Low Delay Filter Banks with Perfect Reconstruction

Filterbänke mit niedriger Systemverzögerungszeit und exakter Rekonstruktion

Von Gerald Schuller

*Beitrag aus dem Institut für Theoretische Nachrichtentechnik und Informations-
verarbeitung der UNIVERSITÄT HANNOVER*

gewidmet Herrn Professor Dr.-Ing. Peter Noll zum 60. Geburtstag

Abstract:

The design of modulated filter banks with a low system delay and with perfect reconstruction will be shown. The filter lengths K can be chosen arbitrarily. The well known orthogonal filter banks have a system delay of $K - 1$ samples. The proposed filter banks can reduce this delay to $N - 1$ samples, where N is the number of bands. The design method uses a decomposition or factorization of the polyphase matrix into cascades of simple matrices. Several factorizations with different properties will be shown. A factorization will be introduced which is more general and needs fewer multiplications than previous approaches ($K/2 + N$). The resulting filter banks can have analysis and synthesis frequency responses that can be made different from each other, leading to biorthogonal filter banks. An optimization algorithm for the frequency response of the resulting filter banks will be given. Examples show the feasibility of designing even big filter banks with many bands with low system delay and high stopband attenuation.

Übersicht:

Ein Verfahren zur Konstruktion von modulierten Filterbänken mit kurzer Verzögerungszeit und mit exakter Rekonstruktion wird vorgestellt. Die Filterlänge K kann beliebig gewählt werden. Im Fall der bekannten orthogonalen Filterbänke beträgt die Systemverzögerung $K - 1$ Abtastwerte. Die hier vorgeschlagenen Filterbänke können diese Verzögerungszeit auf $N - 1$ Abtastwerte reduzieren, wobei N die Anzahl der Teilbänder ist. Das Verfahren zur Konstruktion der Filter basiert auf einer Zerlegung oder Faktorisierung der sogenannten Polyphasen-Matrix in Kaskaden einfacherer Matrizen. Mehrere Faktorisierungen mit unterschiedlichen Eigenschaften werden hergeleitet. Es wird eine Faktorisierung gezeigt, welche weniger Multiplikationen als frühere Ansätze benötigt ($K/2 + N$). Die resultierenden Filterbänke können Analyse- und Synthese-Frequenzgänge haben, die voneinander verschieden sind, sogenannten biorthogonale Filterbänke. Ein Optimierungs-Algorithmus für die Frequenzgänge der resultierenden Filterbänke wird vorgestellt. Beispiele zeigen, daß es möglich ist, lselbst große Filterbänke mit vielen Teilbandfiltern mit niedriger Systemverzögerung und hoher Sperrbereichs-Dämpfung zu konstruieren.

Für die Dokumentation:

Filterbänke / Multirate / Digitale Signalverarbeitung / Vorecho-Unterdrückung / Subbandcodierung / Echt-
zeit-Systeme

1. Introduction

Filter banks are used in a variety of applications, like in audio, speech, or video coding, in acoustic echo cancellation, Frequency Division Multiplex and more. They consist of a bank of analysis filters with subsequent downsampling, whose output is used e.g. for coding or manipulating the signal, and a bank of upsamplers followed by synthesis filters, which can reconstruct the signal from the analysis outputs (**Fig. 1**) to various degrees of precision.

It is usually desired that the filters are narrow bandpass filters with little overlap in the frequency domain. For coding applications it is important to have a critical downsampling, i.e. the overall data rate is not increased by the analysis filter bank. E.g. if the same downsampling rate is used for all bands, it should be N, if N bands are used.

Perfect reconstruction filter banks are such that the synthesis filter bank reconstructs the input to the analysis filter bank from its output exact or perfectly, despite the analysis filtering and downsampling. It can be achieved with up to a critical downsampling rate, but in general not at higher downsampling rates. Perfect reconstruction could be obtained by choosing ideal band filters for the analysis and syn-

thesis filter bank. They would be sampled at their Nyquist rate and since they would cover the entire frequency range, perfect reconstruction would result. Ideal band filters are not realizable, but perfect reconstruction can be achieved with realizable filters by designing them suitably, such that the synthesis cancels any aliasing or amplitude distortions which the analysis introduces, so that $\hat{x}(n) = x(n - n_d)$, where n_d is the system delay in samples.

Modulated filter banks have filters with frequency responses which are frequency shifted versions of baseband prototypes, one for the analysis and one for the synthesis filters. They result from multiplying or modulating the baseband impulse response with a modulating function, e.g. a cosine function, which leads to cosine modulated filter banks.
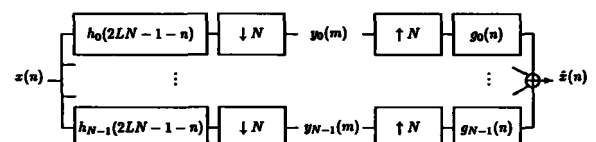

Figure 1: An N-channel filter bank with critical down-
sampling

Modulated filter banks are popular because they usually have a computationally efficient way for their implementation and they are simple to design, because only the two prototypes need to be designed and evaluated.

An important property of filter banks is their overall system delay (n_d), which is the lag between the input of the analysis filter bank and the output of the synthesis filter bank if they are connected directly. Until recently the filter banks investigated and used had a standard system delay, which is the length of the band filters of the filter bank (or the length of the prototype), minus one sample. This means that filter banks with a lower system delay needed to have short filters, resulting in a poor frequency response, or vice versa, filter banks with good frequency responses, close to some desired ideal, needed to be long, resulting in a high system delay. Low delay filter banks don't have this connection between filter length and system delay, they can have a low system delay and also long filters, resulting in good frequency responses. Application examples are real time speech and audio coding systems. In audio coding systems they can also be used for the reduction of so called pre-echoes, audible distortions, caused by the system delay of the used filter bank.

Low delay filter banks were first investigated by Nayebi [15, 16, 17], who showed the possibility of designing low delay filter banks. He treated the general case but had no perfect reconstruction. Later work of the author showed that it is also possible to design perfect reconstruction low delay filter banks [23-26]. This paper will show how to design modulated filter banks (especially with a cosine modulation function) with perfect reconstruction, with variable system delays (ranging from the minimum to the maximum possible delay for a given filter length), and with a structure for a computational efficient implementation.

2. Polyphase Formulation

Fig. 1 shows a filter bank with N bands and critical downsampling. The analysis filtering and downsampling operation can be written as

$$y_k(m) = \sum_{n=0}^{LN-1} x(mN+n)h_k(n) \tag{1}$$

for all integer m, where $y_k(m)$ is the output of the k'th subband channel at the m'th interval. $2LN$ is the length of the filters. The impulse responses don't need to be actually of this size, they can be zero padded. Observe that $h_k(n)$ is the time reversed analysis filter impulse response. h_k can be seen as a "filter vector", because the filtering or convolution operation now resembles a vector multiplication. This has the advantage that it better fits the matrix formulation which will be introduced later. The filter bank may be viewed as processing the input in blocks. For every block of N samples, where m may be viewed as the block index for $x(mN+n)$, N output samples in the variable k are produced: $y_k(m)$; $k = 0, 1, ..., N - 1$. A more convenient way for the representation of this system is the polyphase formulation [6].

