A Hi-Fi Audio Coding Technique for Wireless Communication based on Wavelet Packet Transformation♠

Sansanee Netirojjanakul, Watcharee Veerakachen, and Mongkol Raksapatcharawong

SCORPion♣ Group, Electrical Engineering Department
Kasetsart University, Jatujak, Bangkok 10900 Thailand
E-mail: fengmkr@ku.ac.th http://www.ee.ku.ac.th/scorpion

Abstract

This paper presents a sufficiently low bit rate Hi-Fi audio coding technique with low computation designed for transmitting real-time high-quality audio signal over wireless channel. This technique applies wavelet packet transform to decompose audio signal into subbands to eliminate redundant data using spectral and temporal masking properties. The encoded audio data is framed with some critical field is protected by channel coding to improve noise immunity when frames are transmitted wirelessly. Experimental results show that transparent CD-audio quality can be achieved at 80kbps encoding bit rate. Moreover, the proposed technique still offers near CD-audio quality when frames are transmitted over AWGN channel with BER below 10^{-5}. These encouraging results clearly exhibit the superior features of our technique compared to others such as Ogg/Vorbis and MP3, which are ubiquitously employed nowadays.

Keywords: *hi-fi audio coding, wavelet packet transform.*

1. Introduction

Although wireless communication has played a great role in our lifestyle but to transmit high fidelity (a.k.a. hi-fi) audio signal wirelessly at a reasonable cost is still challenging. Presently available audio coding techniques aim to reduce bit rate and put less concern on complexity and efficient wireless transmission. Such audio CODECs like ISO/MPEG [1,2,3] and Ogg/Vorbis [4] are suitable for non real–time applications and audio archive. On the other hand, this research focuses on a hi-fi (near CD quality) audio coding technique that provides sufficiently low bit rate and complexity (hence, can be processed in real–time). In addition, the technique must tolerate the noise (i.e., bit error rate—BER) in wireless channel at a reasonable degree.

Current hi-fi audio CODECs employ entropy coding such as run-length and Huffman code [3,5,6], where important parameters required to decode are assumed to be error–free. Otherwise, the frame will be discarded making it susceptible to noisy wireless channel. As a result, we identify such parameters and protect them using channel coding that can correct up to 29 bits (one of every seven bits). Perceptual coding technique is used reduce the bit rate based on "human hearing masking" property. In general, perceptual codec consists of five modules which are filer bank, psychoacoustic analyzer, bit allocation, quantizing and encoding, and framing, as shown in Figure 1.

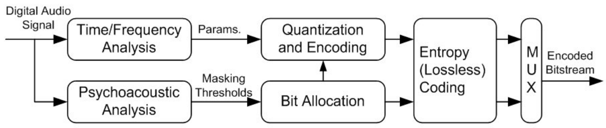

Figure 1 Structure of Perceptual Codec

The output from filter bank (time/frequency analysis) is quantized according to masking thresholds calculated by the psychoacoustic analyzer. In [1,2] use polyphase filter bank which requires 512 coefficients to represent each filter. This can take significant time to encode the signal. In contrast, we use wavelet–based filter bank to transform signal into wavelet domain that analyzes both time and frequency simultaneously. The filter bank uses fewer coefficients and can represent variable sized sub-bands that more accurately match the characteristic of non–stationary audio signal [7]; human can only detect frequency difference at low– or medium–frequency. This hypothesis leads us to believe that the proposed technique should be simpler (can be implemented in hardware). Moreover, MATLAB experiments confirm that, at 80kbps, our wavelet–packet audio codec yields comparable audio quality to that of the 64kbps MP3 and Ogg/Vorbis. It is worth noting that the higher bit rate of wavelet codec partially accounts for channel coding. This makes it wireless transmission ready.

The rest of the paper is organized as follow: Section 2 elaborates on structure, function, and algorithm of the proposed wavelet–packet audio codec. Experimental results are shown in Section 4. Section 5 concludes and presents future work.

♠ This work was supported by National Science and Technology Development Agency (NSTDA) and National Electronic and Computer TEchnology Center (NECTEC) through the project "A Feasibility Study of Wireless Hi-Fi Audio Transmission using Direct Sequence Spread Spectrum (DSSS) Technique", grant number NT–C005–47.

