Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Voice compression using discrete cosine transform and wavelet transform

To cite this article: J A Rolon-Heredia *et al* 2019 *J. Phys.: Conf. Ser.* **1403** 012001

Voice compression using discrete cosine transform and wavelet transform

J A Rolon-Heredia[1], V M Garrido-Arevalo[1], and J Marulanda[2]
[1] Universidad Tecnológica de Bolívar, Colombia
[2] Universidad Tecnológica de Pereira, Colombia

E-mail: jrolon@utb.edu.co

Abstract. Compressing information, such as voice, allows freeing up memory space, better utilizing bandwidth, in other words taking advantage of resources more efficiently in a time where technologies such as Voice over Internet Protocol are booming and the demand for information it requires higher speeds and quality. Therefore, this paper presents the acquisition and digital processing of voice signals, as well as the application of the discrete cosine transform and the wavelet transform using Matlab software version 2017b, licensed by the Technological University of Bolivar. Likewise, a criterion is established in order to eliminate little relevant information without affecting the quality of the output signal for which indicators such as the signal to noise ratio and the mean square error are calculated.

1. Introduction

The compression of data, particularly the human voice, has been a constant problem, because more and more information is available to be sent on the same channel with its respective fixed bandwidth [1-4]. An example of this is cellular networks where many users share the same bandwidth. Compression allows more users to share the system at any given time. However, the compression of the human voice poses significant challenges beyond the introduction of noise, since in addition to the message itself, the voice can transmit feelings and emotions that in most cases can significantly alter the meaning of the message transmitted. That is why the voice compression techniques [5-7] and [8] must allow, in addition to the obvious reduction in the size of the information, that the receiver can identify the full meaning of the message transmitted. Some techniques used in voice compression have also been used in image compression [9-11]. In [12] the use of wavelet transform is proposed to compress ECG signals by parameterizing wavelet mothers, achieving a reduction in the information rate without damaging the relevant diagnostic data when reconstructing the signals. This technique was tested in the MIT-BIH arrhythmia database and it was found that there was low distortion with a good compression ratio, preserving the important forms and amplitudes of the ECG. The described method is flexible to control the quality of the reconstructed signals and the volume of the compressed signals, by establishing an objective quadratic mean error and a priori compression ratio. Taking into account the above, in this paper, a methodology for signal compression will be provided based on eliminating redundancy between neighboring samples, that is, representing the data by the smallest possible number of coefficients within an acceptable loss of visual quality.

2. Transforms

2.1. Discrete cosine transform

The discrete cosine transform (DCT) is a variation of the discrete Fourier transform where the image is broken down into sums of cosines (and not of sines and cosines as in Fourier's), therefore, it is a real transform, it is given by the Equation (1):

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \quad (1)$$

As well as, the inverse transform is defined by Equation (2).

$$f(x) = \sum_{x=0}^{N-1} \alpha(u) C(u) \cos\left(\frac{(2x+1)u\pi}{2N}\right), \quad (2)$$

where: $\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} \text{ para } u = 0 \\ \sqrt{\frac{2}{N}} \text{ para } u = 1,2,\dots,N-1 \end{cases}$. It can be seen in Equation (2) that for $u = 0$, $C(u=0) = \sqrt{\frac{1}{N}} \sum_{x=0}^{N-1} f(x)$.

2.2. Wavelet transform

It is possible to analyze a signal in time and frequency using a technique called multiresolution analysis, this analyzes the signal for different frequencies with different resolutions [13]. This analysis is designed to provide good temporal resolution and poor frequency resolution for high frequencies and good frequency resolution and low time for low frequencies. In addition, it is the basic idea behind the discrete wavelet transform (DWT) and from it is the great advantage that it presents against the Fourier transform with windows. In the latter case, once the scale of our window has been chosen, the entire signal will be analyzed with the same resolution, while in the case of the DWT there is a variable resolution [14]. This type of analysis is based on obtaining the spectral components of a signal as amplitudes of a series of special functions, called wavelets. It is necessary that the set of functions used form a base, which implies that any signal has a single possible decomposition and that this decomposition can then be reversed obtaining the original signal [15]. For the discrete case, the DWT is defined by Equation (3):

$$C[j,k] = \sum_{n \in Z} f[n] \Psi_{j,k}[n], \quad (3)$$

where the mother wavelet is presented in Equation (4):

$$\Psi_{j,k}[n] = 2^{-\frac{j}{2}} . \Psi\left[2^{-j} n - k\right] \; j,k \in Z \quad (4)$$

