The Effect of Different Compression Schemes on Speech Signals

Jalal Karam, and Raed Saad

***Abstract*—**This paper studies the effect of different compression constraints and schemes presented in a new and flexible paradigm to achieve high compression ratios and acceptable signal to noise ratios of Arabic speech signals. Compression parameters are computed for variable frame sizes of a level 5 to 7 Discrete Wavelet Transform (DWT) representation of the signals for different analyzing mother wavelet functions. Results are obtained and compared for Global threshold and level dependent threshold techniques. The results obtained also include comparisons with Signal to Noise Ratios, Peak Signal to Noise Ratios and Normalized Root Mean Square Error.

***Keywords*—**Speech Compression, Wavelets.

I. Introduction

APPLICATIONS to Arabic speech compression involve real time coding of speech for mobile satellite communication, cellular phones, and audio for videophones or video teleconferencing system. Other applications involve vocoding of speech signals for storage, synthesis and transmission [8]. The DWT of a given speech signal concentrates speech energy in few neighboring coefficients allowing natural compression. In this paper we introduce a flexible compression scheme that uses wavelets and their transforms. The flexibility of this new paradigm is attained by observing:

i. The analyzing wavelet used
ii. Decomposition level
iii. Compression ratios
iv. Frame size
v. Measured parameters
vi. Type of threshold used

In this paper, a flexible paradigm depicted in Fig. 5 is introduced to compress Arabic speech signals. The signal is first divided it into different size frames, which are then analyzed using particular mother wavelets up to a level 7 representation using DWT.

The Arabic digits are the focus of compression, namely. Different compression parameters are calculated for these signals and compression ratios are derived for to different types of compression schemes, namely, the Level Dependent and Global Threshold techniques. The following section introduces wavelets and two of their related transforms, namely, the Continuous Wavelets Transform (CWT) and the DWT. While speech compression is discussed in Section 3 with some details, the implementation of the system and compression parameters used in this work are included in Section 4. The last two sections give a detailed discussion on the results obtained, and the conclusion of this paper.

II. Wavelets

A wavelet is a finite energy signal defined over specific interval of time [1]. The main interest in wavelets is their ability to represent a given signal at different. Wavelets are used to analyze signals in much the same way as complex exponentials (sine and cosine) used in Fourier analysis of signals. Unlike Fourier, wavelets can be used to analyze non-stationary, time-varying, or transient signals [9] [10]. This is an important aspect, since speech signals are considered to be non-stationary. A given signal is represented by using translated and scaled versions of a mother wavelet as it is explained below. They are also localized in time and frequency domains [9].

A. Continuous Wavelet Transform

The wavelet transform is a two parameter expansion of a signal in terms of a particular wavelet basis function [1]. Given Ψ(t) called the mother wavelet; all other baby wavelets are obtained by simple scaling and translation of Ψ(t). $\Psi a,t(t) = (1/\sqrt{a})\ \Psi[(t-b)/a]$.

Where a and b are the scaling and the translation parameter respectively. A nice approach to the CWT representation is first to inspect the Fourier transform represented mathematically by:

$$F(\omega) = \int s(t)e^{-j\omega t}dt$$

Replacing the complex exponential in the Fourier transform with Ψa,t(t) yields:

$$C(S,U) = \int \sqrt{a}\ \Psi[(t-b)/a]dt$$

In other words with wavelet transform, reference to frequency is replaced by reference to scale [8], [9][10].

Manuscript received September 23, 2006.

Jalal Karam, is with the Gulf University of Sciences and Technology, Department of Mathematics and Natural Sciences, Hawally 32093, P.O. Box 7207, Kuwait City, Kuwait (phone: +965-779-3890; e-mail: karam.j@gust.edu.kw).

Raed Saad was with Lebanese American University, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: Raed.Saad@lau.edu.lb).

B. Discrete Wavelet Transform

In our application the discrete wavelet transform is applied. By choosing scale and position based on power of 2, CWT is reduced to DWT without any loss in energy. The scaling parameter is discrete and dyadic, $a = 2^{-j}$. The translation is discretized with respect to each scale by using $\tau = k2^{-j}T$ [5]. $\Psi j,k(t) = (2^{j/2})\ \Psi[(2^{j}t\text{-}kT)]$.

