Scalable Embedded Zero Tree Wavelet Packet Audio Coding

Pao-Chi Chang and Jen-Hsin Lin

Department of Electrical Engineering, National Central University
Chung-Li, Taiwan 32045
E-mail: pcchang@mbox.ee.ncu.edu.tw
Fax: 886 3 4255830
Phone: 886 3 4227151 ext. 4466

Abstract **- Multimedia transmission over Internet is getting popular and increasingly important. In particular, scalable coding is desirable for heterogeneous network with varies bandwidths. In this work, we propose a scalable embedded zero tree wavelet packet (Scalable EZWP) audio coding system that is a scalable audio compression system using wavelet packet decomposition and embedded zero-tree coding. We focus on multi-layer low bitrate coding which delivers high perceptual quality. In the base layer, the overlapped audio segment is first transformed by wavelet packet. Then the local significant coefficients are extracted, quantized, and coded by variable length coding. In the enhancement layer and the full band layer, the residual signal that is the difference between the original and the output of the previous layer is coded via EZW with psychoacoustic model and arithmetic coding. The target bit rates for three layers are 16, 32, and 64 Kbps, respectively. The performance of the proposed coding system is only slightly inferior to MPEG-1 layer 3 at 64 Kbps while it provides bitrate scalability that is suitable for multimedia distribution over Internet with heterogeneous networks.**

I. INTRODUCTION

The transmission of audio and video signals over Internet becomes increasingly important at this moment. The bitrate scalability is desirable for transmission over Internet which is composed of heterogeneous networks. The demands for scalable, high quality, low complexity, and low bitrate coding reveal the importance of good and efficient audio compression. In 1995, Karelic and Malah proposed a wavelet-packet based zero-tree coder which was superior to MPEG Layer I [1]. However, the psychoacoustic model was not taken into account. In 1998, an audio coding system with adaptive wavelet packet decomposition and psychoacoustic modeling was designed by Srinivasan and Jamieson [2]. In this work, we propose a scalable embedded zero tree wavelet packet (Scalable EZWP) audio coding system that is a scalable audio compression system using wavelet packet decomposition and embedded zero-tree coding.

In section II, we describe general aspects of the wavelet packet decomposition structure. Section III shows how the psychoacoustic model is applied to this system. Then we present the modifications of embedded zero-tree coding in Section IV. The three-layer scalable EZWP coding is presented in Section V. The simulation results are discussed in section VI. Finally, conclusions are given in section VII.

II. WAVELET PACKETS

The wavelet packet system was proposed by Ronald Coifman [3] to allow a finer and adjustable resolution of frequencies at high frequencies. It also gives a richer structure that allows adaptation to particular signals or signal classes.

To achieve efficient compression, the bandwidth of each subband should match that of critical bands as closely as possible. Therefore, the decomposition structure of our system is chosen as in Fig. 1 [4]. The bandwidth of each subband compared with each critical band is illustrated as Fig. 2. As for wavelet filters, Scalable EZWP system uses the biorthogonal 18-tap FIR filter, which has linear phase property [5]. The audio signal is divided it into short stationary segments to get good compression efficiency. The segment is chosen to be 1024 samples which is about 23.32 ms at 44.1KHz.

III. PSYCHOACOUSTIC MODELING

The human auditory system (HAS) [6] has many useful features as to audio compression. Psychoacoustic model makes it practical via a series of mathematical formulas. The goal of the model is to obtain the minimum masking threshold of each subband. Then we can adjust the quantization resolution such that the quantization error is below these thresholds and thus inaudible.

Scalable EZWP system has the same model as MPEG psychoacoustic model II. The noise-masking thresholds for the critical bands are calculated via a 1024-point FFT. The tonality measure, which ranges from 0 to 1, is based on the predictability of the current frame from the past two frames. The spreading function describes the property of the ear-to-mask noise at a frequency in the neighborhood of a tone. Then, the "just masked" noise level, that is minimum masking threshold, is calculated from spreading function and the tonality index. The absolute threshold of hearing (ATH) and pre-echo control are also incorporated. Finally, the minimum threshold for each subband is extracted.