The parameters τ, s are defined according to the dyadic scale i.e. in powers of 2^n, so that $\tau = 2^j k$ and $s = 2^j$ [15]. That is, the DWT will consist of multiplying the study signal ($f[n]$), by the wavelet on each scale j while the latter travels along the entire time axis. The process is repeated, but with different scales until the desired number of bands or scales is obtained [13]. The analysis can be interpreted as a measure of the similarity between the base functions ($\Psi_{j,k}[n]$) and the study signal ($f[n]$), in their frequency content. Therefore, the calculated coefficients ($C[j,k]$) (Equation (3)) indicate how close the signal is to the wavelet on a given scale. If an orthonormal base is available, the decomposition can be undone and the original signal obtained, which is known as inverse transform and it is defined by Equation (5).

Journal of Physics: Conference Series **1403** (2019) 012001 doi:10.1088/1742-6596/1403/1/012001

$$f[n] = \sum_{j \in Z} \sum_{k \in Z} C[j,k].\Psi_{j,k}[n] \tag{5}$$

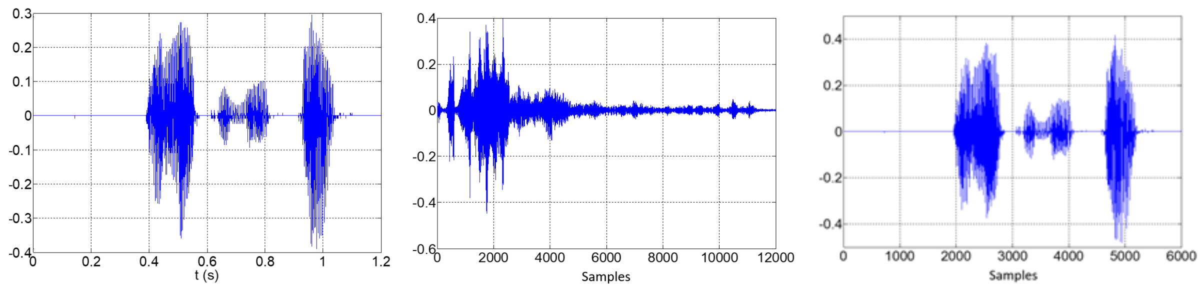
That is, the original signal ($f[n]$) is recovered, from the wavelet base ($\Psi_{j,k}[n]$), and the coefficients obtained by the DWT ($C[j,k]$).

3. Methodology

3.1. Acquisition of the input signal
In Figure 1(a) you can see the input signal as a function of the time that corresponds to the phrase "hello how are you".

3.2. Discrete cosine transform / wavelet transform application
In Figure 1(b) the distribution of the coefficients from the application of the DCT is observed, it is important to note that, compared to the input signal, the DCT has greater amplitude but less data, making the transmission of this, more efficient versus sending the original signal. Figure 1(c) shows the DWT of the input signal with a level of 2; It is important to note that it can be seen that the transform contains less data than the input signal, as happened in the application of the DCT.

3.3. Compression
In the case of study, compression is understood as the elimination of little relevant information, while maintaining an acceptable level of the output signal, this acceptable level refers to making the output signal understandable. To achieve this, it is necessary to eliminate all the terms of the matrix of coefficients that are between ± 30% of the maximum value.



(a) (b) (c)

Figure 1. (a) Input signal, (b) DCT, and (c) DWT.

3.4. Evaluation of the quality of the reconstructed signal
To assess the quality of the output signal, the signal to noise ratio (SNR) that is defined by Equation (6) and the mean square error with respect to the original signal, defined by Equation (7), will be used.

