Comparative analysis of audio coding using wavelet transform and periodized wavelet transform.

M. ROSA ZURERA, F. LÓPEZ FERRERAS, P. JARABO AMORES AND S. MALDONADO BASCÓN.

Dpto. de Teoría de la Señal y Comunicaciones.
Escuela Politécnica de la Universidad de Alcalá.
Ctra. Madrid-Barcelona, km. 33.600
28,871 Alcalá de Henares, Madrid
SPAIN

Abstract:-A comparative analysis of two audio coding methods based on the wavelet transform are presented. First of all, we present a CD monophonic audio coder using the discrete wavelet transform. Also, the algorithm to translate the psycho-acoustic information from the Fourier to the wavelet domain is described. Second, a different audio coder using the periodized wavelet transform is presented. These two schemes for audio coding are compared, focusing on the introduced delay, necessary binary rate to ensure transparent coding and computational complexity. The results show that the audio coder based on the periodized wavelet transform allows audio coding with very low delay, but using a high binary rate. Results using orthonormal wavelet with compact support are presented, and it is shown the trade off between the support of the wavelet function and the compression rate (increasing the support reduces slightly the binary rate, with no improvement in the perceptual quality of the decoded audio signal).

Key Words: Audio, Coding, Wavelet Transform, Periodized, Bit rate. CSCC'99 Proceedings, Pages:2091-2095

1 Introduction

Multimedia applications including audio signals are emerging nowadays. Broadband audio signals need a lot of storage capacity and high binary rates for transmission. As an example, CD monophonic audio signals are obtained by sampling at 44.1 kHz and 16 PCM coding. This motivates the research and development of audio coding.

Several coding strategies have been applied since seventies. Among them, transform coding and subband coding are the most important ones, and have been applied to the development of audio coding standards (MPEG, ATC, etc.).

Since 1993, after publication of the excellent paper by Sinha and Tewfik [3], wavelet transform has been applied to audio coding by several authors. This transform has some characteristics that make it suitable for audio coding purpose. Unfortunately, till now, there is not an easy algorithm to apply psycho-acoustic information to the wavelet domain. The algorithm proposed by Sinha and Tewfik needs long and selective filters to implement the wavelet transform. M. Rosa and al. [8] have proposed a new algorithm to do that. With this algorithm, transform coders using the wavelet transform can be developed using filters that generate wavelets with any compact support, opening new strategies for audio coding.

Another important question when using the wavelet transform for audio coding is the need for audio signal segmentation, in order to ensure the stationary nature of the processed audio segments. This paper analysis to strategies to implement an audio coder using the wavelet transform, including signal segmentation:

1. Overlap-save based algorithm: final conditions of filters in the filter bank after processing an audio segment are saved and applied as initial conditions for the next one.
2. Periodized wavelet-transform: each segment of audio signal is considered as a period of a periodic signal.

The audio coder squeme is described in section 2. For comparative analysis purpose, the same psycho-acoustic model and the same algorithm to translate the psycho-acoustic information to the wavelet domain are used for the two analysed strategies. They are summarised in sections 3 and 4.

2 Coder and decoder schemes.

The following points describe the coder structure:

- First of all, the input audio signal is segmented in frames of 2^N samples each one. The results presented in this paper have been obtained with frames of 1024 samples.
- After that, each frame is decomposed using a wavelet packet decomposition, represented in figure 1, that approximates the critical band decomposition of the human ear.

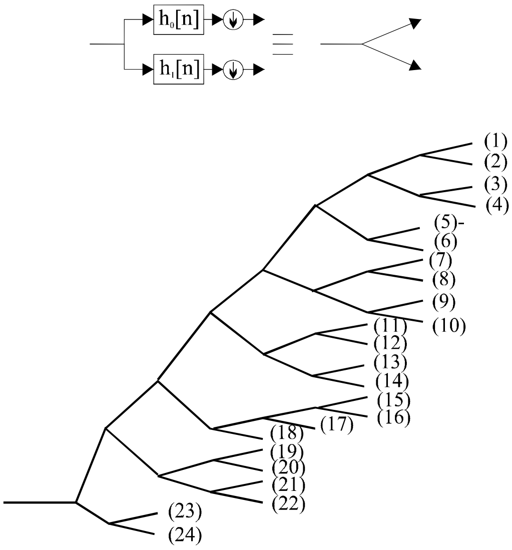

Fig. 1. Wavelet packet decomposition.

