COMPUTATIONALLY EFFICIENT WAVELET PACKET CODING OF WIDE-BAND STEREO AUDIO SIGNALS

Mark Black

Department of Electrical Engineering
The University of Western Ontario
London, Ontario, N6A 5B9, Canada
e-mail: sblack@ee.ryerson.ca

Mehmet Zeytınoğlu

Department of Elect. & Computer Eng.
Ryerson Polytechnical University,
Toronto, Ontario, M5B 2K3, Canada
e-mail: mzeytin@ee.ryerson.ca

ABSTRACT

This paper presents a new audio compressor based on the wavelet packet (WP) decomposition. The major drawback of the present audio compressors is the large computational effort associated with subband decomposition and psychoacoustic modeling. We integrate the psychoacoustic model with the design of the decomposition filterbank which separates the wideband signal into 28 subbands closely approximating the critical bands. The psychoacoustic model exploits noise masking and joint stereo coding to compress the subband signals. We demonstrate that the WP decomposition provides sufficient resolution to extract the time-frequency characteristics of the input signal. The WP based audio compressor provides transparent sound quality at compression rates comparable to the MPEG compressor with less than one third of the computational effort.

1. INTRODUCTION

The ISO/MPEG standard [1] and other wideband audio coders [2] are based on subband decomposition. The subband signals are fed into a psychoacoustic model which identifies redundant audio information. The *psychoacoustic model* and the *decomposition filterbank* determine the computational requirements of the audio coder. The implementation of the filterbank requires considerable amount of computational effort and therefore it continues to attract research for faster implementations [3]. In the case of uniform bandwidth signal decomposition there are efficient techniques such as polyphase decomposition and modulated filterbanks [4]. The major drawback of uniform bandwidth condition is the mismatch of the signal decomposition with the psychoacoustic model which requires non-uniform decomposition of the wideband signal. In the MPEG Layer-I model, the filterbank decomposes the audio signal into 32 equal bandwidth subbands. Efficient implementation of the filterbank is achieved by a polyphase filterbank which however, cannot provide the resolution required by the psychoacoustic model. Therefore, the MPEG coder employs an FFT analyzer which increases the computational load.

In this study, we present an integrated approach to the design of the decomposition filterbank by incorporating the resolution requirements of the psychoacoustic model into the design of the decomposition filterbank. The wavelet transform (WT) has recently been proposed as a new multiresolution decomposition tool [5]. The WT is identified with a dyadic tree filterbank which provides a constant-Q decomposition. We further decompose the outputs of the WT to obtain a wavelet packet (WP) representation. The WP coder when compared to the MPEG Layer-I coder:

- can directly drive the psychoacoustic model;
- requires 1/3 of the computational effort;
- achieves comparable or better data compression;
- achieves transparent sound quality.

2. WAVELET PACKET SUBBAND CODER

The subband decomposition influences coder performance, delay and computational effort. The WP decomposition provides a closer approximation to the critical bands as defined in the psychoacoustic model. Figure 1 depicts the resolution achieved by various decomposition schemes in relation to the critical bands. The WP decomposition has also been successfully used to model human hearing [6] and to compress audio data [7]. The psychoacoustic model uses

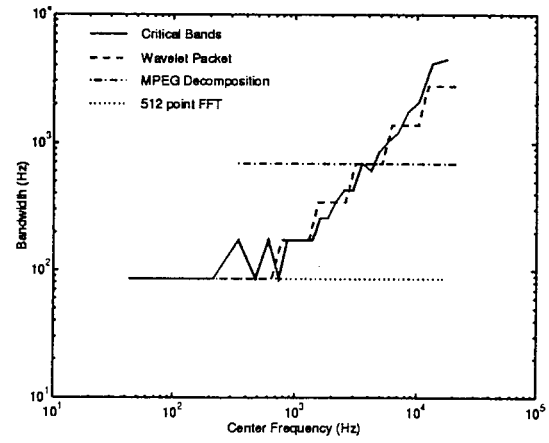

Figure 1: Approximation to *critical bands*.

