ON THE CHOICE OF WAVELET FILTERS FOR AUDIO COMPRESSION

P. Philippe, F. Moreau de Saint-Martin, L. Mainard

C.C.E.T.T.
BP 59, rue du Clos Courtel,
35512 Cesson Sévigné cedex, France.
phlip@ccett.fr

ABSTRACT

We address the issue of choosing an optimal wavelet packets transform for audio compression. We present a comparison method based on a perceptual approach, which provides an entropic bit-rate for "transparent" coding of a given audio signal. The test with different wavelets leads to the conclusion that the most significant synthesis criterion for audio compression is the so-called "coding gain", while frequency selectivity, regularity and orthogonality seem less relevant.

1. INTRODUCTION

Wavelet packets transforms appear as an interesting tool for signal compression. They have been thoroughly studied for image compression (e.g. [1, 2, 3]), but they have been introduced in audio compression more recently [4]. While the decomposition [5, 4] and the adaptivity [6] have been until now the most studied points, we address in this paper the issue of choosing the wavelet filters. We want to know whether some wavelet filters are more efficient than others in audio compression, i.e. whether frequency selectivity, regularity or coding gain are relevant criteria for the synthesis of wavelet filters in audio compression.

A compression allowing a transparent quality is required, i.e. the reconstruction error has to be inaudible. Perceptual models [7] have been developed, that provide the admissible reconstruction error spectrum for transparent coding (masking curve). Based on such a model, the method we present computes an estimation of the bit-rate needed for this transparent coding.

By applying this comparison method to wavelet filters with different characteristics, different synthesis criteria are compared : In this way, the choice of the optimal synthesis criterion can be made, taking into account perceptual criteria.

2. COMPARISON METHOD

Principle : We compare, for different filters, the bit-rates which are necessary to ensure a transparent compression (i.e to keep the error due to the quantization under the masking curve). A usual transform–quantization–coding scheme is used. The quantization is assumed to be uniform, with an adaptive step, varying with time and subband number. An entropic coding stage is added for noiseless significant bit-rate reduction.

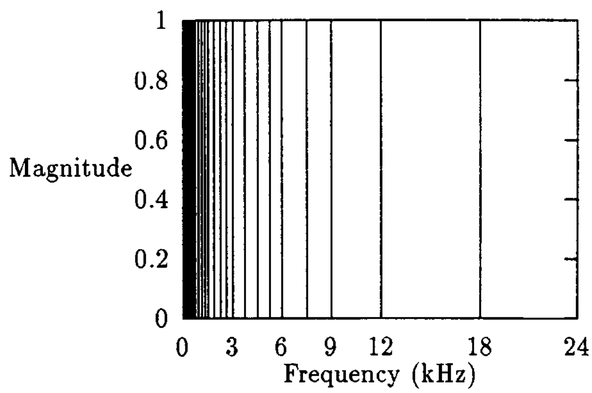

Figure 1: Frequency decomposition following the bark scale.

Transform : We use a dyadic wavelet packets transform. The decomposition tree is chosen so that the equivalent subbands follow the Bark scale : it is similar to the tree chosen in [6]. The nodes of the dyadic tree are chosen so that the 25 subbands of the resulting filter bank are similar to the natural decomposition into 25 Bark channels. This dyadic structure allows a good choice of a time-frequency transform with reference to psychoacoustics knowledge. Fig. 1 presents the ideal decomposition realized by the chosen tree.

Quantization and coding models : With the former decomposition the first issue in the estimation of the characteristic bit-rate for a given filter is the determination of the quantization step in each subband. The quantization noise is assumed white in each subband, with an energy N_i, and the subbands are quantized independently.

The power spectral density of the resulting quanti-

0-7803-2431-5/95 $4.00 © 1995 IEEE

zation noise has to be under the masking curve $\psi(\omega)$ computed with the model 1 from [7] :

$$\sum_{i=1}^{M} N_i \left|H_i(\omega)\right|^2 \leq \psi(\omega) \qquad (1)$$

where $M = 25$ is the number of subbands and the H_i are the synthesis filters of the equivalent parallel filter bank.

