

New Algorithm for Speech Compression Based on Discrete Hartley Transform

Noureddine Aloui¹, Souha Bousselmi², and Adnane Cherif²

¹Centre for Research on Microelectronics and Nanotechnology, Sousse Technology Park, Tunisia

²Innov'Com Laboratory, Sciences Faculty of Tunis, University of Tunis El-Manar, Tunisia

Abstract: This paper presents an algorithm for speech signal compression based on the discrete Hartley transform. The developed algorithm presents the advantages to ensure low bit rate and to achieve high speech compression efficiency, while preserving the quality of the reconstructed signal. The numerical results included in this paper show that the developed algorithm is more effective than the discrete wavelet transform for speech signal compression.

Keywords: Speech signal compression, discrete hartley transform, discrete wavelet transform.

Received April 13, 2015; accepted May 2, 2016

1. Introduction

Speech signal compression is an important field of digital signal processing. It has two main objectives; reduces storage and bandwidth requirements of digital speech signals before storage or transmission via communication channels. In fact, the speech compression methods allow to represent speech data using the minimum number of bits possible, while preserving the perceptual quality. Therefore, speech compression has many practical applications such as mobile telephony, voice over IP and storage [7, 17].

In the past few decades the discrete transformations have emerged as a new tool for signal analysis, such as compression [18, 19], denoising [1, 20], digital filtering [4, 5] and recognition [12, 21]. In the field of speech compression, the discrete transforms provide an important compression ratio and preserve the quality of the decompressed speech signal compared to others existed techniques. In fact, the discrete cosine transform introduced by Ahmed *et al.* [2], especially its modified version [6] has been used by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (CEI) standard to compress speech signal MP3 format (Moving Picture Experts Group: (MPEG1) layer 3). This later has become an industry standard for audio compression. Thus, the discrete wavelet transform (introduced by Grossman and Morlet. [10] gives better compression performances compared to the Global System Mobile (GSM), the Linear Predictive Coding (LPC) [16] and the Code Excited Linear Prediction (CELP) [15]. Due to the effectiveness of the discrete transforms for signal compression, in this paper the Hartley transform is particularly exploited to compress speech signals. This transformation has been introduced by Bracewell [4]. It presents the discrete version of the continuous Hartley

transform, introduced by Hartley [9]. The Hartley transform resembles the Fourier transform, but it presents some advantages.

It allows transforming a real functions to another real functions (it is not necessary to require complex numbers). The discrete Hartley transform is applied in many fields of signal processing such as audio and image [11, 14].

The remaining of this paper is divided into three sections; section 2 covers the Hartley transform theory. Section 3, discusses the principal methodology of speech compression based on the discrete Hartley transform. Then, the performance evaluation of the developed algorithm is carried out in section 4. Finally, section 5 dedicated to the conclusion and some remarks.

2. Signal Processing Based Using Discrete Hartley Transform

This section describes the principle of signal processing using discrete Hartley transform. The first and the second subsection present the Hartley transform theory. The third subsection gives the implementation methodology of the discrete Hartley transform followed by an example of signal processing using discrete Hartley transform.

2.1. Continuous Hartley Transform

The Continuous Hartley Transform (CHT) of a real waveform $x(t)$ is defined as [9]:

$$\Psi(\omega) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} x(t) (\cos(\omega t) + \sin(\omega t)) dt \quad (1)$$

The Inverse Continuous Hartley Transform (ICHT) is obtained using the following equation:

$$x(t) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} \Psi(\omega) (\cos(\omega t) + \sin(\omega t)) d\omega \quad (2)$$

2.2. Discrete Hartley Transform

The DHT is the discrete version of the continuous Hartley transform.