This work was supported by a grant from the National Science and Engineering Research Council (NSERC), Canada.

 0-7803-2431-5/95 $4.00 © 1995 IEEE

a set of strongly overlapping bandpass filters. The bandwidths of these filters are known as the *critical bands*. Figure 1 demonstrates the suitability of the WP decomposition as a front end to the psychoacoustic model. The WP decomposition separates the wideband signal into 28 subbands as depicted in Figure 2. A closer investigation of

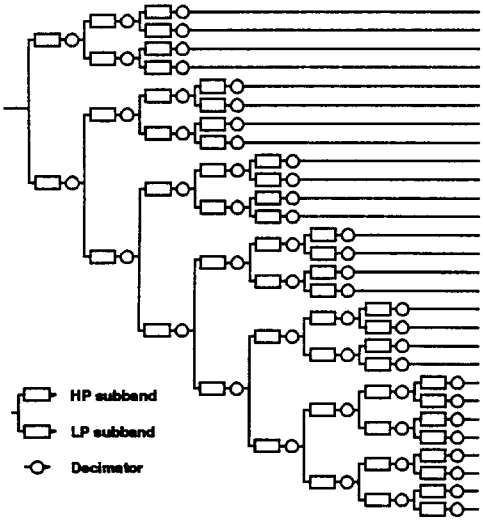

Figure 2: 28 band WP decomposition.

Figure 1 reveals that the 8-stage, 28-band WP decomposition does not exactly achieve the resolution required by the critical bands at center frequencies above 10 kHz. One can introduce an additional stage to achieve a 9-stage, 29-band WP decomposition—as used in [7]. However, for a given filter length this will further increase the total delay. Therefore, we use an 8-stage, 28-band WP decomposition at the expense of introducing minor modifications to the psychoacoustic model to address the issue of lower resolution relative to the critical bands.

2.1. Decomposition Filters

We use 16-tap FIR filters derived from the the Daubechies wavelet. This wavelet is the optimal wavelet among the class of all equal length wavelets if the optimization is carried out with respect to the compression of the audio signal [7]. In [7] Sinha and Tewfik have determined that as the order of the wavelet function is increased the filters derived from the wavelet function achieve a better separation between the subband signals. Consequently, higher compression of the wideband signal is achieved.

Three factors were considered in selecting the filter order: *subband separation*, *computational effort*, and *coder delay*. A longer impulse response sequence for the filters in the decomposition filterbank improves the subband separation while increasing the computational effort and the coding delay. In this study we use 16-tap FIR filters. Even with low order filters the decomposition tree depicted in Figure 2 achieves good separation particularly at lower frequency subbands which are deeper in the decomposition tree. It should also be noted that the polyphase decomposition filterbank employed in the MPEG model results in 16-tap FIR filters. During this phase of this study we also used a filterbank based on 32-tap FIR filters. However, this filterbank achieved only 0.731% higher compression ratio relative to the case where the filterbank was derived from 16-tap FIR filters. The coder delay is a function of the number of decomposition stages, and the order of FIR filters In particular, the overall coder delay Δ is given by:

$$\Delta = N(2^S - 1) \qquad \text{samples} \tag{1}$$

where N is the filter order, and S is the number of WP decomposition stages. With 16-tap FIR filters and 8-stage decomposition filterbank the WP audio coder has a coding delay of just under 100 ms at 44.1 kHz sampling frequency.

