ON THE CODING GAIN OF DYNAMIC HUFFMAN CODING APPLIED TO A WAVELET-BASED PERCEPTUAL AUDIO CODER

N. Ruiz Reyes[1], M. Rosa Zurera[2], F. López Ferreras[2], P. Jarabo Amores[2], and P. Vera Candeas[1]

[1]Departamento de Electrónica, Universidad de Jaén, Escuela Universitaria Politénica,
23700 Linares, Jaén, SPAIN, e-mail: nicolas@ujaen.es
[2]Dpto. de Teoría de la Señal y Comunicaciones, Universidad de Alcalá, Escuela Politécnica
28871 Alcalá de Henares, Madrid, SPAIN, e-mail: manuel.rosa@uah.es

ABSTRACT

This paper evaluates the coding gain of using a dynamic Huffman entropy coder in an audio coder that uses a wavelet-packet decomposition that is close to the subband decomposition made by the human ear. The subband audio signals are modeled as samples of a stationary random process with laplacian probability density function because experimental results indicate that the highest coding efficiency is obtained in that case. We have also studied how the entropy coding gain varies with the band index. The proposed adaptive Huffman coding method gives rise to an average coding gain of approximately 0.25 bits per sample compared to binary coding. A further coding gain can be achieved if time-varying filter banks are used. Experimental results tell us that using a suitable method to translate the psychoacoustic information to the wavelet domain, combined with our adaptive Huffman coding scheme, binary rates of about 64 kbps can be obtained for transparent coding of CD quality monophonic audio signals.

1. INTRODUCTION

Coding of CD quality audio signals has become a key technology in the development of current audio systems. CD quality monophonic audio signals are obtained with sampling frequencies of 44.1 kHz and 16 bits PCM coding. So, it is necessary a binary rate of 705.6 kbps for transmission, justifying the research and development of efficient audio coding techniques in order to reduce this high transmission rate. In many applications, such as high quality audio transmission and storage, the goal is to achieve transparent coding of CD quality audio signals at the lowest possible bit rates.

Most audio coding algorithms are based on: 1) removal of statistical redundancies in the audio signal, and 2) masking properties of the human auditory system to "hide" distortions. Traditional subband and transform coding techniques provide a convenient framework for coding based on both principles. Several of these techniques have contributed to the development of the ISO/MPEG audio coding standards.

The first one, called ISO/MPEG-1 [1], supports sampling rates of 32, 44.1 and 48 kHz, and several operation modes with bit rates ranging from 32 to 448 kbps.

The last one, the ISO/MPEG-4 standard, is composed of several speech and audio coders supporting bit rates from 2 to 64 kbps per channel. ISO/MPEG-4 includes the already proposed AAC standard, which provides high quality audio coding at bit rates of 64 kbps per channel.

Parallel to the definition of the ISO/MPEG standards, several audio coding algorithms have been proposed that use the wavelet transform as the tool to decompose the audio signal. The most promising results correspond to adapted wavelet-based audio coders. Probably, the most cited is the one designed by Sinha and Tewfik [2], a high complexity audio coder that provides low bit rate and high quality audio coding by searching a nearly optimum prototype filter for each audio frame. Others adaptive wavelet-based audio coders achieve similar results with lower complexity by searching the best wavelet packet structure attending to perceptual criteria [3].

To achieve bit rates around 64 kbps, it is necessary to incorporate in our scheme an entropy coding stage. Huffman coding involves the definition of mapping tables attending to the probabilities associated to each input symbol. Here, we propose an adaptive method based on the assumption that subband audio signals can be modeled as samples of a stationary random process with Laplacian probability density function.

2. AUDIO CODER STRUCTURE

In this section, the audio coder structure is described. It works with monophonic audio signals sampled at 44.1 kHz, but can be easily extended for multi-channel audio signals. Each input sample is PCM coded with 16 bits.

The encoder structure is summarized in figure 1.

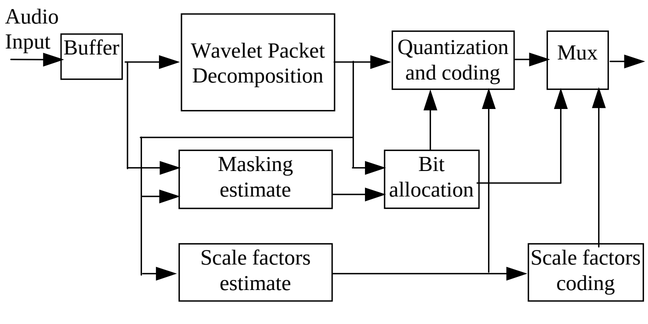

Figure 1. Encoder structure

The main features of the audio coding scheme presented in this paper are:

1. The input signal is analyzed with a filter bank which implements a wavelet packet decomposition adapted to the critical band analysis at the inner ear. It is represented in figure 2.

