2

Low Delay Filter-Banks for Speech and Audio Processing

Heinrich W. Löllmann and Peter Vary

Institute of Communication Systems and Data Processing,
RWTH Aachen University, Germany

2.1 Introduction

Digital filter-banks are an integral part of many speech and audio processing algorithms used in today's communication systems. They are commonly employed for adaptive subband filtering, for example, to perform acoustic echo cancellation in hands-free communication devices or multi-channel dynamic-range compression in digital hearing aids, e.g., [34, 81]. Another frequent task is speech enhancement by noise reduction, e.g., [4, 81]. This eases the communication in adverse environments where acoustic background noise impairs the intelligibility and fidelity of the transmitted speech signal. A noise reduction system is also beneficial to improve the performance of speech coding and speech recognition systems, e.g., [41].

The choice of the filter-bank has a significant influence on the performance of such systems in terms of signal quality, computational complexity, and signal delay. Accordingly, the filter-bank design has to fulfill different, partly conflicting requirements in dependence of the considered application.

One prominent example is speech and audio processing for digital hearing aids. The restricted capacity of the battery and the small size of the chip set limit the available computational power. Moreover, a low overall processing delay is required to avoid disturbing artifacts and echo effects, e.g., [1,75]. Such distortions can occur when the hearing aid user is talking. In this case, the processed speech can interfere with the original speech signal, which reaches the cochlea with minimal delay via bone conduction or through the hearing aid vent. To prevent this, the algorithmic signal delay of the filter-bank used for the signal enhancement must be considerably lower than the tolerable processing delay, i.e., the latency between the analog input and output signal of the system. In addition, a filter-bank with non-uniform time-frequency resolution, which is similar to that of the human auditory system, is desirable to perform multi-channel dynamic-range compression and noise reduction with a small number of frequency bands.

A common choice for many applications is still the uniform DFT analysis-synthesis filter-bank (AS FB). This complex modulated filter-bank can be efficiently realized by means of a polyphase network (PPN) [77] and comprises the weighted overlap-add method as a variant hereof [11, 12]. This scheme is often used for frame-wise processing, e.g., in noise reduction systems of speech coders. In this case, a fixed buffering delay occurs and the additional delay due to overlapping frames can be reduced by an appropriate window design, e.g., [53].

However, the frequency resolution of the uniform (DFT) filter-bank is not well adapted to that of the human auditory system. The non-uniform frequency resolution of the human ear declines for an increasing frequency, which can be described by the Bark frequency scale [84]. Therefore, several authors have proposed the use of non-uniform AS FBs for speech enhancement to obtain an improved (subjective) speech quality [9,16,19,26,27,61]. One rationale for these approaches is that a filter-bank with a non-uniform, approximately Bark-scaled frequency resolution incorporates a perceptual model of the human auditory system. Another reason is that on average most of the energy and harmonics of speech signals are located at the lower frequencies.

One approach to achieve an approximately Bark-scaled frequency resolution is to employ the discrete wavelet (packet) transform, which can be implemented by a tree-structured AS FB, e.g., [9, 19, 26, 27]. Another method is to use frequency warped AS FBs [19, 26, 27, 61]. These filter-banks possess usually a lower signal delay and a lower algorithmic complexity than comparable tree-structured filter-banks.

The allpass transformation is a well-known technique for the design of frequency warped filter-banks [6, 17, 58, 79]. These filter-banks can achieve a Bark-scaled frequency division with great accuracy [73]. This property is of interest for speech and audio processing applications alike and allows to use a lower number of frequency channels than for the uniform filter-bank. A disadvantage of this approach is that the allpass transformation of the analysis filter-bank leads to (increased) aliasing and phase distortions. The compensation of these effects results in a more complex synthesis filter-bank design as well as a higher algorithmic complexity and signal delay. These drawbacks often prevent to exploit the benefits of frequency warped filter-banks for (real-time) speech and audio processing systems.

In this chapter, we discuss alternative design concepts for uniform and frequency warped filter-banks. The aim is to devise a general filter-bank design with the same time-frequency resolution as the conventional uniform and allpass transformed AS FB, but with a considerably lower signal delay.

For these purposes, the design of uniform and non-uniform AS FBs is reviewed in Sec. 2.2, and approaches to achieve a reduced signal delay are discussed. The alternative concept of the filter-bank equalizer (FBE) is introduced in Sec. 2.3. The effects of time-varying coefficients are analyzed, and an efficient implementation of the FBE is devised. A generalization of this concept is given by the allpass transformed FBE, which is presented in Sec. 2.3.6.

Measures to achieve (nearly) perfect signal reconstruction are described, and the algorithmic complexity of different filter-bank designs is contrasted. For applications with very demanding signal delay constraints, a modification of the FBE is proposed in Sec. 2.4 to achieve a further reduced signal delay with almost no loss for the subjective speech quality. In Sec. 2.5, the discussed filter-banks are applied to noise reduction and the achieved performance is investigated. Finally, a summary of this chapter is provided by Sec. 2.6.

2.2 Analysis-Synthesis Filter-Banks

In this section, some design concepts for uniform and non-uniform analysis-synthesis filter-banks are briefly reviewed, which form the basis (and motivation) for our alternative low delay filter-bank design introduced in Sec. 2.3.

2.2.1 General Structure

The general structure of an *analysis-synthesis filter-bank* (AS FB) is shown in Fig. 2.1. The discrete, real input signal $y(n)$ is split into M subband signals $y_i(n)$ by analysis bandpass filters with impulse responses $h_i(n)$ for $i \in \{\,0, 1, \ldots, M-1\,\}$. These subband filters can have different bandwidths $\Delta\Omega_i$ to achieve a non-uniform frequency resolution. The limited bandwidth of the subband signals $y_i(n)$ allows to perform a downsampling. The subsampling rates R_i for each subband can be determined by the general rule

$$R_i \leq \frac{2\,\pi}{\Delta\Omega_i} \quad \text{for} \quad R_i \in \{\,1, 2, \ldots, M\,\} \quad \text{and} \quad \sum_{i=0}^{M-1} \Delta\Omega_i = 2\,\pi\,. \tag{2.1}$$

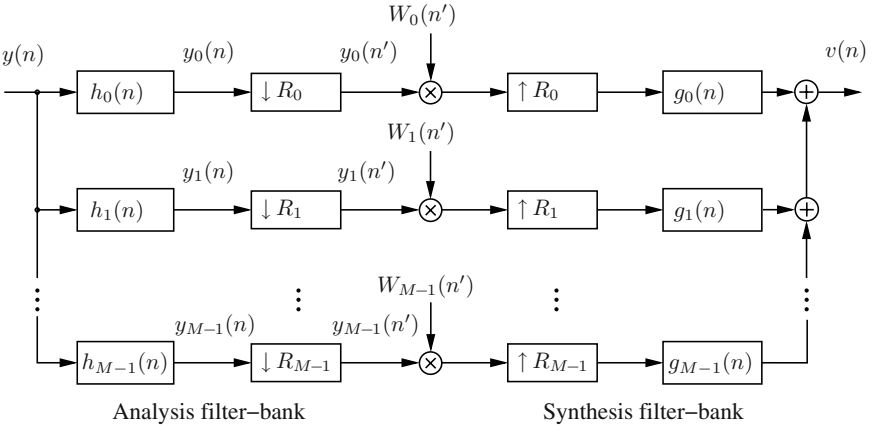


Fig. 2.1. M-channel analysis-synthesis filter-bank (AS FB) with subsampling and spectral weighting.

For a *uniform* filter-bank, the bandwidths of all subband filters are equal and the same subsampling rate $R_i = R$ is taken for each subband signal. *Critical subsampling* is performed if $R = M$. An *oversampled* filter-bank performs non-critical subsampling where $R < M$.

The signal reconstruction is accomplished by the synthesis filter-bank, which consists of the upsampling operations and interpolating bandpass filters with impulse responses $g_i(n)$. A filter-bank achieves *perfect (signal) reconstruction* (PR) with a delay of d_0 samples if

$$v(n) = y(n - d_0) \tag{2.2}$$

for $W_i = 1 \,\forall\, i$. Accordingly, near-perfect reconstruction (NPR) is achieved if this identity is approximately fulfilled.

AS FBs are commonly used for adaptive subband processing as indicated by Fig. 2.1. The spectral gain factors $W_i(n')$ are adapted at a reduced sampling rate based on the downsampled subband signals $y_i(n')$. For example, this filtering technique is frequently used for the enhancement of noisy speech signals, e.g., [31, 80, 81].

