International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:1, 2008

End Point Detection for Wavelet Based Speech Compression

Jalal Karam

***Abstract*—**In real-field applications, the correct determination of voice segments highly improves the overall system accuracy and minimises the total computation time. This paper presents reliable measures of speech compression by detcting the end points of the speech signals prior to compressing them. The two different compession schemes used are the Global threshold and the Level-Dependent threshold techniques. The performance of the proposed method is tested wirh the Signal to Noise Ratios, Peak Signal to Noise Ratios and Normalized Root Mean Square Error parameter measures.

***Keywords*—**Wavelets, End-points Detection, Compression.

I. Introduction

A Speech compression system focuses on reducing the amount of redundant data while preserving the integrity of signals. The different transformation of speech signals to the time-frequency and time-scale domains for the purpose of compression aim at representing them with the minimum number of coding parameters. This paper examines the effect of such representations induced by eliminating un-necessary data present at both ends of the signals. Compression parameters are computed for speech frames at level 5 Discrete Wavelet Transform (DWT) representation of the signals following a Global threshold and a Level-Dependent threshold techniques. In the last few years, the architecture of speech processing paradigms have been perturbed by the introduction of the new powerful analysis tool called wavelets. The theory of wavelets is a product of many independent developments in the fields of pure and applied mathematics, electrical engineering, quantum physics and seismic geology. The interchange between these areas in the last decade produced many new important and vital wavelet applications such as image and signal compression, turbulence, human vision, radar and earthquake prediction [2] to name a few. Section 2 of this paper introduces the details of the mother wavelet used as the analysing function of the speech signals. Section 3 gives a complete discription of the Global thresholding and the Level thresholding schemes used for compreesion. and Section 4 discuss speech compression using wavelets and compare two common threshold approaches.

II. Daubechies Related Wavelet

In [3], it was shown that reversing the coeficients of the filters in the time-domain preserves perfect reconstruction in the filter bank and allows the construction of new scaling and wavelet functions. So, given the filter:

Jalal Karam is the head of the department of mathematics and natural sciences at Gulf University for Science and Technology, Kuwait City, Kuwait, email: karam.j@gust.edu.kw

$$H_0 = [h_0, h_1, ..., h_N]$$

reversing its coefficients produces the filter:
$Rev(H_0) = H_{0r} = [h_N, h_{N-1}, ..., h_0]$.

The alternating flip filter of $Rev(H_0)$ is the filter:

$$H_{1r} = [h(0), -h(1), +h(2), ..., +h(n)].$$

The order flip of H_{0r} is the filter

$$F_{0r} = [h_0, h_1, ..., h_N].$$

Finally, the alternating sign filter of $Rev(H_0)$ is the filter:

$$F_{1r} = [-h_N, +h_{N-1}, ..., +h_0].$$

The filters H_{0r}, H_{1r}, F_{0r} and F_{1r} satisfy all the three time domain conditions of perfect reconstruction. From the coefficients of the wavelet $db6$, the coefficients of the new wavelet are thus derived. Figure 1, shows the scaling and wavelet functions of the derived wavelet. One notes the reflective similarities among the scaling and wavelet functions of $db6$ with those of the derived wavelet. This mother wavelet is employed as the analyzing function for the speech signals used for the experiments conducted in this paper.

III. Speech Compression

The goal of using wavelets to compress speech signal is to represent a signal using the smallest number of data bits commensurate with acceptable reconstruction and smaller delay. Wavelets concentrate speech information (energy and perception) into a few neighboring coefficients, this means a small number of coefficients (at a suitably chosen level) will remain and the other coefficients will be truncated [1]. These coefficients will be used to reconstruct the original signal by putting zeros instead of the truncated ones.

A. Thresholding Techniques

Thresholding is a procedure which takes place after decomposing a signal at a certain decomposition level. After decomposing this signal a threshold is applied to coefficients for each level from 1 to N (last decomposition level). This algorithm is a lossy algorithm since the original signal cannot be reconstructed exactly [6]. By applying a hard threshold the coefficients below this threshold level are zeroed, and the

output after a hard threshold is applied and defined by this equation :-

$$y_{hard}(t) = \begin{cases} x(t), & |x(t)| > \delta \\ 0, & |x(t)| \leq \delta \end{cases} \quad (1)$$

where $x(t)$ is the input speech signal and δ is the threshold. An alternative is soft thresholding at level δ which is chosen for compression performance and defined by this equation :-

$$y_{soft}(t) = \begin{cases} sign(x(t))(|x(t)| - \delta), & |x(t)| > \delta \\ 0, & |x(t)| \leq \delta \end{cases} \quad (2)$$

where equation 1 represents the hard thresholding and equation 2 represents the soft thresholding.