♣ ***SCORPion*** stands for ***S***uperior ***CO***mmunication ***R***esearch and ***P***rototyping for commercializat***ion***

2. Wavelet-packet Audio CODEC

This section explains structure and function of the proposed wavelet–packet codec which takes audio CD samples (44.1ksps @ 16bps PCM) as its inputs. Audio samples are framed at 1,024 samples each [7]. Fewer samples would affect the coding efficiency whereas more samples would impose long coding delay, violating the real–time requirement. Subsequent frames share 16 overlapping samples to reduce discontinuity between reconstructed frames at the decoder. Each frame is windowed by raised cosine filter to avoid sudden change of the signal [7] then proceeds to the encoder shown in Figure 2.

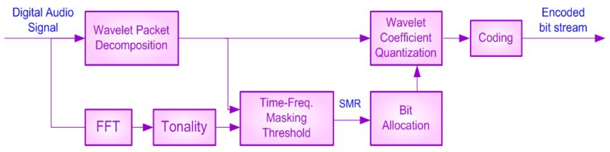

Figure 2 Structure of Wavelet-packet Audio CODEC

Wavelet Packet Decomposition

This part decomposes and transforms the audio frame into wavelet domain. The decomposition tree consists of 29 subbands [8] chosen to resemble the critical band of human hearing as depicted in Figure 3. Each subband will be quantized differently according to the signal–to –mask ratio (SMR) calculated by the Psychoacoustic Model (discussed later).

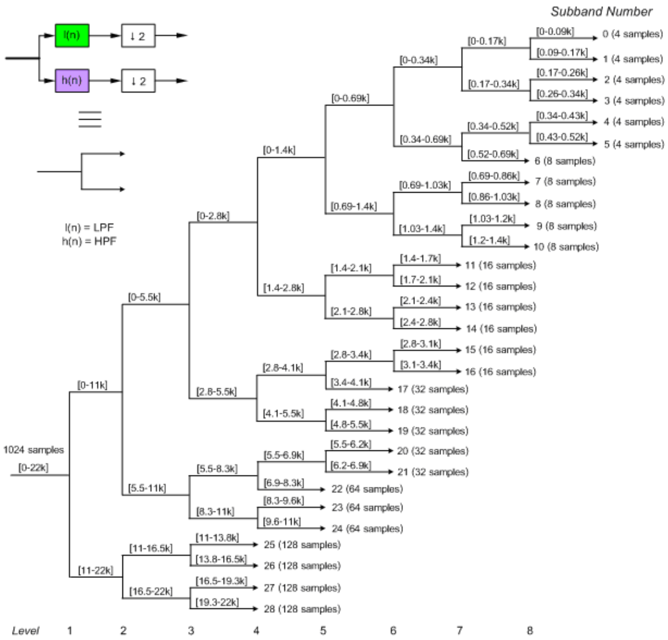

Figure 3 Decomposition Tree for 44.1kbps Audio Frame

To avoid cases where wavelet coefficients are more than the number of samples in time domain, each frame is viewed as periodic. Eq [1] is used to decompose frame and down–sampling by two at each level of the tree when there are N samples (s), L wavelet function coefficient (h), and L scaling function (l).

$$\begin{bmatrix} l_0 & l_1 & l_2 & \cdots & l_{L-1} & 0 & 0 & 0 & \cdots & 0 \\ h_0 & h_1 & h_2 & \cdots & h_{L-1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & l_0 & \cdots & l_{L-3} & l_{L-2} & l_{L-1} & 0 & \cdots & 0 \\ 0 & 0 & h_0 & \cdots & h_{L-3} & h_{L-2} & h_{L-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{L-2} & l_{L-1} & 0 & 0 & \cdots & 0 & l_0 & l_1 & \cdots & l_{L-3} \\ h_{L-2} & h_{L-1} & 0 & 0 & \cdots & 0 & h_0 & h_1 & \cdots & h_{L-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{N\times N} \bullet \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ \vdots \\ s_{N-1} \end{bmatrix}_{N\times 1} = \begin{bmatrix} a_0 \\ c_0 \\ a_1 \\ c_1 \\ a_2 \\ c_2 \\ \vdots \\ c_{N/2-1} \end{bmatrix}_{N\times 1} \qquad (1)$$

Psychoacoustic Analysis

Fundamental concept of the perceptual encoder is to eliminate the audio signals that human cannot perceive because of signal masking or ambiguity. In fact, we need not encode nor be interested in signal components that are below the hearing threshold. Three masking patterns, which are the absolute threshold of hearing, frequency masking, and temporal masking [9], are therefore used to find the appropriate quantizing bits for each subband while minimizing the quantization noise.