2.2.2 Tree-Structured Filter-Banks

Tree-structured filter-banks are used to achieve a uniform or, more commonly, a non-uniform time-frequency resolution. They are mostly realized by the discrete wavelet transform (DWT) or by quadrature mirror filters (QMFs), e.g., [7, 77, 83]. Tree-structured filter-banks can realize an octave-band frequency analysis as depicted in Fig. 2.2. The input signal is split into a lowpass (LP) and highpass (HP) signal which can be each downsampled by a ratio of two. This step can be repeated successively until the desired frequency resolution is (approximately) achieved. This procedure leads to different signal delays for the subband signals which can be compensated by corresponding

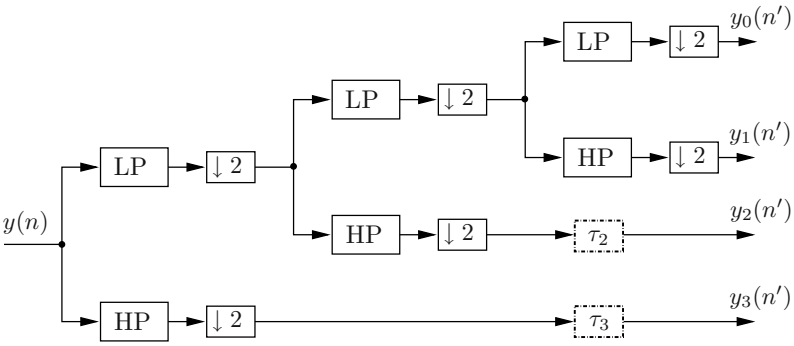


Fig. 2.2. Tree-structured filter-bank with three stages realizing an octave-band analysis.

delay elements. A more flexible adjustment of the frequency resolution can be accomplished by the wavelet packet decomposition, e.g., [7]. The signal delay of a tree-structured AS FB is equal to $(2^J - 1)\, d_\mathrm{s}$ with J marking the number of stages and d_s denoting the signal delay of the underlying two-channel AS FB. Hence, such filter-banks exhibit a high signal delay, especially if subband filters of high degrees are needed to avoid aliasing distortions owing to subband processing, cf. [9, 16, 26].

2.2.3 Modulated Filter-Banks

An important class of filter-banks constitute modulated AS FBs. The individual subband filters are derived by uniform modulation of a single prototype filter which yields a uniform time-frequency resolution, e.g., [12,77]. The input-output relation for fixed spectral gain factors W_i and an input signal of finite energy can be written as [77]

$$V(z) = \frac{1}{R} \sum_{r=0}^{R-1} Y(z\, E_R^r) \sum_{i=0}^{M-1} H_i(z\, E_R^r)\, G_i(z)\, W_i \tag{2.3}$$

with the modulation factor defined by

$$E_R = e^{-j\frac{2\pi}{R}} \,. \tag{2.4}$$

Due to the subsampling operations, the AS FB is a time-variant system even for fixed gain factors W_i. To account for this behavior, we determine the overall transfer function of the filter-bank by a *series* of time-shifted unit sample sequences as input, i.e., $y(n) = \delta(n - d)$ with $d \in \mathbb{N}_0$. Inserting $Y(z) = z^{-d}$ into Eq. 2.3 leads to the transfer function

$$\begin{aligned} T_d(z) &= \frac{V(z)}{z^{-d}} \\ &= \underbrace{\frac{1}{R} \sum_{i=0}^{M-1} H_i(z)\, G_i(z)\, W_i}_{T_\mathrm{lin}(z)} + \underbrace{\frac{1}{R} \sum_{r=1}^{R-1} E_R^{-d\,r} \sum_{i=0}^{M-1} H_i(z\, E_R^r)\, G_i(z)\, W_i}_{\mathcal{E}_\mathrm{A}(z)} \,. \end{aligned} \tag{2.5}$$

The linear transfer function of the filter-bank is given by $T_\mathrm{lin}(z)$. The *aliasing distortions* due to the subsampling operations are represented by $\mathcal{E}_\mathrm{A}(z)$. The AS FB is a linear periodically time-variant (LPTV) system with period R since

$$T_{d+R}(z) = T_d(z) \tag{2.6}$$

according to Eq. 2.5. Therefore, a *linear time-invariant* (LTI) system is obtained if $\mathcal{E}_{\mathrm{A}}(z) = 0$, so that no aliasing distortions occur. Perfect reconstruction according to Eq. 2.2 is achieved, if the transfer function of Eq. 2.5 is given by

$$T_d(z) = z^{-d_0} \quad \text{for} \quad d \in \{\,0, 1, \ldots, R-1\,\} \quad \text{and} \quad W_i = 1 \;\forall\, i \tag{2.7}$$

where the limited set of values for d follows from Eq. 2.6. An example are paraunitary filter-banks which fulfill this condition with a delay of $d_0 = L$ samples where L denotes the degree of the FIR prototype filters [77].[1]

A filter-bank with perfect reconstruction ensures complete aliasing cancellation only if no subband processing is performed, that is, $W_i = 1$ for Eq. 2.5. Therefore, oversampled filter-banks are commonly used for adaptive subband filtering to avoid strong aliasing distortions in consequence of spectral weighting, e.g., [4,19]. In contrast, critically subsampled AS FBs are a typical choice for subband coding systems, e.g., [77,83].

An important realization of a (complex) modulated filter-bank is given by the *DFT filter-bank*. The subband filters have the transfer functions

$$H_i(z) = \sum_{l=0}^{L} h(l)\, E_M^{i\,l}\, z^{-l} \tag{2.8}$$

$$G_i(z) = \sum_{l=0}^{L} g(l)\, E_M^{i\,(l+1)}\, z^{-l}\,; \quad i \in \{0, 1, \ldots, M-1\} \tag{2.9}$$

where $h(n)$ and $g(n)$ denote the finite impulse responses (FIRs) of the analysis and synthesis prototype filter, respectively. The use of linear-phase prototype filters leads to a signal delay of $d_0 = L$ samples, cf. [77].

A common choice for the FIR filter degree is $L = M-1$, but a higher degree can be taken to increase the frequency selectivity of the subband filters. Such a filter-bank can be efficiently implemented by a *polyphase network* (PPN). The analysis filters of Eq. 2.8 can be written as

$$H_i(z) = \sum_{\lambda=0}^{M-1} H_\lambda^{(M)}(z^M) \cdot z^{-\lambda} \cdot E_M^{\lambda\,i}\,; \quad i \in \{\,0, 1, \ldots, M-1\,\} \tag{2.10}$$

with the 'type 1' polyphase components defined by [77]

$$H_\lambda^{(M)}(z) = \sum_{m=0}^{l_M-1} h(m\,M + \lambda)\, z^{-m}\,; \quad \lambda \in \{\,0, 1, \ldots, M-1\,\}\,. \tag{2.11}$$

It is assumed that the length of the prototype filters is an integer multiple of M according to

[1] One property of paraunitary filter-banks is that the sum of the subband energies is equal to the energy of the input signal.

$$L + 1 = l_M \, M \; ; \quad l_M \in \mathbb{N} \tag{2.12}$$

which can be always achieved by an appropriate zero-padding. The synthesis filters of Eq. 2.9 can be expressed by

$$G_i(z) = \sum_{\lambda=0}^{M-1} G_\lambda^{(M)}(z^M) \cdot z^{-(M-1-\lambda)} \cdot E_M^{-\lambda\, i} \; ; \quad i \in \{\, 0, 1, \ldots, M-1 \,\} \tag{2.13}$$

with the 'type 2' polyphase components

$$G_\lambda^{(M)}(z) = \sum_{m=0}^{l_M - 1} g\big((m+1)\, M - \lambda - 1\big)\, z^{-m} \; ; \quad \lambda \in \{\, 0, 1, \ldots, M-1 \,\} \, . \tag{2.14}$$

Fig. 2.3 shows the derived PPN realization of a DFT filter-bank. The subsampling operations can be moved towards the delay elements due to the so-called 'noble identities', cf. [77]. The discrete Fourier transform (DFT) can be computed efficiently by the fast Fourier transform (FFT), e.g., [59]. Hence, this PPN filter-bank implementation possesses only a low computational complexity. The processing scheme of Fig. 2.3 can also be interpreted as weighted

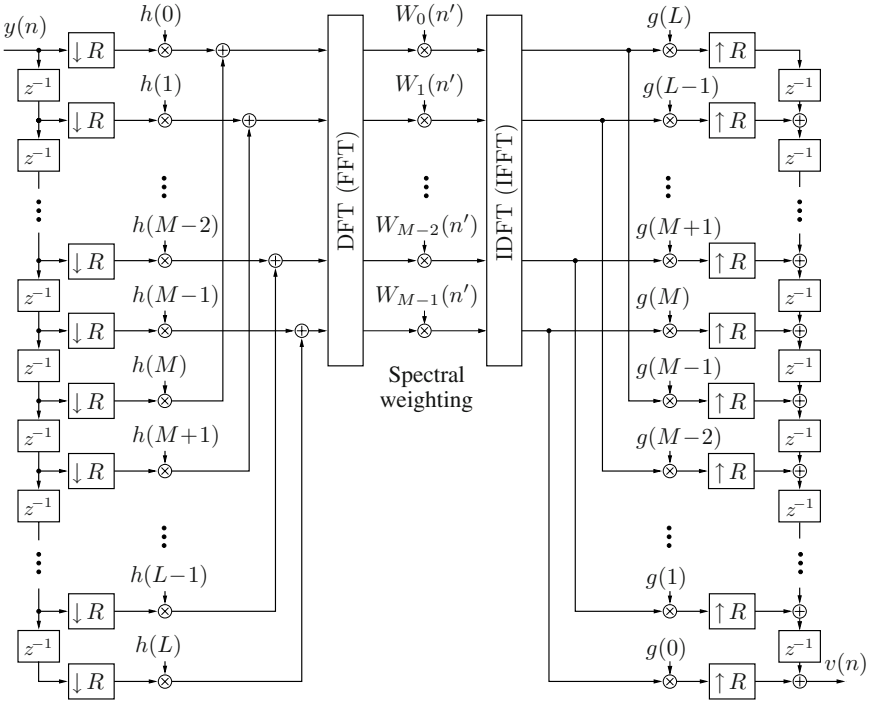


Fig. 2.3. Polyphase network (PPN) realization of a DFT AS FB for $L + 1 = 2\, M$.

overlap-add method [11,12]. The delay chains buffer the samples of the input and output frames whose overlap is determined by the subsampling rate R.