The integer k represents the translation of the wavelet function; it indicates time in wavelet transform. Integer j, however, is an indication of the wavelet frequency or spectrum shift and generally referred to as scale. The DWT transforms a discrete input signal vector into two sets of coefficients the approximation CA containing low frequency information and the detail coefficients CD containing high frequency information. Fig. 1 shows a level 2 DWT decomposition of an input signal s (t) [8] [9] [10].

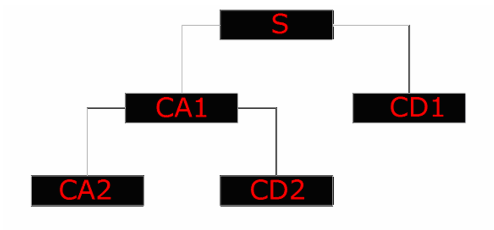

Fig. 1 Level 2 Decomposition

Most commonly used wavelets are categorized into two classes: orthogonal and bi-orthogonal wavelet system. Orthogonal wavelets decompose signals into well behaved orthogonal signal spaces. Biorthogonal wavelets are more complicated and are defined based on a pair of scaling and wavelet function. The wavelet of interest, the one used in this work, is the Daubechies wavelet family, in which carry out very unique compression properties, intended for wavelet coefficients.

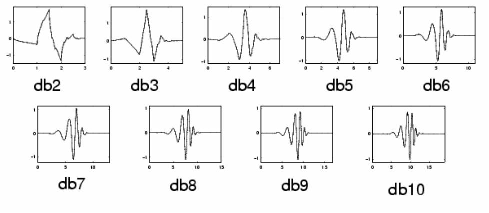

Fig. 2 Plots of Different Daubechies Orthogonal Wavelets

Fig. 2 shows many of the Daubechies orthogonal wavelets. One notes that **db10** is a much smoother function than the rest. In general for the Daubechies orthogonal wavelets db_i; as i increase the wavelet becomes smoother. In this paper, the analyzing functions chosen were db4, db8, db10 and db20. The DWT can be computed using octave band filter bank [6] [9]. The signal is split into two segments via a two-band filter bank, a low pass or lower resolution version, and a high pass one. The lower resolution version is then split again, and so on. This is illustrated in Fig. 1 and 3. The high pass filter HP generates high frequency coefficients containing low energy; these are the detail coefficients of the signal indicated as CDi. Low pass filter LP generates the approximation coefficients; designated by CA. Those coefficients contain most of the energy in the speech signal [10]. For multi resolution analysis CA are decomposed a level further into detail and approximation coefficients. The output of the HP filter is down sampled and fed into a detector to detect all coefficients below a certain threshold and replace them by a zero. The down sampling will retain N/2 of the signal coefficient the ones that are only needed. To reconstruct the signal we apply the Inverse Discrete Wavelet Transform (IDWT) illustrated in Fig. 4.

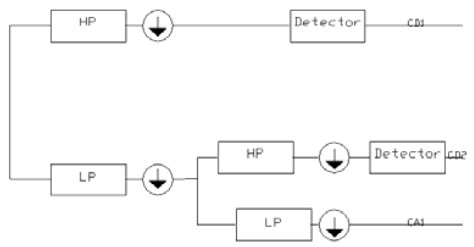

Fig. 3 DWT Decomposition (Analysis)

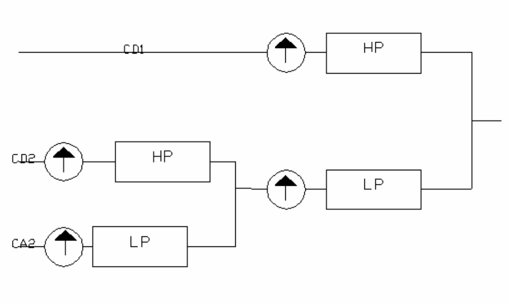

Fig. 4 IDWT Reconstruction (Synthesis)

