Multi Rate Audio Coding Based On Combining Wavelet with DCT Transform

Adnan I. Hussein

Dept. of computer science, College of Education, Dohuk University

AIH2007@gmail.com

Abstract

In this paper an efficient algorithm proposed to encode the audio signals with multirate capability. The algorithm based on combining discrete wavelet with DCT transform for maximum decorrelation. The coefficients of the frame are scaled and encoded using non uniform quantizer. The main features of this algorithm are: low complexity and near transparent audio quality resulted in the range 48 – 64 Kbps for most SQAM signals. The algorithm outperform much better than DWPT with SPIHT algorithm previously.

Keyword : wavelet , DCT , audio , coding , Huffman , and psychoacoustics.

تشفير الأشارة السمعية بمعدل بيانات متعدد بالأعتماد على ربط تحويل المويجة مع تحويل الجيب تمام

عدنان أسماعيل حسين

قسم علوم الحاسبات / كلية التربية / جامعة دهوك

الخلاصة

تم في هذا البحث أقتراح طريقة جديدة وكفوءة لتمثيل الأشارات السمعية بمعدل بينات متعدد. تعتمد الطريقة على أستخدام تحويل المويجة (DWT) مع تحويل الجيب تمام (DCT) لتقليل معامل الأرتباط الى الحد الأدنى. معاملات كل اطار يتم تقييسها وتشفيرها باستخدام ، مكم غير منتظم. من أبرز مزايا هذه الطريقة سهولتها و خلوها من التعقيدات الموجودة في الطرق التقليدية الأخرى، اضافة الى أن الأداء الممتازل عند معدل بيانات مابين 48 - 60 كيلوبت ثانية ، حيث أن الأشارة الناتجة لايمكن تمييزها بسهولة عن الأشارة الأصلية. تبين من خلال البحث ان أداء هذه الطريقة هو الأفضل مقارنة بطريقة SPIHT عند جميع معدل البيانات ولأنواع مختلفة من الأشارات.

Received 1 Oct. 2007 **Accepted 9 Dec. 2007**

1. Introduction

Source coding of wideband audio signals for storage and/or transmission application over band limited channels is currently a research topic receiving considerable attention. Its applications are in the fields of audio production, program distribution and exchange, digital audio broadcasting, digital storage, video conference and multimedia applications. The industrial standard for wideband audio signal with sampling rate at 44.1 KHz which covers the entire audible frequency range of the human hearing system, each sample is quantized into 16 bits, without compression, the bit rate will be 705.6 Kb/sec for one channel. The goal of audio data compression is to get the bit rate as low as possible without perceptible distortion.

Most proposed audio coders are transform coders or subband coders. They mainly include three parts: subband decomposition or transform, dynamic bit allocation and the coding algorithm. First the original audio data is transformed into subband signals; the target bit rate is dynamically allocated among the subbands through a psychoacoustic model; and then each subband signal is encoded to a bit stream [1].

Several of these techniques have contributed to the development of the ISO/MPEG audio coding standards. The first one, called ISO/MPEG-1, supports sampling rates of 32, 44.1 and 48 kHz, and several operation modes with bit rates ranging from 32 to 448 kbps. The last one, the ISO/MPEG-4 standard, is composed several speech and audio coders, supporting bit rates from 2 to 64 kbps per channel. ISO/MPEG-4 includes the AAC, already proposed in ISO/MPEG-2 audio coding standard, which provides high quality audio coding at bit rates of 64 kbps per channel. The techniques presented by ISO/MPEG standards are aimed at constant rate transmission, although MPEG has made some attempts at standardizing scalable compression techniques [2][3][4][5].

In addition to very low bit rate compression, modern audio coding systems have additional features that make the systems more flexible for

different applications [6]. One important feature is scalability. Scalability means that the bit-stream is organized in the form of layers, where a lower quality part of the signal can be decoded without any information about the higher quality part. Scalability is useful when the transmission channel cannot guarantee the full bandwidth to accommodate the complete bitstream. The first idea on scalable audio coding was proposed by Brandenburg and Grill [7]. They also proposed several schemes to build scalable audio coding systems based on the MPEG-2 NBC standard [8].

