ON THE PERFORMANCE OF WAVELETS FOR LOW BIT RATE CODING OF AUDIO SIGNALS

P.E. Kudumakis and M.B. Sandler

Signals Circuits and Systems Research Group
Department of Electronic and Electrical Engineering
King's College London, Strand, London WC2R 2LS
England
tel: ++44 (0)71-873 2041
fax: ++44 (0)71-836 4781
e-mail: panos@orion.eee.kcl.ac.uk

ABSTRACT

The performance of some different wavelet families, including for comparison a well known family of QMFs, is investigated for low bit rate coding of audio signals. For the assessment of the coding gain of these wavelets, both octave and uniform subband coding schemes have been evaluated, using both constant and dynamic bit allocation, with and without entropy noiseless Huffman coding. The influence of complexity of these wavelets, in terms of number of filter coefficients, against the quality of the decompressed audio signals in terms of Segmental-SNR (dB), is presented, at different bit rates. In addition, this evaluation suggests that perceptually transparent quality of monophonic signals can be achieved at 24 kbits/sec (Fs= 8kHz, 3 bits/sample) for speech applications and at 64 kbits/sec (Fs= 48kHz, 1.33 bits/sample) for music related applications, as in digital audio transmission and storage.

1. Introduction

Traditional subband and transform coding techniques, indicate that perceptually almost transparent coding of monophonic compact disk (CD) quality signals (sampled at 44.1 kHz) can be achieved approximately at bit rates of 96 kb/s. Several of these techniques have contributed to the development of the ISO-MPEG [1] audio coding standard. More recent developments include, the adaptive wavelet selection method combined with dynamic dictionary coding [2] and the pitch-synchronous wavelet transform [3], which claim to achieve similar quality at bit rates 64 kb/s with Fs= 44.1 kHz and 21 kb/s with Fs= 8 kHz, respectively. The disadvantage of these two methods is the long coding delay. This factor is important especially for real time practical applications. In comparison to the above techniques, our approach, based on a tree-structure filterbank subband coding combined with entropy noiseless Huffman coding, claims perceptually transparent quality at similar bit rates but with shorter delay.

The approach of the paper is firstly to review the wavelet filters used in the assessment. This is followed by a description of the subband coding schemes that have been evaluated for the assessment of the coding gain of these wavelets and lastly, the performance of these wavelets under these coding schemes is discussed.

2. Wavelets Used in the Assessment

2.1. Orthogonal Wavelets

- **DAUB-A** The most popular and frequently used orthogonal wavelets are the original Daubechies wavelets [4], [5]. They are a family of orthogonal wavelets indexed by $N \in \mathsf{N}$, where N is the number of vanishing wavelet moments. They are supported on an interval of length $2N-1$. A disadvantage is that, except for the Haar wavelet ($N=1$), they cannot be symmetric or antisymmetric. Their regularity increases linearly with N and is approximately equal to $0.2075N$ for large N. This construction does not lead to a unique solution if N and the support length are fixed. In fact, this family corresponds to choosing the extremal phase. These compactly supported wavelets with extremal phase and highest number of vanishing moments, compatible with their support width, are the most asymmetric. They are also known as "minimum phase wavelet filters".

- **DAUB-B** Another family is constructed in [4] by choosing, for each N, the solution closest to linear phase (or closest to symmetry). This leads to

0-7803-2431-5/95 $4.00

compactly supported wavelets with the maximum number of vanishing moments but "less asymmetric" compared with the "minimum phase wavelet filters".

- **COIF-A** In [4] the construction of orthonormal wavelet bases is suggested, with vanishing moments not only for the wavelet but also for the scaling function. Their construction was suggested by R. Coifman, and I. Daubechies therefore named them "coiflets". These coiflets are much more symmetric than the previous families but there is a price to pay for this. They have support width $3N-1$, as compared to $2N-1$ for the previous families.

2.2. Biorthogonal Wavelets

It is well known for subband filtering that symmetry and exact reconstruction are incompatible, if the same FIR filters are used for reconstruction and decomposition. As soon as this last requirement is given up, symmetry is possible.

