

# MMSE Design of Modulated and Tree-Structured Filter Banks for Efficient Tradeoffs Between Rate, Distortion, and Decoder Complexity

Karine Gosse, *Member, IEEE*, Tanja Karp, *Member, IEEE*, François Moreau de Saint-Martin, Pierre Duhamel, *Fellow, IEEE*, and Alfred Mertins, *Member, IEEE*

**Abstract**—Designing filter banks for source coding purposes classically relies on perfect reconstruction filters. However, several studies have shown recently that taking the quantization noise into account in the design could yield noticeable reduction of the reconstruction distortion. In particular, a joint optimization of the synthesis filters and quantizers with respect to the output mean squared error (MSE) was proposed, and the resulting scheme was called an minimum MSE (MMSE) filter bank. However, this approach was dedicated to parallel uniform filter banks, which require a very large computational complexity. This paper is concerned with the possibility of providing MMSE filters under the constraint that the filter bank is either of a modulated or a tree-structured kind and thus of low implementation cost. Various approaches for the optimization of the synthesis bank are presented, taking into account the structural constraint. Depending on the number of parameters to be optimized, we obtain various tradeoffs between decoder complexity, transmitted bit rate, and reconstruction distortion.

**Index Terms**—Efficient realizations, filter bank optimization, MMSE filter banks, rate-distortion optimization.

## I. INTRODUCTION

CLASSICALLY, in a subband coding scheme, an  $M$ -band Filter Bank (FB) splits the signal to be encoded into decorrelated subband components, prior to their quantization. This latter step, performing the lossy compression, is often followed by an entropy-coding (or variable-length coding) stage, which is reversible on the decoder side without any loss of information. Designing a coding scheme thus requires the optimization of these various stages, jointly or not. For example, a classical choice includes a perfect reconstruction (PR) filter bank in conjunction with subband quantizers. In such a scheme, only the quantizers are tuned for minimum

output distortion, while the synthesis bank follows directly from the choice of the analysis bank, independently of the bit-rate allocation.

However, PR filter banks only provide a minimal reconstruction error when used in absence of subband quantization and previous studies already aimed at optimizing quantizers and/or the synthesis filters [1]–[10]. The approach in [1] and [10], jointly optimizes the quantizers and the synthesis filters with respect to the output mean squared error (MSE), yielding optimal (non-PR) filter banks having optimal quantization steps in the subbands, here denoted as minimum MSE (MMSE) filter banks. However, these global solutions for the synthesis filters were obtained for general  $M$ -band critically decimated analysis FB's. Optimal synthesis filters solving the minimization problem were not constrained to be of any particular structure, and therefore, even if the analysis FB was structurally constrained to have low arithmetic complexity (tree-structured or modulated), the synthesis filters did not allow for efficient realizations via fast transforms or lattice implementations.

This paper proposes solutions for obtaining MMSE FB's preserving distinctive implementation features of filter banks, *plus* reducing the number of parameters that are adapted according to the second-order statistics of the signal—this makes the optimization easier than in the general formulation. Furthermore, the complexity of the implementation is reduced.

Two main types of solutions are presented: 1) solutions dedicated to modulated filter banks and 2) solutions with tree-structured filter banks. For both cases, efficient algorithms performing the optimization are proposed, and the improvement over PR schemes having optimal subband quantizers is shown by means of rate-distortion curves. Moreover, each case is formulated so that various choices for the tradeoff between rate, distortion, and optimization complexity exist.

Since most real-world signals are not stationary, the practical use of MMSE FB's in actual coding schemes is still a difficult task. In fact, MMSE FB's have to be adapted according to variations of the signal statistics. Hence, the synthesis FB has either to be recomputed at the receiver (very large complexity), or has to be transmitted (at the cost of some bit rate). In both cases, our approach brings some advantages. The way to adapt the MMSE filters to nonstationary signals is, in itself, not the subject of this paper, but is briefly discussed in the last section.

Manuscript received December 17, 1996; revised December 2, 1997. Part of the work described here was previously presented in Proc. EUSIPCO, Trieste, Italy, September 1996; Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 1996; and IEEE ICASSP, 1996.

K. Gosse is with the Centre de Recherche de Motorola, Espace Technologique de Saint-Aubin, F-91193 Gif-sur-Yvette cedex, France.

T. Karp is with the University of Mannheim, B6/26, D-68131 Mannheim, Germany.

F. Moreau de Saint-Martin is with the Ministère de l'Économie, des Finances et de l'Industrie, DGSI/SERICS, Sous Direction Communication Audiovisuelle et Electronique Grand Public, F-75353 Paris 07 SP, France.

P. Duhamel is with Department Signal, ENST, 46, F-75634 Paris cedex 13, France.

A. Mertins was with the University of Kiel, D-24143 Kiel, Germany. He is now with the University of Western Australia, Nedlands, 6907WA, Australia.

Publisher Item Identifier S 1057-7130(98)04671-0.

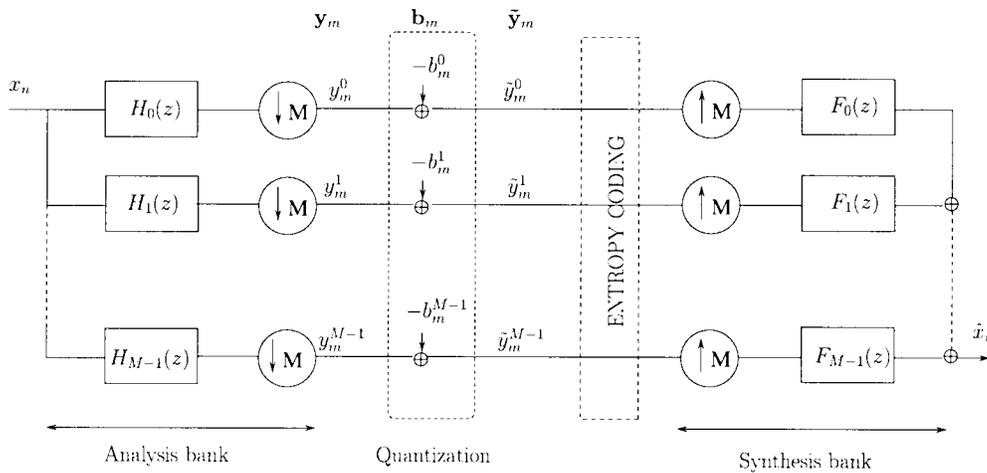


Fig. 1. Compression system including a filter bank and a quantization stage.

### A. Connection with Previous Works

Taking additive noise into account in filter design has already been successfully applied in digital signal processing: it has led to Wiener filters and to MMSE equalizers in the digital communications field. The study in [3] generalizes these scalar solutions by presenting matrix Wiener filters for subband coders. The same idea appears in [7] with FIR filters, and is extended to the 2-D case. Further generalization is shown in [1], where general solutions are given for jointly optimizing parallel, critically sampled, synthesis filters and uniform subband quantizers. In comparison, [3] is mainly an asymptotic study of matrix Wiener filters with infinite length. Yet, in none of these works do the synthesis banks obtained keep the original structure (modulated or tree-structured) of the synthesis filter bank.

In [4]–[6], the quantization noise is taken into account in the special case of nonuniform quantization, for which part of the quantization noise is correlated to the signal, but no joint optimization of synthesis filters and bit allocation was feasible, since this work was restricted to the PR case.

The work in [2] is related to the problem of modulated filter bank design. Optimal synthesis prototype filters are obtained for given subband quantizers. However, the optimization includes neither the choice of the quantizers, nor the tuning of additional degrees of freedom, as will be done here. Note that the optimization of tree-structured filter banks in an MMSE sense has not been addressed before. Part of the work described here was previously presented in [8]–[10].

### B. Outline of the Paper

Section II briefly recalls the problem of optimizing a coding scheme based on parallel, critically sampled analysis filter banks with subbands of equal widths and uniform quantization [1]. In Section III, this optimization problem is modified so as to constrain the synthesis filters to be modulated versions of a low-pass prototype. At first glance, the only tunable parameters in this case seem to be the prototype coefficients. However, this does not allow for a tuning of the synthesis filters' frequency responses according to the spectral density

of each subband signal. On the other hand, this mechanism can be partly restored by inserting a matrix  $\mathbf{A}$  in the subbands after the quantization (see Fig. 2). Various forms for  $\mathbf{A}$  enable to tune the optimization and implementation complexity.

Section IV is dedicated to tree-structured filter banks and their optimization in an MMSE sense. We show how to formulate the problem so as to get an efficient algorithm derived from the one in Section II. As in the modulated case, the proposed method enables us to find flexible tradeoffs between output SNR (signal-to-noise ratio) improvement and optimization complexity. We conclude in Section V with a discussion on source coders based on MMSE filter banks, including the time-adaptation problem.

