Al-Khwarizmi Engineering Journal, Vol. 8, No.4, PP 1-8 (2012)

Al-Khwarizmi Engineering Journal

Speech Compression Using Multecirculerletet Transform

Sulaiman Murtadha **Ali. K. Ibrahim**

Department of Electrical Engineering/University of Baghdad

(Received 20 March 2011; accepted 14 May 2012)

Abstract

Compressing the speech reduces the data storage requirements, leading to reducing the time of transmitting the digitized speech over long-haul links like internet. To obtain best performance in speech compression, wavelet transforms require filters that combine a number of desirable properties, such as orthogonality and symmetry.The MCT bases functions are derived from GHM bases function using 2D linear convolution .The fast computation algorithm methods introduced here added desirable features to the current transform. We further assess the performance of the MCT in speech compression application. This paper discusses the effect of using DWT and MCT (one and two dimension) on speech compression. DWT and MCT performances in terms of compression ratio (CR), mean square error (MSE) and peak signal to noise ratio (PSNR) are assessed. Computer simulation results indicate that the two dimensions MCT offer a better compression ratio, MSE and PSNR than DWT.

Keywords: *Sound, Speech Compression, MCT, DWT.*

1. Introduction

Data compression is the process of converting an input data stream (the source stream or the original raw data) to a data stream with a smaller size (the output, the bit stream, or the compressed stream). A stream is either a file or a buffer in memory. Data compression is popular for two reasons: (1) People like to accumulate data and hate to throw any data away. No matter how big a storage device one has, sooner or later it is going to overflow. Data compression seems useful because it delays this inevitability. (2) People hate to wait a long time for data transfers. When sitting at the computer, waiting for a web page to come in or for a file to download, anything longer than a few seconds is a long time to wait [1]. Speech coding is a lossy type of coding, which means that the output signal does not exactly sound like the input. The input and the output signal could be distinguished to be different. Coding of audio however, is a different kind of problem than speech coding. Audiocodingtries tocodetheaudioin a perceptually lossless way.

This means that even though the input and output signals are not mathematically equivalent, the sound at the output is the same as the input. This type of coding is used in applications for audio storage, broadcasting, and Internet streaming [2]. Speech compression plays an important role in teleconferencing, satellite communications and multimedia applications. However, it is more important to ensure that compression algorithm retains the intelligibility of the speech. The success of the compression scheme is based on the simplicity of the technology and efficiency of the algorithm used in the system. Parametric coding techniques are commonly used methods for speech compression and its application [4]. This paper deals with speech compression of isolated spoken words. A flexible compression scheme that is based on MCT decomposition is used in this work. This paper is organized as follows. Section 2 gives a discussion of the Discrete Wavelet Transforms. Section 3 explains the data base used for the experiment.Section 4 is devoted for computer simulation and results and section 5 concludes the paper.

2. Discrete Wavelet Transform

The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies. The procedure starts with passing this signal (sequence) through a half band digital lowpass filter with impulse response h[n]. Filtering a signal corresponds to the mathematical operation of convolution of the signal with the impulse response of the filter.
The convolution operation in discrete time is defined as follows [4]:

$$L[k] = \sum_{n} x[n].H[2k-n] \qquad \ldots(1)$$

$$H[k] = \sum_{n} x[n].F[2k-n] \qquad \ldots(2)$$

$L[k]$ and $H[k]$ are the outputs of the lowpass and highpass filters.

3. Multicircularlet Transform

One famous multifilter bank is the GHM filter proposed by Geronimo, Hardian, and Massopust [5]. The GHM basis offers a combination of orthogonality, symmetry, and compact support, which cannot be achieved by any scalar wavelet basis [6]. The GHM two scaling and wavelet functions satisfy the following two-scale dilation equations:

$$\begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} = \sqrt{2} \sum_{k} H_k \begin{bmatrix} f_1(2t-k) \\ f_2(2t-k) \end{bmatrix} \qquad \ldots(3)$$

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \sqrt{2} \sum_{k} G_k \begin{bmatrix} f_1(2t-k) \\ f_2(2t-k) \end{bmatrix} \qquad \ldots(4)$$

where H_k for GHM system is four scaling matrices H_0, H_1, H_2, and H_3.Also, G_k for GHM system are four wavelet matrices: G_0, G_1, G_2, and G_3.
The individual coefficients values of these matrices are generated using the following procedure:

1. Apply the 2-D linear Convolution between the G's & H's . This can be achieved as follows:
a) compute $A_{i1} = H_i \otimes H_i$
b) compute $B_{i1} = G_i \otimes G_i$
where i= 0,1,2,3
2. Now compute the 2-D Convolution between the resultant of step 1 & the G's & H's. This can be done through the following way:
a) Compute $A_{i2} = A_{i1} \otimes H_i$
b) Compute $B_{i2} = B_{i1} \otimes G_i$
where i= 0,1,2,3

2. The linear convolution process will be repeated several times. It was found that the optimal results were at the third step. This selection is based upon the evaluation of the application of the resultant coefficients in compression task.
The proposed matrix coefficients A's and B's wear obtained by performing the following computations:
a) Compute $A_i = A_{i2} \otimes H_i$
b) Compute $B_i = B_{i2} \otimes G_i$
where i= 0,1,2,3

The new multifilter Bases functions are denoted by A's and B's which are stated as
A_0, A_1, A_2, A_3 & B_0, B_1, B_2, B_3
The results of the third convolution (basis functions) are:
a) The A's 2x2 matrices are:

$$A_0 = \begin{bmatrix} 1.4561 & 1.4131 \\ -1.0265 & -0.9857 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 1.6896 & 1.5814 \\ 1.4376 & 1.5450 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.0977 & -0.0945 \\ 0 & 0 \end{bmatrix},$$

$$A_3 = 1.0e-005 * \begin{bmatrix} 0.6250 & 0 \\ 0 & 0 \end{bmatrix}$$

b) The B's 2x2 matrices are:

$$B_0 = \begin{bmatrix} 0.0460 & 0.0343 \\ -0.0459 & -0.0\,342 \end{bmatrix},$$

$$B_1 = \begin{bmatrix} 2.7696 & -2.4406 \\ -2.4142 & 2.7225 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 1.6610 & -1.6066 \\ 1.6581 & -1.6038 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -0.0302 & -0.0302 \\ 0.0052 & 0.0052 \end{bmatrix}$$

The results show that at the third convolution the number of zeros in the resultant matrix increased which gives matrices similar in construction to Laurent matrix. Due to the good characteristics of the transformed 1-D signal (speech) by the basis functions obtained from the third convolution, it will be adopted as the transform named "Multicircularlet transform". The AB multifilter bank coefficients are 2 by 2 matrices, and during the convolution step they must multiply vectors (instead of scalars). This means that multifilter banks needs 2 input rows. This transformation is called preprocessing. The most obvious way to get two input rows from a given signal is to repeat the signal. Two rows go into the multifilter bank. This procedure is called "Repeated Row" which introduces oversampling of the data by a factor of 2 [8]. For a given scalar input signal $\{X_k\}$ of length N (N is assumed to be a power of 2 and so is of even length), repeated row preprocessing of this signal is by repeating the input stream with the same stream multiplied by a constant α. So the input length-2 vector are formed from the original as:

$$\begin{bmatrix} X_k \\ a\,X_k \end{bmatrix} \qquad \ldots(5)$$

Where k=0,1,2,…,N-1

Here α is constant, and from the preprocessing scheme of the GHM multifilter bank , α is equal to $(1/\sqrt{2})$. It is found that this factor is very suitable for preprocessing the application of the proposed transform.

After the multifilter bank reconstruction (synthesis) step a postfiltering is applied.

If the preprocessing step, represented by a matrix multiplication,

$$PX = V_0 \qquad \ldots(6)$$

Where P is 2NxN, X is Nx1, and V_0 is 2Nx1, then in detail

$$\begin{bmatrix} 1 & 0 & 0 & \cdots \\ a & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & a & 0 & \cdots \\ & & & \ddots \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} X_0 \\ a\,X_0 \\ X_1 \\ a\,X_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} v_{0,0}^{0} \\ v_{0,0}^{1} \\ v_{0,1}^{0} \\ v_{0,1}^{1} \\ \vdots \end{bmatrix} \qquad \ldots(7)$$

For computing discrete multicircularlet transform, the transformation matrix can be written as follows:

$$W = \begin{bmatrix} A_0 & A_1 & A_2 & A_3 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & A_0 & A_1 & A_2 & A_3 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ A_2 & A_3 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & A_0 & A_1 \\ B_0 & B_1 & B_2 & B_3 & \vdots & \vdots & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & B_0 & B_1 & B_2 & B_3 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & B_0 & B_1 & B_2 & B_3 \\ B_2 & B_3 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & B_0 & B_1 \end{bmatrix} \qquad \ldots(8)$$

Where A_i , B_i are the impulse responses of the AB multifilter bank.