- **COIF-B** These are symmetric biorthogonal bases close to non-symmetric coiflets. In fact, both the analysis and synthesis filters have similar coefficient values to each other and to the corresponding orthonormal coiflet. They have been constructed, using the Laplacian pyramid filter as either analysis or synthesis filter [4].

- **SYMM-A,B,C** The biorthogonal symmetric bases that have been used here can be found in [6].

2.3. Johnston-QMF

The QMF filters defined by Johnston [7] have also been used in the comparison although they are not perfect reconstruction filter banks. This is done because they approximate ideal filters reasonably well, and thus they are good approximations for orthogonal wavelets, with the advantage of having linear phase and being standard filter banks.

3. Experiments and Results

In coding of non-stationary signals, such as sharp attacks, it is useful for the filterbank to use frequency subdivision schemes that approach the critical bands [8] of the human auditory mechanism. However, for the coding of stationary signals the approach is to use a full depth tree decomposition to maximize the coding gain, even if the decomposition does not mimic the human filter. The Wavelet Transform (WT) or octave band decomposition and the Wavelet Packet (WP) or uniform frequency subdivision scheme fit well with these requirements.

Coding scheme	Complexity Mult/tions	Max. No Subbands	Subbands used here
log	2L	log2(L)	10
uniform	Llog2(L)	L/2	32

Table 1: Characteristics of the two different frequency subdivision schemes.

Thus in this section, Wavelet Transforms are compared against Wavelet Packets. The transform frame length is equal to $L = 1024$ samples. Thus, the number of octaves or the number of scalefactors that has to be transmitted to the receiver in the octave band decomposition case is equal to $log_2 L = 10$, while for the uniform case the number of subbands or the number of scalefactors is kept equal to 32, as in [1]. Table 1 summarises the characteristics of these two different approaches. It has also to be noticed that the complexity of the WT is independant of the number of octaves. This is the reason that the WT scheme uses 10 stages of decomposition while the WP uses only 5 stages.

Throughout the simulations that are presented here, the same music signal [9] of duration 8.192 secs has been processed. This has been captured at Fs= 48 kHz, 16 bits/sample PCM and at Fs= 8 kHz, 13 bits/sample PCM.

Figure 1 shows the behaviour of the minimum phase (DAUB-A) family of wavelet filters versus the quality of the decompressed audio signal, in terms of Segmental-SNR (dB), using:

- The WT and WP representation, combined with Constant Bit Allocation (CBA).

- The WP representation, combined with Dynamic Bit Allocation (DBA), based on Psychoacoustic Model-1 as adapted for use with MPEG Layer-2, [1]. However, in our experiments we have not included the tonality in order to keep the overall complexity of the codec low. Another reason is that the Segmental-SNR is more valid to compare these methods since no masking has been taken into consideration. However, the tonality would improve the compression further if included.

- The WP representation, combined with DBA as above and lossless entropy Huffman coding, as in MPEG Layer-3, [1].

Although the octave band decomposition scheme uses the maximum number of subbands compatible with the frame length, its performance is poorer in comparison to 32-uniform band decomposition, when both use CBA. However, the number of subbands is a compromise between coding gain and practical considerations such as complexity and processing delay.

Efficient signal compression results when subband signals are quantized with subband-specific bit allocation, based on input power spectrum and the model of perception. This is shown in figure 1 for the 32-uniform band decomposition scheme using DBA. In this case is also clear that as the number of the wavelet filter coefficients increases the quality of the decompressed signal also increases.

While the Dynamic Bit Allocation strategy exploits some of the human hearing characteristics, further reduction in the bit rate requires getting rid of the statistical redundancies of the signal. That is the ideal case for entropy noiseless Huffman coding. Figure 1 reveals a $7dB$ gain when Huffman coding is used over the 32-uniform band decomposition scheme with DBA.

Figure 2 shows the performance associated with the number of coefficients, of the different families of orthogonal and biorthogonal wavelet filters, which were mentioned in Section 2. For the sake of comparison the QMF-filters are also incuded. A 32-uniform band decomposition scheme with DBA combined with Huffman coding has been used. As the number of filter coefficients increases the performance increases for both orthogonal and biorthogonal wavelets.

