FLEXIBLE FREQUENCY DECOMPOSITIONS FOR COSINE-MODULATED FILTER BANKS

O.A. Niamut and R. Heusdens

Dept. of Mediamatics
Delft University of Technology,
Delft–The Netherlands
E-mail: {O.A.Niamut, R.Heusdens}@ITS.TUDelft.NL

ABSTRACT

We investigate the use of nonuniform cosine-modulated filter banks for audio coding. A rate-distortion framework is employed, similar to the work in [1], to select the filter bank structure from a large library of possible frequency decompositions. A new flexible frequency decomposition algorithm is proposed that jointly optimizes the filter bank structure and the bit allocation over the subband channels. Experimental results for both synthetic and real audio signals are provided. The new algorithm shows significant improvements in comparison with fixed uniform frequency decompositions, but special care has to be taken to reduce the size of the decomposition overhead.

1. INTRODUCTION

In most of the current audio coding standards a cosine-modulated filter bank (CMFB [2]) is employed, using either a polyphase or lapped transform implementation. These filter banks provide a uniform frequency decomposition, i.e. a decomposition where all the subband channels are uniformly spaced in frequency. However, for more efficient coding of audio and speech signals, a larger library of filter bank structures is required in order to adapt the time-frequency resolution of the filter bank to the signal's changing characteristics [3].

A large library of filter bank structures is for instance provided by wavelet packets [4]. Various algorithms have been proposed that choose the optimal wavelet packet basis and corresponding quantizers per time segment, where optimality is defined in a rate-distortion (R-D) sense [5]. The resulting frequency decompositions are no longer restricted to uniform band divisions. On the other hand, for CMFBs only few algorithms exist to obtain time-varying nonuniform frequency decompositions [6, 7]. However, when compared to wavelet packets, CMFBs possess interesting properties for audio coding such as good frequency selectivity and simple design of transition filters.

In this paper, we propose a new algorithm to obtain a rate-distortion optimal frequency decomposition of an audio signal using CMFBs. By combining techniques for the design of nonuniform filter banks and dynamic programming-based R-D optimization, we construct the flexible frequency decomposition algorithm.

The organization of this paper is as follows. In Section 2 some previous methods to obtain time-varying frequency decompositions are discussed. Section 3 describes the new algorithm in detail. In Section 4 some examples are provided and a comparison with fixed uniform decompositions is made. Section 5 contains the conclusions and recommendations for future work.

The research was conducted within the ARDOR project, supported by the E.U. grant no. IST-2001-34095

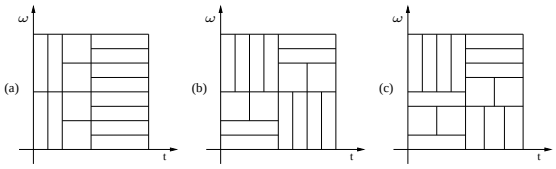

Fig. 1. *Time-frequency tilings as obtained by decomposition algorithms. (a)Window-switching tiling (b)Single Tree tiling (c)Flexible Frequency Decomposition tiling*

2. PREVIOUS WORK

For audio coding, several methods for adapting the time-frequency resolution of the analysis system have been proposed. In [8], the window-switching algorithm is presented. The time-frequency resolution is adapted by switching the analysis block length, typically between a long-duration/high-frequency resolution mode and a short-duration/low-frequency resolution mode. The short window applied to a frame containing a transient will tend to minimize the temporal spread of quantization noise (which results in a reduction of pre-echos). Furthermore, it is desirable to constrain the high bit rates associated with transients to the shortest possible temporal regions only.

Although implemented in most of the current audio coding standards, the window-switching technique has some drawbacks. For instance, special transition windows have to be employed when switching between resolutions. This introduces extra coder delay and the spectral properties of these windows are poor compared to those of the original windows [9]. Moreover, the resulting frequency decompositions are still uniform and therefore limited in their ability to model non-stationary fragments correctly. See Figure 1a for an example of the time-frequency tilings that can be obtained using window-switching.