Let $\{x(n), n=0,1,\dots,N-1\}$ a real valued sequence, the DHT is obtained by [4]:

$$v(k) = N^{-1} \sum_{n=0}^{N-1} x(n) (\cos(2\pi kn/N) + \sin(2\pi kn/N)) \quad (3)$$

The Inverse Discrete Hartley Transform (IDHT) is obtained using the following Equation:

$$x(n) = \sum_{k=0}^{N-1} v(k) (\cos(2\pi kn/N) + \sin(2\pi kn/N)) \quad (4)$$

2.3. Signal Processing using DHT

Using Equation (3) the DHT can be written in matrix form as follows:

$$\begin{bmatrix} v(0) \\ v(1) \\ \dots \\ v(N-1) \end{bmatrix} = \frac{1}{N} \begin{bmatrix} h_{0,0} & h_{0,1} & \dots & h_{0,N-1} \\ h_{1,0} & h_{1,1} & \dots & h_{1,N-1} \\ \dots & \dots & \dots & \dots \\ h_{N-1,0} & \dots & \dots & h_{N-1,N-1} \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \dots \\ x(N-1) \end{bmatrix} \quad (5)$$

$$= \frac{1}{N} H_{N,N} * x_N$$

Where $H_{N,N}$ is square Hartley matrix of order N :

$$h_{kn} = \cos(2\pi kn/N) + \sin(2\pi kn/N) \quad (6)$$

where k and n are integers from 0 to $N-1$.

As the Hartley matrix is invertible, the IDHT can be obtained by computing the inverse Hartley matrix:

$$x_N = H_{N,N}^{-1} * v_N \quad (7)$$

The achieved inversion error E_r of Hartley matrix can be calculated as follows:

$$E_r = \|H_{N,N} * H_{N,N}^{-1} - I_{N,N}\| \quad (8)$$

Where $I_{N,N}$ is a square identity matrix of order N .

Figure 1 illustrates the inversion error of Hartley matrix for different sizes. Here, the inverse matrix is computed using the function “Inv” of MATLAB. As observed, the inversion error of Hartley matrix for different sizes is about 10^{-13} . These values can be reduced by using small matrix sizes.

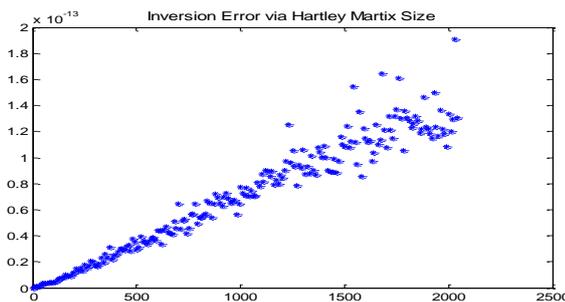


Figure 1. Inversion error variation vs hartley matrix size.

- Example: Let $x(n)$ the original signal (MATLAB sample data “leleccum.dat”) analyzed by frames of 64 samples and $H_{N,N}$ the square Hartley matrix of order 64.

After performing the forward and inverse DHT the quality of the reconstructed signals is evaluated in term of Root Mean Square Error (RMSE) using the following Equation:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (x(n) - y(n))^2} \quad (9)$$

where $y(n)$ is the reconstructed signal.

The following Figure illustrates a typical plot of the original signal “leleccum.dat”, the transformed (Figure 2-a) and the reconstructed (Figure 2-b) versions.

The calculated RMSE between the original and the reconstructed signals is equal to 5.09×10^{-15} . Hence, it is clear that the DHT can allow analyzing signals with near perfect reconstruction.