## II. GENERAL FRAMEWORK FOR MMSE FILTER BANKS

In a subband coding scheme, as depicted in Fig. 1, the quantizers introduce some distortion in the subbands, whose amount depends on the bit-rate allocation. The problem solved in [1] is to compute both the quantization steps and the synthesis filters so as to minimize the mean squared distortion  $D = \mathcal{E}[|\hat{x}_n - x_{n-n_0}|^2]$  ( $\mathcal{E}$  stands for mathematical expectation,  $\hat{x}_n$  for the reconstructed signal, and  $x_{n-n_0}$  for the stochastic input signal delayed by  $n_0$  samples) under a bit-rate constraint. We therefore need an expression of the reconstruction error as a function of the synthesis filters and quantizers. As a result, rate-distortion curves for classical PR filter banks having their quantization steps tuned can be compared to those of MMSE FB's in which *both* the quantization steps and the synthesis filters are tuned. Note that in order to enable a comparison to the classical PR filter bank compression schemes, all analysis FB's considered in this paper are chosen in such a way that their PR synthesis counterpart exists, even if this property is not required in the MMSE framework.

### A. The Optimality Criterion

For the sake of conciseness, some steps in distortion calculations are omitted in the following development since they are already described in [1].

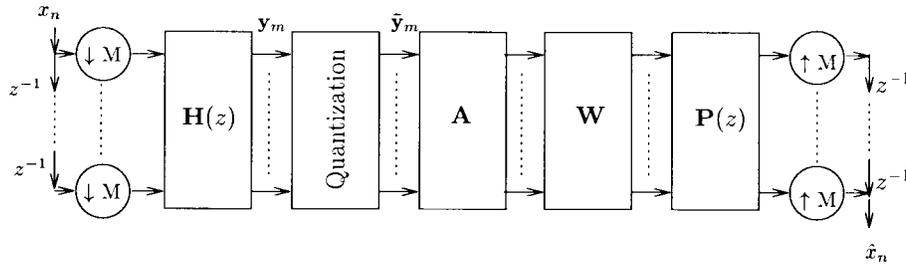


Fig. 2. Polyphase implementation of an  $M$ -band coding system including an analysis filter bank  $H(z)$ , a quantization stage, a combination matrix  $A$ ; a modulation matrix  $W$ , and the prototype  $P$ .

The criterion to be minimized is chosen as the mathematical expectation of the squared reconstruction error  $|\hat{x}_n - x_{n-n_0}|^2$ . Since the reconstructed signal  $\hat{x}_n$  is cyclostationary of period  $M$  due to the up-sampling operation ( $M$  is also the number of subbands), the output MSE is

$$D = \frac{1}{M} \sum_{i=0}^{M-1} \mathcal{E}[|\hat{x}_{mM+i} - x_{mM+i-n_0}|^2], \quad (1)$$

$D$  can be expressed by means of the following variables:

- synthesis filters coefficients vectors  $\mathbf{F}_i^T$

$$\mathbf{F}_i^T = (f_i^0 \cdots f_i^{M-1}, f_{i+M}^0 \cdots f_{i+M}^{M-1} \cdots f_{i+(K-1)M}^{M-1}). \quad (2)$$

The components of  $\mathbf{F}_i^T$  appear to be the coefficients of the  $i$ th polyphase components of the synthesis filters.  $K$  denotes the length of these polyphase components, and  $f_i^j$  the  $i$ th coefficient of the  $j$ th synthesis filter,  $i, j = 0, \dots, M-1$ ;

- autocorrelation matrix  $\mathbf{R}_{\tilde{y}\tilde{y}}$  of the quantized subband signals

$$\mathbf{R}_{\tilde{y}\tilde{y}} = \mathcal{E}\{\tilde{\mathbf{y}}_m \tilde{\mathbf{y}}_m^T\} \quad (3)$$

with  $\tilde{\mathbf{y}}_m^T = (\tilde{y}_m^T, \tilde{y}_{m-1}^T, \dots, \tilde{y}_{m-K+1}^T)$  and  $\tilde{\mathbf{y}}_m = (\tilde{y}_m^0, \tilde{y}_m^1, \dots, \tilde{y}_m^{M-1})^T$  (according to Fig. 1); note that by construction,  $\mathbf{R}_{\tilde{y}\tilde{y}}$  contains cross-correlation terms between all quantized subband signals;

- cross correlations between the vector  $\tilde{\mathbf{y}}_m$  and the delayed input sample  $x_{mM-n_0+i}$ ;
- the variance of the input signal  $\sigma_x^2$ .

The resulting expression for  $D$  is given by the following equation:

$$D = \frac{1}{M} \sum_{i=0}^{M-1} (\mathbf{F}_i^T \mathbf{R}_{\tilde{y}\tilde{y}} \mathbf{F}_i - 2\mathbf{F}_i^T \mathcal{E}\{\tilde{\mathbf{y}}_m x_{mM-n_0+i}\}) + \sigma_x^2. \quad (4)$$

Modeling the quantization noise in subband  $k$  as an additive noise ( $b_m^k = y_m^k - \hat{y}_m^k$ ), the MSE can be rewritten in terms of the autocorrelation matrix of the quantization noise,  $\mathbf{R}_{bb} = \mathcal{E}\{\mathbf{b}_m \mathbf{b}_m^T\}$ , and the cross correlation matrix of the noise and the subband signals,  $\mathbf{R}_{by} = \mathcal{E}\{\mathbf{b}_m \tilde{\mathbf{y}}_m^T\}$  where  $\mathbf{y}_m$  and  $\mathbf{b}_m$  are defined in the same way as  $\tilde{\mathbf{y}}_m$ . Note that  $\mathbf{R}_{bb}$  and  $\mathbf{R}_{by}$  have the same structure as  $\mathbf{R}_{\tilde{y}\tilde{y}}$  before and contain cross correlations

of noise and signals from various subbands, Thus

$$D = \underbrace{\sigma_x^2 + \frac{1}{M} \sum_{i=0}^{M-1} (\mathbf{F}_i^T \mathbf{R}_{yy} \mathbf{F}_i - 2\mathbf{F}_i^T \mathcal{E}\{\mathbf{y}_m x_{mM-n_0+i}\})}_{D_f} + \underbrace{\frac{1}{M} \sum_{i=0}^{M-1} [\mathbf{F}_i^T (\mathbf{R}_{bb} - 2\mathbf{R}_{by}) \mathbf{F}_i + 2\mathbf{F}_i^T \mathcal{E}\{\mathbf{b}_m x_{mM-n_0+i}\}]}_{D_b}. \quad (5)$$

The MSE is thus divided in two terms. The first error term  $D_f$  is due to the non-PR property of the filter bank. It cancels if the synthesis filters, represented here by vectors  $\mathbf{F}_i$ , form a PR filter bank with the given analysis bank. The second term,  $D_b$ , is due to additive quantization noise. At very high bit rates as well as in absence of quantization,  $\tilde{\mathbf{y}}_m$  approaches  $\mathbf{y}_m$ , and  $D_b$  tends to zero. Therefore, at high bit rates, the optimized synthesis filter bank converges to a filter bank providing PR with the analysis bank. Note that  $D_f$  encompasses aliasing and linear distortion. We did not distinguish between both, despite the fact that aliasing may be more perceptible. This is a deliberate choice, made for simplicity. Elaborating on our approach using perceptually weighted criteria instead of the MSE may lead to improved subjective coding performance.

### B. Quantization Noise

In [1], two additive noise models were introduced: the classical white input-independent process and a more accurate colored-noise model [11]. Since the latter did not improve the performance of the resulting MMSE scheme, the subband quantization noise resulting from uniform quantization is assumed here to be uncorrelated across the subbands and inside each subband. Under this assumption, matrix  $\mathbf{R}_{by}$  is zero,  $\mathbf{R}_{bb}$  is diagonal, with diagonal terms  $\sigma_{b^k}^2$ , and the distortion is

$$D = \underbrace{\sigma_x^2 + \frac{1}{M} \sum_{i=0}^{M-1} (\mathbf{F}_i^T \mathbf{R}_{yy} \mathbf{F}_i - 2\mathbf{F}_i^T \mathcal{E}\{\mathbf{y}_m x_{mM-n_0+i}\})}_{D_f} + \underbrace{\frac{1}{M} \sum_{i=0}^{M-1} \mathbf{F}_i^T \mathbf{R}_{bb} \mathbf{F}_i}_{D_b}. \quad (6)$$

Under the high resolution assumption, noise variances  $\sigma_{b^k}^2$  in subband  $k$  are classically related to quantizers  $q_k$  by  $\sigma_{b^k}^2 = q_k^2/12$ .

### C. Bit-Rate Constraint

The optimization is undertaken with the constraint of a fixed total bit rate  $R_T$

$$\sum_{k=0}^{M-1} R_k = R_T \quad (7)$$

where  $R_k$  denotes the bit rate in subband  $k$ . In the following, the optimization is carried out upon the bit rates, and a rate-distortion relation is required to express the reconstruction distortion as a function of these variables. For source coding purposes, two situations are of particular interest.

