Al-khwarizmi Engineering Journal

Al-Khwarizmi Engineering Journal, vol.1, no.1,pp 52-60, (2005)

Speech Signal Compression Using Wavelet And Linear Predictive Coding

Dr. Tarik Zeyad **Ahlam Hanoon**

Electrical Engineering Dept. / College of Engineering / University of Baghdad

Abstract

A new algorithm is proposed to compress speech signals using wavelet transform and linear predictive coding. Signal compression based on the concept of selecting a small number of approximation coefficients after they are compressed by the wavelet decomposition (Haar and db4) at a suitable chosen level and ignored details coefficients, and then approximation coefficients are windowed by a rectangular window and fed to the linear predictor. Levinson Durbin algorithm is used to compute LP coefficients, reflection coefficients and predictor error. The compress files contain LP coefficients and previous sample. These files are very small in size compared to the size of the original signals. Compression ratio is calculated from the size of the compressed signal relative to the size of the uncompressed signal. The proposed algorithms where fulfilled with the use of Matlab package.

Keyword: Wavelet, Speech Coding, Linear Predictive Coding, Levinson Durbin.

1. Introduction

Compression can be achieved by reducing redundancy. It is the process of reduction in the amount of signal space that must be allocated to a given message set or data sample set. This signal space may be in a physical volume, such as data storage medium: an interval of time, such as the time required to transmit a given message set [1][2].

Bit rate reduction or data reduction are all terms which mean basically the same thing it means that the same information is carried using smaller quantity or rate of data.[3][4] Compression is the process of converting an input data stream [The source stream or the original raw data] into another data stream [The output or the compressed data] that has a smaller size.

2. The Proposed Speech Signal Compression Algorithm

The proposed compression system divided into four headlines:

2.1 <u>Preprocessing Part</u>:

- Acquisition of the received signal in an IF mode, in our work the signals are recorded using microphone in different places and conditions, so a number of signals are recorded.
- Using Haar and db4 wavelet decomposition algorithm to decompose the received speech signal.
- After that the decomposed frames are segmented into 256 samples per segment. Accordingly the time or frequency variations are viewed in small intervals. By this way the processing will be more accurate.
- Now each frame is multiplied by finite – duration window. This process is called windowing. Rectangular window having duration of one-pitch periods. This produces output spectrum very close to that of the vocal tract impulse response. The equation of a rectangular window is:

$$w_r(n) = \begin{cases} 1 & 0 < n < N-1 \\ 0 & otherwise \end{cases} \qquad (1)$$

where $\omega_r(n)$ is the signal samples inside the window.

Now the data is ready to an efficient feature extraction part.

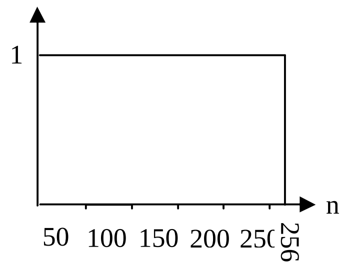

Fig. (1) Rectangular window of 256-sample length.

2.2 The Feature Extraction Part:

The feature extracting helps to reduce noise as well as the signal component redundant to the process of compression. Discrete wavelet transform is more popular in the field of digital signal processing. To parameterize the speech signal it is first decomposed in a dyadic form. The decomposition processed only on the low frequency branches, which is the approximation coefficients since it is more intelligible part as well as most of the information in this part, while the details coefficients are ignored since they contain noise. All data operations can be performed using just the corresponding wavelet coefficients and hence approximation coefficients fed to the next stage (linear predictor stage).

2.3 Linear Prediction Based Coefficients Calculation:

The linear prediction (LP) provides parametric modeling techniques, which is used to model the spectrum as an autogressive process. These parametric models are basically used in compression system.

The ability of linear prediction as applied to speech, however, lies not only in its predictive function but also in the fact that it gives a very good model of the vocal tract, which is useful for both practical and theoretical purposes for representing speech for low bit transmission or storage.

2.3.1 Autocorrelation method for calculating the LPC

Autocorrelation method was used for many reasons: 1) It requires less amount of storage. 2) The number of multiplication needed for computation is small. 3) The number of sample will be as large as several pitch periods to ensure reliable results. 4) The number of poles will depend on the sampling rate, it is known that for every 1KHz sampling rate two poles (one conjugate pole) will be needed.