Algorithm to find masking threshold is shown below:

1) Each frame is windowed by Hanning filter to reduce spectrum spread and then transformed using FFT. Power spectrum is calculated for each point (X_k) which is used to find the spectral flatness measure—SFM in the next step. SFM is used to determine the noiselike or tonelike nature of the signal.

2) Calculate the masking energy offset for each sub-band and subtract it from the energy of that subband ($E_{sb}(i)$). This offset depends on characteristic of the signal if it is similar to tone or vice versa. In [10], tone–masking–noise and noise–masking–tone are lower than $E_{sb}(i)$ about 14.5+i dB and 5.5dB. Hence, the offset O_i in dB can be calculated by

$$O_i = \tau(14.5 + i) + (1 - \tau)5.5 \qquad (2)$$

Where tonality coefficient (τ) is

$$\tau = \min\left(\frac{SFMdB}{SFMdB\max}, 1\right) \qquad (3)$$

Where $SFMdB_{MAX}$ = –60dB, if $SFMdB$ = –60dB then signal is entirely tonelike else if $SFMdB$ = 0dB then signal is completely noiselike.

3) The spreading of bark energy $SP_F(i)$ is the convolution of energy in the subband that is considered tone in (2) using spreading function. The result is averaged over all coefficients within the critical band.

4) Temporal masking is based on temporal spreading energy in the critical band, which is the convolution between the temporal energy sequence (C^2_{ij}) and the resemble version of the linear temporal spreading function [11] within the critical band. The temporal

spreading energy complies $SP_T(i,j) \geq C^2_{ij}$, thus, we can define the temporal masking factor as

$$\beta_{ij} = \frac{SP_T(i,j)}{C^2_{ij}} \geq 1 \tag{4}$$

Here β_{ij}=1 means that C_{ij} is not masked by adjacent coefficients whereas β_{ij}>1 means otherwise and the temporal–frequency masking (M_{ij}) can be estimated by adding $SP_F(i)$ with β_{ij}.

5) The absolute threshold of hearing ($ATH_{SPL}(f)$) is the average sound pressure sound pressure level (SPL) below which the human ear does not detect any stimulus. This threshold is represented in dB as

$$ATH_{SPL}(f) = 3.64 f^{-0.8} - 6.5 e^{-0.6(f-3.3)^2} + 0.001 f^4 \tag{5}$$

Where f is frequency in kHz.

6) Compare M_{ij} with $ATH_{SPL}(f)$ for each subband and select the higher value to represent the masking threshold at that frequency.

7) Find the minimum masking threshold for each subband that will be used to calculated the signal–to–mask ratio ($SMR_{sb}(i)$) of that subband.

Bit Allocation for Wavelet Coefficients of the Subband

To allocate bits for the subband, we find the mask–to–noise ratio (in dB) MNR=SNR–SMR. The subband with minimum MNR is allocated bits first because lower MNR value reflects lower noise masking. This reduces quantizing error and minimizes noise that might be heard in the reconstructed audio signal.

Quantizing the Wavelet Coefficients

Wavelet coefficients are normalized before quantizing to improve accuracy. We define 64 scale factors to be selected as appropriate for each subband. The scaled coefficients are uniformly quantized according to the number of bits received from bit allocation algorithm. Experimental results suggest that the quantized values should be threshold adjusted to be all positive because they yield better reconstructed audio signal.

Data Framing

Frame structure is illustrated in Figure 4. The 24–bit sync field (010101…) is used for frame synchronization. The 203–bit field signifies the number of bits allocated to each subband. The following (maximum) 174–bit field stores scaling factors for each subband. The last variable bits field contains the quantized wavelet coefficients.

It is clear that bit errors during transmission through wireless channel would significantly affect the decoding performance. However, within the frame, the 203–bit field that provides bit allocation is the most sensitive to errors because the decoder would not know the correct bit allocation of the subbands. This may lead to frame skipping or incomplete decoding. As a result, we protect this field using BCH(7,4) code which can correct one bit out of 7 bits (4 bits represent the bit allocation for the subband and 3 bits are for error correction), and up to 29 bits can be corrected.

The scaling factor is 6–bit data for each subband and is sent only for the subband with non–zero bit allocation. The length of quantized wavelet coefficient data depends on binary encoding of all subbands. We pad bit 0's to the frame to keep the bit rate constant and to ease the decode process at the decoder.