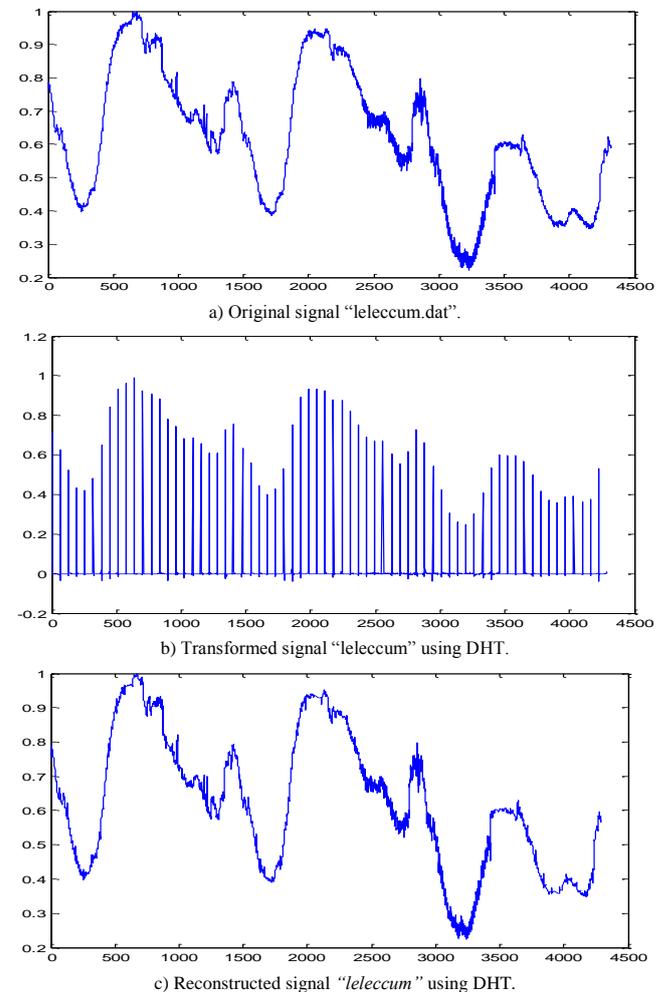


Figure 2. Original transformed and reconstructed “leleccum.wav” signal using DHT.

3. Speech Signal Compression using DHT

The basic idea behind speech signal compression using DHT is linked to the representation of a signal after applying the DHT. So, after performing the DHT

to the speech signal, the most of the obtained coefficients have a negligible or zero magnitudes.

For that, the speech compression is achieved by truncating the small values of coefficients

(insignificant data), while keeping a good quality of the reconstructed signal.

The Figure below illustrates the different steps involved in speech signal compression using DHT.

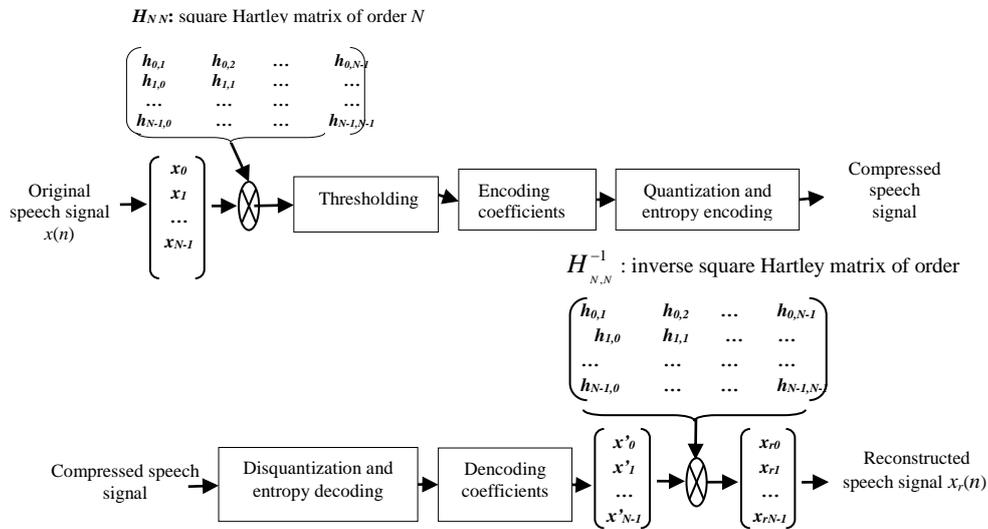


Figure 3. The flowchart of the speech compression algorithm using DHT.

- **Step 1:** In this step, the Hartley matrix order (N) is chosen and the DHT is performed on the original speech signal frames. Here, the Hartley matrix order must be equal to the frame length.

$$v_N = \frac{1}{N} H_{N,N}(n) * x_N \tag{10}$$

N is integer from 0 to $N-1$.