III. Wavelet Compression

In many applications, speech signals are either stored for later use or transmitted over some media. In both cases, one is interested in reducing the size of the signal because of the cost, time and other benefits. The search arises for a mechanism to compress the signal and obtain lower bit rates. Different techniques were implemented each having its advantages and disadvantages. In general, they can be classified into two types, lossy and lossless. The Hufmann coding, for example, is considered to be lossless compression. The original data will be totally restored without any modifications. Lossy compression does not completely retain the original signal; consequently some of the information is lost. For speech signals, this loss is acceptable since we are interested only in recognizing the signal. Wavelets compression technique is considered to be lossy where the reconstructed signal is not an exact match of the original signal. One of the most important advantages of wavelet transform is that it concentrates speech formation (energy and perception) into a few neighboring coefficients [3]. Also when applying the DWT to a given speech signal many coefficients of small values (depending on level we choose) are thus considered insignificant. The retained coefficients will still have the larger percentage of energy in the signal. The process of compressing the digit signals is discussed in the next section which also contains the different mechanisms to be used during the process.

A. Data Base

The signals are chosen from the Arabic digit speech data base. The digits were spoken by different speakers and recorded in the studio at the Lebanese American University, Byblos. The small size studio is designed to minimize noise and equipped with a multi directional microphone made by Neumann to collect the speech signal with the best quality. The signals are then recorded and transformed into wave sounds (.wav). The tools used are the PRO-Tools Control 24 Dig-Design Device. This device is a computerized digital mixer.

B. Choosing the Decomposition Level

The DWT on a given signal, the decomposition level can reach up to level $L = 2^K$, where K is the length of the discrete signal. Thus we can apply the transform at any of these levels. But in fact, the decomposition level depends on the type of signal being analyzed. For the processing of speech signals, decomposition up to scale 7 is adequate [7]. In this paper, level 5 DWT is obtained for every signal and comparisons were made with level 6 and 7 decompositions.

C. Choosing Appropriate Wavelets

The type of wavelet is of high importance for such experiments. It directly affects the Signal to Noise Ratio (SNR) of the output signal. Choosing the appropriate wavelet will maximize the SNR and minimizes the relative error. As mentioned earlier Daubechies wavelets have good compression property for wavelet coefficients [1], giving better SNR ratios. Wavelets with more vanishing moments provide better reconstruction quality. Daubechies wavelets are developed with maximum regularity; the number of zero moments is maximized, leading to the best wavelet family for compression. The selected members of this orthogonal Daubechies family are db4, db8, db10 and db20.

D. Choosing the Frame Size

Dividing the speech signal in different frame sizes is used to examine their effect on the overall compression performance, since framing aims to improve the compression ratios effect. Framing aim to improve the compression ratios obtained. Three frame sizes are tested in this paper (20ms, 0.25s, and 0.5s). The frames obtained are analyzed separately being considered a vector in its own right.

E. Threshold Techniques

The coefficients obtained after applying DWT on the frame concentrate energy in few neighbors. Thus we can truncate all coefficients with “low” energy and retain few coefficients holding the high energy value. The two thresholding techniques are implemented according to the following algorithms [3].

1. Global Threshold

The global threshold technique works by retaining the wavelet transform coefficients which have the largest absolute value. For a given speech signal, the global threshold algorithm first divides the speech signal into frames of equal size **F**. Then the wavelet transform of each frame is computed. Usually with length **T** > **F**. These coefficients are sorted in an ascending order and the first **L** coefficients are retained. In practice, these coefficients along with their positions in the wavelet transform vectors are stored or transmitted [3][5]. For these reasons, 2.5***L** coefficients are used to represent the original **F** samples distributed as follows: 8 bits for the amplitude and 12 bits for the position leading to 2.5 bytes. The Compression Ratio **CR**, can then be defined by: **CR**=**F / 2.5*L.**

Each frame is reconstructed by replacing the missing coefficients by zeros.