A frequency-varying decomposition method based on wavelet packets (WP) is disclosed in [10], where the *Single Tree* algorithm jointly finds the WP basis and bit allocation that are optimal in a rate-distortion sense. A Lagrange optimization technique is em-

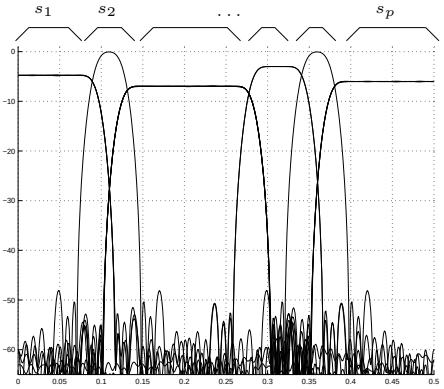

Fig. 2. *A decomposition S_k is a collection of adjacent (nonuniform) subband channels $s_1, \ldots, s_p$.*

ployed that searches along the convex hull of the R-D curve to determine the jointly optimal WP basis and corresponding quantizer choices. However, the use of wavelet packets in the Single Tree algorithm has several drawbacks. First of all, the frequency decompositions are limited to dyadic intervals (i.e. binary decompositions) only. Figure 1b shows an example of a tiling that can be obtained, while Figure 1c shows a tiling that cannot be achieved with the Single Tree algorithm. Secondly, carefully designed filters are needed at the segment boundaries [11] when the Single Tree is combined with time-segmentation algorithms. Moreover, the subband filters have poor frequency responses due to the cascaded implementation of the WP filter bank.

Some work on a frequency-varying CMFB has been reported in [12]. However, this algorithm starts from a decomposition that resembles the critical band structure. Within each critical band, only binary decompositions are possible.

3. FLEXIBLE FREQUENCY DECOMPOSITION

Given an M-channel uniform CMFB, we want to minimize the total distortion over all possible frequency decompositions and all possible ways of quantizing the corresponding subband signals such that the total required bit rate does not exceed a certain target rate R_t. If we limit ourselves to the case where every possible decomposition consists of subband channels having a bandwidth that is an integer multiple of a predefined minimum bandwidth (i.e. the bandwidth of the filters of the underlying uniform CMFB), this problem becomes the frequency equivalent of the flexible time segmentation algorithm proposed in [1].

To state the problem more formally we introduce some notation. Let $S = \{S_1, \ldots, S_{2^{M-1}}\}$ be the set of all possible frequency decompositions, where $S_k = \{s_1, \ldots, s_p\}$ is a collection of adjacent (nonuniform) frequency intervals. Figure 2 shows an example of such a decomposition. Furthermore, assume that we are given a set of quantizers $\{q_n\}$ to quantize the subband samples in a decomposition and let $Q = \{Q_l, \ldots, Q_N\}$ denote the set of all possible ways of quantizing the different decompositions S_k, where $Q_l = \{q_1(s_1), \ldots, q_p(s_p)\}$. The problem that we want to solve can then be expressed as

$$\min_S \min_Q D(S_k, Q_l) \qquad (1)$$
$$\text{subject to} \quad R(S_k, Q_l) \leq R_t.$$

Clearly, Eq. 1 can be solved by introducing a Lagrange multiplier $\lambda \geq 0$ and solving the unconstrained minimization problem

$$\min_S \min_Q J(\lambda) = \min_S \min_Q \sum_{i=1}^{p} J_i(\lambda, s_i, q_i(s_i)), \qquad (2)$$

where we assume that rate and distortion are additive over the subband channels.

Solving Eq. 2 directly would require an exhaustive search of computational complexity $\mathcal{O}(2^M)$. However, if we can assume that the different subband channels are mutually uncorrelated, the search for the optimal quantizer strategy given a particular decomposition can be done on a channel-by-channel basis, that is,

$$\min_Q \sum_{i=1}^{p} J_i(\lambda, s_i, q_i(s_i)) = \sum_{i=1}^{p} \min_{q_i(s_i)} J_i(\lambda, s_i, q_i(s_i)). \qquad (3)$$

This assumption is the key step in reducing the search complexity since we now can solve Eq. 2 using the dynamic programming technique [13], which results in a computational complexity of $\mathcal{O}(M^2)$.

