revised May 26, 2016
Accepted Jun 14, 2016

Keyword:

Adaptive thresholding
DWAT
DWT
Speech compression

ABSTRACT

This paper proposes a new adaptive speech compression system based on discrete wave atoms transform. First, the signal is decomposed on wave atoms, then wave atom coefficients are truncated using a new adaptive thresholding which depends on the SNR estimation. The thresholded coefficients are quantized using Max Lloyd scalar quantizer. Besides, they are encoded using zero run length encoding followed by Huffman coding. Numerous simulations are performed to prove the robustness of our approach. The results of current work are compared with wavelet based compression by using objective criteria, namely CR, SNR, PSNR and NRMSE. This study shows that the wave atoms transform is more appropriate than wavelets transform since it offers a higher compression ratio and a better speech quality.

Copyright © 2016 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Bousselmi Souha,
Department of Physics, Faculty of Sciences of Tunis,
Farhat Hached University,
El Manar, PB 2092, Belvedere, Tunisia.
Email: bousselmi.souha2008@gmail.com

1. INTRODUCTION

The improvement in the speech compression field is mainly related to the need of rapid and efficient techniques for data storage and transmission.The purpose of any compression technique is to present the speech signal using few bits while preserving the quality of the reconstructed signal. Speech compression is based on reducing the redundancy between samples; it has become ubiquitous in many applications such as mobile telephony and voice over IP. In literature, the speech compression algorithms are split into two main categories: lossless compression and lossy compression. The first one provides an exact reconstruction of the original signal; however, it cannot achieve low data rates. We can mention the Run Length Encoding and the Huffman coding as the most known algorithms in this category. The second one cannot obtain an exact reconstruction, but it provides high compression ratios [1].

In general, the existing speech compression algorithms combine between both of them in order to increase as much as possible the compression ratio. There are different techniques of audio compression, namely the direct speech compression, parameter extraction and transformation methods. The direct compression consists of extracting the significant information in temporal domain to approximate the original signal [2]. Parameter extraction methods extract the parameters of the signal such as linear predictive compression [3]. Transform compression technique (e.g., discrete cosine transforms [4], wavelet transforms [5]) convert the signal from the time domain to another parsimonious domain. Among them, the wavelet transform is the most popular one since it was used in many signal processing applications, [6]-[9].

Many comparative studies have been proven that wavelet outperforms the DCT (Discrete cosines transform) which is utilized by the MPEG standard, FFT (Fast Fourier Transform) [4] and LPC (Linear Predictive Coding) [10]. In recent years, new multi-scale transform called wave atoms has been emerged, it has been included in many articles in the field of image processing [11]-[14], this transform is well suited for representing the images, data, due to its directionality and sparsity compared with the discrete wavelet transform. Sparsity is the important criterion that can be considered in speech compression, in contrast of directionality, which is considered for only 2D or higher dimensional signal [15]. In this context, we have proposed a new compression system to explore the use of wave atoms in the field of speech compression; we have also proposed an adaptive threshold, which depends on the SNR estimation to preserve the quality of the reconstructed signal.

This article is structured as follows. The next section describes the discrete wave atoms transform. The section III gives more detail about the proposed speech compression system. Then, simulation results are presented in section IV. Finally, we conclude this work with section V.

2. DISCRETE WAVE ATOMS TRANSFORM

In [10], a signal is considered as oscillatory model when it can be described as the function below:

$$f(x) = \sin(N_g(x))h(x) \tag{1}$$

x, is coordinate. g, and h are C^{∞} scale function. h has a compact support in $[0,1]^2$ and N is a large constant. Fourier series decomposes a function having a finite duration or which is periodic into a sum of oscillating function, namely sines and cosines. In Fourier transform, sparsity is missed due to discontinuities, which is known as Gibbs Phenomenon. It needs an important number of coefficients to reconstruct a discontinuity with minimal loss of accuracy. For getting sparse solution of signal *f*, wave atoms were proposed by Demanet and Ying in [16],[17].

Theorem: For f be of the form (1).Assume g has no critical points. Then f can be represented to accuracy ε in by the largest $C_\varepsilon N$ wave atoms coefficients in absolute value, where for all $M > 0$, there exists $C_M > 0$ such that $C_\varepsilon \le C_M \varepsilon^{-1/M}$.