- **Step 2:** After performing the DHT, a manual thresholding is applied to the obtained coefficient vectors (v_N). It consists in truncating coefficients below a fixed threshold (T) using the following Equation:

$$v_{new}(n) = \begin{cases} v(n) & \text{if } |v(n)| \geq T \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $v_{new}(n)$ are the thresholded coefficients.

- **Step 3:** The thresholded vectors of coefficients $v_{new}(n)$ contain many string values of zeros. So, DHT based compression is achieved by efficiently encoding them. There are many techniques to perform the encoding such as Run Length Encoding (RLE).

In this research work, the string values of zeros are encoded by two values. The first value indicates the start sequence of zeros while the second represents the number of consecutive zeros in the v_{new} vectors.

- **Example:**

Let $v_{new} = \{1, 3, 0, 0, 0, 0, 0, 0, 0, 5, 2, 0, 0, 0, 0, 0, 0, 7\}$ a vector of 20 thresholded coefficients.

If each coefficient is encoded using 16 bits, so v_{new} needs $16 \times 20 = 40$ bytes. However, if the previous encoding method is applied, the encoded vector will be $v_{comp} = \{1, 3, 0, 8, 5, 2, 0, 7, 7\}$, then v_{new} need only $16 \times 9 = 18$ bytes.

- **Step 4:** In this step, quantization followed by entropy coding are applied to the encoded vector of coefficients. There are many methods to quantize coefficients such as vector, scalar and uniform quantization. The entropy coding can be performed using Huffman or arithmetic methods. In this work, uniform quantization and entropy coding are applied to compress speech signal using DHT.

4. Results and Discussions

This section evaluates the speech compression methodology based on DHT previously discussed. Two tests were performed to demonstrate the effectiveness of the developed algorithm. The first test is an evaluation in the speech compression field. While the second, presents a comparative study of performances with the algorithm based on Discrete Wavelet Transform (DWT). Here, the compression performances are evaluated using CR, SNR, PSNR and NRMSE given by the following Equations:

- **Signal to Noise Ratio (SNR):**

$$SNR = \frac{\sum x(n)^2}{\sum |x(n) - y(n)|^2} \tag{12}$$

- **Peak Signal to Noise Ratio (PSNR):**

$$PSNR = 10 \log_{10} \left(\frac{N x(n)^2}{\|x(n) - y(n)\|^2} \right) \tag{13}$$

- **Normalized Root Mean Square Error (NRMSE):**

$$NRMSE = \sqrt{\frac{(x(n) - y(n))^2}{(x(n) - \mu_x(n))^2}} \tag{14}$$

- **Compression Ratio (CR):**

$$CR = \frac{\text{Size of original audio signal}}{\text{Size of compressed audio signal}} \quad (15)$$

where, $x(n)$ and $y(n)$ are respectively the original and the reconstructed audio signal. N is the length of the reconstructed audio signal and $\mu_x(n)$ is the mean of the signal.

4.1. Evaluation Results of the Developed Algorithm

Table 1 lists the obtained results when the developed algorithm implemented using MATLAB software. Here, some TIMIT speech signals (sx10.wav to sx20.wav) are used. These signals are coded using 16-bits with 16 KHz sampling frequency.

Table 1. Compression performances using the developed algorithm based on DHT.

TIMIT speech files	CR	SNR	PSNR	NRMSE	Original sizes in kilobytes (KB)	Sizes in kilobytes (KB)
sx10.wav	7.321	20.332	39.372	0.096	90.400	12.347
sx11.wav	7.578	21.486	39.270	0.084	72.601	9.580
sx12.wav	7.804	20.390	40.637	0.095	61.601	7.892
sx13.wav	3.443	20.342	36.934	0.096	69.400	20.152
sx14.wav	6.844	18.032	38.619	0.125	84.601	12.361
sx15.wav	4.944	16.708	37.842	0.146	88.400	17.878
sx16.wav	6.659	19.879	39.610	0.101	65.601	9.851
sx17.wav	5.726	22.063	38.639	0.078	61.201	10.687
sx18.wav	6.308	19.847	39.137	0.101	86.800	13.759
sx19.wav	10.355	19.378	40.430	0.107	109.201	10.544
sx20.wav	4.783	20.618	37.966	0.093	110.601	23.123