The optimal frequency decomposition is now found recursively. Let $J_{k,l}$ denote the Lagrangian cost for encoding the frequency range $s_{k,l} = [\frac{\pi}{M}k, \frac{\pi}{M}l)$. Then, at each iteration i, the best frequency decomposition of the interval $[0, \frac{\pi}{M}i)$ is found by solving

$$J^*_{0,i} = \min_{0 \leq k \leq i} (J^*_{0,k} + J_{k,i}), \quad i = 1, \ldots, M, \qquad (4)$$

where $J^*_{0,i}$ is the minimum cost for coding the interval $[0, \frac{\pi}{M}i)$. Figure 2 illuminates this procedure. After having found $J^*_{0,M}$ we can easily determine the optimal frequency decomposition by backtracking all the optimal split positions.

Obviously, if we do not know the right λ in advance, we have to repeat the aforementioned procedure for different values of λ in order to determine the optimal λ (i.e. the one that gives rise to $R = R_t$). Since the rate is a convex function of the distortion, efficient algorithms exist to find the optimal λ in a few iterations, e.g. the bisection method [10].

The computation of the Lagrangian costs for solving Eq. 4 can become very complex. In general, if we replace two adjacent subband channels by two double-bandwidth channels, the perfect reconstruction property is lost so that the other channel filters have to be modified as well and thus the subband signals. A complete signal transformation is then necessary for each and every possible decomposition, 2^{M-1} in total, which is unacceptable in most applications.

If the subband merging technique presented in [7] is employed, we can reduce the number of required signal transformations to only one, since the other decompositions can be derived by a simple post-processing of the subband signals of the underlying uniform CMFB. This is the main reason for applying this technique to the design of nonuniform frequency decompositions.

It is important to note that the merging operation does not reduce the number of channels by itself. For example, merging 2 adjacent channels results in 2 double-bandwidth channels, each having a different time localization. See Figure 4 for an example. As a result, in order to find the optimal bit allocation for a particular frequency interval $s_{k,l}$ we need different quantizers for the subband channels that constitute the interval under consideration.

Summarizing, the flexible frequency decomposition algorithm can be implemented as follows:

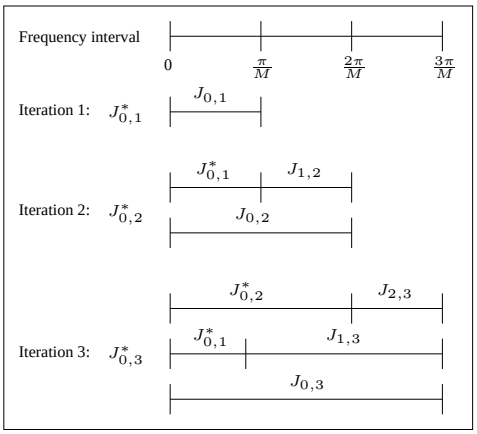

Fig. 3. *Dynamic Programming is employed to search iteratively for the optimal decomposition*

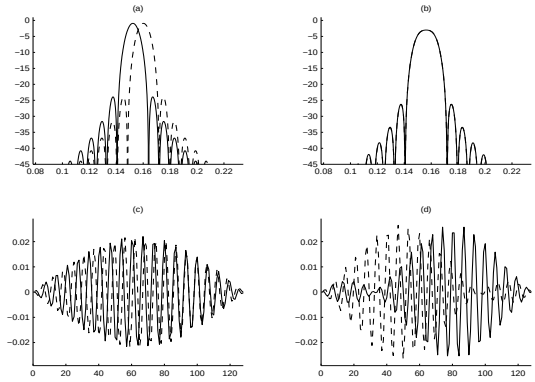

Fig. 4. *Resolution switching for 2 filters with subband merging (a)Magnitude response of unmerged filters (b)Magnitude response of merged filters (c)Time localization of unmerged filters (d)Time localization of merged filters*

---

1. For $k \in \{1, 2, \ldots, M\}$, compute every possible decomposition S_k of the frequency interval $[0, \frac{\pi}{M}k)$.
2. For every decomposition S_k, compute all possible ways $Q_l = \{q_1(s_1), \ldots, q_p(s_p)\}$ of quantizing the i subband samples and record the resulting distortions and bit rates.
3. For an initial value λ, find the optimal decomposition $S^*_{0,i}$ of the interval $[0, \frac{\pi}{M}k)$, resulting in the minimum cost $J^*_{0,i}$ for $i = 1, \ldots, M$, where $J^*_{0,0} = 0$.
4. Find the optimal value of λ, that corresponds to the target rate R_t, using the bisection algorithm [10].