For speech enhancement applications, the time-varying spectral gain factors $W_i(n')$ of the filter-bank might cause audible artifacts due to so-called 'block-edge effects'. This can be avoided by non-critical subsampling ($R < M$) and a dedicated prototype filter design, e.g., [31,81].

2.2.4 Frequency Warped Filter-Banks

2.2.4.1 Principle

A frequency warped digital system can be obtained by replacing the delay elements of its transfer function by allpass filters

$$z^{-1} \rightarrow H_{\mathrm{A}}(z)\,. \tag{2.15}$$

For this *allpass transformation*, a causal, complex allpass filter of first order is considered, whose transfer function is given by

$$H_{\mathrm{A}}(z) = \frac{z^{-1} - a^*}{1 - a\,z^{-1}}\,; \quad |a| < |z|\,; \quad |a| < 1\,; \quad a = \alpha\,e^{j\,\gamma} \in \mathbb{C}\,; \quad \alpha, \gamma \in \mathbb{R} \tag{2.16}$$

with $\mathbb{C}$ marking the set of all complex numbers and $\mathbb{R}$ marking the set of all real numbers. The asterisk denotes the complex-conjugate value. One possible implementation of this allpass filter is shown in Fig. 2.4. The frequency response reads

$$H_{\mathrm{A}}(z = e^{j\,\Omega}) = \frac{e^{-j\,\Omega} - a^*}{1 - a\,e^{-j\,\Omega}} = e^{-j\,\varphi_a(\Omega)} \tag{2.17a}$$

$$\varphi_a(\Omega) = 2\,\arctan\left(\frac{\sin\Omega - \alpha\,\sin\gamma}{\cos\Omega - \alpha\,\cos\gamma}\right) - \Omega\,. \tag{2.17b}$$

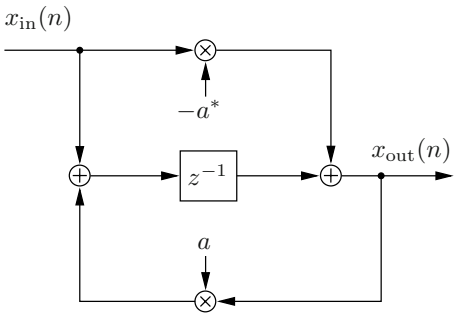


Fig. 2.4. Realization of an allpass filter of first order.

This concept can be extended to allpass filters of higher order [35,36], but the allpass transformation of first order is of most interest here.[2]

The allpass transformation is a bilinear transformation and allows to alter the frequency characteristic of a digital system without changing its coefficients. This property is exploited for the design of variable digital filters [10,70,71]. The cutoff frequency of these filters is adjusted by the allpass coefficient, whereas the shape of the frequency response (e.g., number and magnitude of the ripples or the stopband attenuation) is not changed. The allpass transformation can also be employed to perform short-term spectral analysis with a non-uniform frequency resolution or to construct non-uniform digital filter-banks [6,17,58,79].

The allpass transformation of the analysis filters of Eq. 2.8 yields the warped frequency responses

$$H_i(z = e^{j\,\varphi_a(\Omega)}) = \sum_{l=0}^{L} h(l)\, E_M^{i\,l}\, e^{-j\,\varphi_a(\Omega)\,l} \tag{2.18}$$

$$= \widetilde{H}_i(e^{j\,\Omega})\,; \quad i \in \{0, 1, \ldots, M-1\} \tag{2.19}$$

due to Eq. 2.15 and Eq. 2.17. Hence, the allpass transformed filter-bank is a generalization of the uniform filter-bank, which is included as special case for $a = 0$. The allpass transformation causes a *frequency warping*

$$\Omega \rightarrow \varphi_a(\Omega) \tag{2.20}$$

where the course of the phase response $\varphi_a(\Omega)$ is determined by the allpass coefficient a. The effect of this allpass transformation is demonstrated in Fig. 2.5. For a real and positive allpass coefficient $a = \alpha > 0$, a higher frequency resolution is achieved for the lower frequency bands and vice versa for the higher frequency bands. The opposite applies if $\alpha < 0$. Thus, the frequency resolution can be adjusted by a single coefficient without the requirement for an individual subband filter design, which is sometimes needed for the construction of non-uniform filter-banks (cf. Sec. 2.2.5). A complex allpass transformation is of interest, if a more flexible adjustment of the frequency resolution is desired, cf. [35].

The allpass transformation allows to design a non-uniform filter-bank whose frequency bands approximate the *Bark frequency scale* with great accuracy [73]. The frequency resolution of the human auditory system is determined by the so-called 'critical bands'. The mapping between frequency and critical bands can be described by the critical band rate with the unit 'Bark'. An analytical expression for the Bark frequency scale is given by [84].

[2] An allpass transformation of order N maps the unit circle N-times onto itself which causes a comb filter structure. This comb filter effect is undesirable for the design of filter-banks and can be avoided by additional allpass filters at the price of an increased algorithmic complexity and a higher signal delay, cf. [35,36].

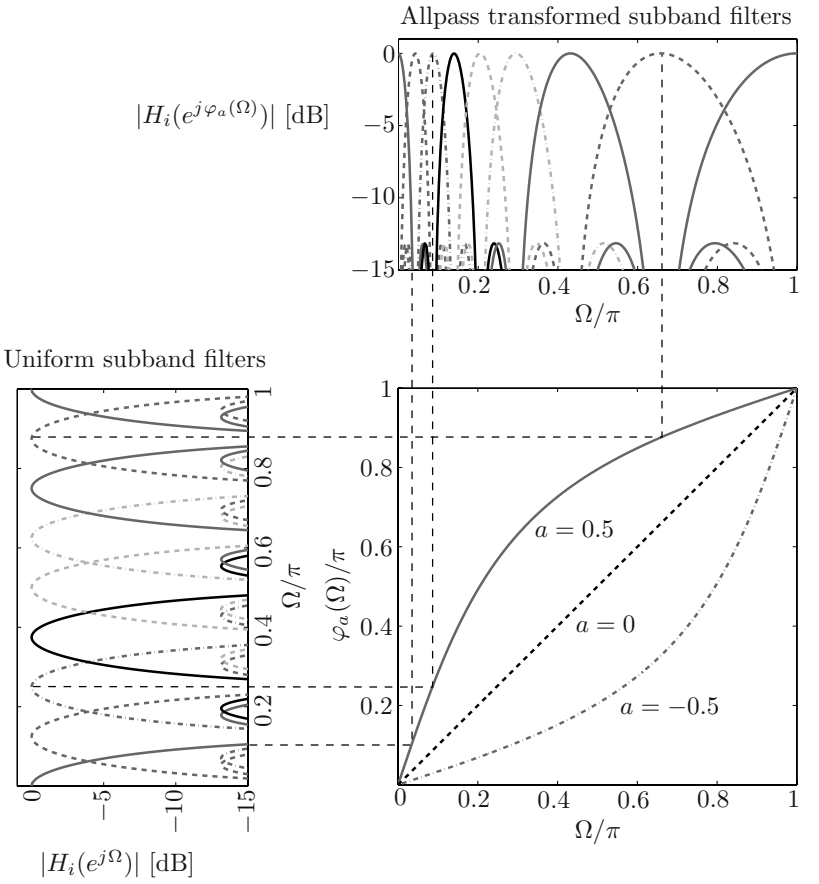


Fig. 2.5. Allpass transformation of subband filters for $M = 16$ and a rectangular prototype filter of length $L + 1 = M$.

$$\frac{\xi(f)}{\text{Bark}} = 13 \cdot \arctan\left(\frac{0.76\,f}{\text{kHz}}\right) + 3.5 \cdot \arctan\left(\left(\frac{f}{7.5\,\text{kHz}}\right)^2\right) . \tag{2.21}$$

Fig. 2.6 illustrates that such a frequency division can be well approximated by means of an allpass transformed filter-bank. A filter-bank with approximately Bark-scaled frequency bands can also be realized by the wavelet packet decomposition [9]. However, the obtained tree-structured filter-bank has a significantly higher signal delay and a higher algorithmic complexity than a comparable allpass transformed filter-bank.