2.2. Filter Implementation

For efficient implementation of the filterbank we employ *lattice* filters. Several researchers have developed techniques for efficient implementation of lattice filters. However, the majority of these techniques reduce the computational effort by a factor of two relative to transversal filter implementation [8, 9, 10]. The lowpass and highpass filter pair used in each WP decomposition stage represents a paraunitary perfect reconstruction quadrature mirror filterbank (PR-QMF). By exploiting the symmetry of the paraunitary QMF we reduce the computational effort by 75% relative to a transversal filter. Consider a paraunitary PR-QMF where

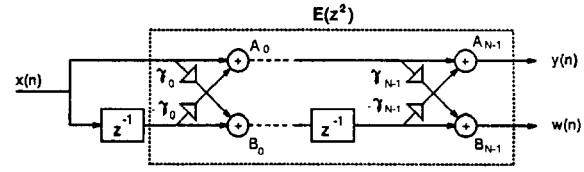

Figure 3: Lattice Filter

the impulse response sequence of the the N*th* order lowpass FIR filter is $\{b_0, \ldots, b_{N-1}\}$. Such an PR filterbank can be implemented as a cascade of N lattice sections as shown in Figure 3 [8]. The input-output relations of the individual lattice sections are recursively given:

$$A_{m-1}(z) = \frac{1}{1+\gamma_m^2}[A_m(z) + \gamma_m B_m(z)]; \tag{2}$$

$$B_{m-1}(z) = \frac{z}{1+\gamma_m^2}[-\gamma_m A_m(z) + B_m(z)]. \tag{3}$$

We solve for the lattice coefficients using equations (2), (3) and $b_{N-1} = \gamma_{N-1}$. Because the lowpass and highpass filters used in the WP decomposition constitute a PR pair the lattice coefficients exhibit the symmetry $\gamma_m = \gamma_{m-1}$, $m = 1, 3, \ldots, N-1$ where we assumed that N is an even integer. The lattice matrix E will then contain z^2 terms only. Therefore, we can interchange the order of filtering and decimation shown in Figure 4(a) [8]. Figure 4(b) depicts the resulting configuration. Table 1 lists the number of DSP operations[1] required by each filter.

[1] We define addition, multiplication, and multiply-accumulate operations as DSP operations which can be executed in identical instruction cycles on a digital signal processor

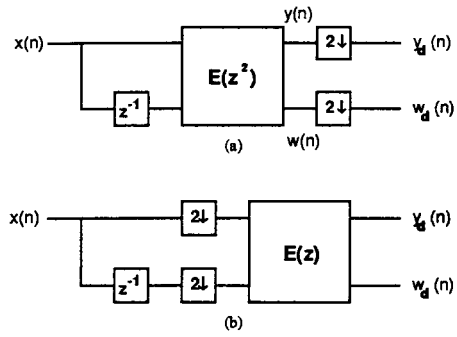

Figure 4: (a) Conventional Lattice, (b) WP Lattice.

Filter Type	DSP Operations
Transversal FIR	$2N$
PR Lattice	$N+3$
Reorganized PR Lattice	$N/2+2$

Table 1: Computational effort requirements.

3. PSYCHOACOUSTIC MODEL

The WP coder is based on the psychoacoustic model described in [1]. The use of a psychoacoustic model is necessary to identify the redundant audio information present in a given analysis block. The analysis provided by the model is then used in requantization of the subband signals.

3.1. Masking

Masking is the effect whereby a signal, the maskee, is rendered partially, or completely, inaudible by the presence of a nearby strong signal, the masker. Figure 5 depicts masking by a sinusoid S_2. The shaded region corresponds to the power levels of nearby signal components that will masked by the presence of S_2. In this particular case, the weaker signal S_3 is completely inaudible. The signal S_1 is only partially masked. The perceptible portion of the signal lies above the masking curve, allowing the coder to raise the noise floor in the subbands containing the masked signals, thereby requiring fewer bits to represent the corresponding subband signals. As the masking level is a function of the signal power, the psychoacoustic model first computes the signal power within each critical band.