IV. EMBEDDED ZERO-TREE CODING

Shapiro proposed the wavelet transform based embedded zero-tree coding for images in 1993 [7]. The "self-similarity" of wavelet coefficients is the key point to make zero-tree algorithm efficient in coding significant map. In the case of audio signal, wavelet packet coefficients of some subbands are highly correlated, too. Furthermore, the perceptual characteristics also support the embedded property.

The first step to perform zero-tree scanning is to determine the tree structure of coefficients, that is based on the relationship between ancestors and descendents. In this work, we consider the harmonics of audio signal. Assuming a coefficient is insignificant, it exists a relatively high probability for most instruments that the coefficients in its harmonics are also insignificant. We also notice that most of the energy of wavelet coefficients concentrate in low frequency bands. In order to generate zero-tree root (ZTR) symbol more easily, we define several high-energy bands to be the roots of the sub-trees. Fig. 3 shows the harmonic sub-tree structure used in the proposed system.

V. THREE-LAYER SCALABLE EZWP AUDIO CODING

The architecture of the proposed system including the encoder and the decoder is showed in Fig.4. It is a scalable audio compression system using wavelet packet decomposition and embedded zero-tree coding. In the base layer, the overlapped audio segment is first transformed by wavelet packet. Then the local significant coefficients are extracted, quantized, and coded by variable length coding. In the enhancement layer and the full band layer, the residual signal that is the difference between the original and the output of the previous layer is coded via EZW with psychoacoustic model and arithmetic coding. The target bit rates for three layers are 16, 32, and 64 Kbps, respectively.

A. Base layer

The base layer provides the minimum audio performance for browsing or indexing an audio sequence. The base layer algorithm is shown in Fig. 5. The wavelet coefficients are passed through local-significant-peak-search algorithm. Relatively high energy coefficients in the neighborhood are extracted as local significant coefficients. The local significance is obtained by using a 3-point sliding window to scan through wavelet coefficients, and then keeping the maximum values of sliding windows. All other coefficients are set to zero. The searching range is limited within the first to the 14^{th} wavelet packet bands to keep the bitrate low. Having been extracted, the peaks are then quantized by 16-level uniform quantizers with different step sizes in different bands. An example of the search result is shown in Fig. 6. Although the psychoacoustic model is not explicitly used, many of the non-peaks which are coded as zeros will be masked by neighboring peaks. Finally, the quantized values are coded by run length coding and Huffman coding for low implementation complexity.

B. Enhancement layer and full band layer

In the enhancement layer and the full band layer, the residual signal that is the difference between the original and the output of the previous layer is coded via EZW with psychoacoustic model and arithmetic coding. The searching range of enhancement layer is extended to the 20^{th} wavelet packet band, and all of the bands for full band layer. For better perceptual quality, we incorporate the psychoacoustic model into the EZW coding. The successive approximation quantization for residual wavelet coefficients is not terminated until the quantizaton noise is below the masking threshold.

VI. SIMULATION RESULTS

We perform simulations in C language on Pentium-II 400 personal computers. All test audio signal are 5-second long monophonic with CD quality which

implies the sampling rate is 44.1 KHz with 16 bits per sample. The evaluation criteria of audio quality include segmental SNR (SSNR) objectively or listening test subjectively.

SSNR of the reconstructed audio in each layer is shown in Fig. 7. The SSNR of base layer is roughly 10 dB while the enhancement layer is about 16 dB and the full band layer is about 21 dB. The bitrate of each layer is shown in Fig. 8. The bitrate is not a constant function of time. However, the averages are roughly 16 Kbps, 32 Kbps, and 64 Kbps, respectively.

Listening tests were performed to compare the proposed system with audio coders with the same bitrate available on Internet. The result of listening test is shown in Table 1. The preference ratio records the percentage of listeners who favor the proposed system. At low bitrates, e.g., 16 or 32 Kbps, the proposed coding system has good perceptual quality. At 64 Kbps bitrate, the quality of the proposed system which is not optimzed in many details is only slightly inferior to MPEG-1 layer-3.

VII. CONCLUSION

We have presented the scalable EZWP audio coding system in this paper. It provides bitrate scalability which is very suitable for Internet transmission with different bandwidths. In all layers, the performance of the proposed system is comparable to existing systems at the same rate.