These simulation results are obtained by using a fixed threshold value equal to 0.0015 and applying the Equation (11). In addition, the speech signals are analyzed by frames of 1024 samples and the Hartley matrix of order 1024. The resulted error of matrix inversion is equal to 7.234×10^{-14} when the inverse is computed using function “*Inv*” of MATLAB.

According to the above table, it can be seen that the developed algorithm for speech compression gives a high compression ratio and preserving the quality of the reconstructed signals.

The Figure below illustrates the plot of the original the speech signal “sx19.wav” (TIMIT Database), the transformed signal Figure 4-b and the reconstructed signal Figure 4-c using the developed algorithm based on DHT. As seen, the reconstructed speech signal version is similar to the original. Here, the used threshold value is equal to 0.0015. The obtained compression ratio is 10.35 and the signal to noise ratio is 19.378dB.

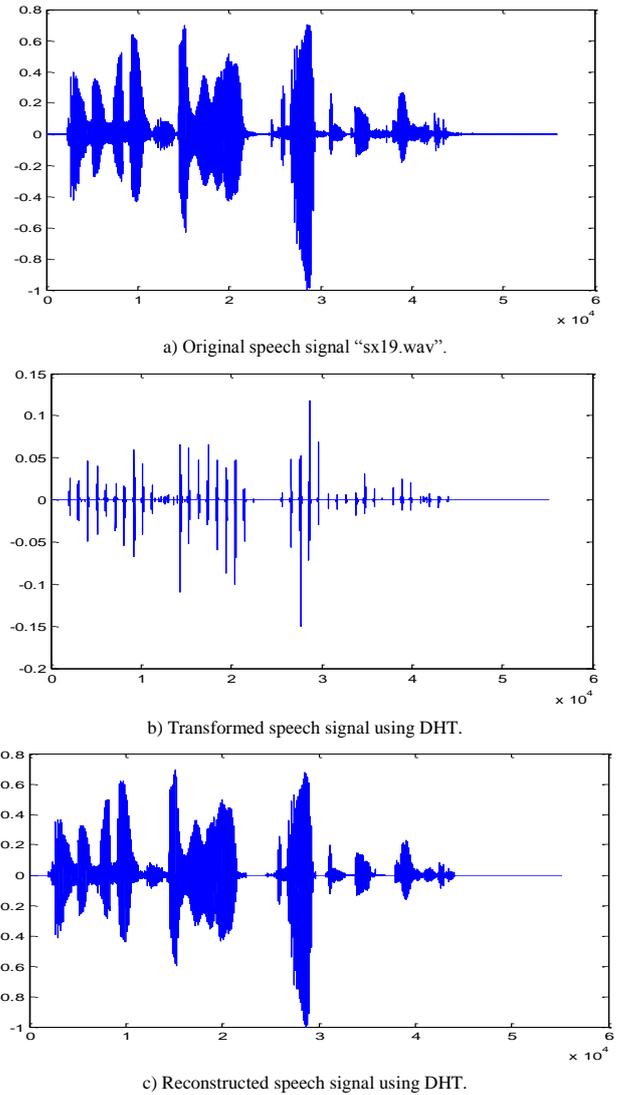


Figure 4. Original speech signal “sx19.wav”.

4.2. Comparative Study of Performances

To evaluate the developed algorithm based on DHT in the field of speech compression, a comparative study of performance with the algorithm based on Discrete Wavelet Transform (DWT) given in [10, 13, 18] is carried out.