2. Wavelet based audio coder

Parallel to the definition of the ISO/MPEG standards, several audio coding algorithms have been proposed that use the wavelet transform as the tool to decompose the signal due to the advantage of high time-frequency resolution it provides [9]. As mentioned in [10], wavelets are particularly suitable for scalable coding because their multi-resolution property can be directly employed for bandwidth scalability.

Many wavelet based algorithms proposed in literature [9][11].The basic idea behind discrete DWT-based subband coders is to quantize and encode efficiently the coefficient sequences associated with each stage of the wavelet decomposition level. Irrelevancy is exploited by transforming frequency-domain masking thresholds to the wavelet domain and shaping wavelet-domain quantization noise such that it does not exceed the masking threshold. Wavelet-based subband algorithms also exploit statistical signal redundancies through differential, run-length, and entropy coding schemes.

3. Wavelet Transform

The Wavelet Transform (WT) is a technique for analyzing signals. It was developed as an alternative to the short time Fourier Transform (STFT) to overcome problems related to its frequency and time resolution properties. More specifically, unlike the STFT that provides uniform time

resolution for all frequencies the DWT provides high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies.

The DWT analysis can be performed using a fast, pyramidal algorithm related to multirate filter banks. As a multirate filterbank the DWT can be viewed as a constant Q filterbank with octave spacing between the centers of the filters as shown in figure (1). Each subband contains half the samples of the neighboring higher frequency subband. In the pyramidal algorithm the signal is analyzed at different frequency bands with different resolution by decomposing the signal into a coarse approximation and detail information. The coarse approximation is then further decomposed using the same wavelet decomposition step. This is achieved by successive highpass and lowpass filtering of the time domain signal and is defined by the following equations:

$$C(k) = \sum_n x(n)\, h(2k - n) \qquad (1)$$

$$d(k) = \sum_n x(n)\, g(2k - n) \qquad (2)$$

Where $C(k)$, $d(k)$ are the outputs of the lowpass filters (h), and highpass filter (g) respectively after subsampling by 2. Because of the downsampling the number of resulting wavelet coefficients is exactly the same as the number of input samples [12]. Wavelet packet (WP) or DWPT representations, on the other hand, decompose both the detail and approximation coefficients at each stage of the tree [11].

A filter bank interpretation of wavelet transforms is attractive in the context of audio coding algorithms. Wavelet or wavelet packet decompositions can be tree structured as necessary (unbalanced trees are possible) to decompose input audio into a set of frequency subbands tailored to some application. It is possible, for example, to approximate the critical band auditory filter bank utilizing a wavelet packet approach.

Moreover, many coefficient finite support wavelets are associated with a single magnitude frequency response QMF pair; therefore, specific subband decomposition can be realized while retaining the freedom to choose a wavelet basis that is in some sense “optimal”.

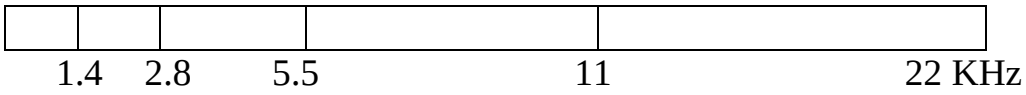

Figure (1) : Subband decomposition of audio signal associated with four level discrete wavelet transform.

4. Related works

Lu and Pearlman investigated a rate-scalable DWPT-based coder that applies set partitioning in hierarchical trees (SPIHT) to generate an embedded bit stream. The coder achieves nearly transparent quality at 55–66 kb/s. The system is also capable of delivering lower rate service from the same bitstream [1].

As an indication of how SPIHT reduces the bit rate of audio signals, Table (1) lists initial results for the eight test signals (Sound Quality Assessment Material (SQAM)) obtained from [13]. The signal content of the files tested is also given in Table (1). Since this set of results is for complete reconstruction combined with bit allocation using the MPEG masking model, the sound quality of the synthesized files were the same as the original. The objective results given are the Segmental Signal to Noise Ratios (SNR) of the synthesized signals.

5. Proposed algorithm

In this paper a low-complexity scalable audio coder system based on combining wavelet with DCT transform. The goal of this work is to design and implement a scalable coder that provide transparent quality at

minimum bitrate as possible with capability reconstructing the signal with multiple level of quality.