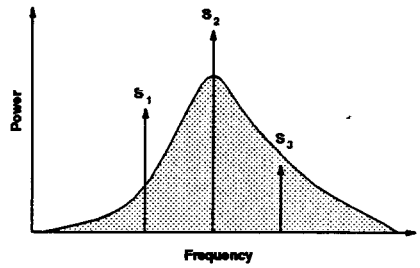

Figure 5: Masking from a Single Tone

Let z_i be the frequency on the bark scale and let $X(z_i)$ be the subband signal at z_i. The psychoacoustic model [1] uses the maximum self-masking level $v_s(z_i)$:

$$v_s(z_i) = -2.025 - 0.175 z_i \tag{4}$$

and the shape of the masking level $v_f(z_i, z_j)$ as a function of the frequency of separation[2] $\Delta z = z_i - z_j$:

$$v_f(z_i, z_j) = \begin{cases} 17\Delta z - 0.4X(z_i) + 11, & -3 \le \Delta z < -1; \\ (0.4X(z_i) + 6)\Delta z, & -1 \le \Delta z < 0; \\ -17\Delta z, & 0 \le \Delta z < 1; \\ -17\Delta z + 0.15(\Delta z - 1)X(z_i), & 1 \le \Delta z < 8. \end{cases} \tag{5}$$

Let $M(z_i, z_j)$ represent the masking threshold at frequency z_i from signal $X(z_j)$:

$$M(z_i, z_j) = X(z_j) + v_s(z_j) + v_f(z_i, z_j). \tag{6}$$

Masking is taken to be additive, so the total masking threshold $M_T(z_i)$ is given

$$M_T(z_i) = 10\log_{10}(10^{T_q(z_i)/10} + \sum_{j=1}^{28} 10^{M(z_i,z_j)/10}) \tag{7}$$

where $T_q(z_i)$ is the *threshold in quiet*. Table 2 lists the computational effort for the MPEG and WP model noise masking calculations expressed as the number of DSP operations required per sample. The term "L" is the number of iterations used to calculate the logarithm.

Operation	DSP Operations
MPEG 512 point FFT	86
MPEG Power Calculation	$10.6 + 8L$
MPEG Sound pressure level	$8 + L$
MPEG Noise power	$2.7 + 0.81L$
MPEG Total	$107.3 + 9.81L$
WP Sound pressure level	$11.7 + 1.3L$

Table 2: Masking threshold computation.

3.2. Joint Stereo Mode

The joint stereo mode (JSM) exploits the statistical dependencies between the left and right channel signals to compress the audio signal more than it is possible using monophonic signal processing only. The psychoacoustic model states that above 2 kHz the audio information is carried by the signal envelope and that we can eliminate the temporal fine structure of the audio signal without introducing noticeable distortion [1]. The ISO/MPEG standard employs JSM only in Layer-III and also leaves the use of the JSM as an option. Since the JSM is applicable only to the upper frequency subbands, the hierarchical structure of the WP coder allows us to combine the left and right channel signals

[2] v_f becomes negligible small for $\Delta z \notin [-3, 8)$.

before the decomposition thus reducing signal processing for the upper frequency bands by half.

The WP coder has four JSM modes. It allows monophonic substitution in zero to three regions corresponding to the frequency bands [2.7, 5.5], [5.5, 11], [11, 22] kHz. We decided against the use of JSM below 2.76 kHz to provide a transition band which minimizes signal distortion due to JSM. The WP coder benefits in two ways from the incorporation of JSM: (1) higher compression rate; and (2) reduction in computational effort. Figure 6 illustrates the

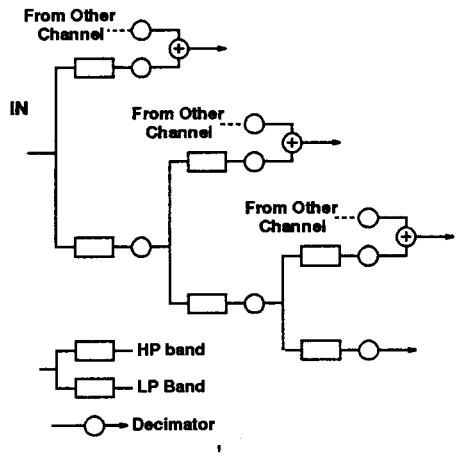

Figure 6: Modification to WT for JSM.

implementation of the JSM. We combine the left and right channel signals within the decomposition structure as requested by the JSM mode. For these signals we perform only a monophonic decomposition further into the decomposition hierarchy. Table 3 lists the four JSM modes and the computational efforts required by each. It should be noted that in the WP case the figures in Table 3 represent average computational efforts. The uneven sampling dyadic of the multirate system causes very uneven computational loading. The use of interrupt based input/output in the implementation of the WP algorithm on a digital signal processor handles the uneven computational load concern.

