OPTIMAL WAVELET PACKETS FOR LOW-DELAY AUDIO CODING

Pierrick Philippe1 François Moreau de Saint-Martin1 Michel Lever1 Joël Soumagne2

1 CCETT, BP 59, rue du Clos Courtel, 35512 Cesson Sévigné cedex, France.
2 SUPELEC, 2 rue Ed. Belin, 57078 Metz cedex 3, France.

ABSTRACT

We address the issue of choosing an optimal wavelet packets decomposition for low-delay audio bit-rate reduction. The decomposition tree and the choice of the wavelet filters for each 2-band cell of the tree are optimized. For a given filterbank we measure the entropy of the symbols resulting from a quantization stage ensuring transparent compression. This measure is the criterion to be minimized under a delay constraint. The optimization is carried out with simulated annealing techniques among a large family of filters. We show that there is no loss choosing the wavelet filters among a family of filters that have been optimized with respect to coding gain for AR(1) models. This provides efficient wavelet packets for low bit-rate low delay audio compression.

2. THE WAVELET PACKETS SUBBAND DECOMPOSITION

Wavelet packets transformation is implemented with iterated two-channel bi-orthogonal or orthogonal filter banks. Without quantization perfect reconstruction of the signal after the analysis-synthesis operations is ensured. Choosing a well suited transformation consists in determining the structure of the decomposition tree (i.e. the bandwidth of each subband) and the corresponding set of wavelet filters. Thus a large number of degrees of freedom are available since all tree structures and all filters combinations are possible.

A major problem is that iterating the filtering and the decimation operations generates a reconstruction delay: the reconstructed signal is a translated version of the original. The delay generated by a cell is related to the length of the filters and to the decimation rate of the cell, so that the delay can be huge, e.g. 30000 samples. We need therefore to take the delay into account in the wavelet packet filter bank design: we focus here on the realization of low delay filterbanks, ensuring nevertheless good compression efficiency.

Usually, authors use wavelet or wavelet packets subband splitting schemes with the same filters along iterations. A solution to achieve low delays is therefore to use short filters or to limit the frequency resolution of the transformation. Authors have also proposed to reduce progressively the filters length when the decimation factors increase. For instance C. Galand [6] reduces the overall reconstruction delay of the tree cascaded filterbank by dividing the filters length by a factor of 2 at each stage. This approach is a good way to reduce the reconstruction delay but may be suboptimal.

In this paper we use different filters for each cell of the tree-structured filterbank. The choice of the filters is based on an estimate of the bit-rate which is necessary for a transparent perceptual coding for standard signals with a given filter bank (this bit-rate is later called entropy associated to a filter bank).

3. ENTROPY ASSOCIATED TO A FILTERBANK FOR A GIVEN SIGNAL

The estimation of the entropy associated to a filterbank was first presented in [4]. The principle is to measure the entropy of the quantized samples under the constraint that the reconstruction noise is kept inaudible.
The purpose is then to minimize the entropy associated to the quantization operation under a perceptual constraint derived from the psychoacoustic model 1 of [7]. The reconstruction error has to be under a so-called masking curve, i.e.:

$$\forall \omega \sum_{k=0}^{M-1} N_k|H_k(\omega/M_k)|^2 \leq \psi(\omega) \quad (1)$$

where $H_k(\omega)$ denotes the frequency response of the $k$th subband synthesis filter, $N_k$ the quantization noise power, $M_k$ the downsampling factor and $\psi(\omega)$ the power spectral density of the masking curve.

The entropy is then computed as:

$$R = \sum_{k=0}^{M-1} \frac{N}{M_k} \Phi \left( \frac{N_k}{E_k} \right) \quad (2)$$

$E_k$ denotes the variance of the subband signal, and $N$ the frame size. $\Phi(\cdot)$ is the entropy of the quantization steps under the assumption that the signal fits a Laplacian probability density function (which is an accurate estimation of the subband audio signal statistics), as computed in [4].