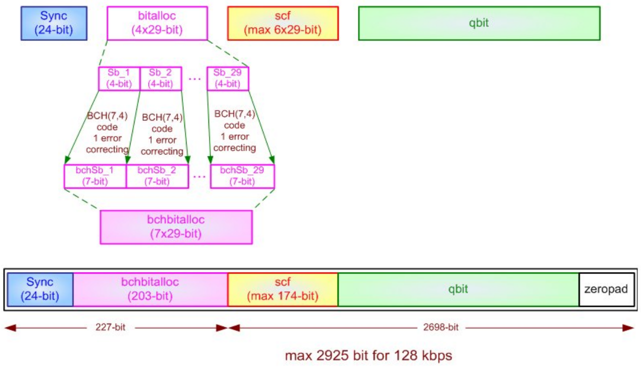

Figure 4 Frame Structure of Wavelet-Packet Audio Coder

3. Experimental Results

The proposed wavelet–packet audio codec is realized as m–files and simulated using MATLAB software. We adjust parameters such as structure of the decomposition tree, frame size, number of wavelet coefficients, etc. The suitable set of parameters is selected to optimize among decoded audio quality, encoded bit rate, and computation complexity.

To measure audio quality, we input mono audio–CD signal (at 705.6kbps) to the proposed encoder/decoder. The signal–to–noise ratio (SNR) is calculated and used as a preliminary audio quality index. We later on play the decoded audio signals in the control room among 29 attendees to measure the hearing satisfaction and bit rate, compared to the original audio–CD and audio signals decoded by the MP3 and Ogg/Vorbis software decoder. This experiment was held at NECTEC on July 9th, 2004.

To measure noise performance, the encoded data is sent through AWGN channel (using software) at various BERs and is reconstructed at the decoder. We perform the same for both MP3 and Ogg/Vorbis CODECs and use hearing satisfaction to compare noise immunity.

It is widely accepted that SNR cannot truly represent audio quality under perceptual codec and our results confirm this concept. As shown in Table 1 is the SNR comparison among wavelet packet audio CODEC at bit rate of 96 kbps, MP3 and Ogg/Vorbis at bit rate of 64 kbps measured through MATLAB simulation. Low SNR yields low audio quality but high SNR may or may not represent high audio quality. Hence, hearing satisfaction is used instead. We select ten audio–CD samples; 8 songs and 2 human voices and input them to the encoder and try to minimize the encoding bit rate. The lowest achievable bit rate is 80kbps which provides comparable audio quality to that of 64kbps MP3 and Ogg/Vorbis CODECs. Experiment performed at NECTEC concludes that more than 66% of the attendees cannot distinguish (more than half of the samples) between the original audio–CD samples and the decoded wavelet–packet audio samples.

Table 1 SNR Comparison of the proposed CODEC and other schemes

Music Samples	Wavelet	MP3	OGG/Vorbis
Alone.wav	21.65	22.9	21
Along Comes a Woman.wav	16.0657	16.35	14.82
Hero.wav	21.87	19.8	20.24
Morning.wav	20.05	18.93	21.27
Piano Sonata in C, K545.wav	36.81	25.87	26.65
What Kind of Man Would I Be.wav	16.19	18.34	17.9
Without You.wav	19.97	18.67	18.77
You're the Inspiration.wav	18.05	19	18.68
news-man.wav	26.31	24.45	21
news-woman.wav	29.93	24.95	22.5

Noise performance of the considered audio CODECs is summarized in Table 2. Clearly, MP3 and Ogg/Vorbis are designed for storage of which low encoded bit rate is great importance. The MP3 decoder can tolerate BER of less than 10^{-6} (at this point the audio quality becomes acceptable) which is not easily achieved in wireless communication. The Ogg/Vorbis decoder is even worse, it cannot decode without frame drops at BER of 10^{-7}. On the other hand, the wavelet-packet decoder can tolerate BER up to 10^{-5} with acceptable audio quality. At BER of 10^{-4} or more, it still can decode the audio signal without frame drops but one may hear sound like water dropping every once in a while. This better noise immunity clearly benefits from BCH coding in the bit allocation field.