3.1. Reduction of Algorithmic Complexity

Several steps can be undertaken to reduce the complexity of the algorithm. For instance, instead of considering every possible combination of subband filters, we can limit the number of adjacent channels merged to powers of 2. As shown in [7], this restriction results in orthonormal nonuniform CMFBs, assuming that the underlying uniform CMFB is also orthonormal. Orthonormal filter banks are desirable, since in the quantization distortion can then be evaluated in the frequency domain only, so that the inverse filter bank operation is not needed at the encoder.

A second reduction in complexity is obtained by setting an upperbound on the number of adjacent channels that are merged. However, this restriction does not necessarily lead to a severe degradation of performance, because the time-frequency localization of filters obtained by merging a large number of subbands is suboptimal.

3.2. Coding of Side Information

The decoder has to be informed about the selected filter bank structure. This structure can be represented as a binary sequence of length $M - 1$, where a one denotes a split between adjacent subband filters and a run of m zeros denotes that $m + 1$ adjacent subband filters are merged. As shown in [14], the information rate for such sequences is close to 1 bit/sample, even if we restrict the maximum number of channels to be merged significantly. Such a decomposition overhead is clearly unacceptable. However, initial coding experiments showed that using simple Huffman coding of the runlengths of ones and zeros already reduces the overhead by a factor 5, resulting in an overhead rate of 0.2 bit/sample.

4. EXPERIMENTAL RESULTS

The flexible frequency decomposition algorithm was implemented in a generic CMFB-based audio codec. The M subband samples were scaled by a single scale factor (the largest absolute sample value). A normalized quantizer was employed, where the quantizer resolution for quantizing the subband signals was varied according to the allocated number of bits.

Figure 5 demonstrates the algorithm performance for a 1st-order AR signal with $\rho = 0.9$. The subband samples from a 16-channel filter bank are coded at a target rate R_t of 24 bits using 8 different quantizer resolutions. Clearly, the use of a variable frequency decomposition results in a better modelling of the signal and a higher SNR. In the example given, pre-echos are reduced significantly.

Table 1. *A comparison between fixed uniform decomposition and variable nonuniform decomposition. Average segmental SNRs are presented. The first column shows the results for a fixed uniform decomposition coded at* 1.5 *bit/sample. The second contains the SNRs for variable decompositions coded at* 1.5 *bit/sample, while the last column presents the result for a fixed uniform decomposition coded at* 1.7 *bit/sample.*

Fragment	**Fixed** (1.5)	**Variable** (1.5)	**Fixed** (1.7)
Castanets	11.7	17.7	13.3
Suzan Vega	19.4	22.8	21.6
German Male	24.8	27.71	27.6

Several audio fragments taken from the SQAM [15] reference disc were coded using the aforementioned coding scheme and compared for both fixed uniform and variable nonuniform

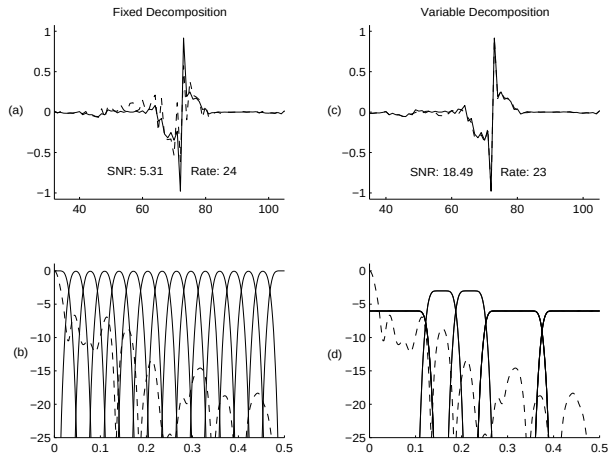

Fig. 5. *A comparison between fixed uniform and variable nonuniform decomposition. (a)Original (solid) and reconstructed (dashed) signal for uniform decomposition (b)Uniform filter bank and signal magnitude response (c)Original (solid) and reconstructed (dashed) signal for nonuniform decomposition (d)Nonuniform filter bank and signal magnitude response*

decompositions. The filter bank used to obtain the uniform frequency decomposition and applied in the subband merging algorithm was a 512-channel uniform CMFB. The target rate was set to 1.5 bit/sample for both cases, resulting in a decomposition overhead of 0.2 bit/sample.