The input is grouped into consecutive blocks of length N which is written as a vector $x(m)$

$$x(m) = [x(mN), ..., x(mN + N - 1)].$$

Now a z-transform can be performed on the vector

$$X(z) = \left[\sum_{m=0}^{\infty} x(mN) \cdot z^{-m}, ..., \sum_{m=0}^{\infty} x(mN + N - 1) \cdot z^{-m} \right].$$

The output of the analysis filter bank is also written as a vector

$$y(m) = [y_0(m), ..., y_{N-1}(m)]$$

and z-transformed

$$Y(z) = \left[\sum_{m=0}^{\infty} y_0(m) \cdot z^{-m}, ..., \sum_{m=0}^{\infty} y_{N-1}(m) \cdot z^{-m} \right].$$

This has the advantage that the analysis filtering and downsampling operation can be expressed very conveniently, if we define the polyphase matrix as

$$[P_a(z)]_{n,k} = \sum_{m=0}^{2L-1} h_k(mN + n) z^{-(2L-1-m)}$$

$n, k = 0, ..., N - 1$, where $[.]_{n,k}$ denotes the element at the n'th row and k'th column, or k is the frequency index and n the time index. The analysis filtering and downsampling operation now is

$$Y(z) = X(z) \cdot P_a(z).$$

The synthesis upsampling and filtering operation can be expressed in the same way. The synthesis polyphase matrix is

$$[P_s(z)]_{k,n} = \sum_{m=0}^{2L-1} g_k(mN + n) z^{-m}$$

so that the output of the synthesis filter bank can be written as

$$\hat{X}(z) = Y(z) \cdot P_s(z).$$

The analysis and synthesis filter bank can now be seen as just a multiplication with square matrices with polynomial elements, as illustrated in **Fig. 2**. Perfect reconstruction requires the product of $P_a(z)$ and $P_s(z)$ to be just a delay, which means the synthesis polyphase matrix for a given analysis polyphase matrix is

$$P_s(z) = P_a(z)^{-1} \cdot z^{-d}$$

where d is the delay of the signal in blocks of length N to make the synthesis filters causal, i. e. such that $P_s(z)$ has no positive powers of z. The system delay now consists of this delay of d blocks to make the polyphase matrix causal, and of the so called blocking delay ($N - 1$ samples), which results from the grouping of the input samples into blocks of length N. The system delay therefore is $n_d = d \cdot N + N - 1$.

3. Modulated Filter Banks

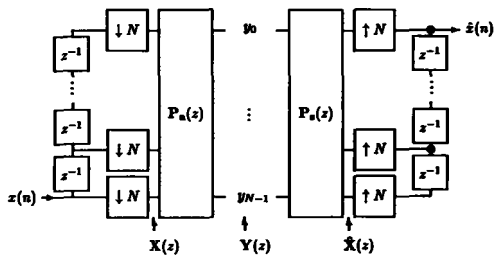Modulated filter banks have certain symmetries in their impulse responses that can be used for the design and effi-

Figure 2: A polyphase representation of an N-channel filter
bank with critical downsampling

cient implementation of the filter bank. The important point
here is that their polyphase matrices can be written as a pro-
duct of a sparse "filter matrix" with polynomial elements
($F_a(z)$, $F_s(z)$), and a transform matrix (T) with real or com-
plex elements. Both can be implemented efficiently. This
means the polyphase matrices can be written as

$$P_a(z) = F_a(z) \cdot T$$
$$P_s(z) = T^{-1} \cdot F_s(z).$$

The form of $F_a(z)$ and $F_s(z)$ is obtained by computing the
expression $F_a(z) = P_a(z) \cdot T^{-1}$ and $F_s(z) = T \cdot P_s(z)$. Assume
e. g. a modulated filter bank like

$$h_k(n) = h(n) \cdot \cos\left(\frac{\pi}{N}(k+0.5)(n+0.5+n_0)\right)$$

$$g_k(n) = h'(n) \cdot \frac{2}{N} \cdot \cos\left(\frac{\pi}{N}(k+0.5)(n+0.5+n_0)\right)$$

$$k = 0,...,N-1,$$

$$n = 0,...,2LN-1.$$

The factor $2/N$ is just a normalization. $h(n)$ and $h'(n)$ are
the analysis and synthesis baseband prototype filters res-
pectively. They are usually lowpass filters. Now take a
Discrete Cosine Transformation T type 4, which is defined
as

$$[T]_{n,k} = \cos\left(\frac{\pi}{N}(k+0.5)(n+0.5)\right), \quad 0 \le n,k \le N.$$

This type of transform is suitable in this case, because it
fits the modulation function. For $n_0 = -N/2$ now observe that

$$F_a = \begin{bmatrix} 0 & P_0(z)z^{-1} & P_N(z) & & 0 \\ & & \cdot\cdot & \cdot\cdot & \\ P_{N/2-1}(z)z^{-1} & & 0 & & P_{N+N/2-1}(z) \\ P_{N/2}(z)z^{-1} & & & & -P_{N+N/2}(z) \\ & \cdot\cdot & & \cdot\cdot & \\ 0 & P_{N-1}(z)z^{-1} & -P_{2N-1}(z) & & 0 \end{bmatrix}.$$

This is a filter matrix with a *diamond* structure, where

$$P_k(z) = \sum_{m=0}^{L-1} h(m2N+k)(-1)^m z^{-2(L-1-m)}. \tag{2}$$

The synthesis filter matrix $F_s(z)$ has the same structure as
$F_a(z)$

$$F_s = \begin{bmatrix} & & P'_{N/2-1}(z) & P'_{N/2}(z) & & & \\ & \cdot\cdot & & & & \cdot\cdot & \\ P'_0(z) & & & & & & P'_{N-1}(z) \\ P'_N(z)z^{-1} & & & & & & -P'_{2N-1}(z)z^{-1} \\ & \cdot\cdot & & & & \cdot\cdot & \\ & & P'_{N+N/2-1}(z)z^{-1} & -P'_{N+N/2}(z)z^{-1} & & & \end{bmatrix}$$

with

$$P'_k(z) = \sum_{m=0}^{\infty} h'(m2N+k)(-1)^m z^{-2m}. \tag{3}$$

For the case of $n_0 = 0$ the following form is obtained,

$$F_a = \begin{bmatrix} P_0(z)z^{-1} & & 0 & & -P_N(z) \\ & \cdot\cdot & & \cdot\cdot & \\ & & P_{N/2-1}(z)z^{-1} & -P_{N+N/2-1}(z) & & 0 \\ 0 & & -P_{N+N/2}(z) & P_{N/2}(z)z-1 & \\ & \cdot\cdot & & & \cdot\cdot \\ -P_{2N-1}(z) & & 0 & & P_{N-1}(z)z^{-1} \end{bmatrix}$$

with $P_k(z)$ as defined in equation (2). This is a Filter Matrix
with a *bi-diagonal* structure. The synthesis filter matrix is,

$$F_s = \begin{bmatrix} P'_0(z) & & & & -P'_{2N-1}(z)z^{-1} \\ & \cdot\cdot & & \cdot\cdot & \\ & & P'_{N/2-1}(z) & -P'_{N+N/2}(z)z^{-1} & \\ & & -P'_{N+N/2-1}(z)z^{-1} & P'_{N/2}(z) & \\ & \cdot\cdot & & & \cdot\cdot \\ -P'_N(z)z^{-1} & & & & P'_{N-1}(z) \end{bmatrix}$$

More filter matrices can be computed, e. g. for other n_0
or for other modulation functions, like for DCT's of diffe-
rent types. The design method which will be described is
capable of designing filter banks with more different n_0 than
shown in the preceding two examples and also for different
modulation functions. But to illustrate the method the shown
cosine modulation type will be used in the following (DCT 4).