- 1) Subband signals are split into blocks, and the number of bits used for a given block depends on the dynamic range  $d_k = \max_y(y^k) - \min_y(y^k)$  of these given signals.
- 2) Entropy coding is performed in each subband. The order 1 entropy of the quantized signals, defined in subband  $k$  as

$$H_k = - \sum_{j \geq 1} p_j^{(k)} \log_2 p_j^{(k)} \quad (8)$$

may also be considered as bit-rate measure ( $p_j^{(k)}$  denotes the occurrence probability of the  $j$ th quantization step in subband  $k$ , and  $H_k$  represents the amount of information brought by a realization of the quantized signal in subband  $k$ , or, in other words, a lower bound of the bit rate reached in practice after entropy coding.) In this case, the optimization is carried out under the constraint of a given entropy budget  $H_T = \sum_{k=0}^{M-1} H_k$  (i.e., entropy-constrained optimization).

In the cases mentioned above, noise variances  $\sigma_{b^k}^2$ , bit rates  $R_k$  (or entropies  $H_k$ ), and variances  $\sigma_{y^k}^2$  of subband signals  $y^k$  are related by

$$\sigma_{b^k}^2 = c_k \sigma_{y^k}^2 2^{-2R_k} \quad \text{or} \quad \sigma_{b^k}^2 = c_k \sigma_{y^k}^2 2^{-2H_k}. \quad (9)$$

The positive parameter  $c_k$  depends on the signal statistics in subband  $k$  as well as on the type of quantization and/or encoding process used.

For uniform quantizers,  $R_k$  and  $q_k$  verify  $R_k = \log_2(d_k/q_k)$ , with  $d_k$ , the dynamic range of the signal as defined above. Alternatively,  $d_k$  may be chosen as  $c_k \cdot \sigma_{y^k}$ . If the bit rate measure is the order 1 entropy of the signals, we have  $c_k = \pi e/6$  under the assumption of Gaussian subband signals, and  $c_k = e^2/6$  under the assumption of Laplacian signals [12]. Note that up to multiplicative constants, the distortion expression is common to all bit-rate measures, thus enabling the use of the same optimization algorithm in all cases.

### D. The MMSE Solutions

The optimization is carried out by optimizing the parameters iteratively, since each optimization (bit allocation, filters) is

simple if the other parameters are fixed. Starting with a PR filter bank and assuming white and uncorrelated quantization noise, we perform the following.

*Step 1:* For the given synthesis filters, the noise term  $D_b$  has to be minimized over the set of quantizers. Expressing the diagonal terms  $\sigma_{b^k}^2$  of  $\mathbf{R}_{bb}$  in (6) by (9), the distortion caused by quantization noise is

$$D_b = \sum_{k=0}^{M-1} c_k \sigma_{y^k}^2 \left( \sum_{n=0}^{N-1} f_n^{k^2} \right) 2^{-2R_k}. \quad (10)$$

The bit rates  $R_k$  are thus obtained by minimizing the Lagrangian functional  $\mathcal{L} = D_b + \lambda(\sum_k R_k - R_T)$ , in which  $R_k$  is constrained to be positive ( $R_k = \rho_k^2$ , see [1] and [13]).

*Step 2:* For the obtained set of quantizers, we now optimize the synthesis filters, setting  $\partial D / \partial \mathbf{F}_i = 0$  in (6)  $\forall i$ ,  $0 \leq i \leq M-1$ , which amounts to solving a set of linear equations. In fact, the distortion is a quadratic function of the synthesis coefficients, such that the new filter coefficients given by

$$\mathbf{F}_i = (\mathbf{R}_{yy} + \mathbf{R}_{bb})^{-1} \mathcal{E}\{\mathbf{y}_m x_{mM-n_0+i}\} \quad (11)$$

necessarily form an optimum (assumption: invertible matrix  $[\mathbf{R}_{yy} + \mathbf{R}_{bb}]$ ). The computation of  $\mathbf{F}_i$  relies on an estimation of matrix  $\mathbf{R}_{yy}$  that will be detailed in the simulation parts.

Steps 1 and 2 are performed in turn, always using the updated values obtained in the previous step as constants. Convergence is ensured by the fact that each step reduces the distortion and thus increases the signal-to-noise ratio (SNR).

Indeed, assume that iteration  $n$  of Step 1 results in bit rates  $R_k^{(n)}$ , filters  $\mathbf{F}_i^{(n-1)}$  and corresponding  $\text{SNR}^{(n)}$ , a function  $f$  of  $[R_k^{(n)}, \mathbf{F}_i^{(n-1)}]$ , and now apply Step 2. Denote  $\text{SNR}^{(n+1)} = f[R_k^{(n)}, \mathbf{F}_i^{(n)}]$ , the distortion with the new optimal synthesis filters  $\mathbf{F}_i^{(n)}$ .  $\text{SNR}^{(n+1)} < \text{SNR}^{(n)}$  is clearly impossible, since  $\mathbf{F}_i^{(n-1)}$  would be a better choice for Step 2, and  $\mathbf{F}_i^{(n)}$  would not be an optimum. Thus,  $\text{SNR}^{(n+1)} \geq \text{SNR}^{(n)}$ .

For the same reasons, the SNR necessarily increases after Step 1. The consecutive SNR values obtained form an increasing and upper-bounded series that converges to a (possibly local) maximum. However, we cannot guarantee that it converges toward the global one. Initializing the process with a PR filter bank, the procedure either improves the classical PR solution, or it stops in Step 2.

### E. Spectral Interpretation of the Improvement Brought by MMSE FB's

The MMSE solution for general  $M$ -band critically sampled synthesis banks with individually operating synthesis filters is mainly a Wiener-type solution with optimized quantizers in the subbands. From the classical Wiener filter theory, we know the following. In the presence of white noise, the frequency response of a Wiener filter with an infinite number of coefficients has a higher attenuation at all frequencies than the inverse filter. Hence, frequency bands with a low signal level being coarsely quantized will be more attenuated by the Wiener synthesis filters than bands with finer quantization. In subbands where the spectral density of the input signal is partly zero, this will result in a passband of reduced width. Globally,



Thus, the distortion is a quadratic function of the synthesis prototype coefficients. Hence, provided that the matrix to be inverted has full rank, optimal vectors  $\mathbf{P}_{i\downarrow}$  are given by

$$\mathbf{P}_{i\downarrow} = [\mathbf{W}_{*,i}^T \mathbf{A}^T \mathbf{R}_{\tilde{y}\tilde{y}} \mathbf{A} \mathbf{W}_{*,i}]^{-1} \mathbf{W}_{*,i}^T \mathbf{A}^T \mathcal{E}\{\mathbf{y}_m x_{mM-n_0+i}\},$$

$$i = 0, \dots, M-1. \quad (15)$$

*b) Optimization of Diagonal Gain Factors:* By developing the distortion as a function of scalar gain factors  $\alpha_{i,i} = [\mathbf{A}]_{i,i}$  ( $0 \leq i < M$ ), the criterion appears to be quadratic in terms of these coefficients. The analytical expression of these dependencies requires the definition of the matrices and vectors in (15a), shown at the bottom of the page.

With the notations in (15a), (14) now reads.

$$D = \sigma_x^2 + \frac{1}{M} \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} \left( \alpha_{j,j} \alpha_{k,k} \sum_{i=0}^{M-1} \mathbf{P}_{i\downarrow}^T \mathbf{W}_{j,i}^T \mathbf{R}_{\tilde{y}^j \tilde{y}^k} \mathbf{W}_{k,i} \mathbf{P}_{i\downarrow} \right) - 2 \sum_{j=0}^{M-1} \left( \alpha_{j,j} \sum_{i=0}^{M-1} \mathbf{P}_{i\downarrow}^T \mathbf{W}_{j,i}^T \mathcal{E}\{\mathbf{y}_m^j x_{mM-n_0+i}\} \right) \quad (16)$$

and optimal gain factors are solutions of the following set of linear equations:

$$\sum_{k=0}^{M-1} \alpha_{k,k} \left( \sum_{i=0}^{M-1} \mathbf{P}_{i\downarrow}^T \mathbf{W}_{j,i}^T \mathbf{R}_{\tilde{y}^j \tilde{y}^k} \mathbf{W}_{k,i} \mathbf{P}_{i\downarrow} \right) = \sum_{i=0}^{M-1} \mathbf{P}_{i\downarrow}^T \mathbf{W}_{j,i}^T \mathcal{E}\{\mathbf{y}_m^j x_{mM-n_0+i}\},$$

$$j = 0, \dots, M-1. \quad (17)$$

*2) Optimization in a Paraunitary Filter Bank:* Optimizing the prototype filter may not always be necessary, especially when the analysis FB has a large number of subbands and narrow transition bandwidths (see Section II-E). Assuming that the filter bank is paraunitary for  $\mathbf{A} = \mathbf{I}$ , the equations for optimizing  $\mathbf{A}$  simplify due to  $\mathcal{E}\{\|x_{n-n_0} - \hat{x}_n\|^2\} = \mathcal{E}\{\|\mathbf{y}_m - \mathbf{A}\tilde{\mathbf{y}}_m\|^2\}$ .