2. Level Dependent Threshold

The level-dependent threshold technique is derived from the Birge-Massart strategy [5]. This strategy works on selecting the retained wavelet coefficients as follows. Let J_0 be the decomposition level, m the length of the coarsest approximation coefficients over 2, and α be a real greater that 1. At level J_0+1 (and coarser levels), everything is kept. For level J from 1 to J_0, the K_J larger coefficients in absolute value are kept using this formula:

$$K_J = \frac{m}{(J_0 + 1 - J)^{\alpha}}.$$

The value of α used is 1.5 as suggested in [5].

The value of the threshold applied depends on the compression ratio we want to achieve. The task is to obtain higher compression ratios and an acceptable SNR needed to reconstruct the signal and detect it. The signal is reconstructed by applying the Inverse Discrete Wavelet Transform IDWT as it is shown in Fig. 5.

IV. Design and Implementations

The introduced system is depicted in Fig. 5 and implemented and simulated to study its performance using Matlab®. Some of the functions used are, *Wavdec, which* computes the multi-level decomposition of the signal and *wavrec that* reconstructs the signal from the coefficients obtained. Two other important functions are: *wdencmp* returns the coefficients after applying a determined threshold. It also computes the percentage of energy retained and the percentage of truncated zeroes.

A. The Compression Parameters

In this paper, four compression parameters are used. They are defined next along with their mathematical expressions.

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

SNR = 10*log (σ_x^2/ σ_e^2)

Where σ_x^2 is the mean square of the speech signal and σ_e^2 is the mean square difference between the original and reconstructed signals.

- **Peak Signal to Noise Ratio**

$PSNR = 10*\log (NX^2 / ||x-r||^2)$

Where 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 difference between original and reconstructed signals.

- **Normalized Root Mean Square Error**

$NRMSE = \mathrm{sqrt}[(x(n)-r(n))^2/(x(n)-\mu_x(n))^2]$

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

- **Retained Signal Energy**

$RSE = 100*||x(n)||^2 / ||r(n)||^2$

Where ||x(n) || is the norm of the original signal and ||r(n) || is the norm of the reconstructed one. For db orthogonal wavelets the retained energy is equal to the L^2-norm recovery performance.

The speech signals compressed are the Arabic digits "Zero' and "eight". They are depicted in Fig. 6 and Fig. 7 respectively along with their compressed versions. These compressed versions were obtained using "db8" and the size of frame is 0.02s. Different Compression Ratios (CR) were obtained. The compressed speech signal is still audible and you can still recognize the output signal. Different parameters were examined when simulating the code. The 8 KHz sampled signals are divided into frames (0.2ms, 0.25s, and 0.5s) and decomposed up to level 5. Each frame is decomposed separately. At this stage the threshold is applied on the coefficients to truncate whatever unnecessary. The obtained coefficients are then used to reconstruct the output compressed signal. Different results were obtained allowing efficient evaluations and comparisons of the used methods and parameters.

V. Discussion

Average SNR, PSNR, and NRMSE are all measured given the frame size, the type of the mother wavelet ψ(t) and value of the threshold. Also, the percentage of Zeros (%Z) and the percentage of the Energy Retained (%ER) are included. Another set of experiment is done on the Arabic Digit zero. In these set of experiments, the threshold does not apply on the approximation coefficients. We are able to disregard more than 92% of the signal coefficients and still retain about 76 % of the energy in the signal using a frame size of 0.02ms and applying the global threshold. That is, keeping track of 8% of the frame to be reconstructed later. Furthermore, an increase in the length of the frame to 0.25s achieved better results. Less coefficients containing more energy and the SNR is maximized while the NRMSE is minimized. Changing the wavelet also has its effect on the compression ratio. If we use a smooth analyzing wavelet like db10 the percentage of the truncated coefficients decreased. However; they produced better SNR. Thus, these wavelets play the role of producing better SNR but less compressing ratios. On the other hand, un-smooth wavelets such as db4 lead to better compression ratios but resulting with low SNR.