4. Proposed Compression Algorithem

4.1. Implementing Wavelet and Multicirculertlet Transform

The following steps will be followed in the implementation of this algorithm:

1. Record and read the input speech in sampled and digitized form. In this step the input speech will be shown as a one dimension array stored in specified vector.
2. Apply DWT, save the coefficients and write them as an audio file format called WAV format.
3. Apply MCT; 1D & 2D MCT for an input speech are used. At first 1D MCT is applied. For this purpose, each speech after sampling will be resized to the power of two. Then by using a different program, 2D MCT will be applied. In 2D MCT program, the speech samples are divided into different levels, each level contains 2^{N} samples.

Write output Files as a stream in the desired format.

4.2. Applying Multicirculerletlet Transform

MCT contains the following steps:

a) Use Hamming window and different levels by changing Ns.

b) It is possible to apply 1D or 2D MCT.

b.1) For applying 1D MCT:

b.1.i) Compute the length of the speech, the number of total samples contained.

b.1.ii) If the length is to power of two, apply MCTs directly, save the coefficients and write them as a WAV file.

b.1.iii) If the length is not to power of two, refer to the original file, append silence or near zero samples to get a power of two speech samples. Compute the MCT for the entire speech, save the coefficients and write them as a WAV file.

b.2) For applying 2D MCT:

b.2.i) It doesn't matter that the length of the speech is to the power of two or not. Starting N = 2 (each frame equals 4 samples) and increasing N (Maximum N = 6, depending on the original speech bit rate) achieve the desired compressed speech, good compression and good quality.

b.2.ii) Framing is used to cut the long-time speech to the short-time speech signal in order to compute 2D MCT. Each frame contains 2^N samples, so the number of columns is equal to 2^N and the number of rows is the integer of dividing the total number of speech samples into 2^N.

b.2.iii) Compute the MCT for each row, then column wise on the result, finally cascade the rows, save the coefficients and write them as a WAV file.

5. Experiment Results

After applying the transforms to speeches, the compressed and uncompressed speeches will be written as a stream file in the form of WAV for playing them on any software like windows media player or use them on the internet as stream media or even more complicated systems like telephone switches. Also, it is possible to write the results in another form using AU extension. But WAV extension is preferred more than AU.

The test material contains seven speech samples stored as files, the format of these files are wave format. Each file has different size and different number of samples. The type of the digital speech is pulse code modulation (PCM) and the tested speech samples are in the bit rate of 64, 256 and 3072 Kbps. The properties of the tested speech samples are presented in Tables (1, 2, 3, 4, 5, 6, and 7) and Fig. (1).

Table 1,
Properties of Tested Speech Samples.

File Name	File Size (BYTE)	Total Samples
Be	2251016	562743
Ru	4155904	1038965
R01	5266700	1316664
R02	5931500	1482864
R03	7886092	1971512
HE	4815572	1203882

Table 2,
Performance Measures of Speech Signals.

Signal	PSNR		MSE	
	Wavelet	2D MCT	Wavelet	2D MCT
Be	49.8350	52.9273	0.6084	0.3314
Ru	49.6102	52.1699	0.6671	0.3945
R01	51.9419	55.1144	0.3841	0.2003
R02	51.8094	54.3706	0.4113	0.2377
R03	49.8011	52.6932	0.6354	0.3498
HE	50.3298	53.1462	0.5676	0.3151

Table 3,
Compression Performance Results Using Haar DWT.