- The masking threshold is estimated in the frequency domain, using the algorithm described in section 3.
- Then, the masking threshold is translated to the wavelet domain using the algorithm described in section 4.
- After that, the wavelet coefficients of the audio frame are coded attending to the psycho-acoustic information.
- The coded wavelet coefficients and the side information of the coding process are multiplexed and sent to the decoder.

The following points describe the decoder structure:

- First of all, the coded wavelet coefficients and the side information are demultiplexed.
- Then, the coded wavelet coefficients are decoded and the inverse wavelet packet decomposition is calculated to obtain the reconstructed audio frame.
- Finally, adjacent frames are overlapped and added to reconstruct the audio signal, where the introduced distortion remains inaudible.

3 Masking threshold.

Masking threshold is estimated in the frequency domain using an algorithm that is based on Johnston's algorithm [5]. An improvement related to the estimation of the tonality coefficient has been included. This coefficient allows to characterise each critical band component as tonal like or noise like. To estimate the tonality coefficient we use the following properties:

1. Sinusoidal signals have a high time correlation that is also present in the time evolution of its short-time Fourier transform.
2. The unpredictability of noise signals in the time domain.

Attending to these properties, the tonality coefficient has been calculated from the prediction error of the signal short-time spectrum. Linear predictors have been used for the magnitude and phase spectra:

$$\left|\hat{X}_t(e^{j\omega})\right| = 2\left|X_{t-1}(e^{j\omega})\right| - \left|X_{t-2}(e^{j\omega})\right| \quad (1)$$

$$\hat{\varphi}_{X_t} = 2\varphi_{X_{t-1}} - \varphi_{X_{t-2}} \quad (2)$$

Each frequency bin of the short-time Fourier analysis of the audio signal is characterised by a coefficient that is a measure of the normalised prediction error:

$$\gamma(\omega) = 1 - \frac{\left\| X(e^{j\omega}) \right| e^{j\varphi(\omega)} - \left|\hat{X}(e^{j\omega})\right| e^{j\hat{\varphi}(\omega)}\right|}{\left|X(e^{j\omega})\right| + \left|\hat{X}(e^{j\omega})\right|} \quad (3)$$

Finally, the tonality coefficient $\alpha(i)$ of the critical band is defined as the arithmetic mean of the coefficients $\gamma(\omega)$ inside the band.

4 Translating psycho-acoustic information to the wavelet domain.

The algorithm to translate psycho-acoustic information to the wavelet domain is described in [8]. Here we summarise it. It is based on two suppositions:

1. Subband signals are orthogonal.
2. Quantization noise is modelled as white noise and is uncorrelated with subband signal.

Supposition 1 is accomplished when orthonormal filters are used. Supposition 2 is accomplished too when we use a high number of quantization levels. When very few quantization levels are used, the deviation from the theoretical model can be easily corrected.

Finally, the variance of quantization noise added to subband signal can be calculated as:

$$\sigma_i^2 = \underset{\omega \in B_i}{median} \left(\frac{T(e^{j\omega})}{\sum_j \left| F_j(e^{j\omega}) \right|^2} \right) \qquad (4)$$

$T(e^{j\omega})$ is the estimated masking threshold, and $F_j(e^{j\omega})$ is the frequency response of the equivalent filter of the reconstruction branch j-th of the filter bank.

5 Options for wavelet transform implementation.

We have two alternatives to implement the decomposition of signal segments:

1. The first one is an overlap-save strategy. After a segment has been decomposed, the final conditions of the filters in the filter bank are saved as initial conditions for processing the next segment. At the synthesis scheme, we proceed in a similar way.
2. The second one is based the periodization of the audio segment. These properties are used:
   - The result of filtering a periodic signal with a FIR filter is another periodic signal with the same period.
   - The decimation process of order 2 of a periodic signal of period $N=2^\gamma$ give rise to a periodic signal of period $N/2=2^{\gamma-1}$.