All the simulation results are obtained using frames analysis of 1024 samples. After performing the DHT or the DWT, a fixed threshold (manually adjusted) is used to truncate (zeroed) the small values coefficients. The threshold values are fixed at 0.045 for the DWT speech compression algorithm and 0.0015 for the speech compression algorithm based on DHT. In addition, for the DWT speech compression algorithm, five decomposition levels are applied and Haar wavelet is used as mother wavelet.

The following Figures show the results of the comparative study of performances in terms of CR, SNR PSNR NRMSE and data sizes. From the histograms, it is clear that the proposed algorithm outperforms the DWT one in terms of all measured performances.

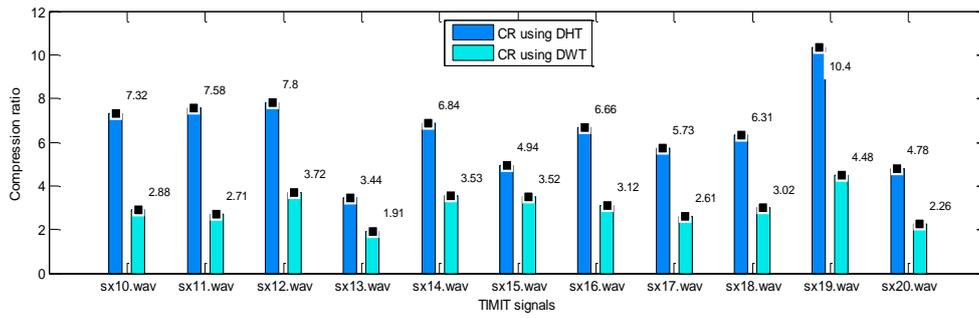


Figure 5. Compression Ratio (CR) variation using DHT and DWT algorithms.

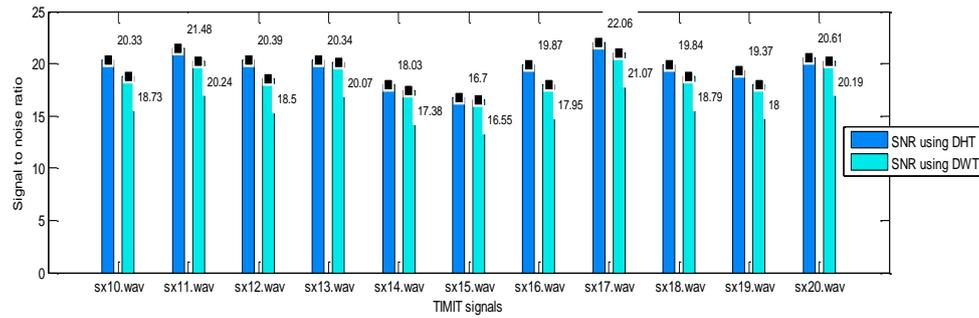


Figure 6. Signal to Noise Ratio (SNR) variation using DHT and DWT algorithms.

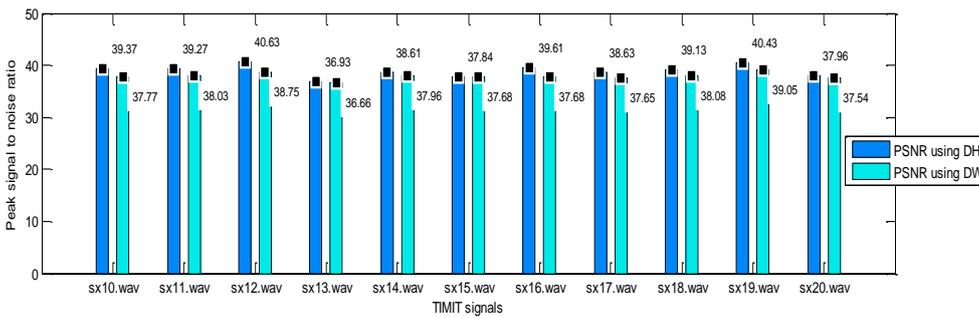


Figure 7. Peak Signal to Noise Ratio (PSNR) variation using DHT and DWT algorithms.