File Name	Input Size	Output Size	CR%	CF	Compression Gain
Be	2251016	281416	12.502	7.999	91.303
Ru	4155904	479528	11.538	8.667	94.785
R01	5266700	588376	11.172	8.951	95.188
R02	5931500	641476	10.815	9.247	96.598
R03	7886092	785800	9.964	10.036	100.155
HE	4815572	551986	11.463	8.724	96.072
Average			11.242	8.937	95.120

Table 4,
Compression Performance Using 1D MCT.

z	Input Size	Output Size	CR%	CF	Compression Gain
Be	2251016	652788	29.000	3.448	53.761
Ru	4155904	1127974	27.141	3.684	56.637
R01	5266700	1299708	24.678	4.052	60.769
R02	5931500	1492908	25.169	3.973	59.913
R03	7886092	1971556	25.000	4.000	60.205
HE	4815572	1273926	26.454	3.780	57.750
Average			26.240	3.823	58.241

Table 5,
Compression Performance Using 2D MCT N = 1.

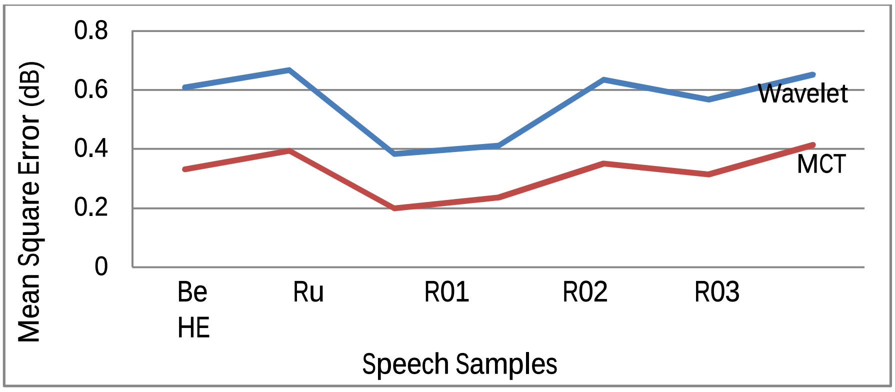
File Name	Input Size	Output Size	CR%	CF	Compression Gain
Be	2251016	584788	25.979	3.849	58.538
Ru	4155904	1027974	24.735	4.043	60.668
R01	5266700	1259708	23.918	4.181	62.127
R02	5931500	1372908	23.146	4.320	63.552
R03	7886092	1771556	22.464	4.452	64.851
HE	4815572	1173926	24.378	4.102	61.301
Average			24.103	4.158	61.887

Table 6,
Compression Performance Using 2D MCT N = 2.

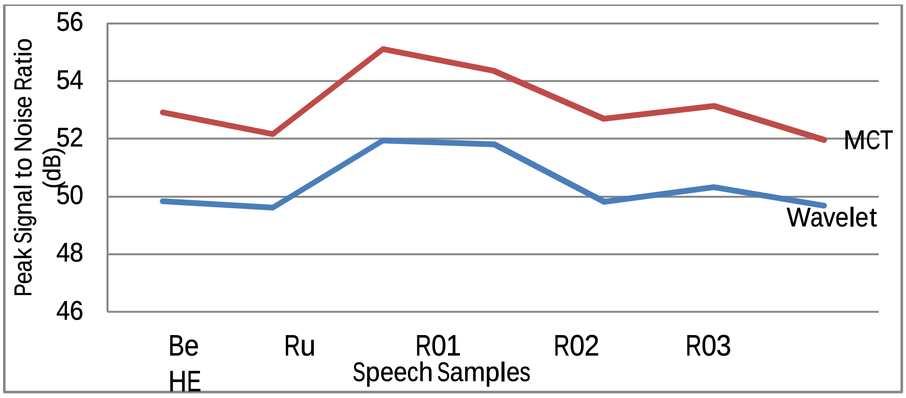
File Name	Input Size	Output Size	CR%	CF	Compression Gain
Be	2251016	273416	12.146	8.233	91.555
Ru	4155904	459528	11.057	9.044	95.635
R01	5266700	528376	10.032	9.968	99.860
R02	5931500	561476	9.466	10.564	102.383
R03	7886092	685800	8.696	11.499	106.066
HE	4815572	501986	10.424	9.593	98.196
Average			10.304	9.817	99.197

Table 7,
Compression Performance Using 2D MCT N = 3.