B. Thresholding methods used in Wavelets Compression

In this section two thresholding algorithms will be introduced and later used in compressing speech signals. These two methods are, Global thresholding and Level dependent thresholding.

C. Global Thresholding

Global thresholding [4] works by retaining the wavelet transform coefficients which have the largest absolute value. This algorithm starts by dividing the speech signal into frames of equal size F. The wavelet transform of a frame has a length T (larger than F). These coefficients are sorted in a ascending order and the largest L coefficients are retained. In any application these coefficients along with their positions in the wavelet transform vector must be stored or transmitted. That is, $2.5L$ coefficients are used instead of the original F samples, 8 bits for the amplitude and 12 bits for the position which gives 2.5 bytes [1]. The compression ratio C is therefore:

$$C = \frac{F}{2.5L} \quad or \quad L = \frac{F}{2.5C} \quad (3)$$

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

D. Level Dependent thresholding

This compression technique is derived from the Birge-Massart strategy [5]. This strategy works by the following wavelet coefficients selection rule :
Let J_0 be the decomposition level, m the length of the coarsest approximation coefficients over 2, and α be a real greater than 1 so :

1) At level J_0+1 (and coarser levels), everything is kept.
2) For level J from 1 to J_0, the K_J larger coefficients in absolute value are kept using this formula :-

$$K_J = \frac{m}{(J_0 + 1 - J)^{\alpha}} \quad (4)$$

The suggested value for α is 1 and was used in [4] [5].

E. Interpretation of the two algorithms

These algorithms are used to compress speech signals and compare the quality of the reconstructed signal with the original. In this section, outlines the steps followed in implementing these algorithms.

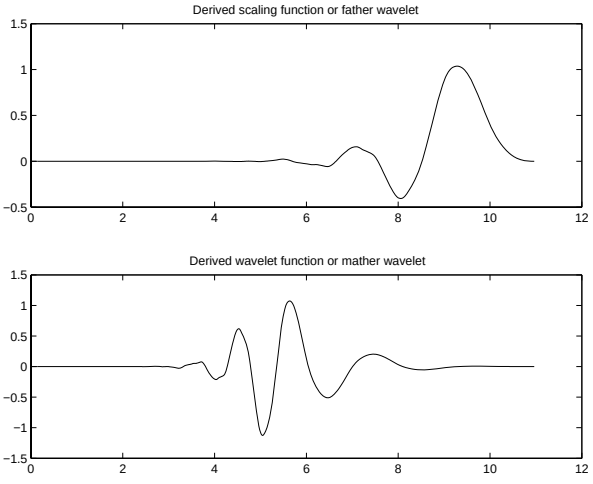

Fig. 1. The scaling and wavelet functions of the derived wavelet.

F. Compression using the Global Thresholding

The following procedure is usually followed to implement the global thresholding to compress speech signals.

1) Divide the speech signal into frames of equal size. In this paper different frame sizes are tested to see how the frame size will affect the performance of the reconstructed signal. Three different frame sizes are examined since wavelet analysis is not affected by the stationarity problem. Expanding the frame length will speed up the processing time which reduces the processing delay.
2) Apply the discrete wavelet transform to each one of these frames separately at the five decomposition levels. This level is chosen since the best performance of the reconstructed signal is obtained at this level.
3) Sort the wavelet coefficients in a ascending order.
4) Apply the global thresholding to these coefficients by choosing the compression ratio and using equation 3 to obtain the non zero coefficients.
5) Keep the retained coefficients and their positions to reconstruct the signal from them.
6) Reconstruct the compressed frames by using the non zero coefficients and their positions and replacing the missing ones by zeros.
7) Repeat steps 2 to 6 to compress all the frames.
8) Insert these reconstructed frames into their original positions to get the reconstructed signal.