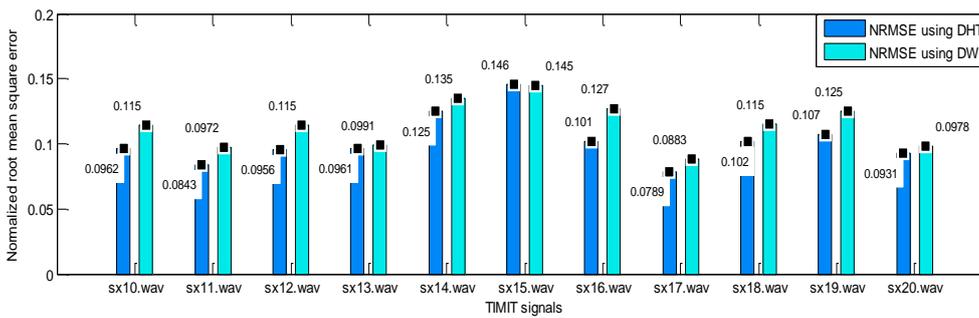


Figure 8. Normalized Root Mean Square Error (NRMSE) variation using DHT and DWT algorithms.

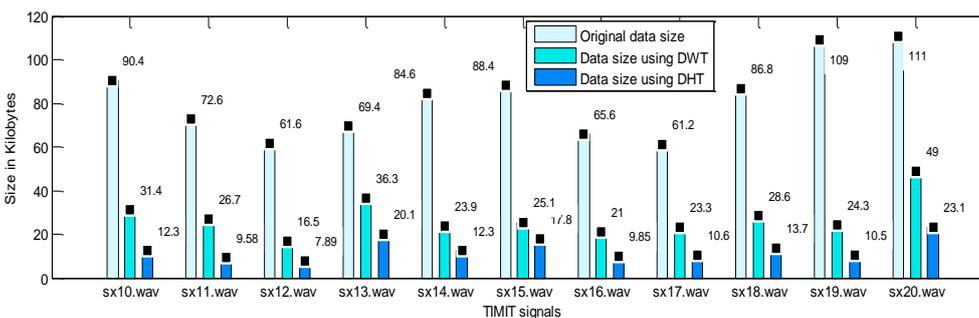


Figure 9. Sizes variation of some TIMIT files using DHT and DWT algorithms.

5. Conclusions

In this paper, a new algorithm for speech compression based on discrete Hartley transform is presented. The obtained results given by this work show clearly the effectiveness of the developed algorithm in the field of speech processing. When the developed algorithm is used to compress different speech signals, it gives a better compression performances (CR, SNR, PSNR, NRMSE) compared to the classical algorithm based on discrete wavelet transform. Finally, the use of this transform proves its effectiveness to compress speech signals.