REFERENCES

[1] Y. Karelic and D. Malah, “Compression of High-Quality Audio Signals Using Adaptive Filterbanks and A Zero-Tree Coder,” *Electrical and Electronics Engineers in Israel*, 1995.

[2] P. Srinivasan and L. H. Jamieson, “High-Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling,” *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 1085-1093, April 1998.

[3] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Information Theory*, vol. 38, pp. 713-718, March, 1992.

[4] D. Sinha and A. H. Tewfik, “Low Bit Rate Transparent Compression using Adapted Wavelets,” *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3463-3479, Dec. 1993.

[5] I. Daubechies, "Ten Lectures on Wavelets," no. 61 in CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[6] E. Zwicker and H. Fastl, Psychoacoustics, Facts and Models (Springer, Berlin, Heidelberg, 1990).

[7] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Trans. Signal Processing, Spec. Issue Wavelets Signal Processing*, vol. 41, pp. 3445-3462, Dec. 1993.

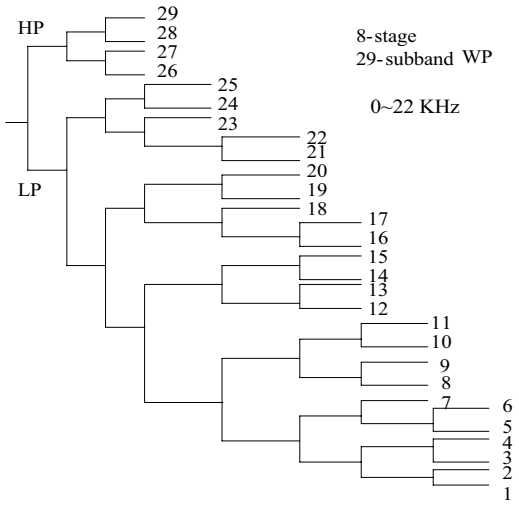


Fig. 1 Decomposition structure of Scalable EZWP system

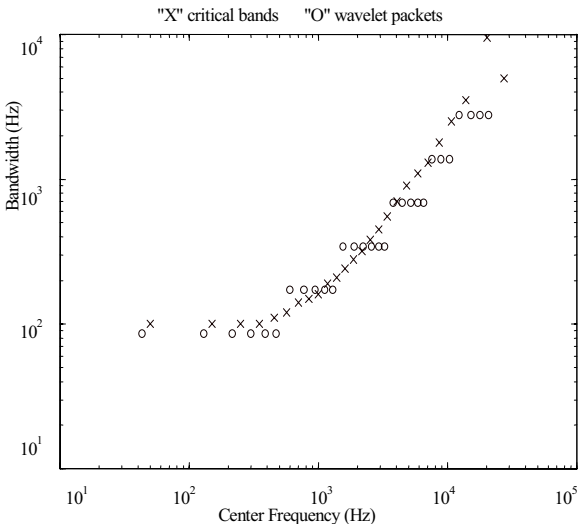


Fig. 2 Bandwidths of wavelet-packet subbands vs. critical bands

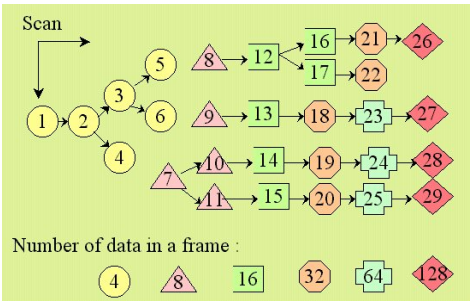
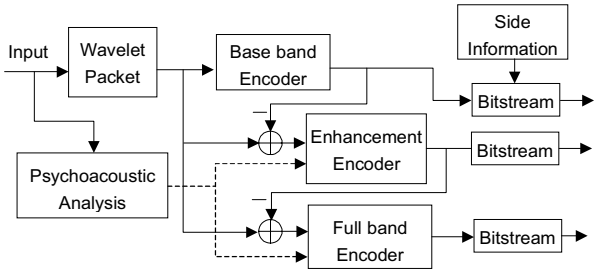
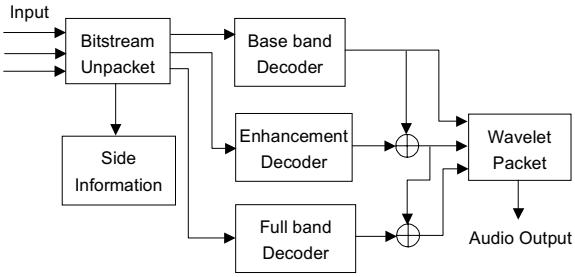