However, the relationship between orthogonal and biorthogonal wavelets is not so clear. It seems that the orthogonal families of wavelets give better results. However, this is not the rule, since a biorthogonal family and the QMFs are clearly superior in terms of Segmental-SNR, although the difference is less than $1dB$. Theoretical results concerning the number of vanishing moments and regularity can be found in [10].

Figures 3 and 4 reveal the performance of the minimum phase (DAUB-A) family of wavelets with 4 and 20 coefficients, with and without Huffman coding when the audio signal has been recorded at 8 and 48 kHz, respectively. These show that Huffman coding is most efficient for low bit rates. Thus, Huffman coding is an inseparable part of any low bit rate codec.

4. Conclusions

It has been shown that the DBA gives rise to the better performance of longer wavelets in terms of Segmental-SNR. It is also shown that some of the biorthogonal wavelets and the Johnston's QMF have better performance than the orthogonal wavelet families. However the difference between them is quite small. Finally, our combination of 32-uniform band decomposition with DBA and Huffman coding results in perceptually transparent quality at 24 kbits/sec (Fs= 8kHz, 3 bits/sample) for speech applications with Segmental-SNR $22.87dB$ and at 64 kbits/sec (Fs= 48kHz, 1.33 bits/sample) for music related applications with Segmental-SNR $25.48dB$, for the music signal in [9].

Acknowledgement: The authors wish to thank Audio Engineering Society Educational Foundation, for supporting this work.

5. References

[1] 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), August 1993.

[2] D. Sinha and A. H. Tewfik, *"Low Bit Rate Transparent Audio Compression Using Adapted Wavelets"*, IEEE Transactions on Signal Processing, Special Issue on Wavelets and Signal Processing, Vol. 41, No. 12, pp. 3463-3479, Dec. 1993.

[3] Gianpaolo Evangelista, *"Pitch Synchronous Wavelet Representations of Speech and Music Signals"*, IEEE Transactions on Signal Processing, Special Issue on Wavelets and Signal Processing, Vol. 41, No. 12, pp. 3313-3330, Dec. 1993.

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

[5] I. Daubechies, *"Orthonormal bases of compactly supported wavelets"*, Comm. Pure Appl. Math., 41 (1988), pp.909-996.

[6] Patrik Andersson, *"Wavelet Transforms and Image Compression"*, MSc Thesis, Chalmers University of Technology, Goteborg, Sweden, 1992.

[7] J. D. Johnston, *"A filter family designed for use in quadrature mirror filter banks"*, in Proceedings of the ICASSP '80 Conference, pp. 291-294, April 1980.

[8] Nikil Jayant, James Johnston and Robert Safranek, *"Signal Compression Based on Models of Human Perception"*, Proceedings of the IEEE, Vol. 81, NO. 10, pp. 1385-1422, Oct. 1993.

[9] Eveleigh and Wilkinson, *"GUD Toons (Good Tunes)"* in Kunta Kinti, The Future Music CD Vol. 4, demo no. 6, in Future Music Magazine No. 14, Dec. 1993.

[10] E. A. B. da Silva, M. Ghanbari, *"On the Coding Gain of Wavelet Transforms"*, in Proceedings of the ISCAS '94 Symposioum, Vol. 3, pp. 193-196, June 1994

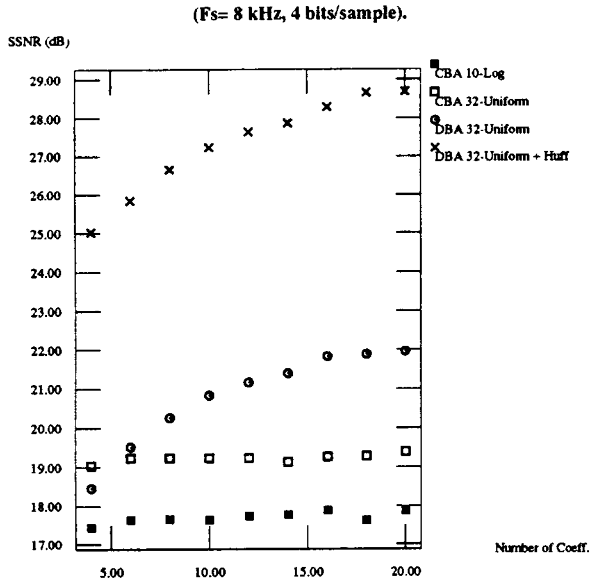

Figure 1: Comparison of log and uniform subband coding schemes using CBA and DBA plus Huffman Coding. The minimum phase (DAUB-A) family of wavelets was used.