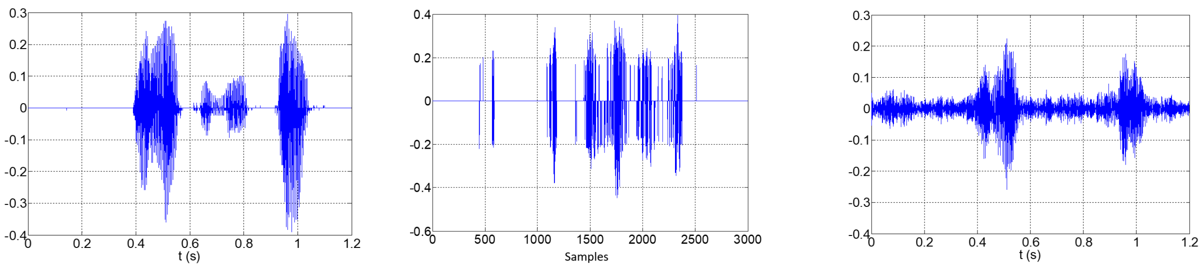
$$\mathrm{SNR} = \frac{\mathrm{Psignal}}{\mathrm{Pnoise}} \tag{6}$$

$$\mathrm{EMC} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{7}$$

4. Results

4.1. System implementation using discrete cosine transform

By applying the IDCT to the signal shown in Figure 1(b), Figure 2(a) is obtained in response, it can be noted that this signal is identical to the input signal (Figure 1(b)) and that there is no offset in them. The mean square error between these two signals has a value of $4.797\text{x}10^{-17}$, which, in practice, is equivalent to zero, therefore it can be indicated that the process carried out does not present information loss and It is of excellent quality. Figure 2(b) shows the compressed signal and with the terms removed. The way in which the amount of information contained in the coefficients before and after truncation is reduced is easily observed, although this reduction of information does not necessarily constitute significant losses in the output signal. In Figure 2(c) the output signal is shown before the input of the trimmed data; In this case, the mean square error is found again, resulting in an error of 3.53%, this being an acceptable error since when reproducing the output signal, it is still understandable.



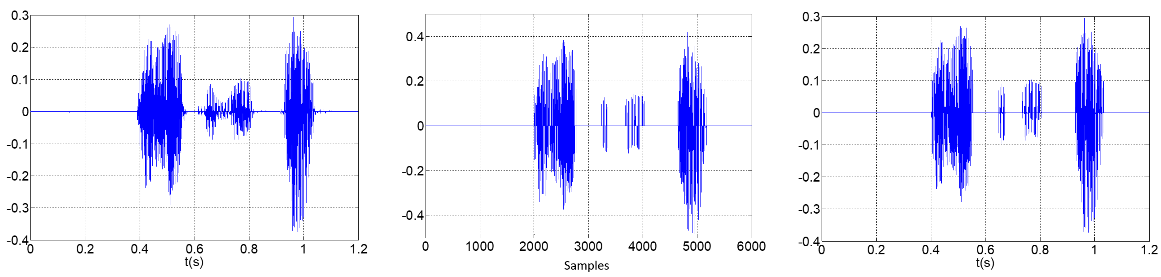
(a) (b) (c)

Figure 2. (a) Output signal DCT, (b) DCT (data cropping) and (c) IDCT (data cropping).

4.2. System implementation using discrete wavelet transform

The signal presented in Figure 3(a) is obtained by applying the IDWT to the signal in Figure 1(b); As expected, this signal has a very low error, similar to that presented in the case of the IDCT without elimination of coefficients.

Figure 3(c) shows the output signal of the system after the wavelet inverse transform has been applied, having as input the signal shown in Figure 3(b) It can be seen that in this case the output signal is much more similar to the signal input that in the case of the DCT (See Figure 2(c)), even in the case of wavelet, the amplitudes of the input and output signal are equal, while in the DCT these amplitudes differ, for this transformed the average error quadratic is 1.52% which in turn is much lower than the system error with DCT.



(a) (b) (c)

Figure 3. (a) Output signal DWT, (b) DWT (data cropping) and (c) IDWT (data cropping).