Fig. 3 Harmonic sub-tree structure

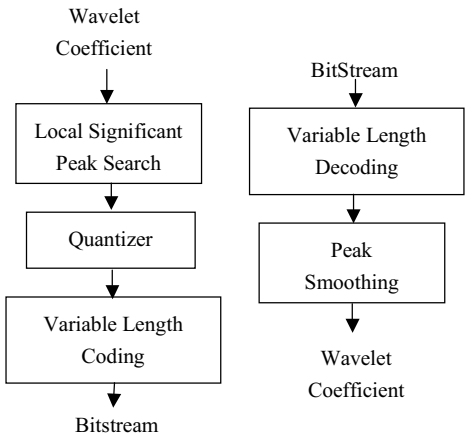


(a) Encoder



(b) Decoder

Fig. 4 Scalable EZWP Codec



(a) Encoder (b) Decoder

Fig. 5 The base layer codec

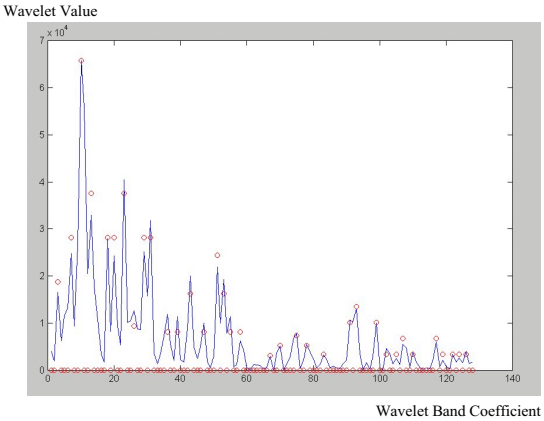


Fig. 6 An example of local significant peak search
Circle : Recorded and quantized value
Solid line : Original wavelet packet coefficients

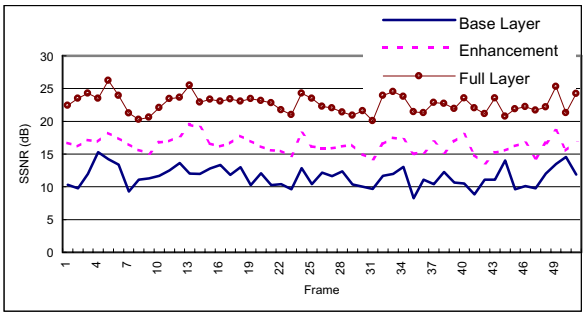


Fig. 7 SSNR of three layer coder

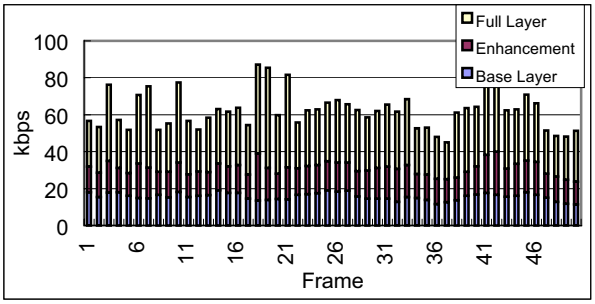


Fig. 8 Bitrate of thre layer coder

Preference Ratio (%)	Base layer 16 Kbps	Enhancement 32 Kbps	Full Band 64 Kbps
chorus	85	83	63
flute	55	63	46
guitar	83	83	71
harpsichord	67	63	46
horn	49	47	29
lute	72	81	69
orches	90	76	74
organ	51	47	34
piano	50	51	40
trumpet	62	68	66
violin	90	90	77

Table 1. Preference ratio of listening test