2.2.4.2 Signal Reconstruction

The allpass transformation of the analysis filter-bank according to Eq. 2.18 leads to phase modifications and a stronger overlap of aliasing components in

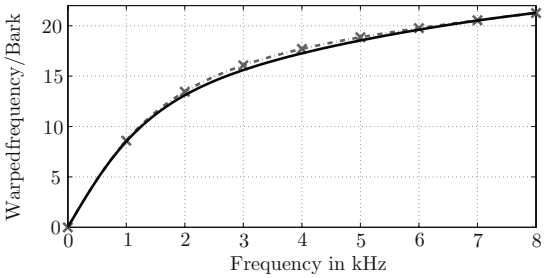


Fig. 2.6. Approximation of the Bark frequency scale: The solid line corresponds to the analytical expression of Eq. 2.21. The dashed line marks the frequency warping for an allpass transformation with $a = 0.576$ and a sampling frequency of 16 kHz.

comparison to the uniform filter-bank. These effects complicate the synthesis filter-bank design. The two main approaches to perform the signal reconstruction in this case are depicted in Fig. 2.7.

The filter-bank structure I uses $L + 1 = M$ filters with transfer functions $P(z, l)$ for $l \in \{ 0, 1, \ldots, L \}$ to compensate the (additional) phase and aliasing distortions caused by the allpass transformed analysis filter-bank. Perfect reconstruction can be achieved by FIR filters. Their coefficients can be determined by analytical closed-form expressions in case of a prototype filter length of $L+1 = M$. This is shown in [72] for critical subsampling and in [21,35] for arbitrary subsampling rates. However, the obtained synthesis subband filters show no distinctive bandpass characteristic. This causes a high reconstruction error if spectral modifications of the subband signals, such as quantization or spectral weighting, are performed [21].

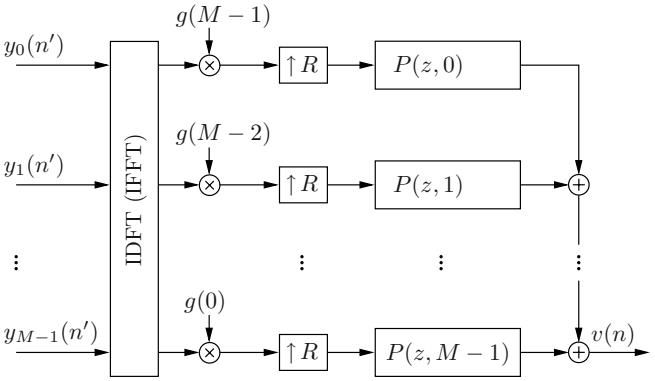
Synthesis subband filters with a distinctive bandpass characteristic are obtained by the filter-bank designs proposed, e.g., in [23, 24, 49] which achieve near-perfect reconstruction. The filters with transfer functions $P(z, l)$ are designed to compensate the phase distortions due to the frequency warping.[3] The aliasing distortions are limited by the higher stopband attenuation of longer subband filters ($L + 1 \gg M$) and a lower subsampling rate R.[4]

A similar principle is used for the synthesis filter-bank structure II of Fig. 2.7, which, however, uses only a single compensation filter with transfer function $P(z, L)$. The allpass transformation is applied to the analysis and synthesis filter-bank, i.e., all delay elements of the uniform filter-bank (shown in Fig. 2.3) are replaced by allpass filters according to Eq. 2.15. If the uniform

[3] If not mentioned otherwise, the more general concept of frequency warping will always refer to an allpass transformation of first order so that both terms are used interchangeably.

[4] Here, the same subsampling rate R is used for each subband signal so that the DFT and IDFT can be executed at a decimated sampling rate. This is not possible for different subsampling rates according to Eq. 2.1.

(a) Synthesis filter-bank structure I



(b) Synthesis filter-bank structure II

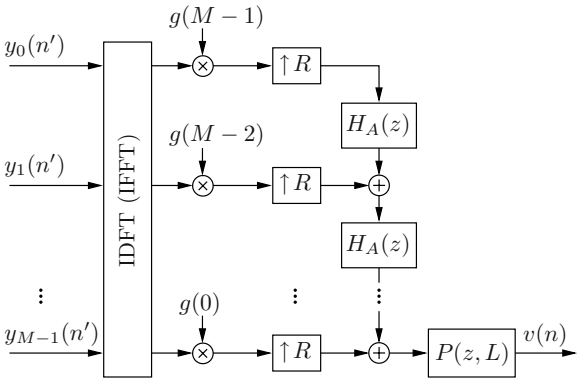


Fig. 2.7. Synthesis filter-bank structures for an allpass transformed analysis filter-bank with prototype filter of length $L+1=M$.

filter-bank fulfills Eq. 2.7 with $d_0 = L$, the allpass transformation leads to the frequency response

$$\widetilde{T}_d(z = e^{j\,\Omega}) = e^{-j\,\varphi_a(\Omega)\,L} + \mathcal{E}_{\mathrm{A}}(e^{j\,\varphi_a(\Omega)})\,. \tag{2.22}$$

The aliasing distortions $\mathcal{E}_{\mathrm{A}}(e^{j\,\varphi_a(\Omega)})$ emerge due to the non-uniform bandwidths of the allpass transformed subband filters. They can be reduced by a lower subsampling rate R and the use of subband filters of higher degrees having narrow transition bands and high stopband attenuations.

The non-linear phase term $\varphi_a(\Omega)\,L$ can cause audible distortions especially for a high prototype filter degree L. Eq. 2.17 reveals that this term corresponds to the frequency response of a cascade of L identical allpass filters termed as allpass (filter) chain. The task of the fixed *phase equalizer* at the filter-bank

output is to compensate these phase distortions (see Fig. 2.7-b). The frequency response of the phase equalizer has to fulfill the general requirement, cf. [48]

$$e^{-j\,d_P\,\varphi_a(\Omega)} \cdot P_{\text{gen}}(e^{j\,\Omega}, d_P) \stackrel{!}{=} e^{-j\,\tau_P\,\Omega}\,; \quad \tau_P \geq 0\,; \quad \tau_P \in \mathbb{R}\,; \quad d_P \in \mathbb{N} \tag{2.23}$$

where $d_P = L$ for the warped AS FB described by Eq. 2.22.

The 'ideal' phase equalizer to fulfill Eq. 2.23 with $\tau_P = 0$ is obviously given by the inverse transfer function of an allpass chain

$$P_{\text{ideal}}(z, d_P) = H_{\text{A}}(z)^{-d_P}\,; \quad |z| < \frac{1}{|a|}\,. \tag{2.24}$$

However, the impulse response of this phase equalizer is infinite and anti-causal (for $a \neq 0$), i.e., $p_{\text{ideal}}(n, d_P) = 0$ for $n > 0$. An approach to realize anti-causal filters is to buffer the input samples in order to process them in time-reversed order [14, 56]. This rather complex technique requires large buffers and leads to a high signal delay.

An alternative approach is to approximate the desired anti-causal phase equalizer of Eq. 2.24 by a causal FIR filter of degree N_P. Its coefficients can be obtained by shifting and truncating the impulse response $p_{\text{ideal}}(n, d_P)$ according to

$$p_{\text{LS}}(n, d_P) = \begin{cases} p_{\text{ideal}}(n - N_P, d_P) & ; \quad n \in \{\,0, 1, \ldots, N_P\,\} \\ 0 & ; \quad \text{else.} \end{cases} \tag{2.25}$$

The transfer function of an inverse allpass chain $H_A(z)^{-d_P}$ is identical to the para-conjugate transfer function of the allpass chain where the z-variable is replaced by z^{-1} and the complex-conjugate filter coefficients are used. Thus, the impulse response of the ideal phase equalizer $p_{\text{ideal}}(n, d_P)$ can be obtained by the time-reversed impulse response of an allpass chain of length d_P with complex-conjugate allpass coefficient a^*. The FIR filter approximation of an IIR filter by truncating its impulse response leads to a least-squares error, e.g., [60]. Thus, the phase equalizer according to Eq. 2.25 is termed as least-squares (LS) phase equalizer. For a complex allpass transformation, a complex output signal $v(n)$ is obtained where the (discarded) imaginary part becomes negligible for a low signal reconstruction error.

A drawback of FIR phase equalizers is that they cause significant magnitude distortions in case of a low filter degree N_P. Hence, this filter degree of the phase equalizer must be high enough to keep phase and magnitude distortions low. Such magnitude distortions are avoided by using an allpass phase equalizer. Its filter degree is determined only by the need to keep the phase distortions due to the warping inaudible. The design of phase equalizers for warped filter-banks is treated in [48] in more detail. It should be noted that the discussed synthesis filter-bank designs for near-perfect reconstruction can also be applied if the prototype filter length $L + 1$ exceeds M.

The signal delay of the considered uniform AS FB is given by $d_0 = L$. The signal delay of the warped filter-bank with LS FIR phase equalizer according to Eq. 2.25 is approximately equal to N_P. This filter degree of the phase equalizer should be considerably higher than the value $d_P = d_0$ so that the warped AS FB with phase equalizer has a significantly higher overall signal delay than the uniform filter-bank. As shown in [49], it is also beneficial to use the LS phase equalizer of Eq. 2.25 for the filter-bank structure I in Fig. 2.7, which leads to (almost) the same signal delay as for the filter-bank structure II with LS phase equalizer.