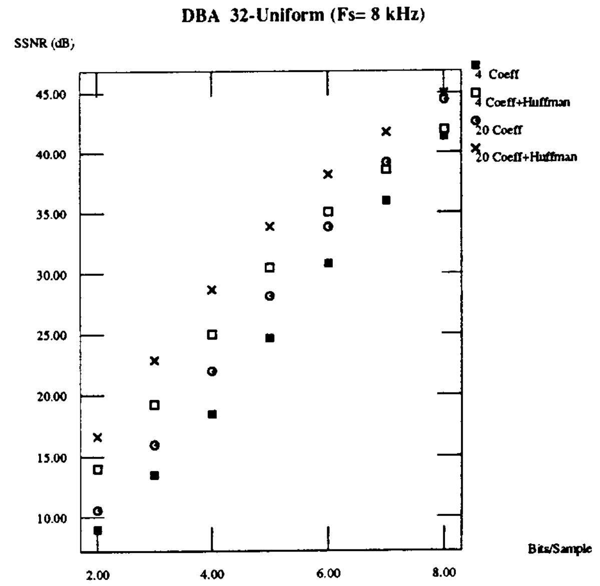

Figure 3: On the performance of the minimum phase (DAUB-A) wavelet family at different bit rates, Fs= 8 kHz.

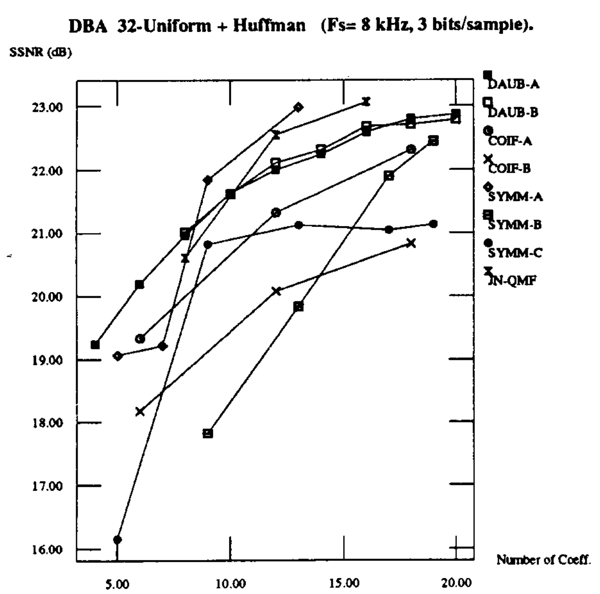

Figure 2: On the performance of different wavelet families, in terms of Segmental-SNR (SSNR).

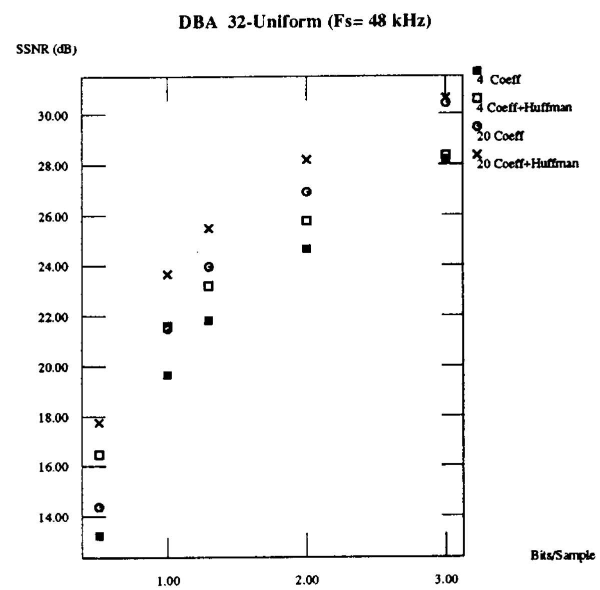

Figure 4: On the performance of the minimum phase (DAUB-A) wavelet family at different bit rates, Fs= 48 kHz.