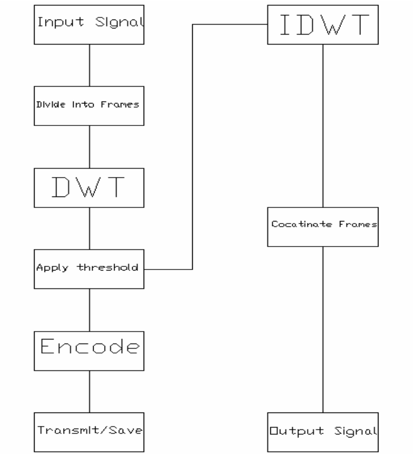

Fig. 5 The Introduced Flexible Paradigm

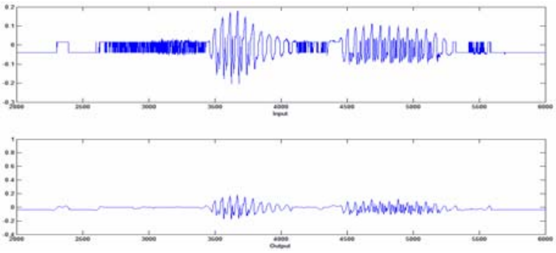

Fig. 6 Arabic Digit zero with compressed version. The CR = 7.65

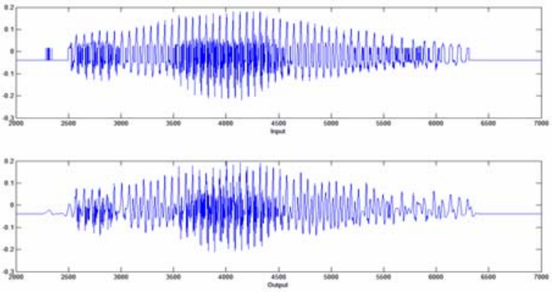

Fig. 7 Arabic Digit Eight with compressed version. The CR = 5.5

VI. Conclusion

In this paper, the performance of the Discrete Wavelet Transform in compressing speech signals is tested and the following points were observed. High compression ratios were achieved with acceptable SNR. No further enhancements were achieved beyond level 5 decomposition. The effect of frame size and the Level Dependent Threshold on the NRMSE is evident while this measurement remains almost constant for all experiments with negligible changes. Increasing the frame size, positively affects the overall performance in both threshold techniques used. Overall global threshold leads to better results than the level dependent threshold technique in the case of SNR and CR. This was the case with and without framing and for both tested digits. It is worthwhile noting that we could not pinpoint the best compression wavelet.

References

[1] Y. T. Chan “Wavelet Basics”, Kluwer Academic Publishers, 1995.

[2] C. Taswell, “Speech Compression with Cosine and Wavelet Packet Near-Best Bases”, ICASSP-96. Conference Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.1, pp: 566 – 568, 1996.

[3] E. Fgee, W. J. Phillips, W.Robertson “Comparing Audio Compression Using Wavelets With Other Audio Compression Schemes”, IEEE, CCECE Conference Edmonton, Alberta, Canada. May 9-12, pp: 698-701, 1999.

[4] N. M. Hosny, S. H. El-Ramly, M. H. El-Said, “Novel Techniques for Speech Compression Using Wavelet Transform”, ICM '99. 11th International Conference on Microelectronics, pp: 225 – 229, 1999.

[5] http://www.mathworks.com/academia MATLAB Wavelet Toolbox 2.

[6] O. Rioul and M. Vetterli, “Wavelets and Signal Processing”, IEEE Signal Process. Mag. Vol 8, pp. 14-38, Oct. 1991.

[7] J. I. Agbinya, “Discrete Wavelet Transform Techniques in Speech Processing,” IEEE Digital Signal Processing Applications Proceedings, IEEE, New York, pp: 514 – 519, 1996.

[8] J. F. Koegel Buford, “Multimedia Systems”, ACM Press, 1994.

[9] G. Strang and T. Nguyen, “Wavelets and Filter Banks”, Wellesley-Cambridge Press, 1996.

[10] J. S. Walker, “Wavelets and their Scientific Applications”, Chapman and Hall/CRC, 1999.