G. Compression Using Level-dependent Thresholding

After the speech signal is divided into equal frame sizes, the following steps are to be followed to implement the level dependent thresholding.

1) Apply the wavelet decomposition to each frame separately.
2) Keep all the coefficients of the last approximation, and use equation 4 to retain coefficients from each detail level.

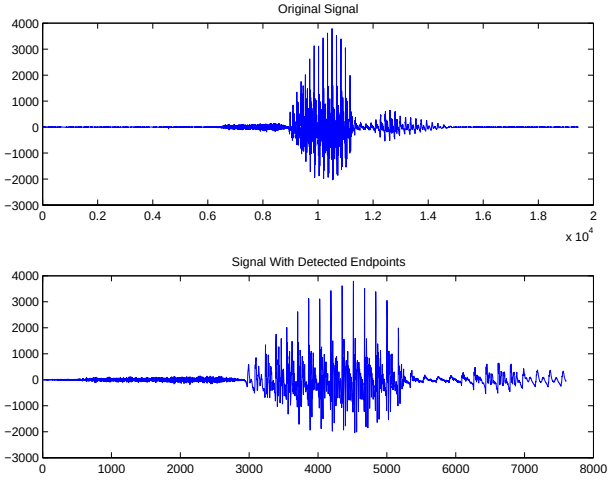

Fig. 2. The signal representing the word ”Seven” in its original form and its end-points detected form.

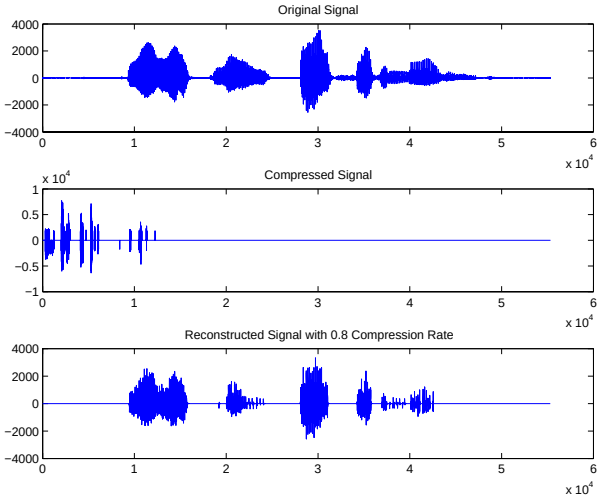

Fig. 3. The Original, Compressed with Level Dependent Threshold at 0.8 compression ratio, and Reconstructed signals of the word ” 4035789 ” all in the time domain.

3) Decompose all the frames and apply step 2 to each one of the frames, then keep the non zero coefficients and their positions using 2.5 bytes as in the global thresholding.
4) Reconstruct each decomposed frame using the non zero coefficients and replace the missing ones by zeros.
5) Insert these reconstructed frames into their original positions to get the reconstructed signal.

Figure 3 shows the speech signal of the word ” 4035789 ” in its original time domain representation, its compressed version using Level Dependent Threshold at 0.8 compression ratio, and the reconstructed signal. Figure 4 shows the same representations for the end-point detected signal.

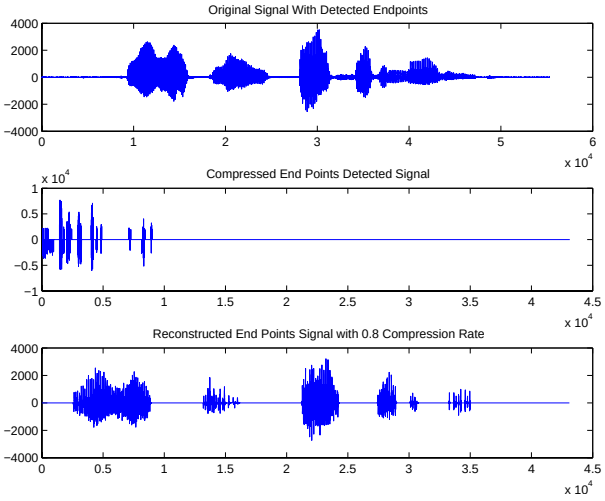

Fig. 4. The Original, Compressed with Global Threshold at 0.8 compression ratio, and Reconstructed signals of the word ” 4035789 ” all in the time domain.