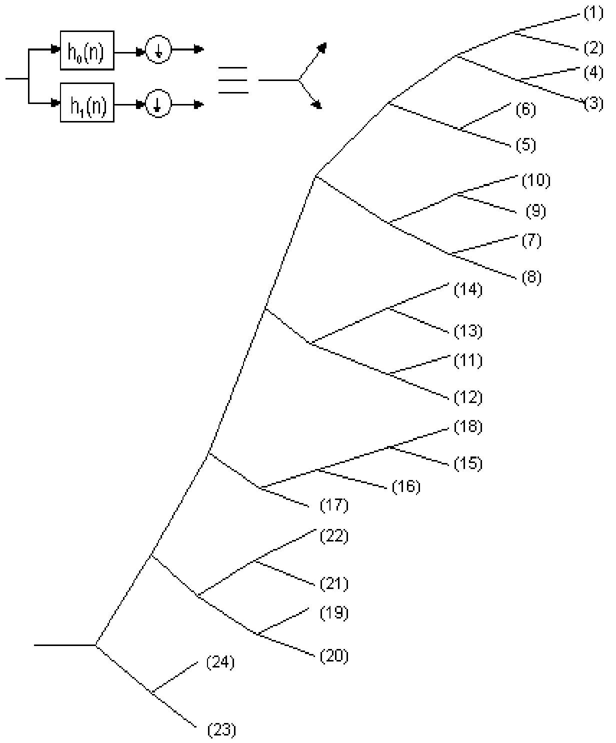

Figure 2. Filter bank for critical band decomposition

2. Due to the fact that audio signals are not stationary, input signals are segmented in frames that can be modeled as windowed samples of a random process. Each frame is 1024 samples sized, that is a suitable duration in order to consider it as a stationary signal in most of the cases.
3. To avoid that the number of wavelet coefficients that characterize an audio frame is higher than the number of samples in the time domain, each frame is interpreted as a periodical signal.
4. To avoid sharp changes in the introduced quantization noise power in the decoded audio signal, adjacent frames overlap 1/64 of their length, and the overlapping samples of each frame are windowed with the square root of a raised cosine function.
5. Parallel to the decomposition of the input signal, a masking threshold in the frequency domain is estimated for each audio frame. Here, we have used the ISO/MPEG-1 psycho-acoustic model 2. This masking threshold is not suitable to be applied directly in the wavelet domain, mainly when short filters are used to implement the wavelet transform. A novel algorithm to translate the psycho-acoustic information from the Fourier to the wavelet domain is included in our coding scheme [4].
6. A set of 15 uniform quantizers is used. The size of the quantization step for a given band is adapted to the scale factor of that band. The scale factors, which are sent to the receiver as side information together with the bit allocation information, are coded using 8 bits log PCM.
7. The quantized wavelet coefficients are entropy coded by the adaptive method proposed in next section and multiplexed with the side information to compose the bit stream transmitted to the decoder.
8. At the decoder, the coded wavelet coefficients and the side information are demultiplexed. Then, coded wavelet coefficients are decoded using the side information and applied to the inverse wavelet decomposition, which reconstructs the audio frame.
9. Each the overlapping samples of the reconstructed frames are windowed again with the squared root of a raised cosine function, and so, after the overlap-add process, the perfect reconstruction property is preserved.

The decoder structure is represented in figure 3.

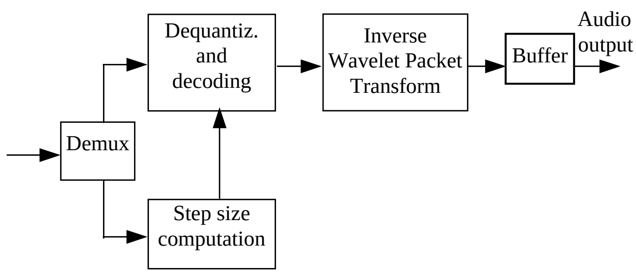

Figure 3. Decoder structure

3. ADAPTIVE HUFFMAN CODING

This section is devoted to a detailed description of our adaptive Huffman coding method. In general, a subband audio signal can be viewed as a zero-mean random signal with a probability density function (p.d.f.) that can be modeled with a generalized gaussian function [5]:

$$f_X(x) = K\, e^{-\left(|x|/\alpha\right)^{\beta}} \tag{1}$$

For β = 1.0 and β = 2.0 we have the special cases of laplacian and gaussian distributions, respectively. We treat to test the hypothesis that subband signals resulting from the wavelet packet decomposition of audio signals can be almost always modeled with laplacian probability density functions. In [5] β has been estimated using the Kolmogorov-Smirnov test and it varies in the range of 0.8 to 1.2 in most cases. Another method to validate the above mentioned hypothesis is checking whether coding gain is obtained using entropy coders adapted to laplacian p.d.f.'s and compare the results with those obtained for other p.d.f.'s.