With white and uncorrelated noise, the output MSE can be written in terms of the original subband signals  $\mathbf{y}_m$  and of the quantized subband signals  $\tilde{\mathbf{y}}_m$  as [8]

$$D = \mathcal{E}\{\|\mathbf{y}_m - \mathbf{A}\tilde{\mathbf{y}}_m\|^2\}$$

$$= \sum_{i=0}^{M-1} \left[ \underbrace{\mathbf{A}_i^T \mathbf{R}_{bb} \mathbf{A}_i}_{D_b} + \underbrace{\mathbf{A}_i^T \mathbf{R}_{yy} \mathbf{A}_i - 2\mathcal{E}\{\mathbf{y}_m^i \mathbf{y}_m^T\} \mathbf{A}_i + \sigma_{y_i}^2}_{D_f} \right] \quad (18)$$

TABLE I  
MODULATED FILTER BANKS OPTIMIZED AND COMPARED IN SECTION III  
(\* THIS NUMBER DOES NOT INCLUDE SUBBAND QUANTIZERS.  
THE LAST COLUMN DESCRIBES THE SYNTHESIS FILTER BANK  
IMPLEMENTATION COST IN MULTIPLICATIONS PER INPUT SAMPLE;  $N = KM$   
DENOTES THE FILTER LENGTH AND  $M$  THE NUMBER OF SUBBANDS

denomination	synthesis FB type	tuned parameters	no. of tuned parameters*	implement. cost of synthesis bank
PR	modulated	quantizers	0	$K + \mathcal{O}(\log_2 M)$
mod. MMSE $\mathbf{A}_{diag}$	modulated	diagonal $\mathbf{A}$ quantizers	$M$	$K + \mathcal{O}(\log_2 M) + 1$
mod. MMSE $\mathbf{A}_{trid}$	modulated	tridiagonal $\mathbf{A}$ quantizers	$3M - 2$	$K + \mathcal{O}(\log_2 M) + 3 - 2/M$
mod. MMSE proto	modulated	prototype quantizers	$N$	$K + \mathcal{O}(\log_2 M)$
mod. MMSE $\mathbf{A}_{diag}$ proto	modulated	diagonal $\mathbf{A}$ prototype quantizers	$N + M$	$K + \mathcal{O}(\log_2 M) + 1$
mod. MMSE $\mathbf{A}$	modulated	plain $\mathbf{A}$ quantizers	$M \cdot M$	$K + \mathcal{O}(\log_2 M) + M$
MMSE	unconstr.	synthesis filters quantizers	$M \cdot N$	$N \cdot M$

where  $\mathbf{A}_i$  is the  $i$ th row of  $\mathbf{A}^T$ ,  $\mathbf{R}_{yy} = \mathcal{E}\{\mathbf{y}_m \mathbf{y}_m^T\}$ , and  $\mathbf{R}_{bb} = \text{diag}(\sigma_{b^0}^2, \sigma_{b^1}^2, \dots, \sigma_{b^{M-1}}^2)$ .

To achieve minimum distortion in an MSE sense, we solve  $\partial D / \partial \mathbf{A}_i^T = 0$  and obtain

$$\mathbf{A}_i^T = \mathcal{E}\{\mathbf{y}_m^i \mathbf{y}_m^T\} (\mathbf{R}_{yy} + \mathbf{R}_{bb})^{-1}. \quad (19)$$

Note that this result somewhat generalizes the results of [4]–[6]. Equation (19) still simplifies if  $\mathbf{A}$  is constrained to be diagonal or tridiagonal, see [8].

### 3) Overall Optimization Algorithm:

- *Step 1:* Optimization of the synthesis system.
  - *Step 1.1:* Synthesis prototype optimization according to (15) (not in the paraunitary case),
  - *Step 1.2:* Matrix  $\mathbf{A}$  optimization, according to (17) in the diagonal, nonparaunitary case or according to (19) in the paraunitary case,
  - Go to Step 1.1 until convergence of Step 1.
- *Step 2:* Bit-rate optimization as described in (10). The synthesis polyphase filters can be calculated from the optimized parameters.
- Go to Step 1 until global convergence of the algorithm.

*4) Implementation Cost of the Different Filter Banks:* The implementation cost of the different types of modulated MMSE filter banks in terms of multiplications per input sampling rate is given in Table I. We there assume that the modulation is performed via a fast transform with

$$\left. \begin{aligned} [\mathbf{W}_{i,j}]_{k,l} &= [\mathbf{W}]_{i+kM, j+lM} \\ [\mathbf{R}_{\tilde{y}^i \tilde{y}^j}]_{k,l} &= [\mathbf{R}_{\tilde{y}\tilde{y}}]_{i+kM, j+lM} \\ \mathbf{y}_m^j &= (y_m^j, y_{m-1}^j, \dots, y_{m-K+1}^j)^T \end{aligned} \right\} \begin{aligned} \forall i, j \in \{0, \dots, M-1\} \\ \forall k, l \in \{0, \dots, K-1\} \end{aligned} \quad (15a)$$

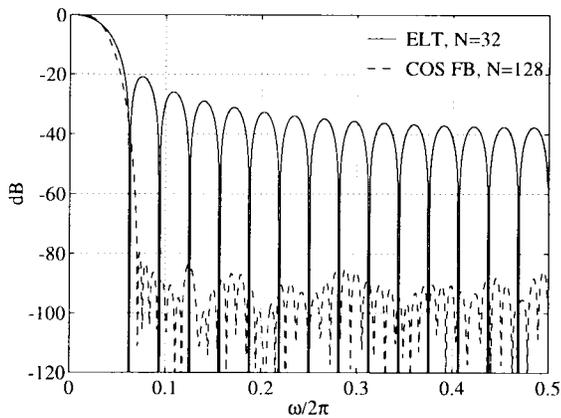


Fig. 3. Prototypes for 8-channel paraunitary modulated filter banks.

$\mathcal{O}(M \log_2 M)$  multiplications every  $M$  samples. It can be easily verified that the implementation cost for all modulated MMSE schemes is significantly lower than for the general MMSE  $M$ -band case denoted as MMSE in Table I.

### C. Simulations

We compare here the performance of various synthesis filter banks, as described in Table I, representing various performance-complexity tradeoffs:

- 1) a PR scheme having optimal subband bit rates (“classical” solution);
- 2) a general MMSE  $M$ -channel filter bank (optimization of all synthesis filters, as in [1]);
- 3) paraunitary schemes combined with 3 different matrices  $\mathbf{A}$  (plain, diagonal  $\mathbf{A}_{\text{diag}}$ , and tridiagonal  $\mathbf{A}_{\text{trid}}$ );
- 4) optimization of the synthesis prototype filter only (matrix  $\mathbf{A}$  is identity);
- 5) joint optimization of the synthesis prototype and  $\mathbf{A}_{\text{diag}}$ .

This comparison is done by means of rate-distortion curves. The unconstrained MMSE solutions are provided in order to show the loss in performance due to the fact that we restrict the optimization result to be a modulated FB.

In all cases, the analysis filters are the same, in order to enable fair comparisons. Two paraunitary analysis filter banks are considered: an 8-band extended lapped transform (ELT) [15] with filters of length  $N = 32$  and a cosine-modulated bank [16] having  $M = 8$  and  $N = 128$ . Both prototype frequency responses are shown in Fig. 3. This choice enables to illustrate the influence of the filter selectivity on the optimization procedure.

The performance is evaluated on the audio signal “The Four Seasons: The Spring” by Vivaldi, CD quality (the first 3 s). The performance measure is the output SNR computed by decomposing, quantizing, and reconstructing the signal. The MMSE FB’s themselves are computed from an estimate of the autocorrelation matrix of the input signal over the whole considered Vivaldi signal.

*Example 1:* Fig. 4 shows the rate-distortion curves obtained for the 8-channel analysis filter bank with  $N = 32$  taps under bit rate constrained optimization. The aim is to illustrate the relative performance of the compared schemes listed above.

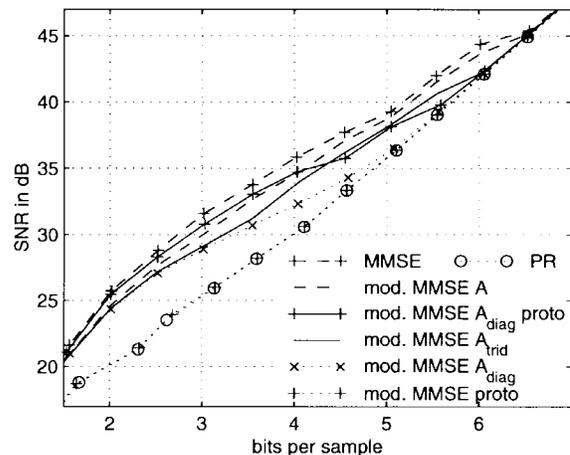


Fig. 4. Measured SNR (dB) versus bit rate for ELT,  $M = 8$ ,  $N = 32$ , and “Vivaldi” signal.