The prediction error is represented by:

$$e(n) = S(n) - \sum_{i=1}^{p} a_i S(n-i) \qquad (2)$$

where $S(n)$ is the value of sample n and $\sum_{i=1}^{p} a_i S(n-i)$ is the predicted value of sample (n).

The prediction parameter (a_i) is known and determined by minimizing the mean square error (MSE) $E[e^2(n)]$. These coefficients are determined by solving (p) (order of LP) with (p) unknowns that obtained by minimizing MSE

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & \vdots & \dots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(P) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(P) \end{bmatrix} \qquad \dots\dots\dots\dots (3)$$

where $R(n)$ is the correlation of the window with it self with shift equals to n ($R(0)$ is the autocorrelation).

This type of matrix is called *Toeplitz matrix* where all elements along a given diagonal are equal and it is very easy to invert it.

2.3.2 Levinson-Durbin Procedure

The L-D algorithm is a recursive algorithm that solves the ($a=R^{-1}R$). It is very computationally efficient. During this algorithm a number of coefficients are generated which are $\{a_i\}$ a set of coefficients and $\{k_i\}$, which is called reflection coefficients. These coefficients can be used to rebuild the set of filter coefficients $\{a_i\}$ and can guarantee a stable if their magnitudes are strictly less than one.

The Levinson Durbin algorithm is summarized by [1]:

$$E^{(0)} = R(0) \tag{4}$$

$$ki = \left\{ R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right\} / E^{(i-1)} \qquad 1 \le i \le p \tag{5}$$

$$a_i^{(i)} = ki \tag{6}$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)} \qquad 1 \le j \le i-1 \tag{7}$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \tag{8}$$

$$a_j = a_j^{(p)} \qquad 1 \le j \le p \tag{9}$$

Fig. (2) Shows the flow chart of the compression of speech signal while fig. (4), illustrates the flow chart of the decompression of speech signal.