Let the index $k \in \mathcal{K}$ represent the subbands in JSM. The decoder replaces the monophonic subband signals M_k, $k \in \mathcal{K}$ with the estimated left and right channel signals:

$$\hat{R}_k = 0.5\, M_k\, P_r/(P_r + P_l), \quad k \in \mathcal{K}; \tag{8}$$

$$\hat{L}_k = 0.5\, M_k\, P_l/(P_r + P_l), \quad k \in \mathcal{K}; \tag{9}$$

where P_r and P_l represent respectively the right and left channel block peak values computed when we make the transition from stereophonic to monophonic signal processing within the decomposition tree.

4. RESULTS AND CONCLUSIONS

Table 3 compares the computational requirements of the MPEG and WP coders. We tested the data compression performance of the WP coders with eight sound files sampled in stereo format at 44.1 kHz. The test files were chosen

Compressor	Operations	Compression
MPEG Layer-I	$375 + 9.8L$	4.98:1
WT with 0 mono bands	$139 + 1.3L$	3.55:1
WT with 4 mono bands	$107 + 1.3L$	5.05:1
WT with 8 mono bands	$91 + 1.3L$	6.96:1
WT with 12 mono bands	$83 + 1.3L$	8.82:1

Table 3: Comparison of the two audio coders.

to be representative of a wide spectrum of audio signals. The length of each test file is 25 s. In this study, we developed a wideband stereophonic audio coder which required less than one third of the computational effort of MPEG Layer-I, psychoacoustic model I. We utilized psychoacoustic masking to minimize the output data rate and JSM to minimize computational effort. The WP based audio compressor provides transparent sound quality at compression rates comparable or superior to that achieved by the MPEG compressor. Distortion due to monophonic processing becomes just perceptible in some audio segments when aggressive JSM was used.

5. REFERENCES

[1] Second draft of proposed standard on information technology of moving pictures and associated audio for digital storage media up to about 1.5 Mb/s. *Doc. ISO/IEC JTC1/SC2/WG11 MPEG 90/001*, 1990.

[2] P. Noll, "Wideband speech and audio coding," *IEEE Commun. Magazine*, pp. 34-44, November 1993.

[3] K. Konstantinides, "Fast Subband Filtering in MPEG Audio Coding," *IEEE Signal Processing Letters*, vol. 1, no. 2, pp. 26-28, February 1994.

[4] P.P. Vaidyanathan, *Multirate systems and filterbanks*, Englewood Cliffs, Prentice-Hall, Inc., 1993.

[5] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intel.*, pp. 674-693, July 1989.

[6] X. Yang, K. Wang, and S. A. Shamma, "Auditory Representations of Acoustic Signals" *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 824-839, March 1992.

[7] D. Sinha and A.H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Tr. on Sig. Proc.*, vol. 41, no. 12, pp. 3463-3479, Dec. 1993.

[8] P.P. Vaidyanathan and P.Q. Hoang, "Lattice Structures for Optimal Design and Robust Implementation of Two-Channel Perfect-Reconstruction QMF Banks," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 81-94, Jan. 1988.

[9] Zhi-Jian Mou and P. Duhamel, "Short-Length FIR-Filters and Their Use in Fast Nonrecursive Filtering," *IEEE Trans. Signal Processing*, vol. 39, no. 6, pp. 1322-1332, June 1991.

[10] Z. Doganata and P. P. Vaidyanathan, "On One-Multiplier Implementations of FIR Lattice Structures," *IEEE Trans. Circuit Syst.*, vol. CAS-34, pp. 1608-1609, Dec. 1987.