These filter matrices could now already be used for the
design of filter banks. To perfectly reconstruct a signal from
a given analysis filter bank the synthesis filter matrix needs
to be the inverse of the analysis filter matrix, multiplied with
a delay z^{-d} to make it causal. But this approach may lead to
IIR synthesis filters, which may not be stable. There would
also be no direct control over the system delay, which is deter-
mined by the additional delay z^{-d}. The goal is now to obtain
FIR analysis and also FIR synthesis filters with the perfect
reconstruction property, to have control over the overall
system delay, and to obtain a structure for an efficient imple-
mentation. This is done by constructing the filter matrices
as a product of several simpler matrices. The simple matri-
ces have an inverse, which is FIR, have different system
delays associated with them, and are sparse with only a few
elements unequal to 1 or 0, which leads to an efficient
implementation.

The design process then consists of choosing the matri-
ces for the desired properties (system delay, filter length) and
then to optimize the resulting product or structure for the
desired frequency response. These simple matrices are
described in the following.

• *Coefficient Matrices* – The first one has a diamond struc-
ture,

$$F := \begin{bmatrix} 0 & & d_0 & d_N & & 0 \\ & \cdot^{\cdot} & & & \cdot^{\cdot} & \\ d_{N/2-1} & & & & & d_{N+N/2-1} \\ d_{N/2} & & & & & d_{N+N/2} \\ & \cdot_{\cdot} & & & \cdot^{\cdot} & \\ 0 & & d_{N-1} & d_{2N-1} & & 0 \end{bmatrix}.$$

Its inverse is easy to compute and has the same structure.

The second one has a bi-diagonal structure,

$$C_i := \begin{bmatrix} c_0^i & & & & & 1 \\ & \cdot_{\cdot} & & & \cdot^{\cdot} & \\ & & c_{N/2-1}^i & 1 & & \\ & & 1 & c_{N/2}^i & & \\ & \cdot^{\cdot} & & & \cdot_{\cdot} & \\ 1 & & & & & c_{N-1}^i \end{bmatrix}$$

with real or complex coefficients. Because these matrices are used in a product or cascade, the columns or rows of all used matrices can be normalized, except for one, which is the F or E_0 matrix (which will be shown). This is why the elements on the anti-diagonal can be 1. Its inverse also has the same structure. Observe that the inverse has no longer 1's on the anti-diagonal, so that the number of multiplications necessary to implement the inverse is higher.

• *Standard Delay Matrix* – It increases the filter length and the system delay.

$$D(z) := \begin{bmatrix} z^{-1} & & & & \\ & \cdot_{\cdot} & & & \\ & & z^{-1} & & 0 \\ 0 & & 1 & & \\ & & & \cdot_{\cdot} & \\ & & & & 1 \end{bmatrix}.$$

Its inverse needs a multiplication by z^{-1} to make it causal.

$$z^{-1} \cdot D(z) := \begin{bmatrix} 1 & & & & \\ & \cdot_{\cdot} & & & \\ & & 1 & & 0 \\ 0 & & z^{-1} & & \\ & & & \cdot_{\cdot} & \\ & & & & z^{-1} \end{bmatrix}.$$

• *Zero-Delay Matrices* – They increase the filter length but not the system delay. The first type is

$$E_0(z) := \begin{bmatrix} 0 & & & & & e_N^0 \\ & \cdot_{\cdot} & & & \cdot^{\cdot} & \\ & & 0 & e_{N+N/2-1}^0 & & \\ & & e_{N+N/2}^0 & e_{N/2}^0 z^{-1} & & \\ & \cdot^{\cdot} & & & \cdot_{\cdot} & \\ e_{2N-1}^0 & & & & & e_{N-1}^0 z^{-1} \end{bmatrix}.$$

Observe that its inverse is causal, so that no multiplication with a delay is necessary.

$$E_0^{-1}(z) := \begin{bmatrix} \hat{e}_0^0 z^{-1} & & & & & \hat{e}_N^0 \\ & \cdot_{\cdot} & & & \cdot^{\cdot} & \\ & & \hat{e}_{N+N/2-1}^0 z^{-1} & \hat{e}_{N+N/2-1}^0 & & \\ & & \hat{e}_{N+N/2}^0 & 0 & & \\ & \cdot^{\cdot} & & & \cdot_{\cdot} & \\ \hat{e}_{2N-1}^0 & & & & & 0 \end{bmatrix}$$

with

$$\hat{e}_j^0 = \frac{-e_{N-1-j}^0}{e_{N+j}^0 e_{2N-1-j}^0}, \quad j = 0...N/2-1,$$

$$\text{and } \hat{e}_{N+j}^0 = \frac{1}{e_{2N-1-j}^0}, \quad j = 0...N-1.$$

For $E_i(z)$, $i > 0$, the elements on the anti-diagonal are 1 (again a normalization). The second type of zero-delay matrices has the same basic structure, but the roles of the matrix and its inverse are switched.

$$G_i(z) := \begin{bmatrix} g_0^i z^{-1} & & & & & 1 \\ & \cdot_{\cdot} & & & \cdot^{\cdot} & \\ & & g_{N/2-1}^i z^{-1} & 1 & & \\ & & 1 & 0 & & \\ & \cdot^{\cdot} & & & \cdot_{\cdot} & \\ 1 & & & & & 0 \end{bmatrix}.$$

Its inverse is

$$G_i^{-1}(z) = \begin{bmatrix} 0 & & & & & 1 \\ & \cdot_{\cdot} & & & \cdot^{\cdot} & \\ & & 0 & 1 & & \\ & & 1 & -g_{N/2-1}^i z^{-1} & & \\ & \cdot^{\cdot} & & & \cdot_{\cdot} & \\ 1 & & & & & -g_0^i z^{-1} \end{bmatrix}.$$

A product of these matrices has to have the shape of one of the filter matrices, i. e. it must have a bi-diagonal or diamond shape and the distribution of the even and odd powers of z must be as in the filter matrices. This ensures that the resulting polyphase matrix leads to a modulated filter bank. The following products or cascades have this property.