For the rate-distortion measure according to (9), we used  $c_k = 12, \forall k$ , in agreement with the observed Vivaldi signal dynamics, in order to avoid any overload noise. The best performance is of course obtained by optimizing all synthesis filters individually: the SNR improvement over PR is  $\geq 5$  dB for bit rates of 1.5–4.5 b/s (bits per sample). However, optimizing the main diagonal of  $\mathbf{A}$  improves the SNR by  $\geq 3$  dB for low bit rates up to 3.5 b/s. This improvement, obtained at very low cost, reduces quickly for higher bit rates. It can be kept  $\geq 3$  dB for bit rates up to 4.5 b/s by optimizing the tridiagonal structure of  $\mathbf{A}$ . Optimizing the full matrix  $\mathbf{A}$ , which results in an increased implementation cost, yields further improvement. This can be explained by the fact that the analysis filters have a moderate attenuation in the stopband ( $\approx 40$  dB), so that nonadjacent subband signals are still correlated. Concerning the prototype optimization, it turns out that “mod MMSE  $\mathbf{A}_{\text{diag}}$  proto” is slightly better than “mod MMSE  $\mathbf{A}$ ” up to 4 b/s, see Table I for the definition of the various cases. For rates  $> 4$  b/s, its performance is close to the “mod MMSE  $\mathbf{A}_{\text{trid}}$ ” case. The “mod MMSE  $\mathbf{A}_{\text{diag}}$  proto” scheme seems to be a good choice, since the resulting implementation complexity is very low (one additional multiplication per subband).

*Example 2:* Fig. 5 gives the rate-distortion curves when using an 8-channel cosine-modulated filter bank with filters of length 128 instead of the ELT in Example 1. This figure intends to show the influence of filters selectivity on MMSE performance.

In this case, all schemes yield almost the same rate-distortion curve. Due to the high stopband attenuation of the prototype, the subband signals are better decorrelated than in the ELT case and no further improvement can be achieved by combining them with a tridiagonal or plain matrix  $\mathbf{A}$ . Therefore, for such analysis filters, introducing a diagonal matrix on the synthesis side is the best solution for improving the output SNR of the filter bank, because it enables the use of the existing low complexity implementations of modulated filters. Note that the only MMSE-based scheme that does not lead to noticeable improvements is the one in which the prototype *only* is optimized (i.e.,  $\mathbf{A} = \mathbf{I}$ ).

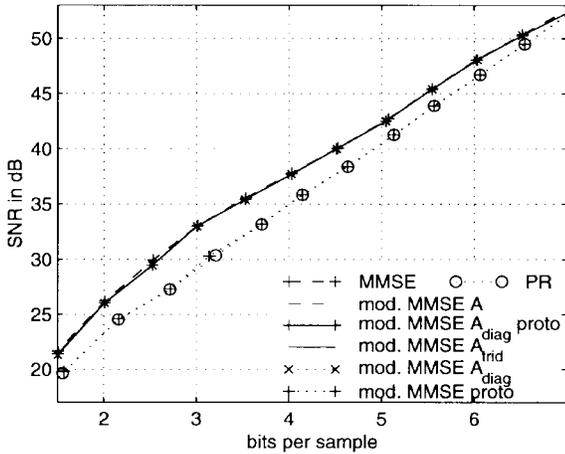


Fig. 5. Measured SNR (dB) versus bit rate for COS FB,  $M = 8$ ,  $N = 128$ , and “Vivaldi” signal.

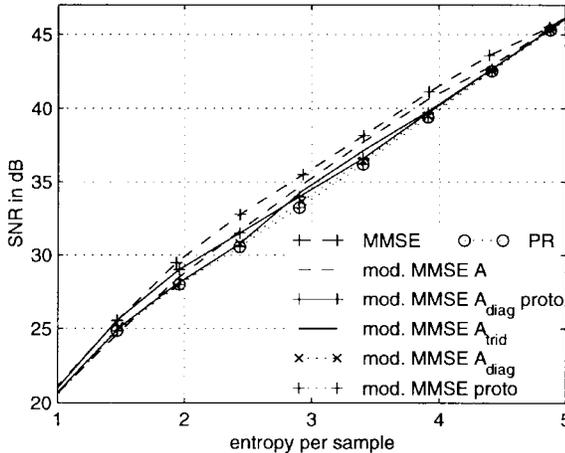


Fig. 6. Measured SNR (dB) versus entropy for ELT,  $M = 8$ ,  $N = 32$ , and “Vivaldi” signal.

*Example 3:* In this example, the bit rate measure is the order 1 entropy of the subband signals, thus yielding an entropy-constrained optimization of filters and quantizers. Fig. 6 shows the curves obtained in this case with the ELT analysis.

We see that the improvement of MMSE schemes over PR ones is reduced. The same conclusions as above can still be given for plain and tridiagonal matrices  $\mathbf{A}$ . Note that introducing a diagonal matrix  $\mathbf{A}$  is not sufficient any more to significantly improve the performance over a PR scheme. Globally, we observe a shrinking of the SNR axis while switching from the bit rate constraint to the entropy constraint, compare, e.g., Figs. 4–6.

An entropy-constrained optimization was also carried out on the selective  $N = 128$  modulated filters. However, no improvement appeared in that case. The reasons for this fact are certainly manifold. First, we are dealing with asymptotic performance. Then, increasing the selectivity (narrow transition bandwidth and high stopband attenuation) of the analysis filters reduces the potential gain brought by MMSE solutions. This shows that a good frequency selectivity is a crucial criterion for the choice of PR filters. It improves the

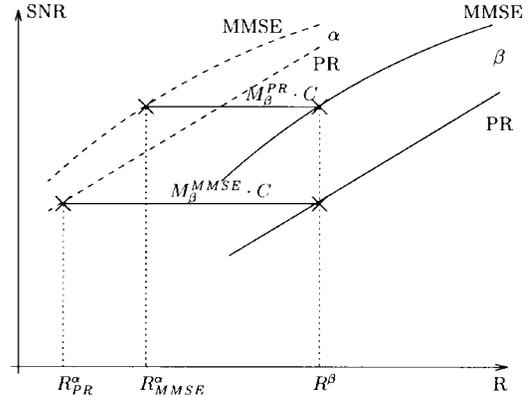


Fig. 7. Translation of rate-distortion curves when the quantization-encoding model changes;  $\alpha$  model (dashed lines), and  $\beta$  model (solid lines).

overall system performance and brings the scheme closer to optimality.

#### D. Some Remarks

1) *Signal Duration:* In the simulations, we used a rather short extract of the “Vivaldi signal (3 s) assuming that the signal is more or less stationary over this duration. For longer signals, the last section of the paper gives some hints on the time adaptation of the MMSE FB’s, according to the second order statistics of the input signal. Moreover, in [1] it has been shown that the improvement in the SNR is kept when using a sequence of 13 s for the estimation of the autocorrelation matrix  $\mathbf{R}_{yy}$  describing now the long term characteristics of the input signal. The same qualitative results can also be expected in the modulated case. Furthermore, an improvement in the SNR can still be achieved, when modeling the autocorrelation matrix as an AR(2) process or when coding synthetic AR signals [1].

2) *Bit-Rate Allocations:* In Examples 1 and 2, the bit allocations resulting from MMSE optimization differ from their PR counterparts. Indeed, the simulations showed that the latter tends to transmit all subbands (even with a very small bit rate) whereas the MMSE FB’s do not transmit subbands in which the signal variance is small. On the other hand, they encode those with signals of greater variances with increased accuracy since they allocate the whole bit rate to these subbands.

3) *Translation of Rate-Distortion Curves:* Another way of explaining the loss of performance observed with the entropy-constrained optimization is illustrated by the rate-distortion curve translation, see Fig. 7.

Consider two models,  $\sigma_{b_k^\alpha}^2 = \alpha \sigma_{y_k}^2 2^{-2R_k^\alpha}$  and  $\sigma_{b_k^\beta}^2 = \beta \sigma_{y_k}^2 2^{-2R_k^\beta}$ , for the quantization stage. For fixed distortion,  $\sigma_{b_k^\alpha}^2 = \sigma_{b_k^\beta}^2$ , we have  $R_k^\alpha = R_k^\beta + \frac{1}{2} \log_2(\alpha/\beta)$ . Thus, the rates differ by a constant  $C = \frac{1}{2} \log_2(\alpha/\beta)$ . Assume that the same number of subbands,  $M_\alpha^{\text{PR}} = M_\beta^{\text{PR}}$ , are transmitted in the PR case with both models  $\alpha$  and  $\beta$ . Then, the total channel bit rate for the PR case with the  $\alpha$  model is  $R^\alpha = R^\beta + M_\beta^{\text{PR}} C$ , whereas the output SNR remains the same as in the  $\beta$  case. This shows that the rate-distortion curve for model  $\alpha$  is

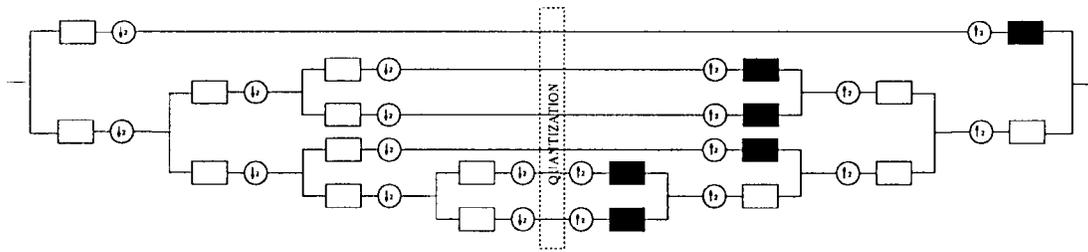


Fig. 8. Example of a 6-band tree-structured filter bank.

obtained by translating the rate-distortion curve corresponding to model  $\beta$  by  $M_{\beta}^{PR} \cdot C$  b/s.