For the optimization of the quantization steps, the relationship between the injected noise and the resulting bit-rate has to be known. Assuming that the subband signal is Laplacian (see appendix A), the bit-rate resulting from a uniform quantization followed by an entropic coding can be approximated as a function :

$$R = \sum_{i=1}^{M} \Phi\left(\sqrt{\frac{12N_i}{E_i}}\right) \qquad (2)$$

where N_i denotes the local energy of the quantization noise and E_i the one of the signal, and where Φ can be computed explicitly :

$$\Phi(u) = -(1-e^{-u})\log_2(1-e^{-u}) - e^{-u}\log_2 \sinh u + \frac{u}{(\log 2)\sinh u} \qquad (3)$$

This model allows the minimization of the bit-rate while keeping the transparency, providing the quantization steps in each subband : the resulting entropic bit-rate is then estimated.

Implementation : The algorithm works in a few fast steps. Applying the wavelet transform to an audio signal allows the local energies in the subbands E_i to be computed. The final bit-rate R as defined in (2) is minimized, while maintaining the noise imperceptible with constraint (1) : the constraint is turned into a penalty so that a standard conjugate gradient algorithm provides the optimal quantization steps. The bit-rate is approached by the entropy of the quantized values.

Discussion : This method provides an estimation of the bit-rate needed for a transparent compression of an audio signal, although the bit-rate needed for the side information (Huffman tables,...) and the potential benefit of some interesting techniques (e.g. vector quantization,...) are not taken into account.

The bit-rate increase due to the side information and the potentiality of bit-rate reduction through vector quantization techniques are likely the same for all similar filter bank transforms, so that it is probably significant to compare them without taking into account these elements.

Delay, time-resolution and complexity are are not taken into account in the comparison. Nevertheless, since they are closely related to the lengths of the filters, an independant control of those characteristics is still possible.

This comparison method proved to be discriminant : the difference between entropic bit-rates for different wavelet filters is significant (usually at least a few kbit/s).

3. RELEVANT SYNTHESIS CRITERIA FOR AUDIO COMPRESSION

The use of wavelet transforms for audio compression gives rise to many questions : What kind of decomposition has to be used? Has the transform to adapt to the signal? Which wavelet filters are well suited for audio compression? We address the latter point, which has not been studied up till now : all authors use Daubechies filters, while many other filters are available and may be efficient.

The discussion arises around the synthesis criteria : selectivity, regularity, coding gain, orthogonality, phase linearity... It is difficult to say which filters provide the best transform for audio compression. Using the method, different wavelet filters are compared. Orthogonal wavelet filters have been synthesized with Rioul [8] and Onno [2] algorithms. Biorthogonal wavelet filters have been borrowed from literature (listed in [9]) or synthesized with an algorithm derived from [10]. Filters' lengths are between 4 and 120. For each length, we get a large panel of filters with very different characteristics. Then the entropic bit-rates of the wavelet filters with some classical properties are compared : frequency selectivity is drawn in Fig. 3, "coding gain" in Fig. 4 and regularity in Fig. 2.

Regularity : O. Rioul [3] has shown that the regularity of the wavelet filters is an important property for image compression. As the audio compression transform also uses iterated filter banks, we could expect the regularity to influence the bit-rate. However there is no correlation between regularity and bit-rate, as shown in Fig. 2.

Selectivity and Coding Gain : Frequency selectivity is an intuitive criterion for audio compression filter banks design : it corresponds to the idea of coding in frequency subbands, and of injecting the quantization noise independently in different subbands. Anyway, a closer look at formula (2) shows that the relevant value is not the injected noise energy in itself, but the ratio of it to the energy of the signal : we need a transform with nice statistical properties.

A classical and simple way of studying the statistical properties of a filter bank is the so-called "coding gain", introduced for orthogonal filter bank transforms [11], and used with auto-regressive signal models [1]. See appendix B for a general presentation (including the biorthogonal case). The definition of coding gain has no direct physical meaning for audio compression, because it is based on an SNR maximization, which is a bad model compared to perceptual coding techniques, and because many approximations do not hold in the context of audio compression.

The results of Fig. 4 are very interesting : the correlation between the bit-rate estimation and the coding gain is very strong, and much stronger than the correlation between the frequency selectivity and the bit-rate estimation (Fig. 3). As a conclusion, the coding gain seems to be a much more relevant criterion than the selectivity for audio compression using wavelet packets.

Orthogonality and Phase Linearity : In the case of dyadic wavelets it is well known that orthogonality and phase linearity can not be hold simultaneously (for filter length larger than 2). The orthogonality is a traditional property of the transforms used in signal processing, but for audio compression it is not necessary. E.g. in image compression, the quantization maximizes the SNR, so that the energy preservation is very useful; in audio compression, SNR and energy preservation are not really relevant. Looking for biorthogonal linear-phase wavelets seems therefore to be interesting. The results of our comparison method with biorthogonal wavelets are similar to the ones of the orthogonal wavelets because the estimation of the bit-rate does not take into account the increase of quality due to the linear-phase property. Biorthogonal wavelet transforms might therefore be an interesting tool for audio compression applications.