The devised concepts for phase equalization are also effective if spectral weighting is performed, cf. [48]. For speech and audio processing, a perfect equalization of the warped phase is not required due to the insensitivity of human hearing towards minor phase distortions, cf. [80,84]. A design example for an allpass transformed AS FB is given later in Sec. 2.5.

2.2.5 Low Delay Filter-Banks

One approach for the design of uniform and non-uniform AS FBs with low delay is to use the *lifting scheme*, which has been originally proposed for the construction of 'second generation wavelets' [15,76]. A single zero-delay lifting step is shown in Fig. 2.8. The new analysis lowpass filter after one lifting step is given by

$$H_0^{(1)}(z) = H_0(z) + H_1(z)\, B(z^2)\,. \tag{2.26}$$

Correspondingly, this procedure can be applied to the analysis highpass filter termed as dual lifting step and so on. The lifting steps for the analysis filters are followed by inverse lifting steps at the synthesis side. By this, the degree of the subband filters is increased without increasing the overall signal delay of the filter-bank, cf. Fig. 2.8. The application of this scheme to the design of (uniform) cosine modulated AS FBs with low delay is proposed in [37,38]. In [23,24], the lifting scheme is applied to the allpass transformed AS FB. The higher aliasing distortions due to the frequency warping are reduced by improving the stopband attenuation of the subband filters. The lifting scheme is used to increase the filter degree while constraining the signal delay of the filter-bank. However, the adding of further lifting steps shows no improvement

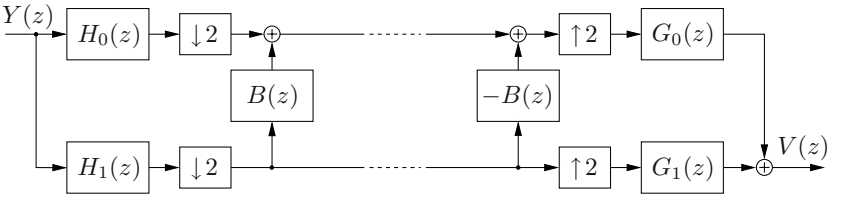


Fig. 2.8. Single zero-delay lifting step for a two-channel AS FB.

after some stages. Therefore, only a limited enhancement of the stopband attenuation and the associated aliasing cancellation can be achieved. Moreover, the analysis and synthesis filter-bank proposed in [23, 24] are operated at the non-decimated sampling rate, which causes a very high computational complexity.

The use of the *warped discrete Fourier transform* (WDFT) is an alternative approach to design a low delay filter-bank with warped frequency bands, e.g., [22, 51, 61]. The WDFT is a non-uniform DFT and calculated by the rule

$$\widetilde{Y}(i) = \sum_{n=0}^{M-1} y(n) \left(\frac{e^{-j\,\frac{2\pi}{M}\,i} - a^*}{1 - a\,e^{-j\,\frac{2\pi}{M}\,i}} \right)^n \quad ; \quad i \in \{\,0, 1, \ldots, M-1\,\}\,. \tag{2.27}$$

In contrast to the DFT (which is obtained for $a = 0$), the frequency points of the WDFT are non-uniformly spaced on the unit circle. The WDFT filter-bank evolves by replacing the (I)DFT in Fig. 2.3 by the (I)WDFT so that the signal delay remains the same. In this process, the center frequencies of the subband filters are shifted, but their bandwidth remain the same. This effect is illustrated in Fig. 2.9. In contrast to the allpass transformed filter-bank (see Fig. 2.5), the spectrum of the WDFT exhibits 'spectral gaps' due to the uniform bandwidths of the subband filters. This complicates the signal reconstruction, which is reflected by an ill-conditioned WDFT matrix for values of about $|a| > 0.2$ and $M > 40$, cf. [22, 61]. The numerical difficulties for the calculation of the inverse WDFT (matrix) become even more pronounced

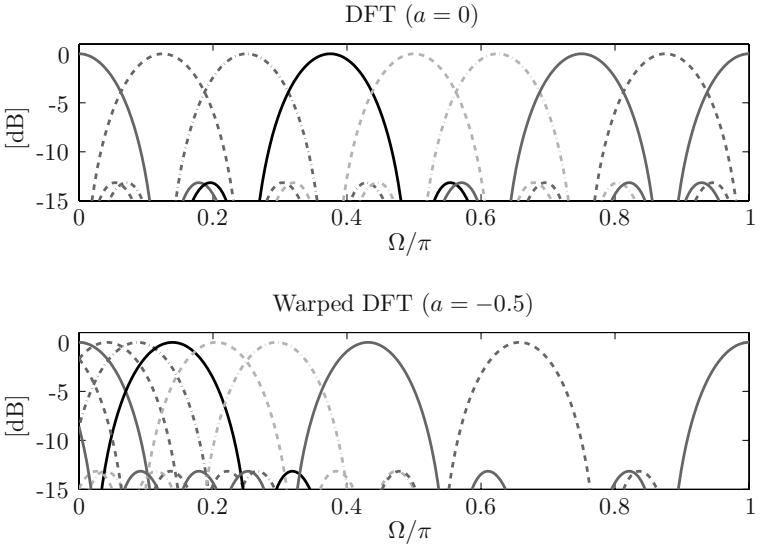


Fig. 2.9. Magnitude responses of the subband filters for a DFT and warped DFT with $M = 16$ channels.

with regard to a practical implementation with finite precision arithmetic, for instance, on a (fixed-point) digital signal processor (DSP).

There are other ways to derive a non-uniform low delay filter-bank from a uniform filter-bank. One approach is to combine an appropriate number of cosine modulated subband filters termed as 'feasible partitioning' [16, 43]. Another method is to use two different uniform filter-banks for the upper and lower frequency bands which are linked by a 'transition filter', e.g., [13, 18]. A good approximation of the Bark scale is difficult to achieve by this approach. The subband filters of these filter-banks need to have a relatively high filter degree to achieve a sufficient stopband attenuation in order to keep aliasing distortions low, especially if subband processing takes place. This causes still a comparatively high signal delay which depends, among others, on the permitted aliasing distortions.

Many designs of uniform and non-uniform filter-banks allow to prescribe an (almost) arbitrary signal delay, e.g., [16, 18, 40, 69]. However, it is problematic to achieve simultaneously a high stopband attenuation for the subband filters as well as a low signal delay. Hence, there is a trade-off between a low signal delay on the one hand and low aliasing distortions (high speech and audio quality) on the other hand, cf. [16].

A low signal delay *and* an aliasing-free signal reconstruction can be achieved by means of the *filter-bank summation method* (FBSM) depicted in Fig. 2.10. The FBSM can be derived from the filter-bank interpretation of the short-time DFT, e.g., [12]. A drawback of this filter-bank structure is its high computational complexity as no downsampling of the subband signals can be performed. Therefore, the AS FB is considered to be more suitable than the FBSM for real-world applications such as speech enhancement [19]. Moreover, the computational complexity of the FBSM is significantly increased, if we apply the allpass transformation to achieve a Bark-scaled frequency resolution.

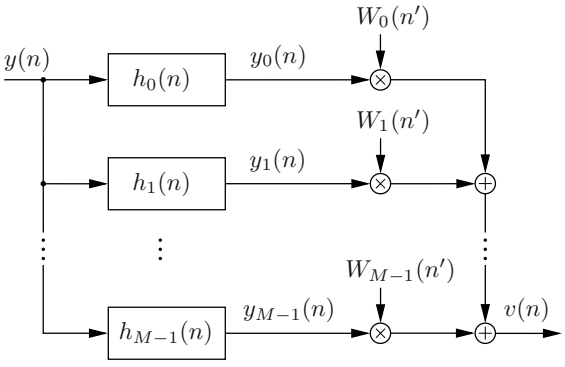


Fig. 2.10. Filter-bank summation method (FBSM) with time-varying spectral gain factors adapted at a reduced sampling rate.

In the following, we derive a uniform and warped low delay filter-bank without the high algorithmic complexity of the FBSM.

2.3 The Filter-Bank Equalizer

An alternative filter-bank concept to that of the conventional AS FB will be devised to perform adaptive subband processing with a significantly lower signal delay.

2.3.1 Concept

The FBSM of Fig. 2.10 is considered. As for the AS FB, the real input signal $y(n)$ is split into the subband signals $y_i(n)$ by means of M bandpass filters. The adaptation of the time-varying spectral gain factors $W_i(n')$ can be performed by the same algorithms as for the AS FB. This adaptation is based on the subband signals $y_i(n)$ and executed at intervals of R sample instants with n' defined by[5]

$$n' = \lfloor n/R \rfloor \cdot R\,; \quad R \in \mathbb{N}\,. \tag{2.28}$$

The operation $\lfloor . \rfloor$ provides the greatest integer which is lower than or equal to the argument.