$$F_a(z) = \left(\prod_{i=1}^m C_i \cdot D^2(z) \right) \cdot F \cdot D(z) \cdot \left(\prod_{i=1}^n G_i(z) \right) \tag{4}$$

for the analysis filters, where $n \geq 0$, $m \geq 0$ are now constants, and

$$F_s(z) = \left(\prod_{i=0}^{n-1} G_{n-i}^{-1}(z) \right) \cdot D^{-1}(z) \cdot z^{-1}$$

$$\cdot F^{-1} \cdot \left(\prod_{i=0}^{m-1} D^{-2}(z) \cdot z^{-2} \cdot C_{m-i}^{-1} \right) \tag{5}$$

for the synthesis filters. The roles of the analysis and synthesis filter matrices can also be switched, since they are inverse to each other. The parameters m and n can be chosen freely and determine the filter length and the overall system delay. The resulting length of the impulse response is $K = m2N + 2N + nN$, and the delay is $n_d = m2N + 2N - 1$ samples. The number of multiplications necessary for the implementation of the analysis filter matrix is the number of coefficients of the simple matrices which are unequal to 1 or 0, and is $mN + 2N + nN/2$ or $K/2 + N$, and for the synthesis filter bank it is $m2N + 2N + nN/2 = K/2 + N + mN$. **Fig. 3** illustrates this structure.

This formulation represents a very broad class of filter banks, and many well known classes of filter banks are a subset of it. Orthogonal or paraunitary filter banks with standard system delay result if F and C_i are restricted to be orthogonal (i. e. rows or columns are orthogonal) and no zero-delay matrices are used ($n = 0$). This leads to the ELT of Malvar [13, 14]. Note that the ELT does not normalize the C_i matrices to ones on the anti-diagonal, so that it needs more multiplications for the analysis part (as many as the synthesis part). The TDAC filter bank of Princen and Bradley [1] results if only the F and the D matrix are used ($n = 0$ and $m = 0$) and F is restricted to be orthonormal (i. e. $F^{-1} = F^{T}$). If n is set to 0 and the C_i matrices are not restricted to be orthogonal, then so called biorthogonal filter banks with a standard system delay are obtained, and if n is not zero, low delay filter banks are the result.

A filter bank with the lowest possible delay, the minimum system delay, requires a different product. It results if the filter matrices are of the form

$$F_a(z) = \prod_{i=0}^{n-1} E_i(z) \qquad (6)$$

$$F_s(z) = \prod_{i=0}^{n-1} E_{n-1-i}^{-1}(z) \qquad (7)$$

with $n > 0$. The resulting impulse responses have $N/2$ trailing zeros, so that the length of the non-zero part is not an integer multiple of N. The length of the non-zero part of the impulse response of the analysis and synthesis filter bank therefore is $K = nN + 0.5N$. The delay that is left is the blocking delay of $N - 1$ samples, which is the minimum possible delay. The number of multiplications necessary for the implementation of the analysis or the synthesis filter matrix is $N + n \cdot N/2 = K/2 + 0.75N$.

These products lead to delays that are less or equal to the standard delay and to delays of the size of even multiples of N (except for the minimum delay case). A product or cascade with the following matrices is even more general because it can also be used for delays of odd multiples of N and for higher delays than the standard delay. It is also important to note that it needs fewer multiplications for the implementation than the previous cascade. Observe that this also means that this new cascade is also more efficient than previous approaches like the ELT filter banks. This formulation uses Maximum-Delay-Matrices, which can be interpreted as time reversed Zero-Delay-Matrices.

• *Maximum-Delay Matrices* – They increase the filter length, but especially the system delay. The first type is

$$A_i(z) = \begin{bmatrix} 0 & & & & z^{-1} \\ & \ddots & & & \cdot^{\cdot^{\cdot}} \\ & & 0 & z^{-1} & \\ & & z^{-1} & a_{N/2}^i & \\ & \cdot^{\cdot^{\cdot}} & & & \ddots \\ z^{-1} & & & & a_{N-1}^i \end{bmatrix} .$$

The inverse needs a multiplication with z^{-2} to make it causal.

$$z^{-2} \cdot A_i^{-1}(z) = \begin{bmatrix} -a_{N-1}^i & & & & z^{-1} \\ & \ddots & & & \cdot^{\cdot^{\cdot}} \\ & & -a_{N/2}^i & z^{-1} & \\ & & z^{-1} & 0 & \\ & \cdot^{\cdot^{\cdot}} & & & \ddots \\ z^{-1} & & & & 0 \end{bmatrix} .$$

The second type looks similar, but has the roles of the matrix and its inverse switched. The matrix $B_0(z)$ also uses coefficients on the anti-diagonal of the matrix, the other matrices ($B_i(z)$ for $i > 0$) can again be normalized.

$$B_0(z) = \begin{bmatrix} b_0^0 & & & & & b_N^0 z^{-1} \\ & \ddots & & & & \cdot^{\cdot^{\cdot}} \\ & & b_{N/2-1}^0 & b_{N+N/2-1}^0 z^{-1} & & \\ & & b_{N+N/2}^0 z^{-1} & 0 & & \\ & \cdot^{\cdot^{\cdot}} & & & \ddots & \\ b_{2N-1}^0 z^{-1} & & & & & 0 \end{bmatrix}$$

$$z^{-2} \cdot B_0^{-1}(z) = \begin{bmatrix} 0 & & & & & \hat{b}_N^0 z^{-1} \\ & \ddots & & & & \cdot^{\cdot^{\cdot}} \\ & & 0 & \hat{b}_{N+N/2-1}^0 z^{-1} & & \\ & & \hat{b}_{N+N/2-1}^0 z^{-1} & \hat{b}_{N/2}^0 & & \\ & \cdot^{\cdot^{\cdot}} & & & \ddots & \\ \hat{b}_{2N-1}^0 z^{-1} & & & & & \hat{b}_{N-1}^0 \end{bmatrix}$$

X ⟹ $\boxed{C_1 \cdot D^2}$ ⟹ ··· ⟹ $\boxed{C_m \cdot D^2}$ ⟹ $\boxed{F \cdot D}$ ⟹