   As a consequence of these properties all the subband signals can be considered as segments of periodic signals.

The periodized transform allows us to characterise segments of a long signal, by sets of coefficients of the same length as the processed segments. The main question is whether or not this follows in an improvement of compression rate.

6 Phase distortion analysis and correction.

The options previously presented are based on Mallat's algorithm [6]. It uses as elemental processing block the two band filter bank represented in figure 2.

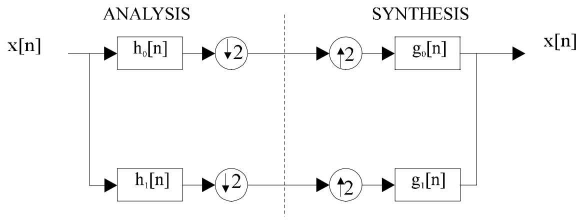

Fig. 2. Elemental processing block to implement the wavelet transform.

To be a perfect reconstruction system, it must be fulfilled:

$$\begin{pmatrix} G_0(z) \\ G_1(z) \end{pmatrix} = \frac{2}{\det(\mathbf{H_m})} \begin{pmatrix} H_1(-z) \\ -H_0(-z) \end{pmatrix}$$

$$\text{where:} \qquad (5)$$

$$\mathbf{H_m} = \begin{pmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{pmatrix}$$

If $h_0[n]$ is the impulse response of a FIR filter extended from n=0 to n=L-1, and the impulse response $h_1[n]=(-1)^n h_0[L-1-n]$, then $g_0[n]$ and $g_1[n]$ are also the impulse responses of FIR filters and $\det(H_m)$ is equivalent to a delay. When orthogonal wavelets are used, the reconstruction stage filters have the following impulse responses:

$$g_0[n] = h_0[-n] \qquad (6)$$

$$g_1[n] = (-1)^{n+1} h_0[n+L-1] \qquad (7)$$

So, if the analysis filters are causal, the synthesis filters are not. We need to use causal filters, so the impulse responses of the synthesis filters are delayed to begin at n=0. As a consequence, the reconstructed signal will be a delayed version of the original one (the delay is L-1 samples).

When the analysis scheme is iterated in both the low and high pass bands, as in the Mallat algorithm [6], a different delay will be applied to the different subbands, resulting in phase distortion if no compensation is introduced. It can be shown that the delay to introduced in a subband signal is $(2^m-1)(L-1)$, where m is the difference of filtering stages between the subband signal and that where most stages are used.

To avoid phase distortion we need to compensate the different delays of the periodic subband signals before applying them to the synthesis stage. When the periodized wavelet transform is used, these delays are implemented as cyclic shifts using *mod N* arithmetic. At the output of the synthesis stage it is obtained the original segment, but cyclic shifted. This shift can be easily corrected.

7 Results.

The proposed options to implement the wavelet packet decomposition in the audio coder have the following properties:

1. When the periodized decomposition is used, each long N audio segment is characterised by N coefficients, and there is no memory of previously processed segments.
2. When the overlap-save strategy is used, the coefficients obtained after processing each audio segment have information of previously processed segments, due to the information contained in the filters initial conditions.
3. The periodization procedure results in an increase of subband signal energy in high frequency bands. So, quantization noise power introduced in these bands is higher than the equivalent one when no periodization is used. The correction of this fact results in an increase of binary rate to achieve transparent coding.
4. The problem described in step 3 can be reduced if adjacent audio frames are overlapped and windowed, to get soft transitions between frames.

The following tables represent the results on audio coding using the two strategies. Four music signals have been used: vocal segment, piano solo, cord instruments and wind instruments, all of then of about 10 seconds duration. The subjective test has been accomplished in an informal way with 20 persons of our research group. Of course, the subjective results are not exhaustive.

It can be observed that the subjective quality of the audio coder, when the periodized wavelet packet decomposition with no overlapping between adjacent frames is used, is very bad. An important improve is obtained when overlapping is included.

All the results have been obtained with minimum phase Duabechies orthonormal filters with 8 vanishing moments and 1024 samples segments.

The quality score of the decoded signal is related with the distortion level, codified as: 5-Imperceptible, 4-Slightly perceptible, 3-Slightly disagreeable, 2- Disagreeable, 1- Very disagreeable.