The impulse response $h_i(n)$ of the ith bandpass filter is obtained by modulation of a prototype lowpass filter with real impulse response $h(n)$ of length $L+1$ according to

$$h_i(n) = \begin{cases} h(n)\,\Phi(i,n) & ;\ i \in \{\,0,1,\ldots,M-1\,\}\,;\ \ n \in \{\,0,1,\ldots,L\,\} \\ 0 & ;\ \text{else}\,. \end{cases} \tag{2.29}$$

The choice for the general modulation sequence and the prototype filter affects the spectral selectivity and time-frequency resolution of the filter-bank. The modulation sequence $\Phi(i,n)$ can be seen as transformation kernel of the filter-bank. In general, it has the periodicity

$$\Phi(i, n + m\,M) = \Phi(i,n)\,\rho(m)\,; \quad m \in \mathbb{Z} \tag{2.30}$$

where $\mathbb{Z}$ denotes the set of all integer numbers. The sequence $\rho(m)$ depends on the chosen transform as shown later. For many transforms (including the DFT) it is given by $\rho(m) = 1\ \forall\, m$.

The input-output relation for the FBSM of Fig. 2.10 can be written as

[5] This definition is more suitable for the following treatment than the common convention $n = R\,n'$.

$$v(n) = \sum_{i=0}^{M-1} W_i(n')\, y_i(n) \tag{2.31}$$

$$= \sum_{i=0}^{M-1} W_i(n') \sum_{l=0}^{L} y(n-l)\, h_i(l)$$

$$= \sum_{l=0}^{L} y(n-l)\, h(l) \sum_{i=0}^{M-1} W_i(n')\, \Phi(i,l) \tag{2.32}$$

for the modulated bandpass filters of Eq. 2.29. The second summation is the spectral transform of the gain factors $W_i(n')$ which yields the coefficients

$$w_l(n') = \sum_{i=0}^{M-1} W_i(n')\, \Phi(i,l)\,; \quad l \in \{\, 0, 1, \ldots, L \,\} \tag{2.33}$$

$$= \mathcal{T}\{\, W_i(n') \,\}\,. \tag{2.34}$$

These $L+1$ *time-domain weighting factors* have the periodicity

$$w_{l+mM}(n') = w_l(n')\, \rho(m) \tag{2.35}$$

due to Eq. 2.30 and Eq. 2.33. The input-output relation finally reads

$$v(n) = \sum_{l=0}^{L} y(n-l)\, h(l)\, w_l(n') \tag{2.36}$$

$$= \sum_{l=0}^{L} y(n-l)\, h_s(l,n')\,. \tag{2.37}$$

The obtained filter-bank structure is a *single* time-domain filter whose coefficients

$$h_s(l,n') = h(l)\, w_l(n')\,; \quad l \in \{\, 0, 1, \ldots, L \,\} \tag{2.38}$$

are the product of the fixed impulse response $h(n)$ of the prototype lowpass filter and the time-varying weighting factors $w_l(n')$ adapted in the short-term spectral-domain.[6] This efficient implementation of the FBSM (which resembles a filter-bank used as equalizer) is termed as *filter-bank equalizer* (FBE) [44, 82]. A sketch of this filter-bank structure is given in Fig. 2.11. A distinctive advantage in comparison to the AS FB is that the output signal $v(n)$ is not affected by aliasing distortions. Moreover, a non-uniform (warped) frequency resolution can be achieved by means of the allpass transformation with lower efforts than for the AS FB as shown later in Sec. 2.3.6. The uniform

[6] For the sake of clarity, the index l instead of the discrete time index n will be used to indicate that $L+1$ filter coefficients are considered.

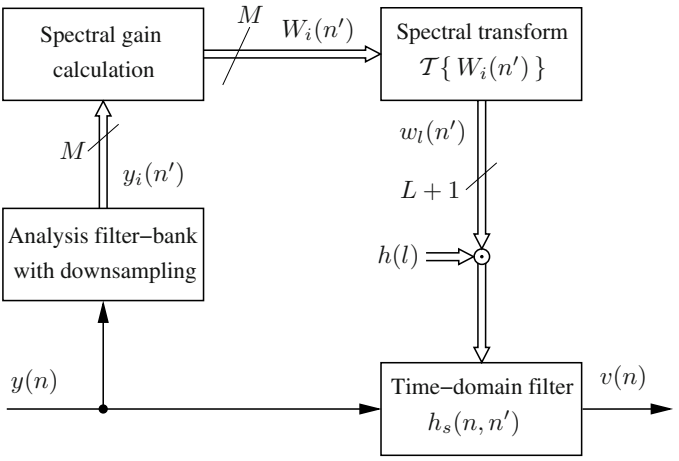


Fig. 2.11. Filter-bank equalizer (FBE) for adaptive subband filtering.

and non-uniform FBE can be used for speech and audio processing with low signal delay, e.g., [45, 68].

A similar approach has been proposed independently in [39] for dynamic-range compression in hearing aids. For acoustic echo cancellation and active noise control applications, a related time-domain filtering approach can be found in [57], where the coefficients are adapted in the uniform frequency-domain. However, the following treatment will show that the concept of the FBE is a much more general and comprehensive low delay filter-bank concept.

2.3.2 Prototype Filter Design

The objective of the prototype lowpass filter design is to achieve perfect reconstruction according to Eq. 2.2. The FBE meets this condition if the following two requirements are fulfilled [44]: Firstly, the general modulation sequence of Eq. 2.29 must have the property

$$\sum_{i=0}^{M-1} \Phi(i,n) = \begin{cases} C & ; \quad C \neq 0\,; \quad n = n_0 \\ 0 & ; \quad n \neq n_0 \end{cases} \quad \text{for} \quad n, n_0 \in \{0, 1, \ldots, M-1\}. \tag{2.39}$$

Secondly, a generalized *M th-band filter* with impulse response

$$h(n) = \begin{cases} \dfrac{1}{C\,\rho(m_c)} & ; \quad n = n_0 + m_c M\,; \quad \rho(m_c) \neq 0\,; \quad m_c \in \mathbb{Z} \\ 0 & ; \quad n = n_0 + m\,M\,; \quad m \in \mathbb{Z} \backslash \{m_c\} \\ \text{arbitrary} & ; \quad \text{else} \end{cases} \tag{2.40}$$

is needed as prototype lowpass filter. Such a filter has equidistant zeros at intervals of M samples and its modulated versions add up to a delay according to

$$\sum_{i=0}^{M-1} H\left(z\,E_M^i\right)\cdot E_M^{i\,n_0} = \frac{M}{C\,\rho(m_c)}\,z^{-(n_0+m_c\,M)}\,. \tag{2.41}$$

The conditions of Eq. 2.39 and Eq. 2.40 can be easily met to achieve perfect reconstruction with a delay of

$$d_0 = n_0 + m_c\,M \tag{2.42}$$

samples. A suitable Mth-band filter according to Eq. 2.40 is given by

$$h(n) = \frac{1}{C\,\rho(m_c)}\,\frac{\sin\left(\frac{2\,\pi}{M}\,(n-d_0)\right)}{\frac{2\,\pi}{M}\,(n-d_0)}\,\mathrm{win}_L(n) \tag{2.43}$$

with the general window sequence defined by

$$\mathrm{win}_L(n) = \begin{cases} \text{arbitrary} & ;\quad 0 \le n \le L \\ 0 & ;\quad \text{else}\,. \end{cases} \tag{2.44}$$

A rectangular window achieves a least-squares approximation error, but other window sequences are often preferred to influence properties of the filter such as transition bandwidth or sidelobe attenuation [59]. Commonly used window sequences are the Kaiser window or the parametric window sequence

$$\mathrm{win}_L(n,\beta) = \begin{cases} \beta + (\beta-1)\,\cos\left(\frac{2\,\pi}{L}\,n\right) & ;\quad 0 \le n \le L\,;\quad 0.5 \le \beta \le 1 \\ 0 & ;\quad \text{else}\,. \end{cases} \tag{2.45}$$

The rectangular window ($\beta = 1$), the Hann window ($\beta = 0.5$), and the Hamming window ($\beta = 0.54$) are included as special cases [63].

The condition of Eq. 2.39 is met, among others, by the Walsh and Hadamard transform (cf. [2]) as well as the *generalized discrete Fourier transform* (GDFT). The transformation kernel of the GDFT reads

$$\begin{aligned} &\Phi_{\mathrm{GDFT}}(i,n) = e^{-j\,\frac{2\,\pi}{M}\,(i-i_0)\,(n-n_0)} \\ &n, n_0 \in \mathbb{Z}\,;\quad i \in \{\,0,1,\ldots,M-1\,\}\,;\quad i_0 \in \{\,0,\,1/2\,\} \end{aligned} \tag{2.46}$$

where Eq. 2.30 applies with $\rho(m) = (-1)^{2\,i_0\,m}$. The DFT is included as special case for $n_0 = i_0 = 0$. For $i_0 = 1/2$, a GDFT filter-bank with oddly-stacked frequency bands is obtained, cf. [12]. A value of $i_0 = 0$ leads to the evenly-stacked GDFT where the above equations apply with $\rho(m) = 1$ and $C = M$.

The Walsh and Hadamard transform are employed, among others, for image processing, cf. [25]. The evenly-stacked GDFT is of interest for speech and audio processing. Thus, this filter-bank type is considered primarily in the following without loss of generality.