The basic idea of the algorithm is to apply wavelet and DCT for maximum decorrelation, then split the coefficients into layers. For full reconstruction all layers must be decoded. For partial reconstructed signal the decoder neglect the layer of very low values. Figure (2) shows the block diagram of the algorithm.

6. Algorithm details

The detail of algorithm explained in the following steps:

Table (1) : Coding result using wavelet transform and SPIHT

Signal	Content	SNR (dB)	Mean Rate (Kbps)
X1	Bass	46.1	167
X2	Electronic Tune	50.9	71
X3	Glockenspiel	46.6	180
X4	Glockenspiel	44.4	201
X5	Harpsichord	31.1	227
X6	Horn	48.0	94
X7	Quartet	43.2	174
X8	Soprano	43.7	162

Step1: The input signal decomposed by four stages DWT using Daubechies filter tap-20 proposed in [1]. The output coefficients arranged in frames of 1024 samples.

Step2: In order to verify psychoacoustic requirements a scaling vector derived from absolute threshold of hearing curve [11] to scale the coefficients of the frame according to their importance to human ear.

Step3: In this work, the signal in wavelet domain classified as stationary, transient, or noise signal. Stationary signal better represented in frequency domain, because a transform like DCT [15] can compact the energy into few coefficients, while the coefficients of transient segment encoded directly, and the noise signal removed by choosing appropriate threshold.

The DCT transform applied to the coefficients of each band in the frame. In order to choose better representation of each segment in the frame, a comparator used to choose best representation based on number of significant coefficients in each representation with respect to some threshold. Five bit transmitted as side information to indicate the type of representation for each band.

Step4: The encoder split the coefficients in to four layers as shown in figure (3). The higher layer is open to span entire range of coefficients for different frames, while the lower layer is chosen to be small enough such that when it removed or neglected by decoder keep the perceptual distortion minimum and in the same time decrease the bit rate significantly, thus a compromise must present. Each layer allocated number of bits that produce inaudible distortion.

Step5: The first step of encoding process is to determine the maximum absolute value in the frame to determine the initial layer, then initial threshold (T_0) taken equal to the minimum value of the layer. Each coefficient classified as significant or not, with respect to the threshold value. If the magnitude of the coefficient larger than or equal to the threshold it classified as significant.

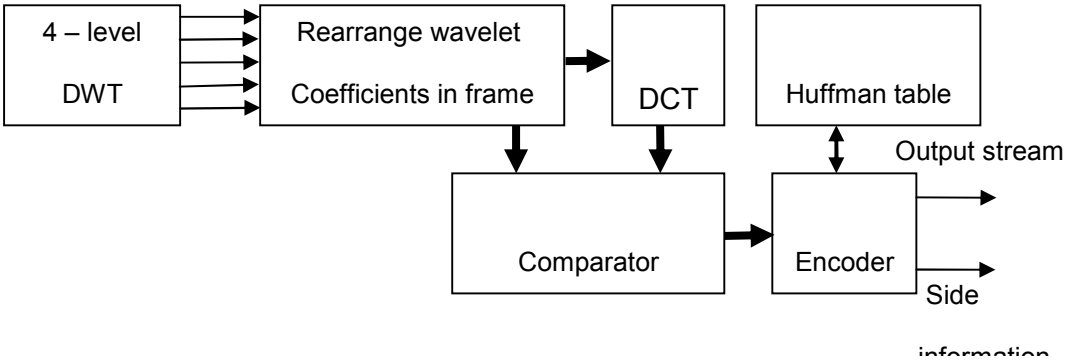

Figure (2) : block diagram of proposed algorithm

The scan process begins by classifying each coefficient as positive, negative, or zero (P, N, and Z). The encoder output positive or negative symbol for each significant coefficient and other stream contain the index of the quantizer according to the following relation:

$$I = round\left(\frac{|C_n| - T_k}{Q_k}\right) \qquad (3)$$

Where C_n is the n^{th} coefficient in the frame, and Q_k is step size quantizer of the layer. After encoding each significant coefficient, it removed from the list, and the scan continued until last significant coefficient encoded. A special symbol used to indicate end of the layer (E). The scan process continued with lower layer until all coefficients in the frame encoded.