Concerning the MMSE case, we have already noted that the number of transmitted subbands is always smaller in the MMSE case than in the PR case ( $M_{\beta}^{MMSE} < M_{\beta}^{PR}$ ). Again, the rate-distortion curve for model  $\alpha$  is obtained by translating the  $\beta$  curve by  $M_{\beta}^{MMSE} \cdot C$  b/s. However, this is less than in the PR case, and, in consequence,  $\alpha$ -curves for PR and MMSE get closer together than  $\beta$  ones; the improvement brought by MMSE optimization in the  $\alpha$  case is reduced. Thus, for the same MMSE filters and bit allocations, we notice a different improvement, brought by MMSE solutions over PR, when changing the noise model.

#### IV. MMSE TREE-STRUCTURED FILTER BANKS

##### A. Problem Statement

In the previous sections, a parallel  $M$ -band filter bank was optimized according to an MMSE criterion. However, in source coding applications also other subband decompositions are of interest. In particular, tree-structured filter banks are now widely studied for image and audio coding. For example, in [17], the decomposition mimics the Bark scale used for audio coding purposes. Unfortunately, such filters do not fit in the previous formulation of the reconstruction distortion.

In this section, the algorithms for finding optimal synthesis filters and subband bit rate allocations are adapted to the general case of tree-structured banks. We assume that the analysis tree can be built by splitting any branch (or subband) with any 2-band PR cell, thus allowing different cells at different levels. This yields the maximum number of adaptable parameters for a coding scheme designer and includes all possible divisions of the frequency axis. Moreover, recent studies [18] have shown the usefulness of these additional degrees of freedom. An example of a 6-band tree-structured filter bank is given in Fig. 8.

Various strategies can be developed in order to carry out an MMSE optimization in a tree-structured filter bank. A first idea would consist in optimizing each filter of the synthesis tree jointly with all quantizers. However, the method presented in Section II for the parallel case cannot be directly applied to tree-structured filter banks, because the distortion is no longer a quadratic function of the cascaded synthesis filters. Therefore, let us consider our problem from a more intuitive point of view, and let us look at the significant degrees of freedom.

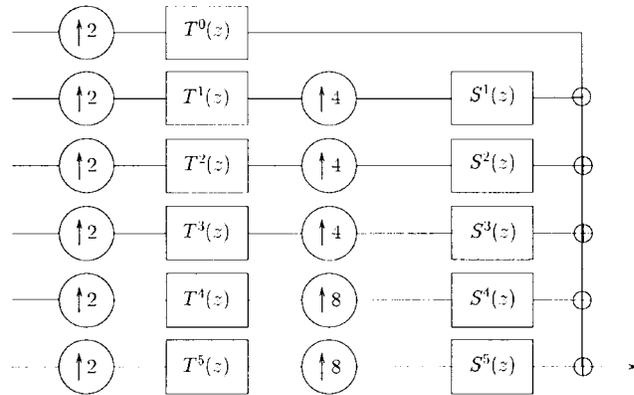


Fig. 9. Equivalent synthesis for the 6-band tree-structured filter bank of Fig. 8.

As it has already been mentioned at the end of Section II, MMSE filters shape the quantization noise. In a tree-structured filter bank, many synthesis filters act on the same frequency band. Therefore, at least in the case of infinite-length filters, optimizing the whole synthesis bank would be redundant. In the case of FIR filters, convergence difficulties are expectable, since any variation of one cell can be compensated by a variation of another cell. In order to shape the quantization noise, we have chosen to optimize a single filter per subband, actually the filter following the quantization/dequantization stage. Two cases occur.

- 1) If the filters are long enough, the resulting degrees of freedom should be sufficient to shape the quantization noise in each band and to obtain good suboptimal solutions. This is described in Fig. 9.
- 2) In some cases, when the filters are very short [18], the number of degrees of freedom may not be sufficient for performing noise shaping. The solution we have chosen is to tune an increased number of parameters: instead of optimizing only the first filter in each branch of the synthesis tree, 2, 3, or more filters in one path are gathered into one equivalent filter of increased length. It is optimized in the same way as before. This is depicted in Fig. 10.

In both cases, we merge all filters that are not tuned into an equivalent one for each branch, say  $S^k$ , the “filter” to be tuned with an MMSE criterion being denoted by  $T^k$ .

The main advantage of dealing with tree-structured filter banks in this way is to enable the use of the general

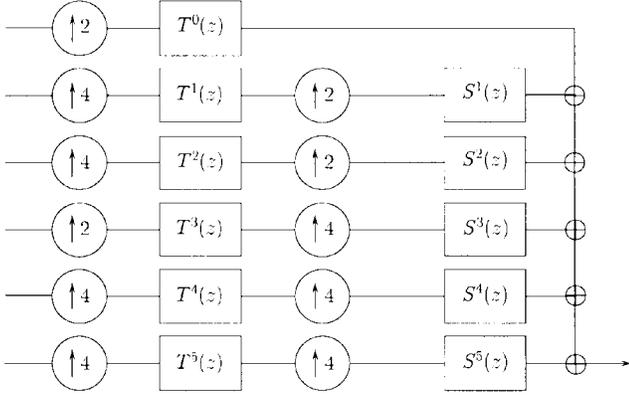


Fig. 10. Equivalent synthesis for the 6-band tree-structured filter bank of Fig. 8.

procedure. The optimization carried out in Section II relies on a distortion function being quadratic with respect to the synthesis coefficients, as well as on a specific relation between distortion and subband bit rates. Both characteristics remain in our formulation of the tree-structured case. Moreover, the computational cost of the optimization remains reasonable.

### B. MMSE Optimization of Tree-Structured Filter Banks

The main difference compared to the initial description is the appearance of an intermediate sampling factor, in branch  $k$ :  $I_k$ . Let  $J_k$  be the total up-sampling factor in subband  $k$ . The quantized signal  $\hat{y}^k$  in the same subband is first up-sampled by a factor  $J_k/I_k$  and filtered by the MMSE synthesis filter  $T^k$  (to be optimized). It is then up-sampled by  $I_k$  and filtered by  $S^k$ , which remains fixed. Thus, the output signal is

$$\hat{X}(z) = \sum_{k=0}^{M-1} S^k(z) T^k(z^{I_k}) \hat{Y}^k(z^{J_k}). \quad (20)$$

Now assume that the analysis bank has been chosen such that in absence of subband quantization it forms a PR system with the synthesis filters  $T^{*k} \cdot S^k$  (this can be easily obtained with 2-band PR cells). As in Section II, the delayed input signal, denoted as  $\check{X}(z) = z^{-n_0} X(z)$ , can be written equivalently as the filtering of the unquantized subband signals  $Y^k(z)$  by the PR synthesis filters  $T^{*k} \cdot S^k$ . The reconstruction error is then

$$\begin{aligned} \hat{X}(z) - \check{X}(z) &= \sum_{k=0}^{M-1} S^k(z) [T^k(z^{I_k}) \hat{Y}^k(z^{J_k}) - T^{*k}(z^{I_k}) Y^k(z^{J_k})]. \end{aligned} \quad (21)$$

Here, all assumptions made in Section II on the quantization noise model are still valid (additive, white, and uniform process). In time domain, the reconstruction error becomes

$$\hat{x}_n - \check{x}_n = \sum_{k,i,j} s_{n-iJ_k-jI_k}^k [-t_j^k v_i^k + (t_j^k - t_j^{*k}) y_i^k]. \quad (22)$$

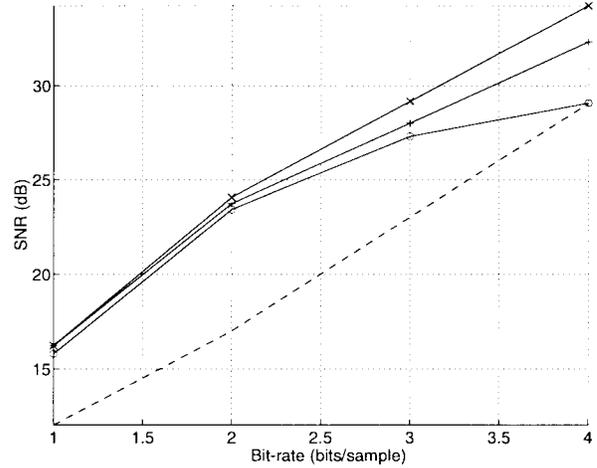


Fig. 11. Measured SNR (dB) versus bit rate for tree-structured filter bank in Example 4 (curve -o-: optimization of 8 parameters per subband; curve +-: optimization of 22 parameters per subband; curve -x-: optimization of 50 parameters per subband). The dashed line corresponds to the PR case.

