

# High Quality Switched Wavelet Packet and Discrete Cosine Transform Audio Coding Technique

*Pao-Chi Chang, Tzung-Hung Wu, and Kong-Yueh Ho*

Department of Communication Engineering, National Central University  
Chung-Li, Taiwan 320

## ABSTRACT

We propose a hybrid coding system that utilizes both Wavelet Packet (WP) and DCT techniques. To process each audio frame, the system selects either WP or DCT as the transform based on the frame flatness measures in the wavelet domain and the frequency domain. If DCT is chosen, all DCT coefficients are quantized by a non-uniform quantizer according to the frequency masking curve. On the other hand, frame data are segmented into 26 fixed subbands when WP is chosen. Then, the system selectively utilizes DCT to promote frequency resolution of each subband based on the subband flatness measure. By quoting optimal bit-allocation for non-ideal filter bank, the masking threshold from psychoacoustic model can be translated into specific criteria in the wavelet domain for quantization. Experiment results show that the proposed system is superior to MP3 and AAC LC profile at 64k bps.

## 1. INTRODUCTION

The transmission of multimedia over network and the storage of multimedia in portable device become increasingly important at the moment. The demands of high quality, low complexity, and low bitrate reveal the importance of good and efficient audio compression. The wavelet packet transform is one of the efficient techniques applied to audio coding [1-2]. In 1993, Sinha and Tewfik proposed an audio coder with adaptive wavelets [3]. They discovered that adaptive wavelets with maximum vanish moments has significant improvement in bit rate. In 1998, an audio coding system with adaptive wavelet packet decomposition was designed by Srinivasan and Jamieson [4]. However, their coders only use wavelet packet with limited frequency resolution without high precision masking threshold in the wavelet domain.

In this work, we propose a Hybrid Wavelet Packet and DCT (HWPDC) audio coding system adopting both wavelet packet and DCT as its transformation kernel to get either finer time resolution or better frequency

resolution. Moreover, we adopt optimum bit allocation algorithm developed by Caini and Coralli [5] that can translate the masking threshold in the frequency domain into the wavelet domain.

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 bit allocation algorithm in Section IV. The whole system architecture is presented in Section V. The simulation results are discussed in section VI. Finally, conclusions are given in section VII.

## 2. WAVELET PACKETS

Discrete wavelet transform is a set of building blocks representing signals or functions. The general statement is as in (1) where scaling functions and wavelet functions are as expansion bases of signal  $g(t)$ .  $j_0$  is the initial scale which may be zero or any integer.  $j$  is the scale of resolution and  $k$  is the translation step.

$$g(t) = \sum_k c_{j_0}(k) 2^{\frac{j_0}{2}} \phi(2^{j_0}t - k) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{\frac{j}{2}} \psi(2^j t - k) \quad (1)$$

$\phi(t)$ : scaling function,  $c_j(k)$ : scaling coefficient

$\psi(t)$ : wavelet function,  $d_j(k)$ : wavelet coefficient

The multiresolution property of wavelet transform is suitable for audio signal processing. The relationship between finer coefficients and coarser ones may be conducted as filtering formulas. Thus, we may exploit filter banks to further analyze it. We choose Daubechies compactly support wavelet as its transform kernel.

The wavelet packet system was proposed by Ronald Coifman [6] 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 to that of critical bands as closely as possible. However, the deeper the decomposition is, the longer the length of effective syntheses filter would be. Therefore, the decomposition

structure [4] as shown in Fig. 1 is chosen, in which the audio signal is divided into short stationary segments to get good compression efficiency. The segment is chosen to be 2048 samples which are about 46.44 ms at 44.1 kHz.

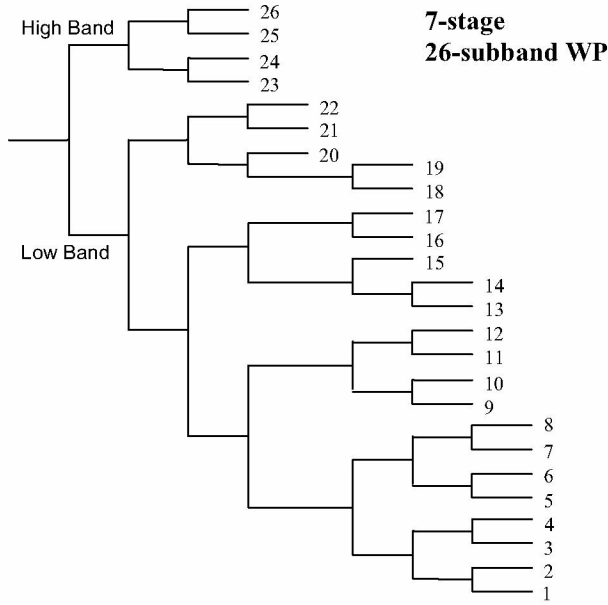


Fig. 1 Decomposition structure of HRPD system

### 3. PSYCHOACOUSTIC MODELING

The human auditory system (HAS) [7] 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.