Step6 : encoder output two streams, first contain the location of significant coefficients, while second stream contain the index values. First stream arranged as group of four symbols and entropy coded using Huffman table. An other table used to encode the second stream. We can

design Huffman table for each layer or use single table for all layers. The question arise which case is best?. Experimental tests show that single Huffman table outperform better than multiple tables because the nature of algorithm cause the probability of the coefficients with low value increased significantly when combined together.

7. Quality measurement

Measuring the sound quality of perceptual audio codec has developed into an art of its own, over the last ten years. Basically, there are three methods: Listening tests, simple objective measurement methods and perceptual measurement techniques.

As a measure of quality, the most popular subjective assessment method is the mean opinion scoring where subjects classify the quality of coders on an N-

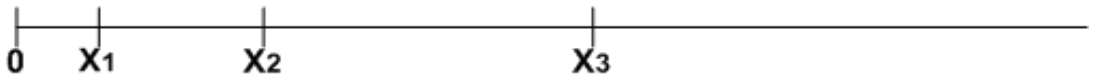

Figure (3) : An example of splitting the coefficients in layers.

point quality scale. The final test is an averaged judgment called the mean opinion score (MOS). Two five point adjectival grading scales are in use, one for signal quality, and other one for signal impairment, and an associate numbering. The 5-point ITU-R impairment scale of table (2) is extremely useful if coder with small impairments have to be graded [16].

Over and over again, people tried to get a measure of encoder quality by looking at parameters such as the signal-to-noise-ratio or bandwidth of the decoded signal. As the basic paradigm of perceptual audio coders relies on improving the subjective quality – by shaping the quantization noise over frequency (and time), leading to an SNR which is lower than is possible without noise shaping – these measurements defy

the whole purpose of perceptual coding. As explained below, to rely on the bandwidth of the encoded signal does not show a very good understanding of the subject. Another approach is to look at the codec output for certain test signal inputs, such as transients or multi-tone signals. While the results of such a test may tell the expert a lot about the codec under test, it is very dangerous to rely solely on such results [17].

8. Experimental results

The algorithm tested on SQAM files available on [13]. All parameters of the coder kept constant for all test signals. Table (3) show the SNR of fully and partially reconstructed signals. Partial reconstruction implemented by neglecting the coefficients of lower layer. The results show that almost all of the SQAM files are coded using a lower mean rate than when SPIHT algorithm. Note the higher SNR results which illustrate the resilience of our algorithm to quantization noise.

The result presented in table (3) for the synthesized signals that are indistinguishable from the original based on two terms SNR and objective tests. It's observed that the proposed algorithm outperforms the result of table (1) by large margin, It outperform by 4.5 – 13 db with lower bit rates.

Table (2) : subjective test score for partially reconstructed signals

Mean opinion score	Impairment scale
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table (3) : Coding result using DWT and DCT transform

		Full reconstruction	Partial reconstruction	
Signal	**SNR**	**Mean Rate (Kbps)**	**SNR**	**Mean Rate (Kbps)**
X1	50.4	144	33.0	60
X2	61.8	26	47.2	13
X3	52.5	122	28.4	30
X4	49.4	174	27.8	56
X5	44.0	224	18.6	62
X6	57.6	66	38.2	32
X7	50.2	184	33.4	66
X8	50.4	124	32.6	52

Also, The algorithm provide superior quality for partial reconstructed signals. To evaluate the performance of the algorithm, we compare these result with those obtained in [14] as shown in table (4). The proposed algorithm outperform by 10 - 27 db (except X4) with less bit rate.

In order to evaluate the performance of our algorithm in worst case, Table (5) shows the test result of subjective quality for partially reconstructed signals. The subjective test implemented by random listeners of ages in the range of 20 – 40 years.

From the result of table (5), we show that the algorithm provides near transparent quality in worst case, and optimal quality achieved in the case of full reconstruction. The good performance of this algorithm at low bit rate can be explained as follows: The dynamic range of wavelet coefficients reflects signal statistics, i.e. Loud signal produce large value of wavelet coefficients and the distortion produced by removing small value coefficients can be masked. In the case of low level signal the distortion is too small to be heard.