Filters' length : As mentioned in the conclusion of the former section, the length of the filters is a criterion which does not appear in the bit-rate estimated by the comparison method, but which has to be taken into account because of delay, time-resolution (pre-echoes), and complexity.

Conclusion : For the synthesis of wavelet filters devoted to audio compression, the relevant criterion to be maximized is the coding gain, for given filters' lengths.

4. CONCLUSION

We proposed a useful method for the choice of the wavelet packets transform for audio compression (and furthermore for every kind of frequency decomposition). It estimates an entropic bit-rate which represents the quality of the filter bank, with relation to real audio signals and according to a perceptual model.

The results of that method are strongly correlated with the so-called "coding gain", although its physical meaning does not hold for audio compression. The "coding gain" is a much more relevant synthesis criterion than the regularity or the frequency selectivity,

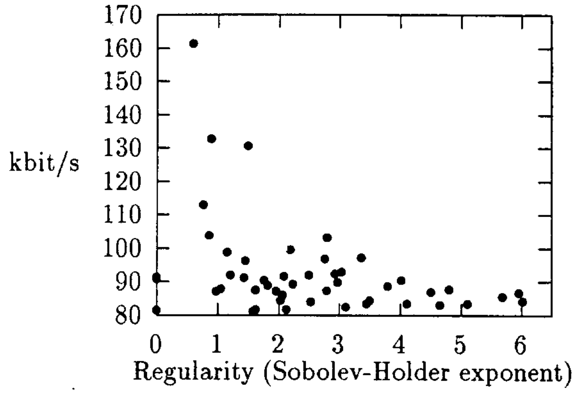

Figure 2: Entropic bit-rate vs regularity

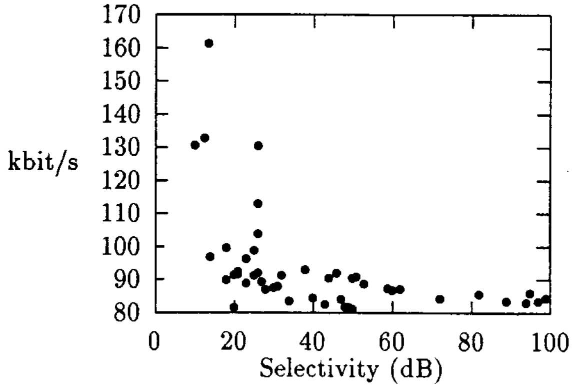

Figure 3: Entropic bit-rate vs frequency selectivity

which are not strongly correlated with the asymptotic bit-rate. Linear-phase biorthogonal wavelets might be interesting when compared with orthogonal ones.

5. REFERENCES

[1] A.N. Akansu and R.A. Haddad, *Multiresolution Signal Decomposition, Transforms, Subbands, Wavelets*, Academic Press, 1992.

[2] P. Onno and C. Guillemot, *Tradeoffs in the Design of Wavelet Filters for Image Compression*, in Proc. VCIP, pp.1536–1547, Cambridge (Massachusetts), 8-11 November 1993.

[3] O. Rioul, *On the Choice on "Wavelet" Filters for Still Image Compression*, ICASSP '93.

[4] M.V. Wickerhauser, *Acoustic Signal Compression with Wavelet Packets*, Wavelets–A Tutorial in Theory and Applications, Academic Press Inc., 1992, pp. 679-700.

[5] J.F. Scholl and D. Rogovin, *Edge effects and Audio Signal Compression with Circular Wavelet Packets*, preprint, Rockwell Science Center, Thousands Oaks, CA 91360, 1994.

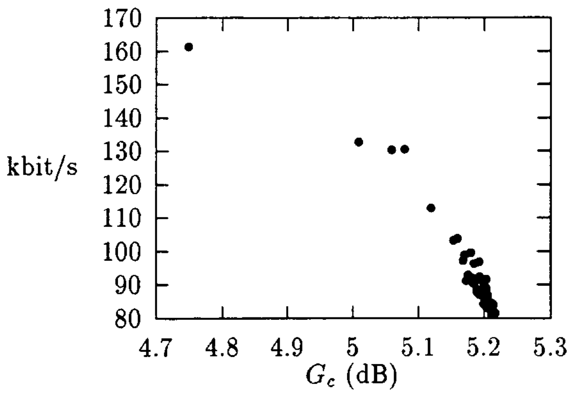

Figure 4: Entropic bit-rate vs coding gain

[6] D. Sinha and A.H. Tewfik, *Low Bit Rate Transparent Audio Compression using Adapted Wavelets*, IEEE Transactions on Signal Processing, Vol. 41, No. 12, December 1993.