This theorem means that wave atoms transform give $O(N)$ coefficients to oscillatory function. Under some accuracy situation, we would need $O\left(N^{3/2}\right)$ curvelet coefficients or $O\left(N^2\right)$ wavelet coefficients. Despite the fact that this conclusion is forced in two or higher dimensional signal, it is easy to revert to one dimensional situation, in which wrapping description and directionality of wave atoms are not considered but the sparsity is preserved. The principle of our study is the assumption that the speech signal obeys this model. Whereas, a speech signal obeys the model mentioned above. We think that wave atoms transform can represent a speech signal more sparsely and improve compression factor while preserving the quality upon reconstruction. Wave atoms are a variant of wavelet packets; they have a high frequency localization that cannot be achieved using a filter bank based on wavelet packets and Curvelet, Gabor atoms. Wave atoms exactly interpolate between Gabor atoms and directional wavelets. The parameter α represent the multi-scale transform properties, from 0 (uniform) to 1 (dyadic). The parameter β measures the wave packet's directional selectivity. Figure 1 exposes the various transforms.

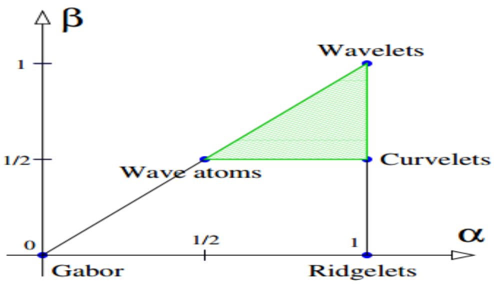

Figure 1. Identification of various transforms as (α, β) families of wave packets [16]

Wave atoms 1D family function is defined as $\varphi_\mu(x)$, with subscript μ = (j, m, n). The indexed point (xμ, ωμ) in phase-space is defined as follows.

$$x_\mu = 2^{-j} n \ , \quad w_\mu = \pi 2^j n \qquad (2)$$
$$C_1 2^j \le \max |m_i| \le C_2 2^j$$

The elements of frame $\varphi_\mu(x)$ are named wave atoms when:

$$\left|\hat{\varphi}(\omega)\right| \le C_M{}^{2-J}(1 + 2^{-j}\left|\omega - \omega_\mu\right|)^{-M} + C_M{}^{2-J}(1 + 2^{-j}\left|\omega + \omega_\mu\right|)^{-M}$$
$$\left|\varphi_\mu(x)\right| \le C_M 2^j (1 + 2^j \left|x - x_y\right|) \qquad (3)$$
$$M > 0$$

In practice, wave atoms are constructed from tensor products of a particular wavelet packet, which satisfies parabolic scaling wavelength that is achieved using decomposition architecture like incomplete wavelet packet as shown in Figure 2 [16].

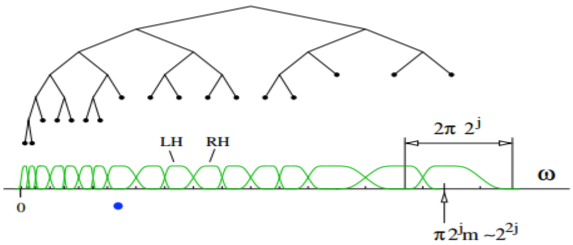

Figure 2. Strategy of wave atoms and corresponding set of subbands [16]

Figure 3 shows the space-frequency domain forms of one-dimensional wave atoms at increasing scales.