2.3.3 Relation between GDFT and GDCT

For speech enhancement, the time-varying spectral gain factors $W_i(n')$ are often calculated by means of a spectral speech estimator, e.g., [5, 20, 50, 52]. For a DFT-based adaptation, the gain factors have the properties

$$\epsilon \leq W_i(n') \leq 1\,; \quad W_i(n') \in \mathbb{R}\,; \quad 0 \leq \epsilon < 1 \tag{2.47}$$

and possess the symmetry

$$W_i(n') = W_{M-i}(n')\,; \quad i \in \{\,0, 1, \ldots, M-1\,\}\,; \quad M \text{ even} \tag{2.48}$$

as the input sequence $y(n)$ is real. The limitation of the gain factors by a lower (often time-varying) threshold ϵ is favorable to avoid unnatural sounding artifacts such as musical noise, cf. [8]. The (I)DFT of the real gain factors of Eq. 2.47 yields time-domain weighting factors $w_l(n')$ corresponding to a (non-causal) zero-phase filter. A linear phase response is obtained for the considered (evenly-stacked) GDFT of Eq. 2.46 with $n_0 = L/2$ and L being even so that the coefficients exhibit the symmetry

$$w_l(n') = w_{L-l}(n')\,; \quad l \in \{\,0, 1, \ldots, L\,\}\,. \tag{2.49}$$

If the used prototype filter has the same symmetry

$$h(l) = h(L-l)\,, \tag{2.50}$$

the time-varying FIR filter of Eq. 2.38 is given by

$$h_s(l, n') = h_s(L-l, n')\,; \quad l \in \{\,0, 1, \ldots, L\,\} \tag{2.51}$$

which implies a *linear phase response*. The GDFT of the gain factors $W_i(n')$ can be computed by the FFT with a subsequent cyclic shift of the obtained time-domain weighting factors by n_0 samples. Instead of the GDFT analysis filter-bank, the DFT filter-bank can be used for the FBE (see Fig. 2.11), because the magnitude of the subband signals is needed only for the calculation of the spectral gain factors.

For the considered GDFT, the weighting factors of Eq. 2.33 are given by

$$w_l(n') = \sum_{i=0}^{M-1} W_i(n')\, e^{-j\,\frac{2\pi}{M}\,i\,(l-n_0)}\,; \quad l \in \{\,0, 1, \ldots, L\,\}\,. \tag{2.52}$$

The substitution $M = 2N$ and exploiting the symmetry of Eq. 2.48 allows the following conversion

$$
\begin{aligned}
w_l(n') &= \sum_{i=0}^{2N-1} W_i(n')\, e^{-j\,\frac{2\pi}{2N}\,i\,(l-n_0)} \\
&= W_0(n') + \sum_{i=1}^{N-1} W_i(n')\, e^{-j\,\frac{\pi}{N}\,i\,(l-n_0)} + W_N(n')\,(-1)^{l-n_0} \\
&\qquad + \sum_{i=1}^{N-1} W_{2N-i}(n')\, e^{-j\,\frac{\pi}{N}\,(2N-i)\,(l-n_0)} \\
&= \sum_{i=0}^{N} W_i(n')\,\nu(i)\cos\left(\frac{\pi}{N}\,i\,(l-n_0)\right) \qquad (2.53a)
\end{aligned}
$$

with

$$
\nu(i) = \begin{cases} 1 & ;\quad i \in \{0, N\} \\ 2 & ;\quad i \in \{1, 2, \ldots, N-1\}\,. \end{cases} \qquad (2.53b)
$$

Eq. 2.53 represents a FBE with $N+1$ channels and the (evenly-stacked) *generalized discrete cosine transform* (GDCT) as modulation sequence[7]

$$
\Phi_{\mathrm{GDCT}}(n,i) = \nu(i)\cos\left(\frac{\pi}{N}\,i\,(n-n_0)\right)\;;\quad i \in \{\,0,1,\ldots,N\,\}\;;\quad n, n_0 \in \mathbb{Z}\,. \qquad (2.54)
$$

For this transformation kernel, the condition of Eq. 2.39 is fulfilled with $M = N+1$ and $C = 2N$.

The relation between GDCT and GDFT FBE has been derived so far without considering the process of the spectral gain calculation. For noise reduction, the spectral gain factors are mostly calculated as (linear or nonlinear) functions of the squared magnitude of the subband signals (spectral coefficients), cf. [4]. This can be expressed by the notation

$$
W_i(n') = f\left(\overline{|y_i(n')|^2}\right)\;;\quad i \in \{\,0,1,\ldots,N\,\}\,. \qquad (2.55)
$$

Only $N+1$ gain factors must be calculated due to the symmetry of Eq. 2.48. The bar indicates that an averaged value (short-term expectation) is mostly taken inherently. Examples are the calculation of the *a priori* SNR by the decision-directed approach [20], or the estimation of the noise power spectral density by recursively smoothed periodograms [54]. The subband signals are complex for the (G)DFT so that

$$
W_i(n') = f\left(\overline{\mathrm{Re}\,\{y_i(n')\}^2} + \overline{\mathrm{Im}\,\{y_i(n')\}^2}\right)\,. \qquad (2.56)
$$

[7] Except for a normalization factor, the DCT-I is obtained for $n_0 = 0$, cf. [66]. For the oddly-stacked GDFT FBE ($i_0 = 1/2$), a similar derivation leads to a modulation sequence which includes the DCT-II for $n_0 = 0$.

It can be assumed that the real and imaginary part are uncorrelated and that both have equal variances and equal probability density functions (PDFs), e.g., [4]. Therefore, almost the same gain factors are obtained by considering the real part of the subband signals only

$$W_i(n') \approx f\left(2\overline{\mathrm{Re}\{y_i(n')\}^2}\right) \tag{2.57}$$

due to the averaging process. Hence, the gain factors calculated by complex DFT values are almost equal to those computed by real DCT values. Accordingly, the replacement of the GDFT of Eq. 2.46 by the GDCT of Eq. 2.54 causes no noticeable differences for the speech enhanced by the FBE.[8]

2.3.4 Realization for Different Filter Structures

The choice of the filter structure plays an important role for digital filter implementations with finite precision arithmetic as well as for *time-varying* filters. Here, only the direct forms of a filter are of interest as they do not require an involved conversion of the time-varying filter coefficients $h_s(l, n')$ such as the parallel form or the cascade form, cf. [59].

The realization of an FIR filter by means of the direct form and transposed direct form is shown in Fig. 2.12. The input-output relations for these two filter forms are given by

$$v_{\mathrm{df}}(n) = \sum_{l=0}^{L} y(n-l)\, h_s(l, n') \tag{2.58}$$

$$v_{\mathrm{tdf}}(n) = \sum_{l=0}^{L} y(n-l)\, h_s(l, n'-l)\,. \tag{2.59}$$

Obviously, the derived FBE according to Eq. 2.37 uses a time-domain filter in the *direct form*.

The input-output relation for the transposed direct form is obtained by inserting Eq. 2.38 into Eq. 2.59 so that

$$\begin{aligned} v_{\mathrm{tdf}}(n) &= \sum_{l=0}^{L} y(n-l)\, h(l)\, w_l(n'-l) \\ &= \sum_{l=0}^{L} y(n-l)\, h(l) \sum_{i=0}^{M-1} W_i(n'-l)\, \Phi(i,l) \\ &= \sum_{i=0}^{M-1} \sum_{l=0}^{L} y(n-l)\, W_i(n'-l)\, h_i(l) \end{aligned} \tag{2.60}$$

[8] In [19], a different comparison between DFT AS FB and DCT-II AS FB has revealed a slightly lower noise suppression for the DCT AS FB.

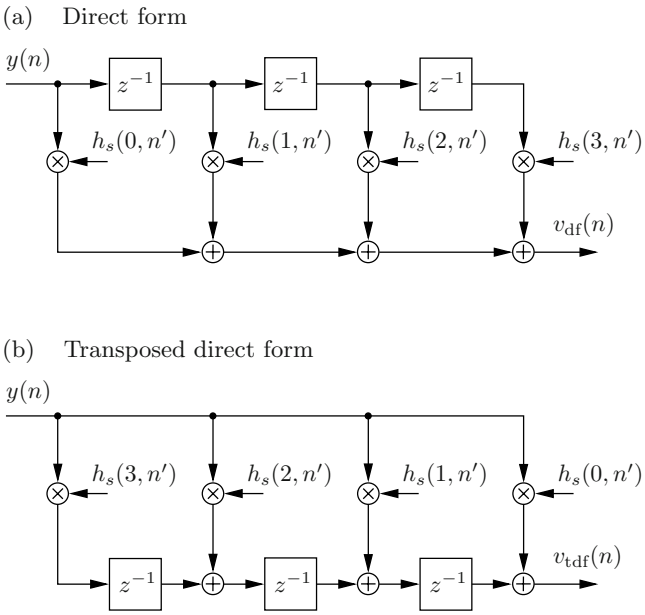


Fig. 2.12. Direct form implementations of a time-varying FIR filter with degree $L = 3$.

due to Eq. 2.33 and Eq. 2.29. The obtained relation for the *transposed direct form* corresponds to the FBSM of Fig. 2.10 with the important difference that the spectral gain factors are now applied *before* the subband filters. Fig. 2.13 shows the derived filter-bank structure. The dash-dotted boxes mark delay elements to account for the signal delay τ_a due to the analysis filter-bank and gain calculation. These delay elements might be omitted for moderately time-varying (smoothed) gain factors to avoid an additional signal delay.