Table 1 shows the resulting average segmental SNRs for three cases. The second column shows the SNR for the uniform decomposition case, while the third column presents the SNR for the nonuniform decomposition, where we did not include the overhead rate. Clearly, a significant improvement in SNR is obtained for all fragments. To compare these result to the case where we could spent an extra 0.2 bit/sample for the fixed uniform decomposition, the last column shows the SNRs. It is clear that for some fragments (e.g. German Male Speech) a further reduction of the overhead rate is necessary.

5. CONCLUDING REMARKS

A new algorithm for rate-distortion optimal frequency decompositions using cosine-modulated filter banks was proposed. The flexible frequency decomposition algorithm jointly optimizes the filter bank structure and the bit allocation over the subband channels. The decomposition overhead was reduced by a simple entropy coder. Experimental results for both synthetic and real audio signals showed that the new algorithm outperforms a fixed uniform frequency decomposition.

The new algorithm is currently being compared to the existing algorithms. Further reduction of the decomposition overhead is necessary to ensure an increase of SNR for all audio signals. Moreover, the incorporation of a perceptual distortion metric that considers both frequency and temporal masking is planned to employ the algorithm in a perceptual audio coder. The flexible frequency decomposition algorithm can then easily be combined with the window-switching technique to increase the adaptive nature of the time-frequency analysis.

6. REFERENCES

[1] Cormac Herley, Zixiang Xiong, Kannan Ramchandran, and Michael T. Orchard, "Flexible time segmentations for time-varying wavelet packets," *IEEE-SP Conference of Time-Frequency and Time-Scale Analysis*, pp. 9–12, October 1994.

[2] R.D. Koilpillai and P.P. Vaidyanathan, "Cosine-modulated fir filter banks satisfying perfect reconstruction," *IEEE Trans. Signal Proc.*, vol. 40, no. 4, pp. 770–783, April 1992.

[3] J. Princen and J.D. Johnston, "Audio coding with signal adaptive filter banks," *IEEE-ICASSP*, vol. 5, pp. 3071–3074, May 1995.

[4] Y.Meyer R.R.Coifman and M.V.Wickerhauser, "Wavelet analysis and signal processing," *Wavelets and their applications*, pp. 153–178, 1992.

[5] C. Herley, Z. Xiong, K. Ramchandran, and M.T. Orchard, "Flexible tree-structured signal expansions using time-varying wavelet packets," *IEEE Trans. Signal Proc.*, vol. 45, no. 2, pp. 333–345, February 1997.

[6] M. Purat and P. Noll, "A new orthonormal wavelet packet decomposition for audio coding using frequency-varying modulated lapped transforms," *IEEE-ICASSP*, vol. a, pp. 1021–1024, April 1996.

[7] O.A. Niamut and R. Heusdens, "Subband merging in cosine-modulated filter banks," *IEEE Signal Processing Lett.*, Accepted for publication.

[8] B. Edler, "Codierung von audiosignalen mit uberlappender transformation und adaptiven fensterfunktionen (in german)," *Frequenz*, vol. 43, no. 9, pp. 252–256, 1989.

[9] S. Shlien, "The modulated lapped transform, its time-varying forms, and its applications to audio coding standards," *IEEE Trans. Speech and Audio Proc.*, vol. 5, no. 5, pp. 359–366, July 1997.

[10] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Proc.*, vol. 2, no. 2, pp. 160–175, April 1993.

[11] C. Herley, J. Kovačević, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3341–3359, December 1993.

[12] M. Purat and P. Noll, "Audio coding with a dynamic wavelet packet decomposition based on frequency-varying modulated lapped transforms," *IEEE-ICASSP*, vol. a, pp. 1021–1024, April 1996.

[13] D.P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[14] S. van de Par R. Heusdens, "Rate-distortion optimal sinusoidal modeling of audio and speech using psychoacoustical matching pursuits," *IEEE-ICASSP*, vol. 2, pp. 1809–1812, May 2002.

[15] "Sound quality assessment material recordings for subjective tests," Technical Centre of the European Broadcasting Union, April 1988.