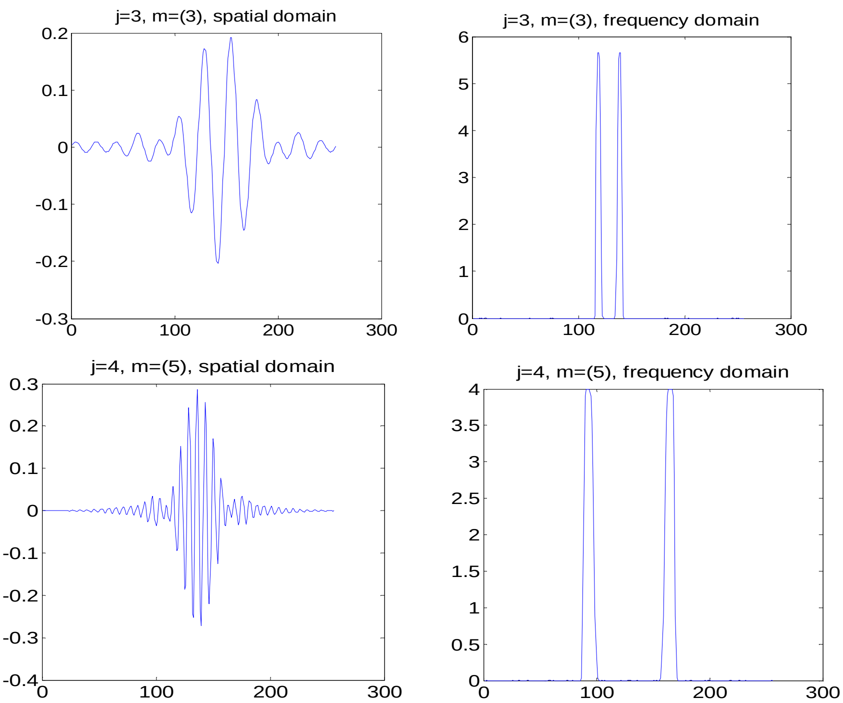

Figure 3. One dimensional wave atoms in space frequency domain at increasing scales

3. ADAPTIVE SPEECH COMPRESSION USING WAVE ATOMS TRANSFORM

The block diagram of the proposed compression system is illustrated by Figure 4. The different steps of the system are explained in the following paragraphs.

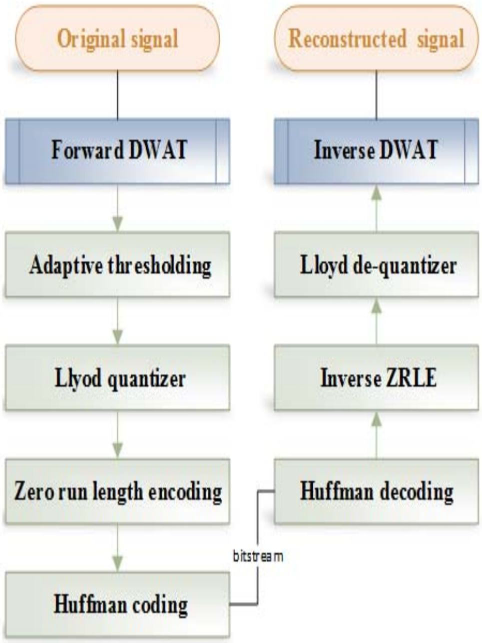

Figure 4. Block diagram of adaptive speech compression using DWAT

3.1. Discrete wave atoms transform

The first step of our approach consists in decomposing the speech signal using DWAT. The particularity of this transform is to convert the temporal representation of a signal into a time-frequency representation. This domain transformation reduces the redundancy and decorrelates the signal's samples, thus, decreases the bitrate of transmission. Wave atoms concentrate speech information into a few coefficients as shown in Figure 5 [3]. Therefore, after applying the wave atoms transform of a signal, many coefficients will either be zero or have negligible magnitudes.

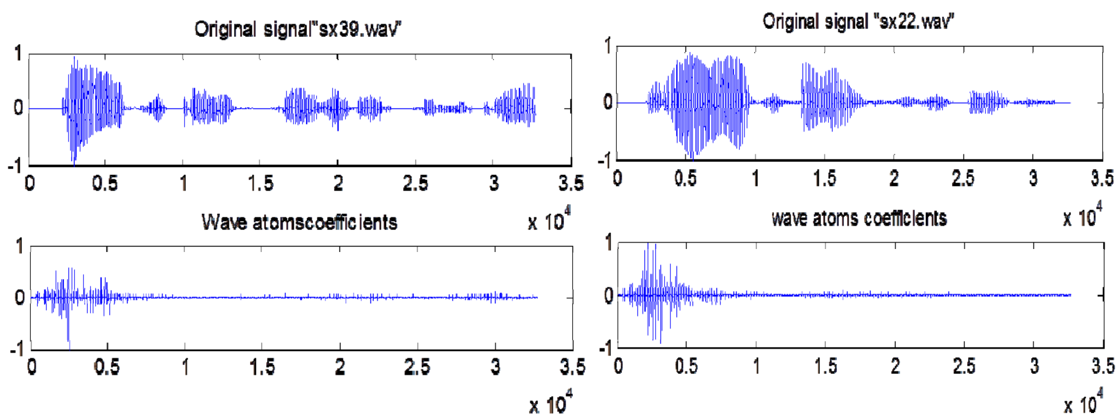

Figure 5. Normalized Wave atoms coefficients of speech signals

3.2. New adaptive thresholding

Thresholding is the most important step in compression based transform; it consists of rejecting the coefficients of the DWAT transform inferior to a given threshold. There are different methods of thresholding, such as the hard thresholding and the soft thresholding which are the commonly used methods. In this work we have used the hard threshold given in this equation:

$$C_{\mathrm{Re}} = \begin{cases} C_{\mathrm{Re}} & if \left| C_{\mathrm{Re}} \right| \geq T \\ 0 & otherwise \end{cases} \tag{4}$$

The choice of threshold T is very delicate; a bad choice of threshold can degrade the signal after reconstruction. There is no suitable threshold for all signals due to the diversity of speech signals. Thus, we have introduced a new adaptive thresholding process which allows the adjustment of the threshold according to the desired speech quality. The flow chart of the adaptive threshold process is shown in Figure 6.

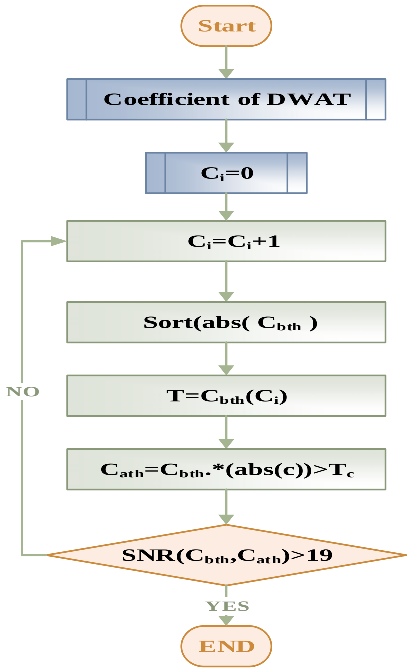

Figure 6. Flow chart of adaptive thresholding

Ci: Retained coefficients.
C_{bth}: Coefficients before thresholding.
C_{ath}: Coefficients after thresholding.

3.3. Lloyd-Max scalar quantization

After thresholding, a quantization process is performed. It deals with the approximation of the retained coefficients of DWAT with a finite set of values. There are two methods of quantization: The scalar quantization and the vector quantization. In general, quantization causes a relative distortion of the signal, which can be minimized by the use of the Lloyd-Max scalar quantizer.

3.4. Encoding

To achieve the speech compression, we have encoded the quantized coefficients using a particular Run Length Encoding suitable for our vector. This type of encoding codes only the runs of zeros with two bytes. The first byte indicates the start of a sequence of zeros and the second one represents the number of zeros. [18] This step is followed by a Huffman coding in order to eliminate any redundancy caused by quantization. To reconstruct the speech signal, we have reversed the different stages (Wave atoms, quantization, coding).

4. TEST AND RESULTS

In this section, a MATLAB program has been developed to implement the speech compression codec based on DWAT as described in this paper. To evaluate the efficiency of the developed algorithm, a comparative study between the DWAT and DWT algorithm is performed using objective criteria; CR, SNR, PSNR and NRMSE. In all simulations, only source speech signals extracted from the TIMIT Database are exploited [19].

- Compression ratio (CR)

$$C = \frac{Length(x(n))}{Lenght(cWC)} \tag{5}$$

cWC, is the length of the compressed wave atoms transform vector.

- Signal to noise ratio (SNR)

$$SNR = 10\log_{10}(\frac{\delta_x^{\ 2}}{\delta_e^2}) \tag{6}$$

Where $\delta_x^{\ 2}$, is the mean square of the speech signal and δ_e^2 is the mean square difference between the original and reconstructed signal.

- Peak signal to noise ratio (PSNR)

$$PSNR = 10\log_{10}\frac{NX^2}{|x-r|^2} \tag{7}$$

N, is the length of the reconstructed signal, X is the maximum absolute square value of the signal x and $|x-r|^2$ is the energy of the error between the reconstructed and original signal.

- Normalized root mean square error (NRMSE)

$$NRMSE = \sqrt{\frac{\left(x(n)-r(n)\right)^2}{(x(n)-\mu_x(n))^2}} \tag{8}$$

$x(n)$, is the speech signal, $r(n)$ is the reconstructed signal, and $\mu_x(n)$ is the mean of the speech signal.