Experimental results come to confirm our hypothesis, as it will be shown in next section. In figure 4 we show an example of wavelet coefficients histogram for a given subband audio signal together with its modeling using laplacian and generalized gaussian distributions.

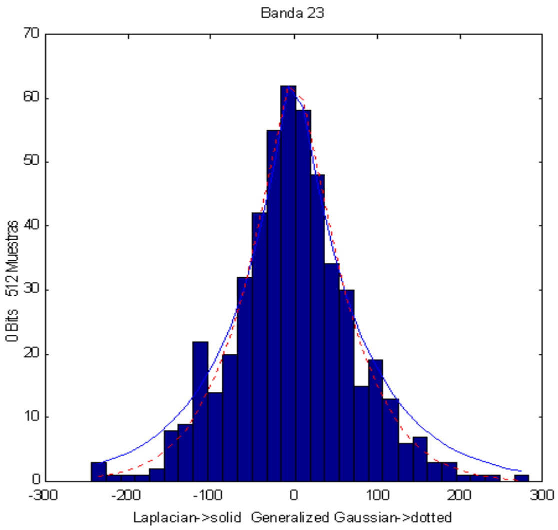

Figure 4. Example of wavelet coefficients histogram and its modeling using laplacian and generalized gaussian p.d.f.'s (α = 0.0745 and β =1.2953).

Once a probability density function is assumed for wavelet coefficients, the next stage in our dynamic Huffman coding technique involves computing the probability for each quantization level at each subband audio signal. Assuming laplacian p.d.f. for subband *k-th* with quantization step size q_k and laplacian parameter α_k, the probability of the quatization level *j-th* (mid-tread quantizer) is defined by:

$$p_j = \frac{1}{2} e^{-\left(\frac{j q_k}{\alpha_k}\right)} \sinh\left(\frac{q_k}{\alpha_k}\right) \qquad j \neq 0$$
$$p_0 = 1 - e^{-\left(\frac{q_k}{2\alpha_k}\right)} \tag{2}$$

The binary rate resulting from uniform quantization followed by Huffman coding can be approximated by the following function [6]:

$$R = \sum_{k=1}^{M} \frac{1}{D_k} \Phi\left(\frac{q_k}{2a_k}\right) \tag{3}$$

Where D_k is the sample rate of subband k and Φ can be explicitly computed as

$$\Phi(u) = -(1-e^{-2u})\log_2(1-e^{-2u}) - \\ - e^{-u}\log_2(\sinh(u)) + \frac{u}{(\log 2)\sinh(u)} \tag{4}$$

The code words for each subband audio signal are obtained from the above probabilities using the Huffman coding algorithm given in [7]. The bit rate estimation is based on the statistical model of the subband signals.

4. EXPERIMENTAL RESULTS

To check the performance of the proposed audio coder, we have obtained some subjective and objective results. Five music samples considered hard to encode have been used, and we have made sure that the set covers a wide variety of signals.

4.1. Objective results

Table 1 shows the entropy coding gain that our method gives rise when using minimum phase Daubechies filters with 32 coefficients rise for both laplacian and generalized gaussian distributions. It also shows the final bit rates achieved for transparent coding.

Test signal	Case I: Laplacian model	Case II: Gener. Gaussian model
Drums	1,58 / 0,23	1,75 / 0,06
Guitar	1,48 / 0,23	1,70 / 0,01
Piano	1,42 / 0,18	1,65 / -0,05
Saxo	1,44 / 0,15	1,68 / - 0,09
Pop	1,49 / 0,25	1,76 / -0.02

Table 1. Objective results: Final bit rate / Huffman coding gain both in bits per sample.

From table 1, it can be remarked the following results:

1. The laplacian distribution is closer to the actual wavelets coefficients distribution that the generalized gaussian one.
2. The coding gain using adaptive Huffman coding for laplacian modeling is around 0.25 bits per sample.
3. The bit rates are near to 64 kbps for most of the test signals.