The output signal is cyclostationary of period  $J = \max_k J_k$ , and the overall distortion is thus given by

$$\begin{aligned} D &= \mathcal{E} \left\{ \frac{1}{J} \sum_{n=0}^{J-1} |\hat{x}_n - \check{x}_n|^2 \right\} \\ &= \frac{1}{J} \sum_k \frac{J}{J_k} \mathcal{E} \{ |v^k|^2 \} \|s^k * t_{\uparrow I}^k\|^2 \\ &\quad + \frac{1}{J} \sum_{n,k_1,k_2,i_1,i_2,j_1,j_2} \mathcal{E} \{ y_{i_1}^{k_1} y_{i_2}^{k_2} \} \\ &\quad \cdot s_{n-i_1 J_{k_1}-j_1 I_{k_1}}^{k_1} s_{n-i_2 J_{k_2}-j_2 I_{k_2}}^{k_2} \\ &\quad \cdot (t_{j_1}^{k_1} - t_{j_1}^{*k_1}) (t_{j_2}^{k_2} - t_{j_2}^{*k_2}) \end{aligned} \quad (23)$$

where  $t_{\uparrow I}^k$  denotes “ $t^k$  up-sampled by  $I_k$ .”

The MMSE optimization consists of minimizing the overall distortion under bit rate constraint taking into account the various sampling rates:  $\sum_k (R_k/J_k) = R_T$ . The MMSE algorithm is straightforwardly derived from the parallel case.

### C. Simulations

We compare, by means of rate-distortion curves, tree-structured PR schemes to tree-structured MMSE ones. In both cases, the quantizers are optimized.

Our MMSE formulation allows us to choose the number of synthesis coefficients to be tuned since one degree of freedom is the number of filtering stages to be optimized. Increasing this number of parameters improves the rate-distortion performance of the corresponding scheme, but it also increases the optimization complexity, while the implementation complexity remains almost identical. The curves presented here highlight the tradeoffs between performance and optimization complexity. The simulation context is similar to the one of the modulated case described in Section III-C.

*Example 4:* Fig. 11 shows the SNR improvement brought by the MMSE optimization of a full tree of depth 3,  $M = 8$  subbands, with Daubechies filters of length  $N = 8$ , and “Vivaldi” signal.

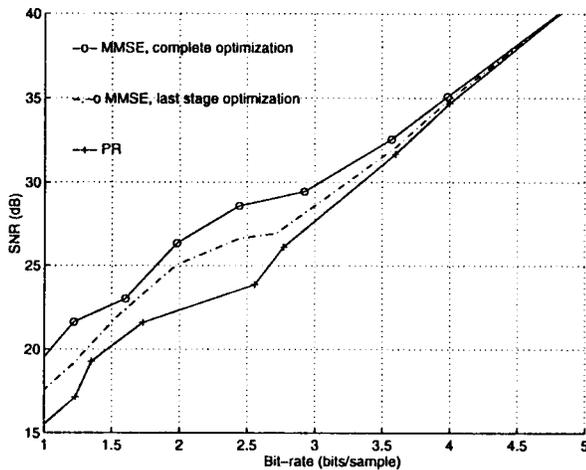


Fig. 12. Measured SNR (dB) versus bit rate for tree-structured filter bank in Example 5.

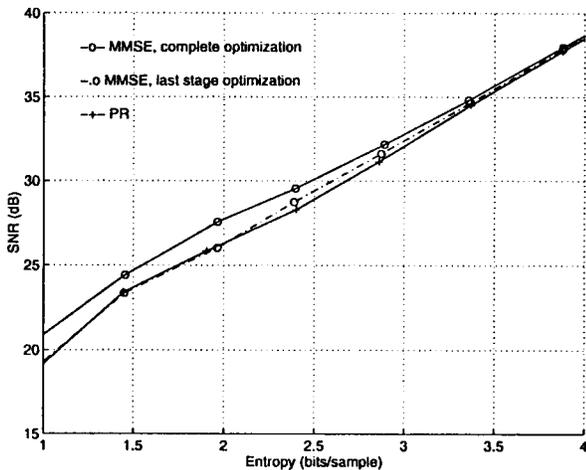


Fig. 13. Measured SNR (dB) versus entropy for tree-structured filter bank in Example 5.

By optimizing the last stage of the tree (8 parameters per subband), the gain brought by the MMSE scheme over the PR one is more than 2 dB in the bit rate range of 1–4 b/s. This can be improved by optimizing 22 degrees of freedom per subband or even the whole equivalent tree (50 parameters per subband). In these cases, the additional SNR improvement mainly occurs at higher bit rates (above 2.5 b/s). This illustrates the tradeoff between complexity and performance. Below 2 b/s, optimizing the last branch of the tree seems to be the best solution (less complex optimization, but comparable rate-distortion performance). Above 2.5 b/s, a system designer can choose the tradeoff according to his/her needs.

*Example 5:* Figs. 12 and 13 show the rate-distortion curves with bit rate constrained and entropy-constrained optimization, respectively, for another tree of depth 6,  $M = 7$  subbands, Daubechies filters of length  $N = 4$ , and an AR(2) model of the “Vivaldi” signal.

As in Example 4, the MMSE optimization improves the output SNR with respect to PR schemes. Again, two performance/complexity tradeoffs are illustrated: 1) the full optimization of the tree and 2) the optimization of the last stage only, bringing less improvement, but at a lower cost. For entropy-constrained optimization a gain of 2 dB is still

achievable in a wide range of bit rates when optimizing the whole synthesis filter bank, see Fig. 13.

The maximal improvement for tree-structured filter banks is of course obtained when optimizing the whole equivalent synthesis bank for  $I_k = 1$  (parallel synthesis bank). If the length of all filters in the tree is  $N$ , for a fully spread tree of depth  $J$  with  $2^J$  subbands, the number of parameters to be optimized is  $(1 + J)[1 + (N - 1)(2^J - 1)]$ . If  $I_k$  is reduced to  $I_k = J_k/2$  (optimization of the last stage in the tree), the number of parameters decreases to  $(1 + J)N$ .

## V. DISCUSSION

The simulations show the flexibility and efficiency of the MMSE approach. However, several questions arise; the most crucial ones are the improvements to be expected when introducing these techniques in a real coding scheme and the way to adapt them to signal statistics variations.

### A. Improvement to be Expected in Practice Thanks to MMSE Solutions

It is easily shown that the MMSE optimization consists of making the reconstructed signal and the reconstruction error uncorrelated. If the analysis filter bank is orthogonal and ideally decorrelates the input signal, and if the quantization is optimal, which implies that the quantization noise and the quantized signal are uncorrelated, then the perfect reconstructing synthesis filter bank is optimal up to a multiplicative coefficient per subband. This point partly explains why SNR improvements vary depending whether the bit rate or the order 1 entropy is estimated: with entropy coding, the optimal bit allocation leads to less correlation in the PR case and thus reduces the SNR improvement. It complements the analysis of Section III-D based on the translation of the rate-distortion curves while switching from bit rate to entropy. The gain to be expected in practical situations lies in between these two extreme situations.

Note that the use of subband coding for audio or video coding should also take into account perceptual characteristics of the human hearing or vision, considering a perceptual distortion criterion rather than the pure MSE. Work in this direction is complementary to the one presented here, since the schemes we describe may also yield interesting performance/complexity tradeoffs if optimized according to different perceptual criteria. Suggestions to the choice of quality criteria for audio coding are given in [19] and [20] where the MSE is replaced by a perceptual Mean Weighted Error (MWE) and brought a noticeable improvement over the reference PR system.

### B. Possible Approaches for Time Adaptation

Whether and how the MMSE filters should vary with the signal statistics is an open issue. There are actually two limiting factors concerning this problem: the coder complexity, and the necessary bit rate for the transmission of the corresponding side information. We list below possible ways of addressing this issue.

- Optimize the MMSE FB with reference to an “average” situation (low cost, but will such a solution be still efficient?).