9. Conclusion

In this paper a new method for audio coding presented. The algorithm exploits the properties of wavelet and DCT to get optimum or near optimum signal representation. The transformed coefficients divided into layers for multirate Delivering purpose. The results show that near transparent audio quality resulted in the range 48-64 Kbps. The

performance of the algorithm compared with two other schemes based on SPIHT. Its obvious from the results that the proposed algorithm outperform better than these algorithms for many reasons:

1. Discrete wavelet transform are used in the proposed algorithm while packet transform wavelet packet are used in other coders to decompose the signal into 29 subbands
2. The DCT representation in the proposed algorithm are not used with all frames (especially with those frame contain transient signals), thus not all subbands uses IDCT transform in decoding process, which increase the speed of signal reconstruction.

Table (4) : Coding result using MLT and SPIHT presented in [14]

		Full reconstruction	**Partial reconstruction**	
Signal	**SNR**	**Mean Rate (Kbps)**	**SNR**	**Mean Rate (Kbps)**
X1	55.5	145	16.7	53
X2	64.2	31	19.2	14
X3	49.4	60	17.9	25
X4	54.1	110	21.8	47
X5	45.8	183	7.6	65
X6	61.1	68	23.3	33
X7	55.5	180	20.1	65
X8	54.2	140	21.4	47

Table (5) : subjective test score for partially reconstructed signals

Signal	**Test result**
X1	5
X2	4.5
X3	4.25
X4	4.5
X5	4.5
X6	5
X7	5
X8	5

3. SPIHT is too much complex because it split the coefficients in many layers depending on the number of bit required to encode maximum value in the frames. While proposed decoder split the coefficients in maximum of four layers which result in simpler and faster decoder.

10. References

[1] Zhitao Lu and William A. Pearlman, "**An efficient, low-complexity audio coder delivering multiple levels of quality for interactive applications**," in 1998 IEEE Second Workshop on Multimedia Signal Processing, 1998, pp. 529–534.

[2] ISO/IEC 11172-3, "**Information technology - Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s - Part 3**", 1992.

[3] Peter Noll, "**Mpeg digital audio coding,**" IEEE Signal Processing Magazine, vol. 14, no. 5, pp. 59–81, Sept. 1997.

[4] H. Purnhagen and N. Miene, "**Hiln - the mpeg-4 parametric audio coding tools**," in Proceedings of ISCAS 2000, 2000, vol. 3, pp. 201–204.

[5] K. Brandenburg and M. Bosi, "**Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding**," Journal of Audio Eng. Soc., vol. 45, no. 1/2, Jan./Feb. 1997.

[6] ISO/IEC JTC1/SC29, CD 14496-3, **"Information Technology –- Coding of Audiovisual Objects: Part 3"**, Audio, Oct. 1997.

[7] K. Brandengburg and B. Grill, **"First Ideas on Scalable Audio Coding,"** 97th AES Convention, San Francisco, November 10-13, 1994.

[8] B. Grill and K. Brandenburg, **"A Two- or Three-Stage Bit Rate Scalable Audio Coding System,"** 99th AES Convention, New York, October 6-9, 1995.

[9] Peter Lee "**Wavelet Filter Banks in Perceptual Audio Coding**," M. Sc. Thesis, Waterloo, Ontario, Canada, 2003

[10] M. B. sandler, etc., "**Audio Coding for Mobile Multimedia Communications**," IEEE Colloquium on the future of Mobile Multimedia communications, pp. 11/1-11/9, Dec. 1996.

[11] T. Painter, "**Perceptual Coding of Digital Audio**," PROCEEDINGS OF THE IEEE, VOL. 88, NO. 4, APRIL 2000.

[12] S.G Mallat "**A Theory for Multiresolution Signal Decomposition: The Wavelet Representation**" IEEE.Transactions on Pattern Analysis and Machine Intelligence, Vol.11,1989,674-693

[13] Mpeg web site at "**http://www.tnt.unihannover.de/project/mpeg/audio**," .

[14] M. Raad, A. Mertins, and I. Burnett, "**Audio compression using the MLT and SPIHT**,"

Available at http://www.elec.uow.edu.au/staff/wysocki/dspcs/papers/025.pdf

[15] David Salomon, "**Data Compression, The Complete Reference**," Springer-Verlag, New York, Inc., 2004.

[16] G. Stoll, and F. Kozamernik, "**listening tests on Internet audio codecs",** EBU Technical review, June, 2000.

[17] K. Brandenburg, and H. Popp, "**An introduction to MPEG layer-3**", EBU Technical review, June, 2000.