Table 2 Noise Performance Comparison

CODEC	Bit rate	BER_{MAX}	Remark
MP3	64kbps	10^{-6}	Audio quality is unacceptable at BER higher than 10^{-6}
Ogg/Vorbis	64kbps (max)	–	Audio skips (frames dropped) even at BER lower than 10^{-7}
Wavelet-packet	80kbps	10^{-5}	Audio is better than the two even at BER higher than 10^{-5}

We compare computation complexity of the wavelet packet audio CODEC and MP3 based on the number of multiply-and-accumulate (MAC) operations required per audio frame (1024 samples) by their associated encoders. Evidently, the proposed scheme takes less than half of that required by the MP3, 37962 versus 81920!

4. Conclusion and Future Work

This paper presents wavelet–packet transformation technique used to encode high quality audio signal for wireless communication. Currently, the encoded bit rate of 80kbps, header included, yields transparent audio–CD quality. This technique is potentially less computation compared to the ubiquitous MP3 and Ogg/Vorbis which provide similar audio quality at 64kbps. Consequently, the wavelet–packet audio codec should easily be realized in hardware and meets the real–time requirement of high quality audio communication. In conjunction with high noise immunity, we believe that the proposed technique is suitable for wireless application such as wireless microphone, wireless speaker, and so on. Last but not least, it is design and developed by Thai research group which we can use with no royalty fee (MP3 is licensed whereas Ogg/Vorbis is open–source).

We currently implement the codec in hardware using FPGA (shown in Figure 5) and expect the prototype to be ready in 2Q05. The result should confirm its real–time performance and all other claims made here. We are also working towards lower bit rate (64kbps) wavelet–packet audio codec that would gain more momentum because it can match the current PCM bit rate at much better audio quality.

5. Acknowledgements

We thank NECTEC/NSTDA for their funding and SCORPion group members who always help each others to overcome all the obstacles.

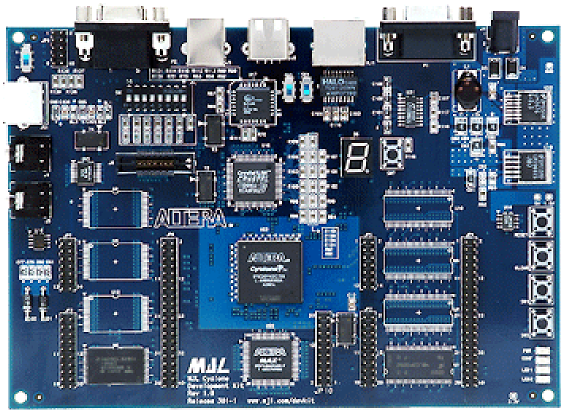

Figure 5 An FPGA Prototype of the CODEC

6. References

[1] D. Pan, "A tutorial on mpeg/Audio Compression", *IEEE Multimedia*, pages 60–74, 1995.

[2] K.B. Brandenburg and G. Stoll, "ISO–MPEG–1 audio: A generic standard for coding of high-quality digital audio", *J. Audio Eng. Soc.*, 42(10):780–792, October 1994.

[3] ISO/IEC 11172-3, Information Technology–Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbps*,* August 1993.

[4] http://www.xiph.org/ogg/vorbis/

[5] W. Kurt Dobson, J. Jack Yang, Kevin J. Smart, and F. Kathy Guo, "High Quality Low Complexity Scalable Wavelet Audio Coding," *ICASS' 97*, pp. 327–330.

[6] N. Ruiz et al., "New Algorithm for Searching Minimum Bit Rate Wavelet Representations with Application to Multi resolution–Based Perceptual Audio Coding." *ICPR'00*, Vol.3, pp 286–289, September 2000.

[7] H.E. Dong-Mei and GAO Wen, "Wideband Speech and Audio Coding Based on Wavelet Transform And Psychoacoustic Model.", *ISCSLP'98*, pp.261–264, 1998

[8] A.H. Tewfik and D. Sinha, "Low–bit Transparent Audio Compression using Adapted Wavelets", *IEEE Transactions on Signal Processing*, 41(12), pp 3463–3479, Dec. 1993.

[9] B. Cavagnolo and J. Bier, "Introduction to Digital Audio Compression"

[10] J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria" *IEEE Journal on Selected Areas in Communications*, 6(2), pp 314–323, February 1988.

[11] Benito Carnero and Andrzej Drygajlo, "Perceptual Speech Coding and Enhancement Using Frame–Synchronized Fast Wavelet Packet Transform Algorithms" *IEEE Transactions on Signal Processing*, 47(6), pp 1622–1635, June 1999.