The test results of the proposed algorithm are summarized in Table1.

Table 1. Performance evaluation of the proposed algorithm using TIMIT speech files

Audio file	Algorithm	CR	SNR	PSNR	NRMSE
sx27.wav	DWAT	10.1198	19.1324	39.8416	0.1105
sx11.wav	DWAT	12.4593	19.1370	36.7158	0.1104
sx12.wav	DWAT	9.8699	19.0968	37.3253	0.1110
sx37.wav	DWAT	12.4121	19.1279	37.6441	0.1106
sx57.wav	DWAT	10.4757	19.1354	35.7199	0.1105
sx26.wav	DWAT	8.4150	19.1210	36.4544	0.1106
sx243.wav	DWAT	12.3281	19.1394	39.0011	0.1104

From the above table, it is obvious that our approach offers a high compression ratio. The SNR is in average of 19 dB that is high enough to certify a good quality of the reconstructed signal. We can as well remark that by using the adaptive thresholding we have got a uniform SNR, in contrast in [4] the threshold value is set manually, which engenders a not uniform SNR that varies between 10dB and 22dB. Hence the quality of reconstructed signal is not assured for all speech signals.

To evaluate the efficiency of our approach a comparative study is established with other studies based on DWT released in [4], [20], and [8]. For the DWT algorithm, we have used the Daubechies orthogonal wavelet db10 and we have applied five decomposition levels and a global thresholding. Given the acoustic differences between male and female we have effectuated comparison tests on three voices from each gender. These results are in Table 2 and Table 3.

Table 2. The performance results of the DWAT and DWT algorithms for female voices

Audio file	Algorithm	CR	SNR	PSNR	NRMSE
sx69.wav	DWAT	9.8108	19.1299	35.4186	0.1105
	DWT	7.7978	18.0038	34.7076	0.1258
sx84.wav	DWAT	10.2915	19.1377	37.1492	0.1104
	DWT	8.2026	17.8513	36.5450	0.1281
sx210.wav	DWAT	13.3856	19.1320	36.2084	0.1105
	DWT	6.3563	18.1233	36.3855	0.1278

Table 3. The performance results of the DWAT and DWT algorithms for male voices

Audio file	Algorithm	CR	SNR	PSNR	NRMSE
sx156.wav	DWAT	12.9313	19.1000	38.0939	0.1109
	DWT	8.9302	17.9074	38.3127	0.1272
sx229.wav	DWAT	9.5145	19.1656	35.0638	0.1101
	DWT	6.5588	17.9410	33.8392	0.1268
sx289.wav	DWAT	9.8997	19.1141	37.0677	0.1107
	DWT	6.1226	17.9807	35.8804	0.1262

Throughout Table 2 and Table 3, it is clearly shown that the proposed system rates are better than DWT for male and female voices. In fact, it has improved the CR, PSNR, SNR parameters; while decreasing the NRMSE. Despite, the wavelet filter optimization used to improve the speech compression using DWT as given by [8], [21], [22] they cannot achieve the compression ratio obtained by the proposed algorithm.

5. CONCLUSION

In this paper, a new adaptive speech compression algorithm using discrete Wave Atoms is presented. The evaluation of performance using objective criteria such as CR, SNR, PSNR and NRMSE shows that the developed algorithm presented in this paper gives a high compression ratio without destruction the quality of the reconstructed speech signal. A comparative study between our DWAT and the DWT methods demonstrates that the proposed algorithm increases the compression factor by 2.5 to 7 without sacrificing the speech intelligibility nor the Signal to noise ratio.

REFERENCES

[1] K. Sayood, "Introduction to Data Compression, Third Edition," *Morgan Kaufmann Publishers*, 2006.

[2] W. Guo, *et al*., "A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals," *Journal of Sound and Vibration*.

[3] P. Vankateswaran, *et al*., "An efficient time domain speech compression algorithm based on LPC and sub-band coding techniques," *Journal of Communication*, vol/issue: 4(6), pp. 423–428, 2009.

[4] G. Rajesh, *et al*., "Speech Compression using Different Transform Techniques," *IEEE International Conference on Computer and Communication Technology (ICCCT)*, pp. 146-151, 2011.