⟹ $\boxed{G_1}$ ⟹ ··· ⟹ $\boxed{G_n}$ ⟹ $\boxed{T}$ ⟹ Y

Y ⟹ $\boxed{T^{-1}}$ ⟹ $\boxed{G_n^{-1}}$ ⟹ ··· ⟹ $\boxed{G_1^{-1}}$ ⟹

⟹ $\boxed{z^{-1}D^{-1}F^{-1}}$ ⟹ $\boxed{z^{-2}D^{-2}C_m^{-1}}$ ⟹ ··· ⟹ $\boxed{z^{-2}D^{-2}C_1^{-1}}$ ⟹ $z^{-d} \cdot$ X
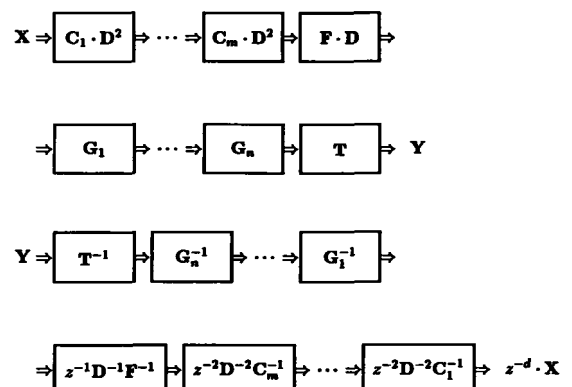
Figure 3: The first structure, the analysis filter bank above, the synthesis filter bank below

with

$$\hat{b}_{N/2+j}^{0} = \frac{-b_{j}^{0}}{e_{N+j}^{0} b_{2N-1-j}^{0}}, \quad j = 0 \ldots N/2 - 1,$$

and $\hat{b}_{N+j}^{0} = \dfrac{1}{b_{2N-1-j}^{0}}, \quad j = 0 \ldots N - 1.$

The coefficients a_j^i and b_j^i can again be real or complex numbers. For $i > 0$ the elements on the anti-diagonal are 1 because of the normalization ($b_j^i = 1$ for $i > 0$ and $j \geq N$).

The following products or cascades result in modulated filter banks with a system delay of the size of even multiples of N

$$F_a(z) = F \cdot D(z) \cdot \left(\prod_{i=1}^{m} A_i(z) \right) \cdot \left(\prod_{i=1}^{n} G_i(z) \right) \qquad (8)$$

if m is even (including 0), and

$$F_a(z) = F \cdot D(z) \cdot \left(\prod_{i=1}^{m} A_i(z) \right) \cdot \left(\prod_{i=1}^{n} E_i(z) \right) \qquad (9)$$

if m is odd ($n \geq 0$, $m \geq 0$). The synthesis filter matrices are

$$F_s(z) = \left(\prod_{i=0}^{n-1} G_{n-i}^{-1}(z) \right) \cdot \left(\prod_{i=0}^{m-1} z^{-2} \cdot A_{m-i}^{-1}(z) \right)$$
$$\cdot D^{-1}(z) \cdot z^{-1} \cdot F^{-1} \qquad (10)$$

for even m, and

$$F_s(z) = \left(\prod_{i=0}^{n-1} E_{n-i}^{-1}(z) \right) \cdot \left(\prod_{i=0}^{m-1} z^{-2} \cdot A_{m-i}^{-1}(z) \right)$$
$$\cdot D^{-1}(z) \cdot z^{-1} \cdot F^{-1} \qquad (11)$$

for odd m.

The resulting filter length is $K = (m + n)N + 2N$, the system delay is $m2N + 2N - 1$, and the number of multiplications necessary for the implementation of the analysis or the synthesis filter matrix is $2N + (n + m) \cdot N/2$ or $K/2 + N$. Observe that for the synthesis filter matrix this number is lower than with the first product or cascade.

For a system delay of the size of odd multiples of N the following product can be used

$$F_a(z) = \left(\prod_{i=0}^{m-1} B_i(z) \right) \cdot \left(\prod_{i=1}^{n} E_i(z) \right) \qquad (12)$$

if m is even and

$$F_a(z) = \left(\prod_{i=0}^{m-1} B_i(z) \right) \cdot \left(\prod_{i=1}^{n} G_i(z) \right) \qquad (13)$$

if m is odd ($n \geq 0$, $m > 0$). The resulting synthesis filter matrices are

$$F_s(z) = \left(\prod_{i=0}^{n-1} E_{n-1}^{-1}(z) \right) \cdot \left(\prod_{i=1}^{m} z^{-2} \cdot B_{m-1}^{-1}(z) \right) \qquad (14)$$

for even m, and

$$F_s(z) = \left(\prod_{i=0}^{n-1} G_{n-1}^{-1}(z) \right) \cdot \left(\prod_{i=0}^{m} z^{-2} \cdot B_{m-i}^{-1}(z) \right) \qquad (15)$$

for odd m. m is greater than 0 because for the minimum delay case cascade (6) and (7) should be used instead. For $n = 0$ this cascade results in filter banks with a maximum delay, which can also be seen as a time reversed minimum delay filter bank. The resulting impulse responses in this last case (delay of odd multiples of N) have $N/2$ leading zeros. If $n > 0$ they also have $N/2$ trailing zeros. The length of the nonzero part of the impulse responses is $K = mN + 0.5N$ if $n = 0$, and $K = (m + n)N$ if $n > 0$, the system delay is $n_d = m2N + N - 1$, and the number of multiplications necessary for the implementation of the analysis or the synthesis filter matrix is $(m + n)N/2 + N$ or $K/2 + 0.75N$ if $n = 0$ and $K/2 + N$ if $n > 0$. The maximum possible delay of this cascade is $n_d = 2K - 1$, which is about twice as much as the standard delay.

The next step of the design procedure now is to determine the unknown matrix coefficients, so that some desired frequency response is approximated. This is done by numeric optimization. Observe that the optimization does not need to take care of the perfect reconstruction property, because it is already guaranteed by the structure.

3.1 Optimization

The goal here is to minimize an error function, which is a nonlinear function with a scalar (real) output and a multidimensional input. Basically any optimization algorithm for the optimization of such functions can be used, but there are differences in convergence and speed. The algorithm described in the following was found to be relatively reliable in convergence and fast.

Define x to be a vector of the s unknown filter matrix entries, which are to be optimized, and $H(x)$ be the weighted frequency responses of the baseband prototype filters or a band filter, at ℓ frequency samples. Analysis and synthesis frequency responses are both contained in this vector, e. g. as a concatenation, so that it has a length of 2ℓ. Let d be the vector of the weighted desired frequency responses at those frequency samples. As the error function a quadratic distance function is chosen. To optimize the magnitude of the frequency response, the following error function is used,

$$f(x) = \sum_{i=1}^{2\ell} \left| |H_i(x)| - d_i \right|^2$$

$$= \sum_{i=1}^{2\ell} \left| H_i(x) - \frac{H_i(x)}{|H_i(x)|} \cdot d_i \right|^2$$