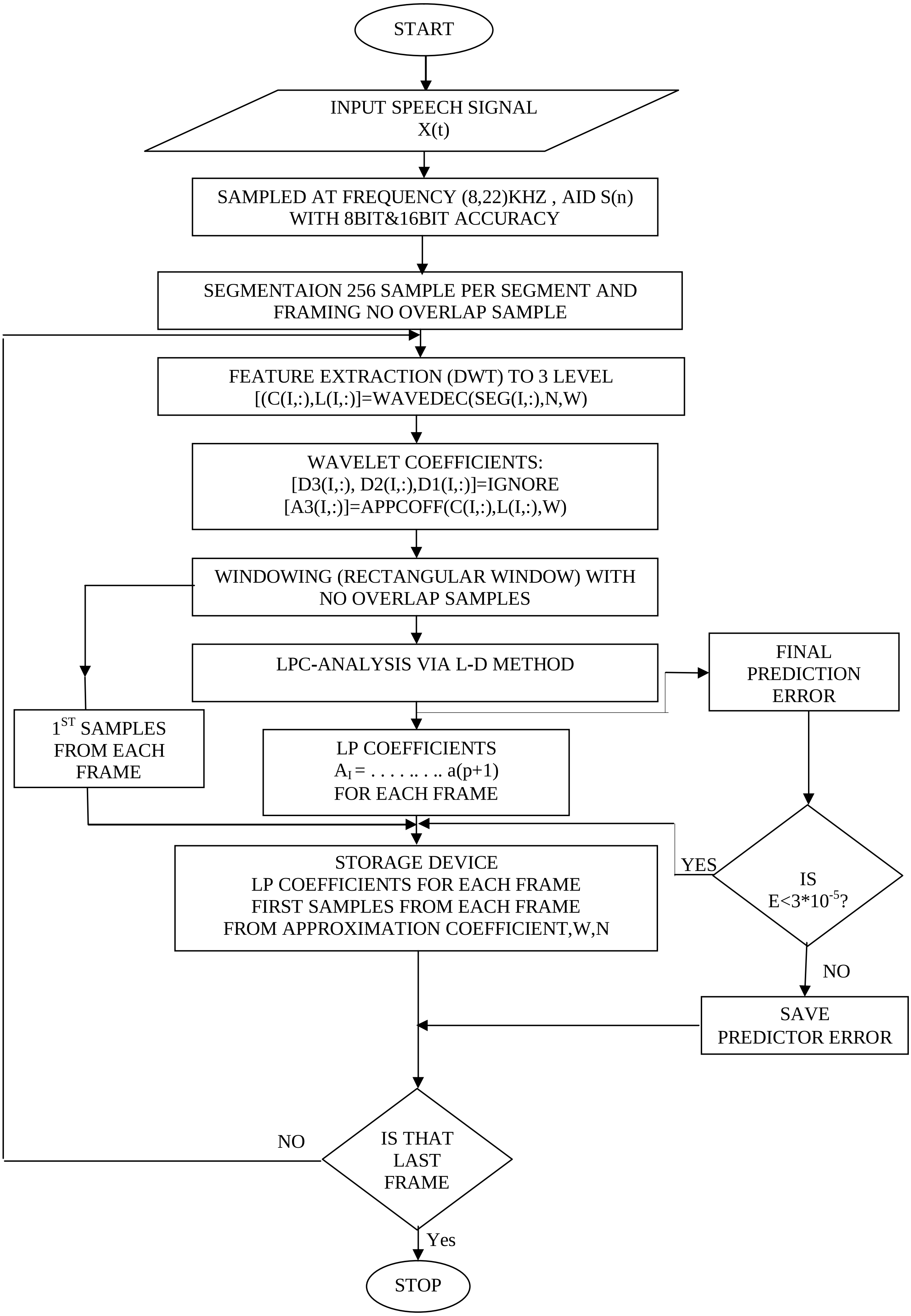

Fig.(2) Flow chart of compression speech signal.