HRPD has the same model as MPEG AAC psychoacoustic model. The noise-masking thresholds for the critical bands are calculated via a 2048-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.

### 4. BIT ALLOCATION

Because the frequency response of the wavelet syntheses filter is not ideal, especially in high decomposition levels,

a major problem in wavlet audio coding is how to allocate the noise. [5] proposes an optimum bit allocation algorithm with nonideal reconstruction filters for  $M$  nonuniform subbands as in

$$R_i = R + \frac{1}{5} \log_2 \frac{\frac{\varepsilon_i^2 \sigma_i^2 \sum_{k=1}^M \frac{h_{ik}}{w_k}}{\mu_i}}{\prod_{j=1}^M \left[ \frac{\varepsilon_i^2 \sigma_i^2 \sum_{k=1}^M \frac{h_{jk}}{w_k}}{\mu_j} \right]^{\mu_j}}, \quad (2)$$

where  $1/\mu_i$  is decimation factor,  $\sigma_i^2$  is variance of signal,  $\varepsilon_i$  is the quantizer performance factor,  $w_k$  is the masking threshold in frequency domain, and  $R_i$  is the average bits per sample of subband  $i$ . The term  $h_{ik}$  is called leakage coefficient that represents the noise energy density injected into  $k$ th frequency subband due to the quantization of  $i$ th wavelet subband.  $R$  is the constraint of average bits per sample.

### 5. SYSTEM ARCHITECHURE

The frame segmentation of HRPD is shown as in Fig. 2. Each frame contains 2048 samples with 128-sample overlapping with adjacent frames to reduce blocking effect. The head and the tail of a frame are multiplied by a 128-point square-root raised-cosine window.

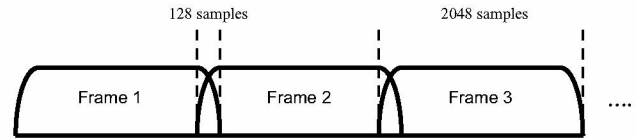


Fig. 2 Frame segmentation

The block diagram of HRPD encoder and decoder is shown in Fig. 3 and Fig. 4, respectively. Each frame data are analyzed by both WP and DCT. The flatness measures in the WP and DCT domains are calculated based on the formula in (3). A suitable transform is then chosen for this frame if its flatness measure is smaller. This switching process is called “Frame Based Mode Decision”, shown as in Fig. 3,

The flatness measure (FM) of a signal  $\mathbf{X}$  is calculated as

$$Flatness(\mathbf{x}) = \frac{\left( \prod_{n=0}^{N-1} \|x(n)\|^2 \right)^{1/N}}{\frac{1}{N} \sum_{n=0}^{N-1} \|x(n)\|^2}, \quad (3)$$

where  $N$  is the frame length and  $X(n)$  are coefficients of WP or DCT. The value of FM indicates the variation degree of signal in that domain. The smaller the value of FM is, the more signal energy concentration in a few coefficients is. Therefore, a more efficient coding may be obtained in this domain.

In addition to the frame based mode decision, a "Subband Based Mode Decision" is used in the WP path, as in Fig. 3. The flatness measure as in (3) is again used in determine the mode. According to the FM, each subband decides whether DCT will be executed or not.

Original music signal is also fed into psychoacoustic model as described in section III in another path to estimate the frequency masking thresholds. Then the bit allocation algorithm, mentioned in section IV, provides the relative noise thresholds for each wavelet subband. WP subband coefficients are then quantized by uniform quantizers with step sizes controlled by thresholds in the wavelet domain.

In Fig. 3, if this frame is more suitable for DCT, all DCT coefficients are quantized by a non-uniform quantizer with the quantization formula [10] shown as

$$Q[k] = \text{int} \left( \left( \text{abs}(dct[k]) * \sqrt[4]{2}^{sc} \right)^{3/4} + 0.4054 \right), \quad (4)$$

where  $sc$  means *scalefactor - common\_scalefactor*. Finally, all the quantized coefficients are encoded with entropy coding and encapsulated into a bitstream.