$$= \sum_{i=1}^{2\ell} |H_i(x) - d_i'|^2.$$

The algorithm that is used to minimize this function is basically the so called method of conjugate directions, which was specialized to this quadratic function, so that the convergence and the speed of convergence could be increased. The idea
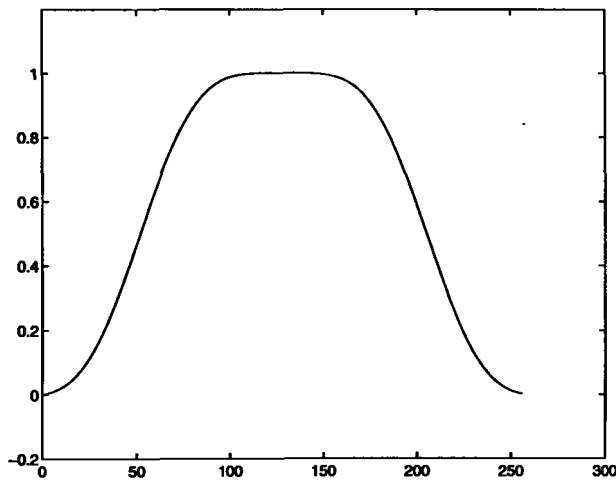
Figure 4: Impulse responses of the baseband prototype for an orthogonal filter bank with 128 bands and 255 samples delay.
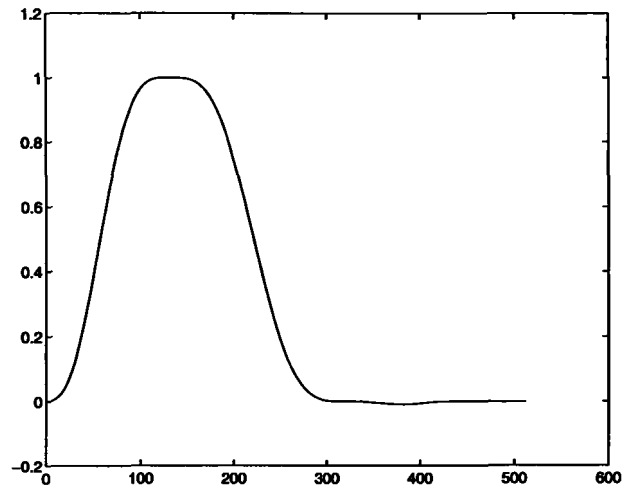

Figure 5: Impulse responses of the baseband prototype for a low delay filter bank, for analysis and synthesis, also with 128 bands and 255 samples delay, but with a filter length of 512 taps.

is to use one-dimensional line minimization for finding the s dimensional minimum, and choosing the directions of the line minimizations carefully. Line minimization can be done e. g. with Newtons method. To illustrate the idea, let x_0 be the starting point of the iteration, and v_i the unit vector in the direction of the line minimization. A step of Newtons method applied to the first derivative of $f(x)$ is

$$x_1 = x_0 - \Delta x$$

$$\Delta x = \frac{\partial f/\partial v_i\big|_{x_n}}{\partial^2 f/\partial v_i^2\big|_{x_n}} \cdot v_i.$$

The derivatives can be computed as

$$\frac{\partial f}{\partial v_i} = 2\,\mathrm{Re}\left\{ (H - d)\frac{\overline{\partial H}^T}{\partial v_i} \right\}$$

$$\frac{\partial^2 f}{\partial v_i^2} \approx 2\,\mathrm{Re}\left\{ \frac{\partial H}{\partial v_i} \cdot \frac{\overline{\partial H}^T}{\partial v_i} \right\}$$

where the overbar means complex conjugate. The s directions of the line minimizations, v_i, are determined by the eigenvectors of the Hessian matrix B of f. Usually it is computationally too expensive to compute the Hessian explicitly. But here the Hessian can be approximated with the first derivative of H, $A = \nabla H^T$, where $a_{i,j} = \partial H_i/\partial x_j$, as $B \approx 2\mathrm{Re}\{AA^T\}$. The following algorithm illustrates the optimization method.

x: Initialization with random numbers
Repeat
    $A = \nabla H|_x^T$
    $v =$ Set of Eigenvectors of $\mathrm{Re}\{A.\overline{A}^T\}$
    For $i = 1$ to s
    {
        Δx: Newton Step in direction v_i
        Repeat
            If $f(x - \Delta x) > f(x)$ then
                reduce Δx
            until $f(x - \Delta x) \le f(x)$
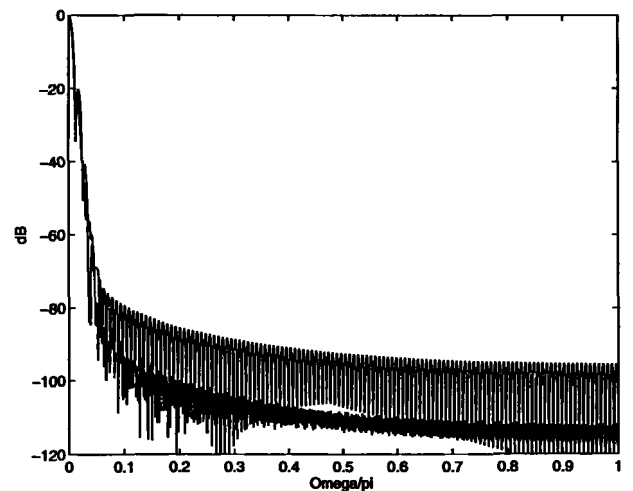   }
    until $|\Delta x|^2 <$ eps;

Figure 6: Magnitude responses of the baseband low delay prototype (the lower curve), identical for the analysis and synthesis filter bank, compared with the orthogonal filter bank.

This minimization process can be started with a random starting point. To make sure that a good minimum was found a second random starting point can be tried. For designing big filter banks (i. e. many bands, long filters), it can be faster to start with a smaller filter bank, with a fraction of the desired numbers of bands. When the optimization for this smaller filter bank is finished, the number of bands can be increased by increasing the size of the filter matrices, e.g. doubling the size and the number of bands by making pairs of coefficients out of each single coefficient. This is then the starting point for the optimization of the bigger filter bank. This process of growing the filter bank can be repeated until the desired size is reached.