- Use a library of MMSE synthesis filters, known at both the transmitter and receiver. The adaptation would then consist in choosing the right filter bank in the library. For instance, an AR(2) model of the signals could help in this choice.
- Really adapt the filters to the signal statistics. The question now is about the length of the signal segments (not too long, not too short).
- The fact that the performance is less sensitive to the change of the filters than to the changes of the bit rate allocation can be taken into account in the approaches above. Thus, the filters could be tuned seldom, while the bit rate allocation should be changed much more frequently.
- Joint optimization of synthesis filters and bit allocation in the Kalman filtering framework. A Kalman filter is a linear MMSE estimator [21]. It allows us to compute an estimate of a signal  $x_n$  thanks to a state equation and an observation equation. The first one describes  $x_n$  recursively as a Gauss–Markov process. The second one relates  $x_n$  to the observed signal  $y_n$ . As shown in [22], a Kalman filter can be used for optimal signal reconstruction for a given bit rate allocation. However, there are two main differences with the MMSE approach: The first one is that Kalman filtering does not consist of a joint optimization of synthesis filters and bit rate allocation. The second point is that the statistical properties of the signal are used through correlation matrices in the MMSE formulation and through a prediction model in the Kalman formulation.

## VI. CONCLUSION

We have shown in this paper how to adapt MMSE optimization procedures for source coding schemes to the special cases of modulated and tree-structured filter banks. In both cases, noticeable SNR improvements are obtained, even with a reduced number of tuned parameters in comparison to the general parallel MMSE filter bank [1]. This allows a reduction of the decoding and optimization complexity as well as a reduction of the number of parameters to be transmitted. For MMSE implementations with preliminary off-line optimization, the work presented in this paper provides complexity reduction of the decoder. For implementations with on-line optimizations, it reduces the cost of the synthesis filter adaptation to variations of the input signal statistics (in bits to be transmitted and/or in optimization complexity). This paper thus proposes several performance-complexity tradeoffs, which can bring significant SNR improvement at low cost.

Moreover, we have presented a thorough discussion on how the MMSE optimization works and on its limits. The analysis emphasizes the influence of the analysis bank, as well as the quantization and coding methods, on the overall performance of MMSE schemes.

Further work in this direction would require us to focus on a given application, and to design a full coding scheme, solving the adaptation problem and taking into account perceptual characteristics of human receptors.

## REFERENCES

- [1] K. Gosse and P. Duhamel, "Perfect reconstruction versus MMSE filter banks in source coding," *IEEE Trans. Signal Processing*, vol. 45, pp. 2188–2202, Sept. 1997.
- [2] A. Dembo and D. Malah, "Statistical design of analysis/synthesis systems with quantization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 328–341, Mar. 1988.
- [3] P. P. Vaidyanathan and T. Chen, "Statistically optimal synthesis banks for subband coders," in *Proc. Asilomar Conf. Signals, Syst., Computers*, Nov. 1994.
- [4] R. Haddad and K. Park, "Modeling, analysis, and optimum design of quantized  $M$ -band filter banks," *IEEE Trans. Signal Processing*, vol. 43, pp. 2540–2549, Nov. 1995.
- [5] N. Uzun and R. A. Haddad, "Cyclostationary modeling, analysis, and optimal compensation of quantization errors in subband coders," *IEEE Trans. Signal Processing*, vol. 43, pp. 2109–2119, Sept. 1995.
- [6] J. Kovačević, "Subband coding systems incorporating quantizer models," *IEEE Trans. Image Processing*, vol. 4, pp. 543–553, May 1995.
- [7] A. Delopoulos and S. Kollias, "Optimal filter banks for signal reconstruction from noisy subband components," *IEEE Trans. Signal Processing*, vol. 44, pp. 212–224, Feb. 1996.
- [8] T. Karp, K. Gosse, A. Mertins, and P. Duhamel, "MMSE filter banks with reduced complexity," in *Proc. EUSIPCO (European Signal Processing Conf.)*, Trieste, Italy, Sept. 1996.
- [9] K. Gosse, T. Karp, P. Duhamel, and A. Mertins, "Modulated filter banks with minimum output distortion in presence of subband quantization," in *Asilomar Conf. Signals, Syst. Computers*, Pacific Grove, CA, Nov. 1996.
- [10] K. Gosse, F. Moreau de Saint-Martin, and P. Duhamel, "Filter bank design for minimum distortion in presence of subband quantization," in *Proc. IEEE ICASSP*, 1996.
- [11] A. B. Sripad and D. L. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Trans. ASSP*, vol. 25, pp. 442–448, Oct. 1977.
- [12] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [13] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Trans. Inform. Theory*, vol. 22, pp. 162–169, Mar. 1976.
- [14] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [15] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, MA: Artech House, 1992.
- [16] T. Q. Nguyen, "Digital filter bank design: Quadratic-constrained formulation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2103–2108, Sept. 1995.
- [17] D. Sinha and A. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Trans. Signal Processing*, vol. 41, pp. 3463–3479, Dec. 1993.
- [18] P. Philippe, F. Moreau de Saint-Martin, M. Lever, and J. Soumagne, "Optimal wavelet packets for low-delay audio coding," in *Proc. IEEE ICASSP*, 1996.
- [19] K. Gosse, O. Pothier, and P. Duhamel, "Optimizing the synthesis filter bank in audio coding for minimum distortion using a frequency weighted psychoacoustic criterion," presented at the *IEEE ASSP Workshop Appl. Signal Processing Audio Acoust.*, New Paltz, NY, Oct. 1995.
- [20] K. Gosse, F. Moreau de Saint-Martin, X. Durot, P. Duhamel, and J. B. Rault, "Subband audio coding with synthesis filters minimizing a perceptual criterion," in *Proc. IEEE ICASSP*, 1997.
- [21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [22] B. S. Chen, C. W. Lin, and Y.-L. Chen, "Optimal signal reconstruction in noisy filter bank systems: Multirate synthesis filtering approach," *IEEE Trans. Signal Processing*, vol. 43, pp. 2496–2504, Nov. 1995.



**Karine Gosse** (M'96) was born in Grenoble, France, on May 18, 1971. She graduated from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1993, where she also received the Ph.D. degree in 1996.

She is currently working for the Centre de Recherche de Motorola, Paris, France, as a Research Engineer on digital communication systems. Her research interests include multirate filtering and source coding, especially audio coding in subbands.



**Tanja Karp** (M'98) was born in Germany in 1969. She received the Dipl.-Ing. degree in electrical engineering and the Dr.-Ing. degree from Hamburg University of Technology, Germany, in 1993 and 1997, respectively.

Since 1997, she has been with Mannheim University as a Research and Teaching Assistant. Her research interests include multirate signal processing, filter banks, audio coding, images coding, and signal processing for communications.



**François Moreau de Saint-Martin** was born in Paris, France, in 1968. He graduated from Ecole Polytechnique, France, in 1991 and Ecole Nationale Supérieure des Télécommunications in 1993, and received the Ph.D. degree from the University of Paris IX Dauphine, France, in 1997.

From 1993 to 1996, he was a member of the Technical Staff at Centre Commun d'Etudes de Télécommunications (CCETT), a research center of the France Telecom Group, Cesson-Sévigné, France.

His work focused on filter banks and wavelets, with application to image compression, region-based image compression, audio compression, and transmission systems. Since late 1996, he has been with the Ministère de l'Economie, des Finances et de l'Industrie, Paris, France, where he is now involved in R&D programs and technical regulation in the areas of audiovisual communication, multimedia, and consumer electronics.

**Pierre Duhamel** (M'87–SM'87–F'98) was born in France in 1953. He received the Ing. degree in electrical engineering from the National Institute for Applied Sciences (INSA), Rennes, France, in 1975, and the Dr.Ing. and Doctorat ès sciences degrees in 1978 and 1986, respectively, both from Orsay University, Orsay, France.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive CCD filters. Later, he worked on fast Fourier transforms and convolution algorithms, and applied similar techniques to adaptive filtering, spectral analysis, and wavelet transforms. Since June 1993, he was a Professor at Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France. He was appointed Head of the Signal Processing Department in 1996. He is now developing studies in channel equalization, including multicarrier systems, and source coding, including joint source/channel coding.

Dr. Duhamel is Chairman of the DSP committee, was Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, and was Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He was a Guest Editor for the special issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING on wavelets.



**Alfred Mertins** (M'95) received the Dipl.Ing. degree from the University of Paderborn, Germany, in 1984, and both the Dr.Ing. degree in electrical engineering and the Dr.Ing. habil. degree in telecommunications from the Hamburg University of Technology, Germany, in 1991 and 1994, respectively.

From 1986 to 1991 he was a Research Associate in the Telecommunications Institute, Hamburg University of Technology, Germany. From 1991 to 1995 he was with the Microelectronics Applications Center, Hamburg, Germany, as a Senior Research Associate. From 1996 to 1997 he was a Senior Lecturer at the University of Kiel, Germany, where he gave courses on telecommunications, stochastic signal analysis, multirate digital signal processing, and wavelets. Since October 1997, he has been with the Visual Communications Research Group, Department of Electrical and Electronic Engineering, University of Western Australia, as a Senior Research Fellow. His research interests include image and video processing, wavelets and filter banks, joint source-channel coding, and digital communications.