References

- [1] Aggarwa R., Singh J., and Gupta V., "Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold," *International Journal of Computer Applications*, vol. 20, no. 5, pp. 14-19, 2011.
- [2] Ahmed N., Natarajan T., and Rao K., "Discrete Cosine Transform," *IEEE Transactions on Computers*, vol. 23, no. 1, pp. 90-93, 1974.
- [3] Aloui N., Barnoussi C., and Cherif A., "New Algorithm for QMF Banks Design and its Application in Speech Compression using DWT," *The International Arab Journal of Information Technology*, vol. 12, no. 1, pp. 86-93, 2015.
- [4] Bracewell R., "Discrete Hartley Transform," *Journal of the Optical Society of America*, vol. 73, no. 12, pp. 1832-1835, 1983.
- [5] Daqrouq K., Abu-Isbeih I., Daoud O., and Khalaf E., "An Investigation of Speech Enhancement using Wavelet Filtering Method," *International Journal of Speech Technology*, vol. 13, no. 2, pp. 101-115, 2010.
- [6] Fan Y., Madisetti V., and Mersereau R., "On Fast Algorithms for Computing the Inverse Modified Discrete Cosine Transform," *IEEE Signal Processing letters*, vol. 6, no. 3, pp. 61-64, 1999.
- [7] Friedrich T. and Schuller G., "Spectral Band Replication Tool for Very Low Delay Audio Coding," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustic*, New Paltz, pp. 199-202, 2007.
- [8] Grossman A. and Morlet J., "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape," *SIAM Journal on Mathematical Analysis*, vol. 15, no. 4, pp. 723-736, 1984.
- [9] Hartley R., "A More symmetrical Fourier Analysis Applied to Transmission Problems," *Proceedings of the IRE*, vol. 30, no. 3, pp. 144-150, 1942.
- [10] Hatem E., Mustafa J., and Mohammed T., "Speech Compression Using Wavelets," <http://site.iugaza.edu.ps/helaydi/files/2010/02/Elaydi.pdf>, Last Visited, 2010.
- [11] John P. and Girija P., "A High Performance Novel Image Compression Technique using Hybrid Transform for Multimedia Applications," *International Journal of Computer Science and Network Security*, vol. 11, no. 4, pp. 119-125, 2011.
- [12] Kodandaram R., Mallikarjun S., Krishnamuthan M., and Sivan R., "Face Recognition using Truncated Transform Domain Feature Extraction," *The International Arab Journal of Information Technology*, vol. 12, no. 3, pp. 211-219, 2015.
- [13] Kumar A., Singh G., Rajesh G., and Ranjeet K., "The Optimized Wavelet Filters for Speech Compression," *International Journal of Speech Technology*, vol. 16, no. 2, pp. 171-179, 2013.
- [14] Mandal J. and Ghosal S., "Separable Discrete Hartley Transform Based Invisible Watermarking for Color Image Authentication," in *Proceedings of the 2nd International Conference on Advances in Computing and Information Technology*, Chennai, pp. 767-776, 2012.
- [15] Nagaswamy S., Comparison of CELP Speech Coder with a Wavelet Method, Thesis, University of Kentucky UKnowledge, 2006.
- [16] Najih A., Ramli A., Ibrahim A., and Syed A., "Comparing Speech Compression using Wavelets with other Speech Compression Schemes," in *Proceedings of Student Conference on Research and Development*, Putrajaya, pp. 55-58, 2003.
- [17] Nikil S. and Chen E., "Audio Compression: Technology and Applications," *AT and T Technical Journal*, vol. 74, no. 2, pp. 23-34, 1995.
- [18] Rajesh G., Kumar A., and Ranjeet K., "Speech Compression using Different Transform Techniques," in *Proceedings of 2nd International Conference on Computer and Communication Technology*, Allahabad, pp. 146-151, 2011.
- [19] Shah D. and Vithlani C., "VLSI-Oriented Lossy Image Compression Approach using DA-Based 2D-Discrete Wavelet," *The International Arab Journal of Information Technology*, vol. 11, no. 1, pp. 59-68, 2014.
- [20] Soon I. and Koh S., "Speech Enhancement Using 2-D Fourier Transform," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 717-724, 2003.
- [21] Trivedi N., Kumar V., and Singh S., "Speech Recognition by Wavelet Analysis," *International Journal of Computer Applications*, vol. 15, no. 8, pp. 27-32, 2011.



Nouredine Aloui is working as Assistant Professor in the Centre for Research on Microelectronics & Nanotechnology (CRMN), Sousse Technology Park, Tunisia. He is a researcher member in the laboratory of Innovation of communicant and cooperative mobiles (Innov'Com), the Higher School of Communication of Tunis (Sup'Com). His research interests embedded systems and signal processing.



Souha Bouselmi is currently pursuing the Ph.D. degree in electronics at the faculty of sciences of Tunis. She is a researcher member in the laboratory of Innovation of communicant and cooperative mobiles (Innov'Com), the Higher School of Communication of Tunis (Sup'Com). Her research interests signal compression.



Adnane Cherif is a Professor at the faculty of sciences of Tunis and responsible in the laboratory of Innovation of communicant and cooperative mobiles (Innov'Com), the Higher School of Communication of Tunis (Sup'Com). His research interests signal processing and renewable energy.