IV. Parameters of Compression

In this paper, the following four compression parameters are used on a subset of the [8] databse containg the digits.

Signal to Noise Ratio:

$$SNR = 10 * log\frac{\sigma_x^2}{\sigma_e^2}$$

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:

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

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:

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

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:

$$RSE = 100 * \frac{||x(n)||2}{||r(n)||^2}$$

Where $||x(n)||$ is the norm of the original signal and $||r(n)||$ is the norm of the reconstructed one. The retained energy is

Signal	CR	%Z	% ER	SNR	PSNR	NRMSE
Original	6.71	96.22	88.34	4.5	28.44	0.32
E-P Dtected	9.32	99.6	91.6	5.12	19.15	0.10

Table 1: Summary of the average results for Level Dependent Threshhold.

Signal	CR	%Z	% ER	SNR	PSNR	NRMSE
Original	8.31	97.37	89.46	7.54	31.41	0.27
E-P Detected	9.62	99.42	92.86	12.41	10.42	0.091

Table 2: Summary of the average results for Global Threshhold.

equal to the L^2-norm recovery performance.

The test set of speech signals [7] are the English digits Zero to nine. The compressed speech signal is still audible and you can still recognize the output signal. Different parameters were examined when simulating the code. The 8Khz sampled signals are divided into frames 0.2ms and decomposed up to level 5 DWT. Each frame is decomposed separately. At this stage the threshold is applied and the remaining coefficients are used in the reconstruct phase.

V. Conclusion

Detecting the end points of speech signals prior to compression was addressed in this paper. This approach proved to be effective and reliable when tested in a Global and a Level Dependent thresholding enviorenments. A reflective mother wavelet associated wih the orthogonal wavelets $db6$ of Daubechies family was used as the analyzing function and Level 5 Discrete Wavelet Transform was performed on each signal. The Measures for various compression parameters were found to be more accurate and robust in the proposed appraoch of compression and in some cases of an order of magnitude better.

References

[1] Feng, Yanhui, Thanagasundram, Schlindwein, S., Soares, F., Discrete wavelet-based thresholding study on acoustic emission signals to detect bearing defect on a rotating machine, Thirteenth International Congress on Sound and Vibration, Vienna, Austria July 2-6, 2006.
[2] Graps, A., An Introduction To Wavelets, IEEE Computational Sciences and Engineering, Volume 2, Number 2, pp: 50-61, Summer 1995.
[3] Karam, J., and Karam, M., “Perfect Reconstruction of Hypermedia Elements Via Filter Banks”, International Journal of Computational and Numerical Analysis and Applications, Vol 5 No. 1, 2004.
[4] Karam, J., A Global Threshold Wavelet-Based Scheme for Speech Recognition, Third International conference on Computer Science, Software Engineering Information Technology, E-Business and Applications, Cairo, Egypt, Dec. 27-29 2004.
[5] Karam, J., Saad, R., The Effect of Different Compression Schemes on Speech Signals, International Journal of Biomedical Sciences, Vol. 1 No. 4, pp: 230 234, 2006.
[6] Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J., Matlab Wavelet Toolbox, Math Works, Natick, MA, 1997.
[7] NIST, TIDIGITS, Speech Discs, Studio Quality Speaker-Independent Connected-Digital Corpus, NTIS PB91-506592, Texas Instruments, Feb. 1991.
[8] NIST, Speech Discs 7-1.1, TI 46 Word Speech Database Speaker-Dependent Isolated-Digital Corpus, LDC93S9, Texas Instruments, Sep. 1991.

Jalal Karam has a Bsc, an Advanced Major Certificate, and an Msc in Mathematics from Dalhousie University in Halifax, Canada. In the year 2000, he finished a PhD degree in Applied Mathematics from the Faculty of Engineering at the Technical University of Nova Scotia. Currently, his is the Head of the Department of Mathematics and Natural Sciences at the Gulf University for Sciences and Technology in Kuwait. Dr. Karam is the Editor - in - Chief of the International Journal of Computational Sciences and Mathematics and a Reviewer for ”Mathematical Reviews” of the American Mathematical Society. He also Chairs the task force of the World Academy of Science in the Middle East and the Near Far East Region.