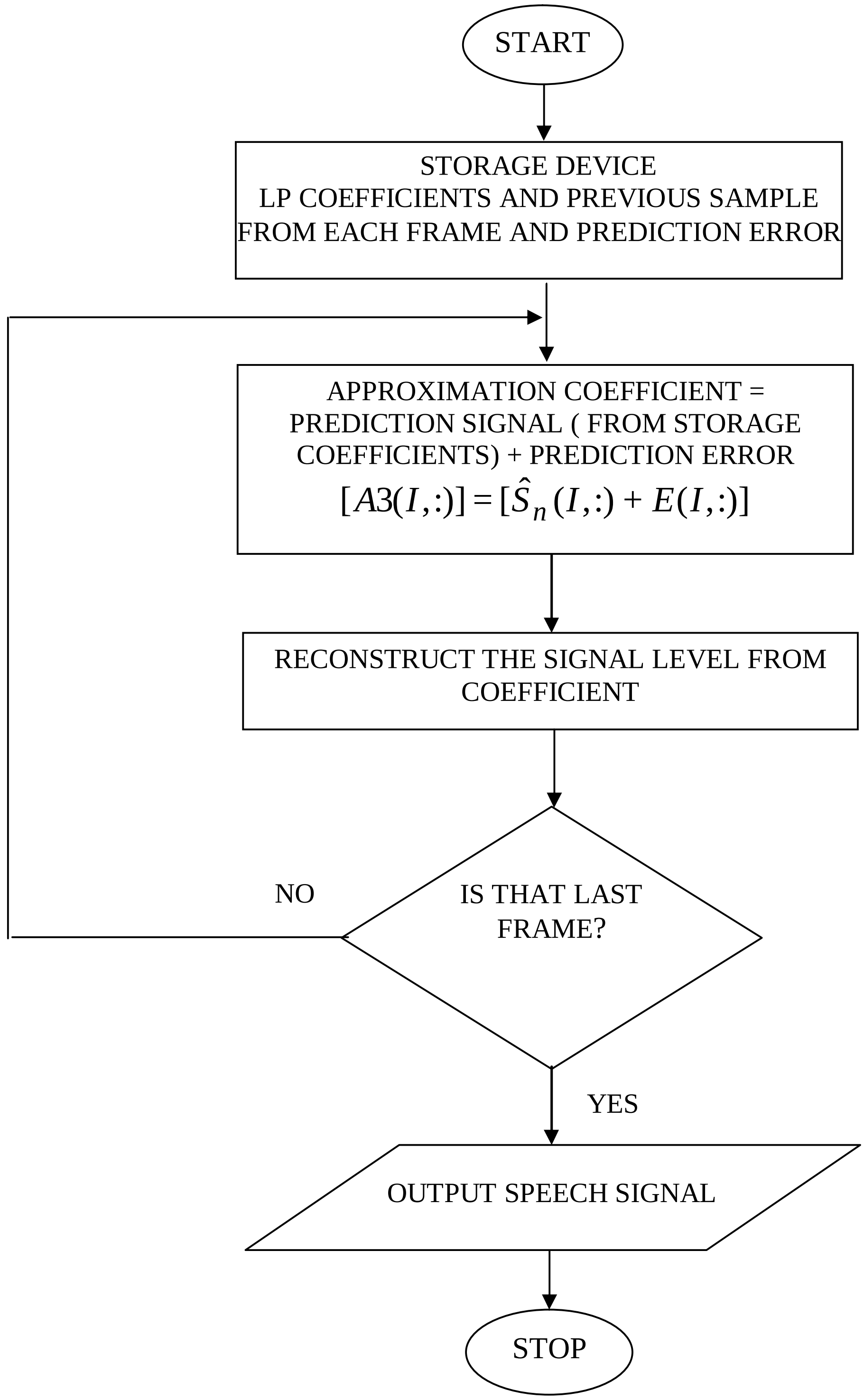

Fig.(3) Flow chart of decompression speech signal.

3. Evaluation Tests of the Algorithm

3.1 Tested Speech Samples:

The test material will contain five speech samples stored in five files, the format of these files are wave format, each file has a different size with respect to the other files of normal Arabic sentences altered by different speakers. The type of the digital speech is pulse code modulation (PCM) and the tested speech samples have 8-bit/samples or 16-bit/samples. The properties of tested wave data are presented in table (1).

Table (1) : Properties of tested wave data

Signal	S_1	S_2	S_3	S_4	S_5
File Name	Allah	aa_1	aa_2	DALEEL	PROG
File Type	Wave file	Wave file	Wave file	Wave file	Wave file
File Size byte	190640	123392	160000	260664	296192
Media Length (s)	23	15	20	14	13
File Format	PCM	PCM	PCM	PCM	PCM
	8 KHz	8 KHz	8 KHz	22 KHz	22 KHz
	8-bit mono	8-bit mono	8-bit mono	16-bit mono	16-bit mono

4.2 Performance Measure

The compression algorithm used in this project has been tested on a number of sounds (about 5-files). Table (2) shows performance measures of speech signals. To evaluate the performance of a compression technique the criteria of measuring distortion in reconstructed sound files are defined. These criteria are necessarily applied. These include signal to noise ratio, peak signal to noise and normalized root mean square error.

The above quantities are calculated using the following formats [1]:

1-Signal to noise ratio

$$SNR = 10\,Log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) \quad (10)$$

σ_x is the mean square of the speech signal and σ_e is the mean square difference between the original signal and reconstructed signal.

2-Peak signal to noise ratio

$$PSNR = 10\log_{10}\frac{(Nx^2)}{\|x-r\|^2} \quad (11)$$

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

3-Normalized root mean square error

$$NRMSE = \sqrt{\frac{(x_{(n)} - r_{(n)})^2}{(x_{(n)} - \mu_{(n)})^2}} \quad (12)$$

$x_{(n)}$ is the speech signal, $r_{(n)}$ is the reconstructed signal and $\mu_{x(n)}$ is the mean of speech signal

.

Table (2) Performance measures of speech signals.

Signal	Wavelet	SNR	PSNR	NRMSE
S_1	Haar db4	5.4328 6.8434	5.9214 5.6851	1.2902 1.3258
S_2	Haar db4	4.2883 4.7211	13.2997 13.1877	1.2686 1.2844
S_3	Haar db4	3.0827 3.3462	10.5618 10.4571	1.1904 1.2048
S_4	Haar db_4	7.2869 11.7732	15.0258 14.7718	1.3448 1.3841
S_5	Haar db4	8.2198 12.9519	16.6317 16.3998	1.3580 1.3948

5. Compression Performance

This section gives the computation of the compression performance of the proposed method in tables (3) and (4) [2].

This table is computed according to the following equations:

Compression ratio = the size of output/the size of input <1 ……..(13)

Compression factor = the size of input/the size of output >1 …… (14)

The expression =100*(1-Compression ratio) …..(15)

Compression gain = 100 log(reference size/compressed size) ……(16)

The first column in this table is the name of the file. The second column is the size of the file before compression and the size of the output file (that represents LPC coefficient and previous samples of each speech signal). The compression ratio is the third column computed by using equation number (13). The fourth column is the compression factor computed by applying equation number (14). The 5[th] column is the expression computed using equation number (15).