3.2 Examples

Fig. 4 shows the impulse response of the baseband prototype filter for an orthogonal filter bank. The length of the filters is 256 taps, and the system delay is 255 samples. It results from the first cascade (equations (4) and (5)) or from the third cascade (equations (8) and (10)), with $m = 0$ and
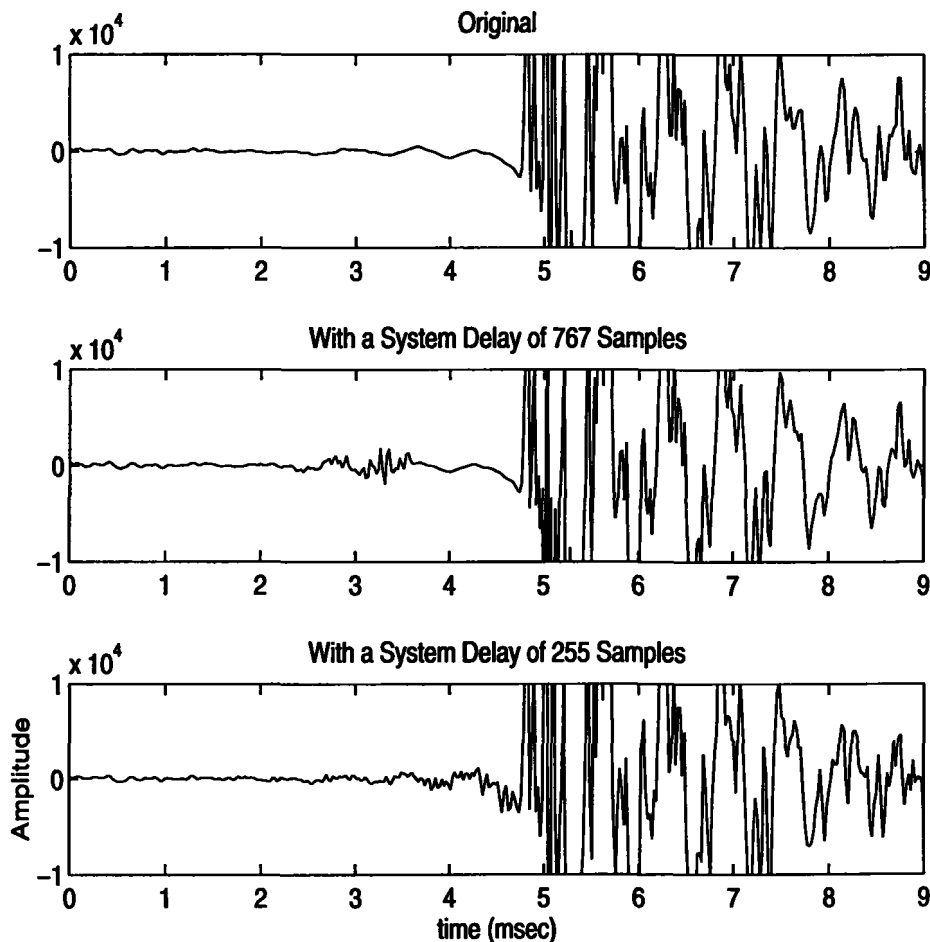
Figure 7: A system delay of 767 samples leads to an audible pre-echo, a system delay of 255 samples has no audible distortions. The pre-echo can be seen at around 3 ms.

$n = 0$, and with F orthogonal. This also is essentially a TDAC filter bank. The impulse response was obtained by using the described optimization algorithm.

Fig. 5 shows the impulse response of a low delay filter bank. It has a filter length of 512 taps, and the same system delay as the ortogonal example, 255 samples. It results from the first cascade (equations (4) and (5)) or from the third cascade (equations (8) and (10)), with $m = 0$ and $n = 2$. This baseband impulse response is identical for the analysis and the synthesis filter bank, except for the sign. The same error function was used for the analysis and synthesis frequency responses for the optimization.

Fig. 6 compares the frequency responses of the two filter banks. It can be seen that the low delay filter bank (the lower curve) has a stopband attenuation which is about 20 dB higher than with the orthogonal filter bank, although it has the same system delay.

Fig. 7 shows an application for low delay filter banks. They can be used to reduce the system delay while keeping the same magnitude responses. The first filter bank has filter lengths of 768 taps and a standard system delay of 767 samples. It can cause audible distortions in audio coding systems, so called pre-echoes, as shown in the middle picture. They arise from increasing the quantization step size when the signal amplitude increases. The analysis and synthesis filters can be seen as spreading the increased quantization error in the time domain, even before the increase of the signal amplitude, as can be seen in the picture at around 3 ms. Low

delay filter banks can influence this distribution of the quantization error, so that its spread before the increase can be made shorter than after the increase. It better matches the psycho-acoustic properties of the ear, and the distortion becomes inaudible [26]. This can be seen in the bottom picture, where a filter bank with filter lengths of 1024 taps but a system delay of only 255 samples was used. Here the distortions are much closer to the amplitude increase of the signal, and are now inaudible.

4. Conclusion

It was shown how to design modulated filter banks, with emphasis on cosine modulated filter banks, with critical downsampling, perfect reconstruction, a broad range of system delays, and with a computationally efficient implementation. Several different cascades with different properties where shown. A new type of cascade or decomposition of the polyphase matrices was introduced, which is more general and more efficient than previous cascades. The system delays which can be realized by these cascades are as low as $N - 1$, which is determined by the downsampling rate and is the theoretical lowest possible value, and as high as $2K - 1$, which results from a filter bank which can be seen as a time reversed version of a minimum delay filter bank. The examples show that the filter quality can be improved significantly by using a low delay filter bank with the same system delay but longer filters instead of a standard delay

filter bank. In the shown case the stopband attenuation was about 20 dB higher. This improved design can be used to increase the coding gain. Alternatively the system delay can be reduced while keeping the same quality of the filters by using low delay filter banks. This can be used to reduce pre-echos and for real time communication applications. The examples also show that it is possible to design big filter banks with these tools, which is also useful for audio coding applications.