Table 1 shows the comparison between the lengths, power, SNR and MSE of each of the signals used and processed in this paper. In this sense it can be seen that it is DCT WDC that has the worst MSE result, however, its SNR is the best of all signals. On the other hand, although the SNRs of the signals with DWT and DWT with DC are low, their MSE are much lower still and the signal length is reduced by half compared to the DCT.

Table 1. Length and power of the signals under studio.

	Length	Power	SNR	MSE
Original Signal	12000	0.0024	-	-
DCT	12000	0.0013	0.0000	0.0000
DCT (data cropping)	12000	0.0008	3.1519	3.3338
DWT	6003	0.0015	0.1036	0.7572
DWT (data cropping)	6003	0.0012	0.4339	1.5203

5. Conclusions

The DCT and wavelet compression techniques were simulated, obtaining the results for each of them. At this point it is important to highlight that the technique that presents the best results is the one used by the avelet transform because in this case the mean square error between the input and the output signal is lower than in the case of the DCT, in addition to this, the wavelet transform does not need an additional block for compression of the data due to its own form, since every time the data is subjected to filters it passes high and passes low the data of minor relevance is eliminated. On the other hand, the mean square errors were calculated for the cases where the transformed and the reverse transformed were applied successively, obtaining very low errors, of the order of $4.3797x10^{-15}$%. In addition to the mean quadratic error, the signal-to-noise ratio indicator was calculated, which allowed us to verify the quality of the output signal, complemented by the figure of the mean square error, resulting in the technique with the best response being the one that uses the discrete wavelet transform.

References

[1] Qian Y and Chen W 2017 Adaptive bayesian compressed sensing based on speech frame signal *9th International Conference on Communication Software and Networks (ICCSN)* (China: IEEE) p 1047
[2] Parkale Y V and Nalbalwar S L 2016 Application of 1 D discrete wavelet transform based compressed sensing matrices for speech compression *Springer Plus* **5(2048)** 1
[3] Sheikh J A, Parah S A, Akhtar S and Bhat G M 2017 Compression and denoising of speech transmission using Daubechies wavelet family *Int. J. Wireless and Mobile Computing* **12(4)** 313
[4] Ram R and Mohnaty M N 2018 Enhancement of speech using deep neural network with discrete cosine transform *Journal of Intelligent & Fuzzy Systems* **35** 1
[5] Sheikh J A *et al.* 2016 On the design and performance evaluation of dwt based compressed speech transmission with convolutional coding *Communications on Applied Electronics* **4(9)** 36
[6] Sheikh J A and Akhtar S 2016 Realization and performance evaluation of new hybrid speech compression technique *International Journal of Modern Electronics and Communication Engineering* **4(3)** 16
[7] Viga R and Chauhanb S 2018 Speech compression using multi-resolution hybrid wavelet using DCT and Walsh transforms *Procedia Computer Science* **132** 1404
[8] Gibson J D 2016 Review speech compression *Information* **7(2)** 1
[9] Tun H M, Moe W K and Naing Z M 2017 Analysis on ECG data compression using wavelet transform technique *International Journal of Psychological and Brain Sciences* **2(6)** 127
[10] Soliman N F, Abd-Alhalem S M, El-Rahman S A, Fouad M and Abd El-Samie F 2016 Speech transmission with COFDM based on different discrete transforms *International Journal of Speech Technology* **19(3)** 565
[11] Wang S *et al.* 2018 Suppression by selecting wavelets for feature compression in distributed speech recognition *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **26** 564
[12] Mohammed A Z *et al.* 2013 A new algorithm for compression of ECG signals based on mother wavelet parameterization and best-threshold levels selection *Digital Signal Processing* **23** 1002
[13] Navarrete Mejía M. and Andrade Mora S. 2006 *Análisis y tratamiento de la señal electrocardiográfica para la detección de parámetros de normalidad bajo la plataforma Labview "Adpan-Ecg"* (Bucaramanga: Universidad Pontificia Bolivariana) p 26
[14] Faundez P. 2000 *Procesamiento digital de señales acústicas utilizando wavelets* (Valdivia: Universidad Austral de Chile) p 59
[15] Cabrera M. and Pérez A 2005 *Análisis de señales biomédicas mediante transformada wavelet* (Buenos Aires: Universidad Nacional San Martín)