[5] O. Yamanaka, *et al*., "Image compression using wavelet transform and vector quantization with variable block size," *Soft Computing in Industrial Applications, SMCia ' 08. IEEE Conference on, Muroran,* pp. 359-364, 2008.

[6] D. Narmadha, *et al.,* "An Optimal HSI Image Compression using DWT and CP," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 4(3), pp. 411– 421, 2014.

[7] H. E. Suryavanshi, *et al.,* "Digital Image Watermarking in Wavelet Domain," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 3(1), pp. 1– 6, 2013.

[8] A. Kumar, *et al.,* "The optimized wavelet filters for speech compression," *Int J Speech Technol* (*Springer),* 2012.

[9] O. Khalifa, *et al*., "Compression using Wavelet Transform," *International Journal of Signal Processing (SPIJ),* vol/issue: 2(5), pp. 17-26, 2008. ISSN (Online): 1985-2339.

[10] M. A. Mawla, *et al.,* "Comparing Speech Compression Using Wavelets With Other Speech Compression Schemes," *Student Conference on Research and Development, Proceedings,* pp. 55-58, 2003.

[11] J. Rajeesh, *et al.,* "Performance analysis of wave atom transform in texture classification," *in Signal, Image and Video Processing,* vol. 8, pp. 1–8, 2014.

[12] F. Foroozan, *et al*., "Wave atom based Compressive Sensing and adaptive beam forming in ultrasound imaging," *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on, South Brisbane, QLD,* pp. 2474-2478, 2015.

[13] Z. W. Qiang, *et al*., "Image Denoising Based on Wave Atoms and Cycle Spinning," *Computational Intelligence and Security (CIS), Eighth International Conference on, Guangzhou,* pp. 310-313, 2012.
[14] Z. Haddad, *et al.,* "Wave atoms based compression method for fingerprint images," *Pattern Recognition*, vol. 46, pp. 2450–2464, 2013.
[15] H. Xu, *et al.*, "ECG data compression based on wave atom transform," *Multimedia Signal Processing (MMSP), IEEE 13th International Workshop on, Hangzhou,* pp. 1-5, 2011.
[16] L. Demanet, *et al.*, "Wave atoms and sparsity of oscillatory patterns," *Applied and Computational Harmonic Analysis,* vol/issue: 23(3), pp. 368-387, 2007.
[17] L. Demanet, "Curvelets, wave atoms, and wave equations," *PHD Thesis in California Institute of Technology*, 2006.
[18] Kinsner, *et al*., "Speech and Image Signal Compression with Wavelets," *IEEE Wescanex Conference Proceedings, IEEE, New York, NY,* pp. 368-375, 1993.
[19] V. Zue, *et al.*, "Speech database development at MIT: TIMIT and beyond," *Speech Communication,* vol/issue: 9(4), pp. 351 – 356, 1990.
[20] H. Ayadi, "speech compression using wavelets," *Electrical & Computer Engineering Department, Islamic. University of Gaza, Palestine,* 2010.
[21] N. Aloui, *et al.*, "A New algorithm for QMF Banks Design and Its Application in Speech Compression using DWT," *International Arab Journal of Information Technology*, 2015.
[22] N. Aloui, *et al*., "Genetic Algorithm for Designing QMF Banks and Its Application In Speech Compression using Wavelets," *I.J. Image, Graphics and Signal Processing,* vol. 6, pp. 1-8, 2013.

BIOGRAPHIES OF AUTHORS

Souha bousselmi was born in Bizerte, Tunisia, in 1986. She received the Master's degree in Electronics from the Faculty of Sciences of Tunis (FST) in 2012. Currently, she is pursuing the Ph.D. degree in Electronics with the Faculty of Sciences of Tunis, in the laboratory of Innovation of communicant and cooperative mobiles (*Innov'Com*), the Higher School of Communication of Tunis (SUPCOM). Her research interests include audio signal compression.

Noureddine Aloui received the master, doctorate degrees from Sciences Faculty of Tunis; he is a Researcher Member in Innov'Com group, Signal Processing Laboratory. His field of interest concerns digital signal processing.

Adnane Cherif received the engineer, master and doctorate degrees from National Engineering School of Tunis (ENIT), in Tunisia.he is a Professor at the Science Faculty of Tunis and responsible in Innov'Com group- Signal Processing Laboratory, Science Faculty of Tunis, Tunisia. His field of interest concerns digital signal processing and speech processing.