To measure an entropy associated to a filterbank we therefore minimize the quantity $R$ in (2) while respecting the constraint (1). Since this entropy takes into account both the statistical properties of the subband signals and perceptual criteria, we therefore can look for transformations optimal in a perceptive coding sense. Since the masking curve depends only on the input signal, filterbanks can be compared through their associated entropy.

4. OPTIMIZATION

The problem is then to choose a decomposition tree and the wavelet filter banks at each elementary cell in the tree. Our strategy consists in considering a large set of filters and in selecting for each cell the best one among this set. The optimization is global and takes into account the delay constraint.

The set of available filters consists of filters designed using Daubechies’ algorithm [8] (maximally flat wavelets), Oron’s algorithm [9] (optimal filters in terms of coding gain over AR(1) signals), Smith-Barnwell algorithm [10] (selectivity) and Rioult’s algorithm [11] (trade-off between regularity and selectivity), and Akansu and Caglar filters [12]. This set of filters allows us to cover a large variety of characteristics and lengths.

The determination of the optimal wavelet filter for each cell is the minimization of the transparent coding bit-rate estimation under the delay constraint. The algorithm consists in applying alternatively two algorithms: An algorithm optimizes the choice of the filters for a given decomposition tree and the other one optimizes the decomposition tree for preselected filters.

4.1. Wavelet filter choice strategy

The optimization strategy consists in a simulated annealing method. For a given tree structure, each node of the tree (each elementary two-channel filter bank) is considered as a location whose status is the selected wavelet filters in the library. In this huge discrete optimization problem, a simulated annealing strategy is well suited.

From a given configuration, a location is selected. The criterion value is computed for different changes of the status of the location. From these possible criterion values, status changes probabilities are computed. The change is carried out randomly following these probabilities.

Let us describe the algorithm: At iteration $k$, the selected filters are $F_k^1, ..., F_k^N$. The temperature is $T_k$. The location which may change is $i$. The entropy (as the objective function) $D$ is computed for several configurations:

$$U(f) = D(F_k^1, ..., F_k^i, f, F_k^{i+1}, ..., F_k^N). \quad (3)$$

Probabilities $P(f)$ are computed as

$$P(f) = \frac{\exp \left( -\frac{U(f)}{T_k} \right)}{\sum_f \exp \left( -\frac{U(f)}{T_k} \right)} \quad (4)$$

The new status for the location $i$ is selected randomly according to the probabilities $P$. $T_k$ decreases to $T_{k+1}$ and a new location $i$ is selected. If $T$ is slowly decreased, under certain conditions, the optimality of the algorithm can be proven. In practice it leads at least to near-optimal configurations.

4.2. Tree structure choice strategy

Given the wavelet filters which are used in all cells, the optimization of the decomposition tree is possible. It is an algorithm with progressive construction of the tree. We start from the basic 1-cell tree. At this point, two bands are possibly split, so that two filter banks have to be compared. Since we can choose the best solution for 3 subbands one can repeat this procedure for the choice of a 4-subband decomposition and so on.

5. RESULTS

For our experiments the optimization was performed using real audio material. The simulations concern the design of subband coder having a delay from 100 to 700 samples. For a 48 kHz sampled audio source, it corresponds to a reconstruction delay in the range of 2 ms to 15 ms.

5.1. On the tree structure

We intend now to look at the best decomposition for a given number of subbands, e.g. 16. The results are presented in a compact way in Fig. 1. The curves are made in the following way: It is the line through points $(f_i, i)$, where $f_i$ is the upper frequency of the frequency band corresponding to subband $i$. The decomposition of an equal bandwidth filter bank would be represented as the diagonal of the rectangle. When the curve grows faster in low-frequencies, this means that the lower-frequency subbands are decomposed more deeply. The results we obtained are that when the length increases (meaning also when the frequency selectivity or the AR(1) coding gain increases), the lower frequency bands are decomposed more deeply.