File Name	Input Size	Output Size	CR%	CF	Compression Gain
Be	2251016	133416	5.927	16.872	122.717
Ru	4155904	279528	6.726	14.868	117.224
R01	5266700	318376	6.045	16.542	121.860
R02	5931500	331476	5.588	17.894	125.271
R03	7886092	385800	4.892	20.441	131.050
HE	4815572	301986	6.271	15.946	120.266
Average			5.908	17.094	123.284



(a) Mean Square Error (MSE)



(b) Peak Signal to Noise Ratio (PSNR)

Fig. 1. Performance Measures of Speech Signals.

It is shown that 1D has the maximum compression ratio percentage. The best compression ratio is in the case of 2D MCT with maximum number of levels, N equal 3.

6. Conclusions

The Haar wavelet transform is the simplest and the fastest one to implement. However, because of its discontinuity, it is *not* optimal to simulate a continuous signal. Based on our experiments. Haar wavelet obtained the worst compression result, which proves the above statement. Db4 found the first continuous orthogonal compactly supported wavelet. Note that this family of wavelets is not symmetric. The advantage *of* symmetry is that the corresponding wavelet transform can be implemented using a mirror boundary condition, which reduces boundary artifacts.

The MCT is a symmetrical transform and easy to implement. The MCT based compression software designed reaches a signal of noise ratio of 55.114db and CR equal to 20.4. The 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.

7. References

[1] D. Salomon, 2007 "Data Compression, the Complete Reference", Fourth Edition, Contributions by, Springer-Verlag London.

[2] "Law bit-rate speech coders for multimedia communication," R.V. Cox and P. Krmn, LEEE

[3] Communications Magazine, pages 34-40, 1996 http://www.bell-labs.com

[4] Shijo M Joseph, Firoz Sha A and Babu Anto P "Speech Compression: A Comparative Study Between Discreet Wavelet and Wavelet Packet Decomposition" International Journal of Computer and Network Security Vol 2. No7 2010.

[5] G. Strange and T. Nguyen, 1996 "Wavelets and Filter banks", Wellesley, Wellesley-Cambridge.

[6] Geronimo, J., Hardian,D.& Massopust, P. " Fractal Function and Wavelet Expansion Based on Several Functions", J. Approx. Theory, Vol. 78, PP. 373-401, 1994.

[7] Burrus, C.S., Goperath, R.A., and Gue, H., "Introduction to wavelet and wavelet transforms", A primer Upper Saddle, NJ(U.S.A.), Prantice Hall, Inc., 1998.

[8] Strela, V. & Walden, A.T. " Orthogonal and biorthogonal multiwavelets for signal denoising and image compression", Proc. SPIE, 3391: 96-107, 1998.

[9] Shi Zhong and Vladimir Cherkassky, "Image Denoising using Wavelet Thresholding and Model Selection", Dept. of ECE, Univ. of Texas at Austin, 2001.

ضغط الكلام باستخدام تحويل الملتي سيكلرليت

سليمان مرتضى عباس علي خليل ابراهيم
قسم الهندسة الكهربائية/ كلية الهندسة/ جامعة بغداد

الخلاصة

ضغط الكلام يقلل متطلبات خزن المعلومات مما يؤدي الى اختصار زمن رسال الكلام الرقمي خلال ضغط الانترنيت. ومن اجل الحصول على اكفأ اداء، استعملت خوارزمية الويفليت والتي تحتاج الى مرشحات ذات خصائص مرغوبة مثل التناظر والتعامد. خوارزمية الملتي سيكلر ليت المشتقة من دالة (GHM) وتستعمل دالة الكونفليوشن الخطية. تم استعمال هذه الخوارزمية الجديدة في مجال ضغط الكلام وذلك للسرعة العالية لهذه الخوارزمية . يناقش هذا البحث تأثير استخدام هذه خوارزمية وخوارزمية الويفليت في هذا التطبيق وكفأءة الاداء حسبت من حيث نسبه الضغط (CR) ، ومعدل مربع الخطأ (MSE)، ونسبه الاشارة الى الضوضاء (PSNR). برامج الحاسبة التمثيلية لهذا التي تؤثر ان خوارزمية الملتي سيركلر ليت ذات البعدين اعطت افضل النتائج .