## 6. SIMULATION RESULTS

We perform simulations in C language on Pentium-4 2.8G personal computers. All testing audio signals are longer than 16 seconds, monophonic with CD quality which implies the sampling rate is 44.1 KHz with 16 bits per sample. The objective audio quality is evaluated by EAQUAL (Evaluation of Audio QUALity), which is based on ITU-R recommendation BS. 1387 [8]. Final output grade is ODG (objective difference grade) with the range of [-4, 0]. -4 represents a very annoying difference between compressed audio sample and the original while 0 means that there is imperceptible difference.

Fig. 5 demonstrates the performance of HWPDP and wavelet packet. WP is more suitable for those songs with almost flat spectrum, such as "bass" and "butter". Fig. 5 shows that the selective DCT indeed improves audio quality significantly. The overall performance is demonstrated in Fig. 6. The proposed HWPDP system is

best for popular music and orchestral music but relatively weak for attack music. Overall, HWPDP audio coding system can provide better quality than MP3 [9], AAC LC [10] at 64k bps.

## 7. CONCLUSION

Neither WP nor MDCT can be the most efficient method to encode all music. It means that both time and frequency information are important in audio coding. In this work, we have presented the hybrid wavelet packet and DCT audio compression system, HWPDP, that can provide either finer time resolution by using wavelet packet or better frequency resolution by using DCT to decompose audio signal. By utilizing "Flatness Measure Mechanism", HWPDP can select a proper transformation accurately and encode all types of music efficiently. As a result, HWPDP can provide superior quality to MP3, AAC LC at the bit rate of 64k bps.

## 8. REFERENCES

- [1] V.P. Devassia, G. Sandeep, N.S. Antony, and T. Thomas, "Audio compression using wavelets and wavelet packets," *Advances in Modelling and Analysis B*, vol.46, pp. 61-72, July 2003.
- [2] M.R. Zurera, N.R. Reyes, P.V. Candeas, and F.L. Ferreras, "Use of the symmetrical extension for improving a time-varying wavelet-packet-based audio coder," *Digital Signal Process Rev.J.*, vol.13, pp. 457-469, 2003.
- [3] 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.
- [4] 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.
- [5] C. Caini and A. V. Coralli, "Optimum Bit Allocation in Subband Coding with Nonideal Reconstruction Filters," *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 157-159, June 2001.
- [6] I. Daubechies, "Ten Lectures on Wavelets," no. 61 in CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia, 1992
- [7] P. E. Kudumakis and M. B. Sandler, "On The Performance of Wavelets for Low Bit Rate Coding of Audio Signals," in *Proc. Int. Conf. Acoust., Speech, Signal Process.* 1995, pp. 3087-3090.
- [8] ITU-R Recommendation BS.1387, Method for Objective Measurements of Perceived Audio Quality, Dec. 1998.
- [9] ISO/IEC 11172-3, "Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio," 1992.
- [10] ISO/IEC, *Final Draft International Standard 14496-3: MPEG-4 Audio*, ISO/IEC JTC1/SC29/WG11 N2503, Oct. 1998.