Quality	**Vocal**	**Piano**	**Wind**	**Cord**
1	90	85	90	95
2	10	15	10	5
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0

Table 1. Quality results for the overlap-save strategy.

Quality	**Vocal**	**Piano**	**Wind**	**Cord**
1	55	90	25	25
2	40	20	15	15
3	5	0	35	45
4	0	0	10	10
5	0	0	15	15

Table 2. Quality results for the periodized wavelet transform without overlapping between frames.

Quality	**Vocal**	**Piano**	**Wind**	**Cord**
1	85	80	85	90
2	15	20	15	10
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0

Table 3. Quality results for the periodized wavelet transform with 1/32 of segment length overlapping.

By the other hand, the following tables show results of binary rates to achieve near transparent coding using Daubechies minimum phase filters that generate wavelets with 4, 8, 12 and 16 vanishing moments.

		Vanishing moments			
CODE		**8**	**16**	**24**	**32**
Vocal	SegSNR(dB)	25.2	25.3	25.4	25.1
	bits/sample	1.905	1.78	1.72	1.71
Piano	SegSNR(dB)	24.3	24.1	23.9	24.0
	bits/sample	2.05	1.92	1.87	1.82
Wind	SegSNR(dB)	25.5	25.3	25.4	25.6
	bits/sample	1.9	1.8	1.75	1.72
Cord	SegSNR(dB)	24.8	25.1	24.9	24.9
	bits/sample	1.93	1.75	1.67	1.63

Table 4. Segmental SNR (dB) and mean number of bits for transparent coding (periodized decomposition with overlapping).

		Vanishing moments			
CODE		**8**	**16**	**24**	**32**
Vocal	SegSNR(dB)	25.4	25.6	25.5	25.3
	bits/sample	1.8	1.67	1.6	1.58
Piano	SegSNR(dB)	24.8	24.6	24.7	24.5
	bits/sample	1.95	1.87	1.81	1.78
Wind	SegSNR(dB)	25.6	25.7	25.7	25.6
	bits/sample	1.82	1.72	1.65	1.62
Cord	SegSNR(dB)	25.3	25.2	25.4	25.3
	bits/sample	1.85	1.73	1.68	1.65

Table 5. Segmental SNR (dB) and mean number of bits for transparent coding (overlap-save strategy).

It can be observed, that the best compression rates are obtained with the overlap-save strategy, and that there is a trade-off between compression rate and coder delay. Another important result is the asymptotic behaviour with the number of vanishing moments of the wavelet generated with Daubechies filters.

8 Conclusions

It has been presented a comparative analysis of two different strategies to implement the subband decomposition of audio signals in a wavelet transform based audio coder. Two important results have been pointed:

1. The trade-off between the decoded signal quality and the introduced delay.
2. The trade-off between the number of vanishing moments and the binary rate.

The bottom line is that near transparent coding with less than two bits sample can be achieve with this coder, using the periodized wavelet transform or the overlap-save strategies.

References:

[1] I. Daubechies, Orthonormal Bases of Compactly Supported Wavelets. *Communications on Pure and Applied Mathematics*, Vol. XLI 909-996 (1988).

[2] M. Vetterli, Wavelet and Filter Banks: Theory and Design". *IEEE Trans. on Signal Processing,* September 1992.

[3] D. Sinha and A. Tewfik, Low Bit Audio Transparent Compression using Adapted Wavelets. *IEEE Trans. on Signal Processing*, December 1993.

[4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Kluwer Academic Publishers.

[5] J. D. Johnston, Transform Coding of Audio Signals Using Perceptual Noise Criteria. *IEEE Journal on selected areas in communications,* vol. 6. No 2, February 1988.

[6] S. G. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. II, No.7.

[7] P.E. Kudumakis and M.B. Sandler. On the Performance of Wavelets for Low Bit Rate Coding of Audio Signals. *ICASSP*, 1995.

[8] M. Rosa Zurera, F. López Ferreras, P. Jarabo and S. Maldonado Bascón, New Method to Translate the Psycho-acoustic Information to the Wavelet Domain in a Wavelet Transform Based Audio Coder. *ECMCS'99.*