[7] ISO/MPEG-Audio, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s*, ISO/IEC 11172-3.

[8] O. Rioul and P. Duhamel, *A Remez Exchange Algorithm for Orthonormal Wavelets*, IEEE Transactions on Circuits and Systems II, Vol. 41, No 8, August 1994.

[9] F. Moreau de Saint-Martin, A. Cohen, P. Siohan, *A Measure of Near-Orthogonality of Biorthogonal Filter Banks*, ICASSP'95.

[10] F. Moreau de Saint-Martin and P. Siohan, *Design of Optimal Linear-Phase Transmitter and Receiver Filters for Digital Systems* Submitted to publication.

[11] N.S. Jayant and P. Noll, *Digital Coding of Waveform*, Prentice-Hall, 1984.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1965.

A. STATISTICAL PROPERTIES OF THE SUBBAND SIGNALS

In section 2 the subband signal probability density function (*pdf*) is assumed Laplacian. In this part, we show the justification.

In general, the subband signal can be viewed as a signal with a generalized Gaussian *pdf*:

$$P_X(x) = a \exp\{-|bx|^{\gamma}\} \tag{4}$$

For $\gamma = 1.0$ and for $\gamma = 2.0$ we have the special cases of a Laplacian and Gaussian *pdf*, respectively.

γ has been estimated with the Kolmogorov-Smirnov test [12]. With real audio signals, γ =0.8 has been obtained. Working with $\gamma = 1$ (Laplacian case) allows explicit computations. For our problem, this lets the entropic bit-rate a little bit smaller than the bit-rate which has been predicted with the Laplacian model, but it does not change the quantization steps computation.

B. ON THE "CODING GAIN"

We extend the reasoning of Noll and Jayant [11] to the biorthogonal case and to wavelet packets decompositions.

Biorthogonal coding gain

The so-called "coding gain" aims to measure the increase of quality (SNR) for a given bit-rate when comparing a simple PCM quantization and a subband quantization. Compared to [11], an extra (small) assumption is necessary in the biorthogonal case : the quantization noise is white and independent from a subband to another, so that as in the orthogonal case the reconstruction error energy in the second case is the sum of the subband quantization error energies [9]. For PCM quantization, the mean square error is $\epsilon\, 2^{-2R}\sigma_X^2$ where R denotes the bit-rate and σ_X^2 the energy of the signal. For subband quantization, it is, in each subband k, $\epsilon_k\, \|H_k\|^2 2^{-2R_k}\sigma_{X,k}^2$. We assume as in [11] that $\epsilon_k \|H_k\|^2 = \epsilon$. The optimal choice of the R_k is shown to be

$$R_k = \frac{1}{2}\log_2 \frac{\sigma_{X,k}^2}{\left(\prod_{k=1}^{M} \sigma_{X,k}^2\right)^{1/M}} + R_{\text{allocated}}$$

We obtain therefore :

$$G_c = \frac{\sigma_X^2}{M\ \left(\prod_{k=1}^{M} \sigma_{X,k}^2\right)^{1/M}}$$

As in [1] we use this formula with an input signal X following an autoregressive model AR(1) (with an autocorrelation $\rho = 0.93$), for single 2-band filter banks.

Wavelets packets coding gain

The latter formula holds for a single 2-band filter bank, i.e. a cell of the wavelet packets transform. When working with wavelet packets transform, an average over the different cells has to be done. Sometimes, the coding gain measures how far the energy of a low-pass (respectively high-pass) signal is concentrated in the low-pass (respectively high-pass) branch. In the orthogonal case, both lead to the same coding gain, so that averaging over the cells does not change anything. In the biorthogonal case, for our decomposition, we have therefore to consider :

$$G_c = 0.75 G_c(\rho) + 0.25 G_c(-\rho)$$

where $G_c(\rho)$ denotes the coding gain for an input signal X following an AR(1) model with autocorrelation ρ.

Acknowledgements : We would like to thank P. Onno and J.B. Rault, C.C.E.T.T. , and J.P. d'Alès de Corbet, University of Paris-Dauphine, for their valuable suggestions and for the fruitful discussions we had.