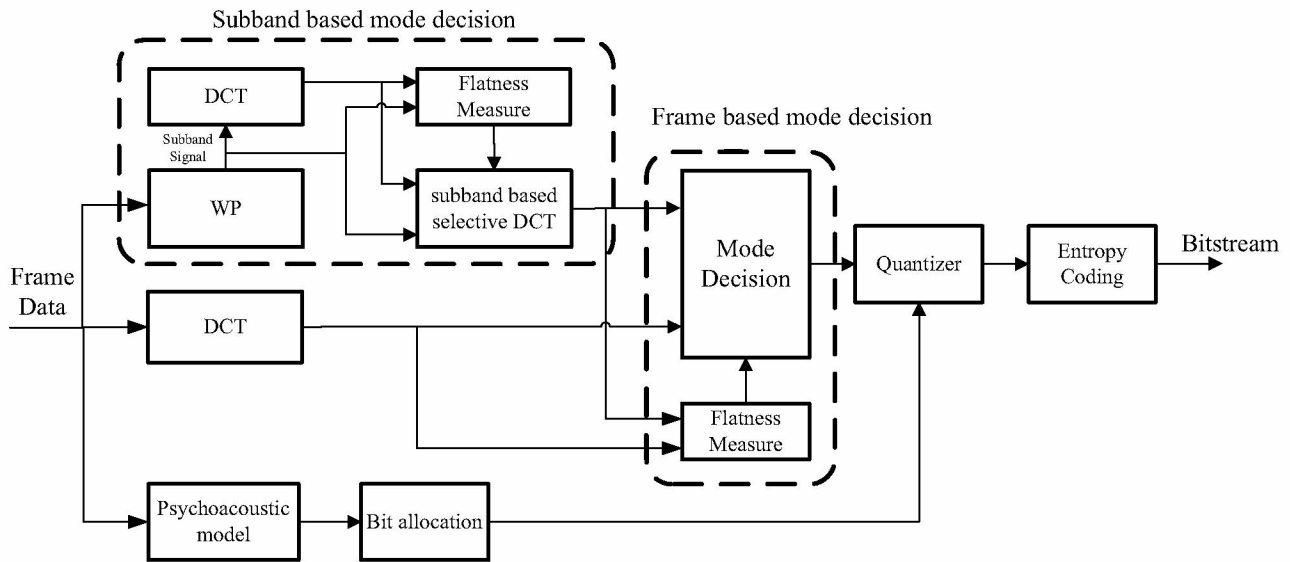


Fig. 3 HWPB encoder block diagram

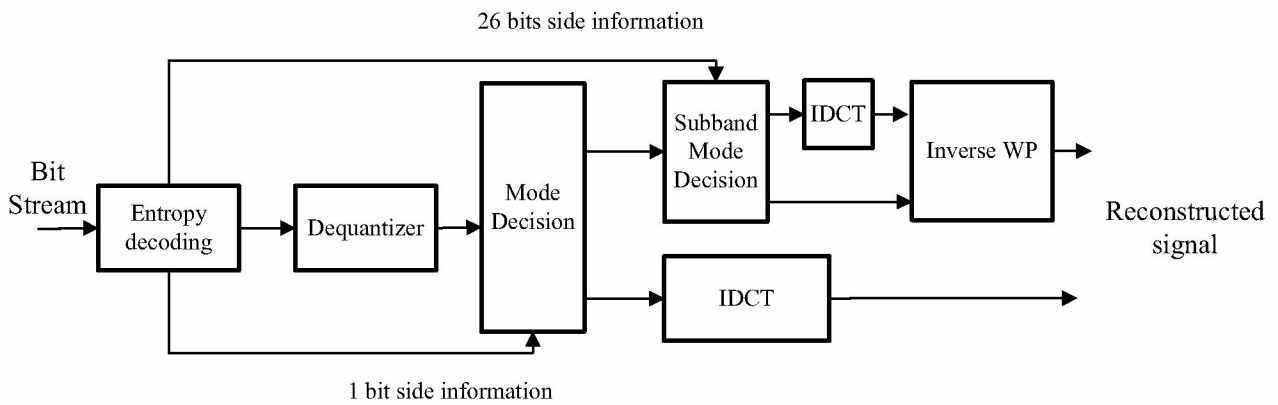


Fig. 4 HWPB decoder block diagram

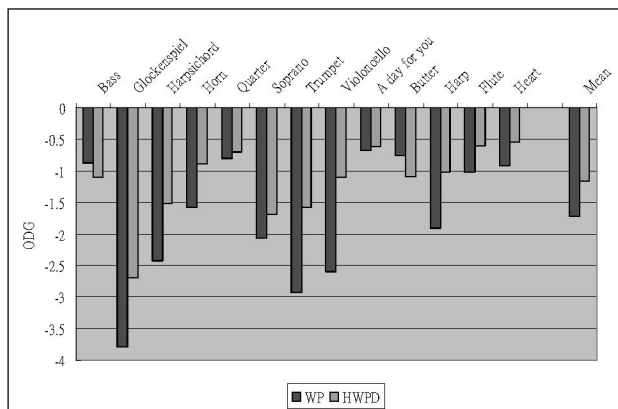


Fig. 5 Comparing WP with HWPB

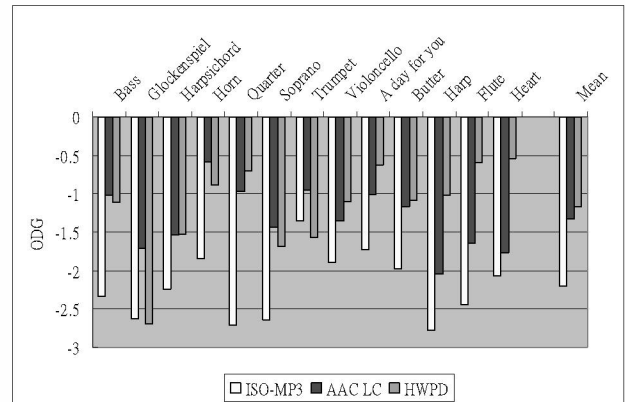


Fig. 6 Overall performance of HWPB