It is an important point to compare the resulting time-frequency decomposition with other classical filter banks. The Fig. 2 shows the normalized subband, which is the
index of the subband divided by the total number of subbands, in function of the frequency. Let us recall that all uniform band filter banks, such as Musicam [7] or as a TDAC [1], would be represented as the diagonal of the rectangle. The decomposition MPAC [13] and optimal wavelet packets are compared with the Bark scale. It is clearly seen that the kind of decomposition we obtain is much closer to the Bark scale than any other previous decomposition.

We then need to choose the number of subbands. This consists in taking into account the side information due to each subbands. The bit-rate of these subband side information is estimated in the range 6–10 bits per frame and per subband. The results are presented in Fig. 3. This shows that the optimal number of subbands is around 16.

5.2. Restricting the filters family

This experiment has been carried out for a given tree structure. For each delay value, we find the corresponding optimal decomposition, i.e. the optimal set of filters. Thus, an entropy is associated to each value of delay, corresponding to the optimal filterbank. Simulations show the results obtained with different sets of filters.

- The first set consists of Daubechies filters [8] whose lengths vary from 2 (Haar filters) to 40. These filters are regular and are the most frequently chosen wavelet filters [5].
- The second set is composed of filters optimal in terms of coding gain, and are designed by Onno's algorithm [9]. In [4] it was proven that these filters are optimal for the wavelet packet decomposition for audio coding, if no delay constraint is added.
- The third set is composed of all (300) available filters.

Fig. 4 shows the results: we have drawn the characteristics bit-rate vs delay for the three previous configurations. As expected, it describes a convex hull, and we can see that the bit-rate decreases when the delay grows up. This is due to the fact that long filters are more efficient than short ones, particularly in terms of frequency localisation or coding gain.

Daubechies' filters already used in previous work [5] are not optimal in this context. Onno's filters perform very well, and it is consistent with [4]: When no delay constraint is considered and when using the same filter along iterations, the most efficient filters are Onno's ones. Furthermore we can also conclude that a mix of filters from different families improve significantly the efficiency of the compression at a fixed delay.

From another point of view, the curves 4 provide a delay-vs-compression chart. There is a trade-off between bit-rate reduction and resulting delay. Better bit-rate reduction can be achieved without delay constraint, and compression efficiency is reduced due to the constraint, but the loss of efficiency is very small when the delay constraint becomes stronger: when the delay is reduced from 700 to 100, the loss of efficiency is around 10%.

6. CONCLUSION

In this paper we have shown how a wavelet packets decomposition can be designed for low bit-rate and low delay audio compression. The design is performed in an objective way giving an optimal solution in a perceptive coding sense. We prove that a delay constraint can be introduced with a very small loss in terms of compression efficiency. It is a new point of great interest that low-delay constraints can be taken into account in wavelet packet filter banks design: Therefore they can be used in usual or low-delay audio compression applications.

REFERENCES


Figure 1. Optimal decompositions made of 16 subbands for different wavelet filters. The filters are optimal with respect to coding gain for AR(1) models [9]. $L$ denotes the length of the filter.

Figure 2. Comparison of the subband-frequency decomposition of different time-frequency representations (Onno's filters [9]).

Figure 3. Estimated bit-rate vs number of subbands: The cost of side information for each subband is estimated as 0–6–10 bits per frame and per subband.

Figure 4. Estimated bit-rate vs delay for three sets of filters. The curve labeled “Daubechies” describes the convex hull obtained using filterbanks using Daubechies wavelets [8], the one labeled “Onno” is corresponding to filterbanks using Onno filters (optimal in coding gain) [9], and “Total” corresponds to a set of filters designed with Smith-Barnwell [10], Daubechies, Onno, Akansu and Caglar [12] and Rioul [11] algorithms.