Finally the expression gain is computed using equation number (16).

The results of compression performance are shown in table (3) by using Haar wavelet transform and in table (4) by using db4 wavelet transform.

Table (3) Compression Performance results when using Haar Wavelet Transform.

File name	Input file size	Output file size	CR %	CF	EX	Compression Gain
S_1	190640	15624	8.19	12.20	91.81	108.6
S_2	123392	10080	8.16	12.24	91.83	108.7
S_3	160000	13104	8.19	12.21	91.81	108.6
S_4	260664	61936	23.76	4.208	76.24	62.413
S_5	296192	56448	19.05	5.247	80.94	71.99
Average			13.47	9.223	86.522	92.04

Table (4) Compression Performance results when using db4 Wavelet Transform.

File name	Input file size	Output file size	CR%	CF	EX	Compression Gain
S_1	190640	18480	9.69	10.316	90.306	101.35
S_2	123392	11928	9.66	10.344	90.33	101.47
S_3	160000	15456	9.66	10.350	90.34	101.5
S_4	260664	73696	28.27	3.537	71.72	54.86
S_5	296192	67032	22.36	4.418	77.36	64.529
Average			15.982	7.793	84.086	84.74

6.Conclusions:

A significant advantage of using wavelets for speech coding is that the compression ratio can be varied. This work shows that wavelet decomposition in conjunction with other techniques such as LPC is promising compression techniques which make use of the elegant theory of wavelets.

Several conclusions can be drawn: -

1- The two proposed techniques work better with clean source materials; noisy sound waveforms produce poor results and need additional processing techniques.
2- Choosing the right decomposition level in the DWT is important for many reasons, for processing speech signals no advantage is gained beyond level 3 usually processing at lower scale leads to a better compression ratio.
3- Linear predictive technique causes time delay and some loss of quality. But they are negligible in terms of cost when compared with the advantages of storage space saving, smaller B.W requirement, lower power consumption and small product size.
4- This proposed method could be classified in the field of symmetrical compression. This case occurs when the compression and decompression use basically the same algorithm but work in opposite directions.

WT and LPC are used together to get more compression ratio, and depending on these ratios the algorithm gives promising results, although each of them can be used individually to compress speech signals but by using WT technique alone the compression ratio can be varied by changing the level of decomposition while the compression ratio is constant when LPC is used.

Keyword: Wavelet, speech coding, linear predictive coding, Levin son Durbin

7. References

[1] Rabiner L.R, and Schafer R.W., (1990), “Digital Processing Of Speech Signals” , Prentice Hall.

[2] Jabber M., (1999), “ Speech Compression And Recognition Using Wavelet Transform”, An M.Sc. Thesis, University of Baghdad, College of Engineering, Electrical Engineering Department.

[3] Markel J.D. and Gray A.H., (1976), “Linear Prediction Of Speech Coding”, Springer Verlag, New York.

[4] Koul M., (1975), “ Linear Prediction Tutorial”, Proceeding of IEEE , Vol.63, No.4, PP.461-580, April.

[5] Belloch G.E., (2001), “Introduction to Data Compression”, Cavnege Mellon University, Web Site, October.

ضغط الملفات الصوتية باستعمال تحويل المويجة والتشفير ذو الأستنتاجات الخطية

د.طارق زياد **احلام حنون**

قسم الهندسة الكهربائية/ كلية الهندسة / جامعة بغداد

الخلاصة:

تم في هذا البحث استخدام التحويل نوع التحويل المويجة واستخدام المرشح ذو معاملات الاستنتاج الخطية لغرض ضغط حجم الملفات التي تحتوي على تسجيلات صوتية. تم أستخدام التحويل المسمى تحويل المويجة نوع Haar و db4 لغرض تنفيذ عملية الضغط الأولى واحتساب نسبة ضغط الملفات بأستخدام كافة الطرق أما بصورة مفردة أو بصوزة مجتمعة. تم احتساب معاملات المرشح باستخدام خوارزمية Levinson Durbin وكانت نسبة ضغط الملفات تعتمد على نسبة خطأ المرشحات تم مقارنة حجم الملفات المضغوطة مع الملفات الأصلية وكذلك أحتساب نسبة الخطأ التي تنتج من عملية الضغط وكانت جميع الملفات التي تم ضغطها هي ملفات قيم.