References:
[1] Princen, J. P.; Bradley, A. B.: Analysis/Synthesis Filter Bank Design Based on Time Domain Alias Cancellation. Trans. on ASSP, (Oct., 1986), pp. 1153–1161.
[2] Vetterli, M.: A Theory of Multirate Filter Banks. Trans. ASSP, (March, 1987), pp. 356–372.
[3] Schuller, G.: Untersuchung von Kurzzeitspektralanalysatoren mit Spiegelfehlerkompensationseigenschaften. Diplomarbeit (Masters Thesis), Technische Universität Berlin, West Germany, 1989.
[4] Vaupel, T.: Transform Coding with Multiple Overlapping Blocks and Time Domain Aliasing Cancellation. Frequenz, (1990) 11–12.
[5] Rao, K. R.; Yip, P.: Discrete Cosine Transform: Algorithm, Advantages, Applications, Boston, MA.: Acad. Press, 1990.
[6] Vaidyanathan, P. P.: Multirate Systems and Filter Banks. New York: Prentice Hall, 1993.
[7] Akansu, A.; Haddad, R.: Multiresolution Signal Decomposition. New York: Academic Press, 1992.
[8] Mau, J.: Perfect Reconstruction Modulated Filter Banks: Fast Algorithms and Attractive New Properties. ICASSP 1993, Minneapolis, pp. III-225–228.
[9] Gopinath, R.; Burrus, C.: Theory of Modulated Filter Banks and Modulated Wavelet Tight Frames. ICASSP 1993, Minneapolis, pp. III-169–172.
[10] Temerinac, M., Edler, B.: LINC: A Common Theory of Transform and Subband Coding. IEEE Transactions on Communications, Vol. 41, (Feb. 1993) NO. 2.
[11] Temerinac, M.; Edler, B.: Overlapping Block Transform: Window Design, Fast Algorithm, and an Image Coding Experiment. IEEE Transactions on Communications, Vol. 43, (Sep. 1995) NO. 9.
[12] Malvar H.: Signal Processing with Lapped Transforms. London, Artech House, 1991.
[13] Malvar, H.S.: Extended Lapped Transforms: Fast Algorithms and Applications. ICASSP 1991, Toronto, pp. 1797–1800.
[14] Malvar, H. S.: Extended Lapped Transforms: Properties, Applications, and Fast Algorithms. IEEE Transactions on Signal Processing, Vol. 40, (Nov. 1992) NO. 11.
[15] Nayebi, K.; Barnwell, T.; Smith, M.: Low Delay Coding of Speech and Audio Using Nonuniform Band Filter Banks. IEEE Workshop on Speech Coding for Telecom. Wistler, Canada Sept. 1991.
[16] Nayebi, K.; Barnwell, T.; Smith, M.: Design of Low Delay FIR Analysis-Synthesis Filter Bank Systems,. Proc. Conf. on Info. Sci. and Sys., March 1991.
[17] Nayebi, K.; Barnwell, T. P.; Smith, M. J. T: Low Delay FIR Filter Banks: Design and Evaluation. Trans. on Signal Processing, (Jan. 1994) pp. 24–31.
[18] Nguyen, T. Q.: Digital Filter Bank Design Quadratic-Constrained Formulation. IEEE Transactions on Signal Processing, Vol. 43 (Sept. 1995). NO. 9.
[19] Nguyen, T. Q.; Heller, P. N.: Biorthogonal Cosine Modulated Filter Bank. ICASSP 96, Atlanta, GA, May 1996.
[20] Nguyen, T. Q.; Koilpillai, R. D.: The Theory and Design of Arbitrary-Length Cosine-Modulated Filter Banks and Wavelets, Satisfying Perfect Reconstruction. IEEE Transactions on Signal Processing, Vol. 44, (March 1996) NO. 3.
[21] Vaidyanathan, P. P.: Causal FIR Matrices with Anticausal FIR Inverses, and Application in Characterization of Biorthonormal Filter Banks. ICASSP 1994, Adelaide.
[22] Vaidyanathan, P. P.; Chen, T.: Role of Anticausal Inverses in Multirate Filter Banks – Part I: System-Theoretic Fundamentals, and Role of Anticausal Inverses in Multirate Filter Banks – Part II: The FIR Case, Factorizations, and Biorthogonal Lapped Transforms. IEEE Transactions on Signal Processing, Vol. 43 (May 1995) NO. 5.
[23] Schuller, G.; Smith, M. J. T.: A General Formulation for Modulated Perfect Reconstruction Filter Banks with Variable System Delay. NJIT 94 Sym. on Appl. of Subbands and Wavelets, March 1994.
[24] Schuller, G.; Smith, M. J. T.: Efficient Low Delay Filter Banks. DSP Workshop, Yosemite, Oct. 1994.
[25] Schuller, G.; Smith, M. J. T.: A New Algorithm for Efficient Low Delay Filter Bank Design. ICASSP 95, Detroit, MI, May 1995.
[26] Schuller, G.: A Low Delay Filter Bank for Audio Coding with Reduced Pre-Echoes. 99th AES Convention, New York, October 6–9, 1995.
[27] Schuller, G.: An Overview Over Filter Banks With Low System Delay Capabilities. European Workshop on Multirate Digital Signal Processing and Applications, Hamburg University of Technology, March 20–21, 1996.
[28] Press, W. H. et al.: Numerical Recipes, Cambridge University Press, 1992.

Dipl.-Ing. G. Schuller
Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung
der Universität Hannover
Schneiderberg 32
D-30167 Hannover

Neues aus Forschung, Industrie und Wirtschaft

Digital-optische Schalter aus Kunststoff – Schlüsselkomponenten in den Telekom-Netzen der Zukunft

Aus einer kürzlich in den USA erstellten Marktanalyse geht hervor, daß auf dem nordamerikanischen Markt mit einer Steigerung des Marktvolumens für optische Schalter von $ 33 Mio. im Jahre 1994 auf $ 687 Mio. im Jahre 2004 gerechnet wird. Eine ähnliche Entwicklung wird für den europäischen und den asiatischen Raum abgeleitet. Um diesen enormen Bedarf in Zukunft decken zu können, ist es erforderlich, daß derartige optische Bauelemente, genau wie elektronische Chips, kostengünstig und in großen Stückzahlen hergestellt werden.

Am Berliner Heinrich-Hertz-Institut haben Wissenschaftler jetzt einen digitaloptischen Schalter aus Kunststoff bis zur technischen Einsatzreife gebracht. Der 25 mm × 3 mm kleine Polymerschalter kann beispielsweise in digital-optischen Vermittlungsstellen zukünftiger Glasfasernetze eingesetzt werden. Dort kann er in weniger als einer tausendstel Sekunde die hochratigen Datenströme von einer Glasfaser auf eine andere schalten. So schnelle Reaktionen sind zur Umleitung des Datenverkehrs in optischen Nachrichtennetzen, bei Leitungsstörungen oder dem Ausfall eines Glasfaserkabels erforderlich. Der Schalter ist weltweit der erste seiner Art in Polymertechnik.

Die Berliner Arbeitsgruppe, unter Leitung von Dipl.-Ing. Norbert Keil, verfügt über langjährige Erfahrung in der Entwicklung optischer Bauelemente auf Polymerbasis und genießt weltweit ein hohes Ansehen. Sie überraschte die Fachwelt bereits im letzten Jahr, auf der CeBIT '95 in Hannover, mit dem weltweit ersten integriert-optischen Raumkoppelfeld in Polymer-Technologie. Die Berliner Neuentwicklung gestattet es nun erstmals, auf eine Temperaturstabilisierung des optischen Bauteils zu verzichten. Außerdem wurde die Ansteuerung wesentlich vereinfacht.

Dem HHI liegen bereits Anfragen aus dem In- und Ausland zwecks Vermarktung und Technologietransfer vor. Darüber hinaus wird an eine eigene Kleinserienfertigung und an die Vermarktung von ersten Prototypen gedacht. Der digital-optische Schalter aus Kunststoff wurde zum Patent angemeldet.

Siemens-Bereich Private Kommunikationssysteme baut Europavertrieb weiter aus

Der Bereich Private Kommunikationssysteme (PN) der Siemens AG, Berlin und München, baut seinen Europavertrieb weiter aus. In diesem Zusammenhang haben PN und die europäische Einkaufskooperation telEurope Systems and Services EWIV, Luxemburg, Mitte September in München eine Absichtserklärung unterzeichnet. Inhalt der Erklärung ist eine künftige intensive Zusammenarbeit beim Vertrieb von Systemen und Produkten der Kommunikationstechnik in Deutschland, Frankreich, Italien, Luxemburg und der Schweiz. Ziel der Kooperation zwischen PN und telEurope ist der nachhaltige Ausbau ihrer Positionen im europäischen Markt.

Mitglieder von telEurope sind die nationalen Einkaufskonsortien Phonet (Deutschland), EG-Tel (Schweiz), TelLux (Luxemburg), Resatis (Frankreich) und Consorzio Telitalia (Italien). Diesen Gesellschaften gehören qualifizierte Installationsunternehmen, Systemhäuser und Fachunternehmen für Kommunikationstechnik an, die in der Lage sind, Produkte und Systeme der Kommunikationstechnik zu vermarkten. Verschiedene Mitglieder von telEurope bieten bereits heute Kommunikationstechnik von Siemens erfolgreich in ihren jeweiligen nationalen Märkten an.