In figure 5 we show how coding gain is changed with the band index. Several important results can be extracted from that figure:

1. Coding gain is increased with the band index, which indicates that laplacian distribution only match properly for high frequency bands, where few filtering stages are used.
2. No improvement is obtained from the first to the tenth subbands, where the coding gain is compensated with the side information necessary at the decoder (α_k), or even coding losses are obtained.
3. Subband 23-th is a special case. A look at figure 2 indicates that this subband contains the highest frequency components, corresponding to those whose power is almost always below the auditory threshold. So, the coder doesn't assign any bits to that subband even when entropy coding is not used.

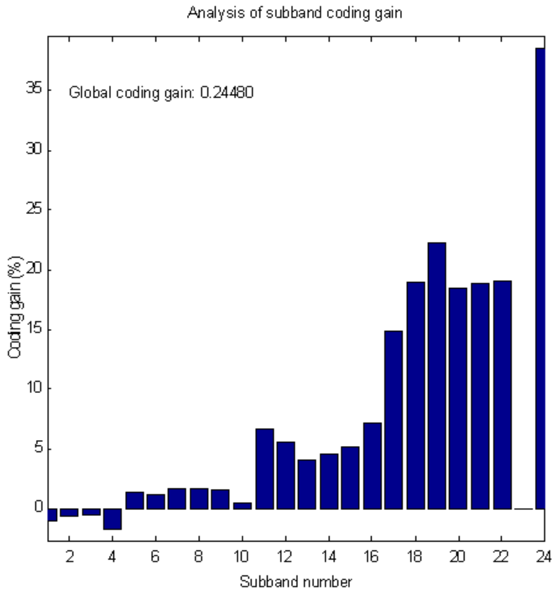

Figure 5. Variation of the coding gain with the band index

The results suggest that a good coding strategy would be the following: binary coding for low frequency subbands and adaptive Huffman coding for the rest.

4.2. Subjective results

We have performed a test for transparency ("double blind test") at a binary rate of approximately 64 kbps with 20 people selected from our research group, all of them aged from 24 to 35 years. The results when using the above mentioned filters are presented in table 2.

Music Sample	Average probability of original music preferred over encoded one	Comments
Drums	0.56	Transparent
Guitar	0.48	Transparent
Piano	0.47	Transparent
Saxophone	0.54	Transparent
Pop	0.51	Transparent

Table 2. Subjective test results: transparency test

The quality is cataloged as 'transparent' because the average probability is around 0.5 for all the test signals. The results confirm that our coder can be considered as transparent at a binary rate of 64 kbps.

Besides, our coder was mostly preferred if it was compared to MPEG-1 layer-3 coder because something like a filtering was found in signals coded with MPEG-1 layer-2 coder. However, listeners found the quality of signals coded with our coder similar to those coded with MPEG-1 layer-3 coder.

5. SUMMARY

We have just presented an adaptive Huffman coding technique included in a wavelet-based perceptual audio coder. Our studies point to laplacian distribution as the best choice for high frequency subband audio signals modeling and indicate that coding gains around 0.25 bits per sample are feasible. A hybrid coding strategy is proposed due to the lack of accuracy of this modeling for low frequency subbands.

Two promising approaches for further bit rate reduction are: 1) vector quantization, 2) model based coding. We are now focused on these two issues. An interesting way to reduce the computational complexity would be the development of a masking model directly in the wavelet domain. This will be another research issue to work on.

Also, it would be interesting to check the performance of our scheme with other entropy coding methods (i.e arithmetic coding).

Other issues to work on are scalable wavelet-based coding, evaluation of different psycho-acoustic models and more extensive subjective quality evaluation.

6. REFERENCES

[1] ISO/IEC 11172-3, "Information technology - Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s" - (Part 3, Audio), 1992.

[2] D. Sinha, A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets", *IEEE Trans. on Signal Processing*, Vol. 41, No. 12, pp. 3463-3479, 1993.

[3] P. Srinivasan, L. H. Jamieson, "High-quality compression using an adaptive wavelet packet decomposition and psychoacoustic modeling", *IEEE Trans. on Signal Processing*, Vol. 46, No. 4, pp. 1085-1093, April 1998.

[4] M. Rosa, F. López, P. Jarabo and S. Maldonado, "New method to translate the psycho-acoustic information to the wavelet domain", *EURASIP Conf. DSP for Multimedia Commun. and Services*, Krakow, Jun 1999.

[5] A. Papoulis, *Probability, random variables, and stochastic processes*, 3/e, McGraw-Hill, 1991.

[6] P. Philippe, F. M. de Saint-Martin and M. Lever, "Wavelet packet filter banks for low time delay audio coding", *IEEE Trans. on Speech and Audio Processing*, Vol. 7, No. 3, pp. 310-322, May 1999.

[7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes in C*, 2/e, Cambridge University Press, 1996.