Switching the coefficients of a digital filter during operation leads to transients which can cause 'filter-ringing' effects.[9] These effects might be perceived by perceptually annoying artifacts. The application to noise reduction revealed that the FBE with time-domain filter in transposed direct form yields a better perceived speech quality than the implementation with the direct form filter. This can be explained by comparing the equivalent FBSMs of Fig. 2.10 and Fig. 2.13: For the transposed direct form, the transients caused by the switching gain factors are smoothed by the following bandpass filters, which is not the case for the direct form implementation.

An alternative method to smooth the FIR filter coefficients independently of the filter form is to perform a kind of 'cross-fading' according to

[9] The term 'filter-ringing' is sometimes used with a slightly different meaning in the context of speech coding.

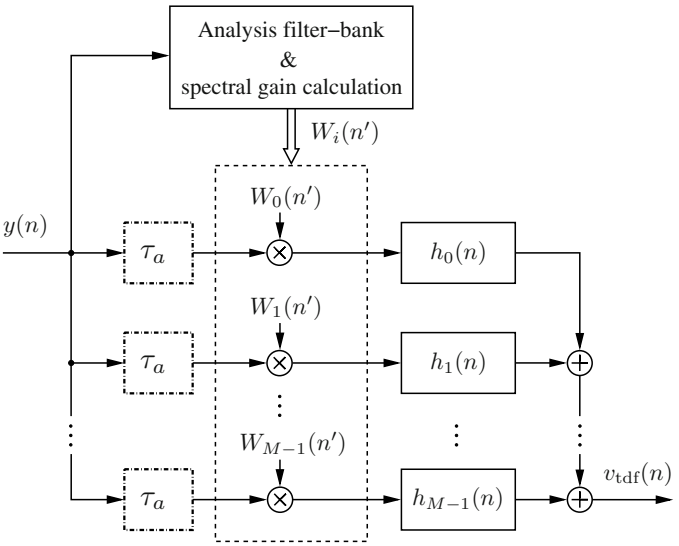


Fig. 2.13. Filter-bank summation method (FBSM) corresponding to the filter-bank equalizer (FBE) with time-domain filter in transposed direct form.

$$\bar{h}_s(l,n) = (1 - c_f(n)) \cdot h_s(l, n' - R) + c_f(n) \cdot h_s(l, n') \tag{2.61a}$$

$$c_f(n) = \frac{n - n'}{R}\,; \quad l \in \{\,0, 1, \ldots, L\,\} \tag{2.61b}$$

with n' is defined by Eq. 2.28. An existing linear-phase property is maintained. The proposed cross-fading method is very effective to avoid audible filter-ringing artifacts and is especially useful if the direct form filter is used.

It should be noted that artifacts due to time-varying spectral gain factors must also be avoided for AS FB systems, e.g., by non-critical subsampling, a dedicated prototype filter design, and smoothing of the gain factors (cf. Sec. 2.2.1).

2.3.5 Polyphase Network Implementation

An efficient polyphase network (PPN) implementation of the FBE shall be developed, which eases the utilization of prototype filters with a long or even infinite impulse response (IIR). This allows, for example, to improve the spectral selectivity of the subband filters in order to reduce the cross-talk between adjacent frequency bins. A low cross-talk can be favorable for some spectral speech estimators since most of them do not consider correlation between the frequency bands, cf. [55].

The FBE is a *time-varying* system. It can be described by the z-transform of the frozen-time impulse response which yields the so-called *frozen-time*

transfer function [42]. The direct form time-domain filter of Eq. 2.38 at sample instant n' has the frozen-time transfer function

$$H_s(z, n') = \sum_{l=0}^{L} w_l(n')\, h(l) z^{-l} \,. \tag{2.62}$$

This transfer function[10] can be expressed by means of the polyphase components of Eq. 2.11 which leads to

$$\begin{aligned} H_s(z, n') &= \sum_{\lambda=0}^{M-1} w_\lambda(n') \sum_{m=0}^{l_M-1} h(\lambda + m\,M)\, z^{-(\lambda+mM)}\, \rho(m) \\ &= \sum_{\lambda=0}^{M-1} w_\lambda(n') \cdot H_\lambda^{(M)}(z^M) \cdot z^{-\lambda} \end{aligned} \tag{2.63}$$

for $\rho(m) = 1$ and l_M defined by Eq. 2.12. The subband signals $y_i(n)$ of Eq. 2.31 are given by

$$y_i(n) = \sum_{l=0}^{L} y(n-l)\, h(l)\, \Phi(i, l) \,; \quad i \in \{\, 0, 1, \ldots, M-1 \,\} \,. \tag{2.64}$$

The z-transform leads to

$$Y_i(z) = Y(z) \sum_{l=0}^{L} h(l)\, z^{-l}\, \Phi(i, l) \,. \tag{2.65}$$

Applying Eq. 2.11 and Eq. 2.30 with $\rho(m) = 1$ results in

$$\begin{aligned} Y_i(z) &= Y(z) \sum_{\lambda=0}^{M-1} \sum_{m=0}^{l_M-1} h(\lambda + m\,M)\, z^{-(\lambda+m\,M)}\, \Phi(i, \lambda) \\ &= Y(z) \sum_{\lambda=0}^{M-1} z^{-\lambda} \cdot H_\lambda^{(M)}(z^M) \cdot \Phi(i, \lambda) \,. \end{aligned} \tag{2.66}$$

The derived PPN implementation of the *direct form* FBE according to Eq. 2.63 and Eq. 2.66 is illustrated in Fig. 2.14. In contrast to the FBE realization of Fig. 2.11, the time-domain filtering and calculation of the subband signals is partly done by the same network. The PPN realization for the oddly-stacked GDFT FBE can be derived in a similar manner (cf. [44]). The same applies for an implementation with type 2 polyphase components.

The transposed direct form of a filter is derived from the direct form representation by transposition of its signal flow graph [59]: Branch nodes

[10] For the sake of brevity, the term transfer function refers to the frozen-time transfer function or the conventional transfer function dependent on the context.

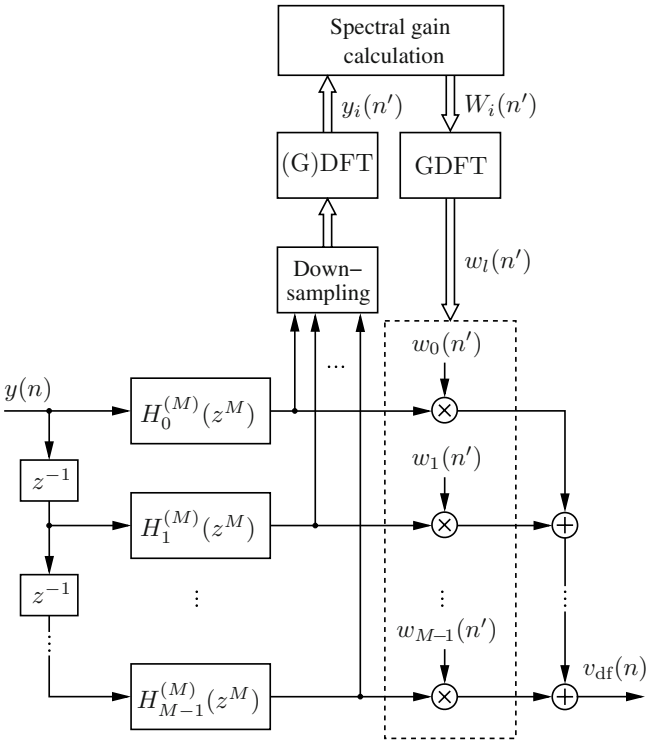


Fig. 2.14. Polyphase network (PPN) implementation of the GDFT FBE for the direct form time-domain filter.

and summations are interchanged as well as the system input and output. All signal directions are reversed. The obtained PPN implementation of the FBE for the *transposed direct form* is shown in Fig. 2.15. Delay elements might be inserted in each branch of the time-domain filter to account for the execution time τ_a to calculate the time-domain weighting factors $w_l(n')$. These weighting factors are calculated by a separate network similar to that of Fig. 2.14 but with the difference that the downsampling is performed directly after the delay elements. Thus, the PPN realization for the transposed direct form requires only a slightly higher algorithmic complexity than the direct form realization, which is discussed in Sec. 2.3.8 in more detail.

A polyphase network decomposition can be performed for FIR filters [3] as well as for IIR filters [78]. The design of IIR Mth-band filters is proposed in [67]. Hence, the developed PPN realization of the FBE enables a realization of Eq. 2.36 for L being infinite, that is, a *recursive* prototype filter.

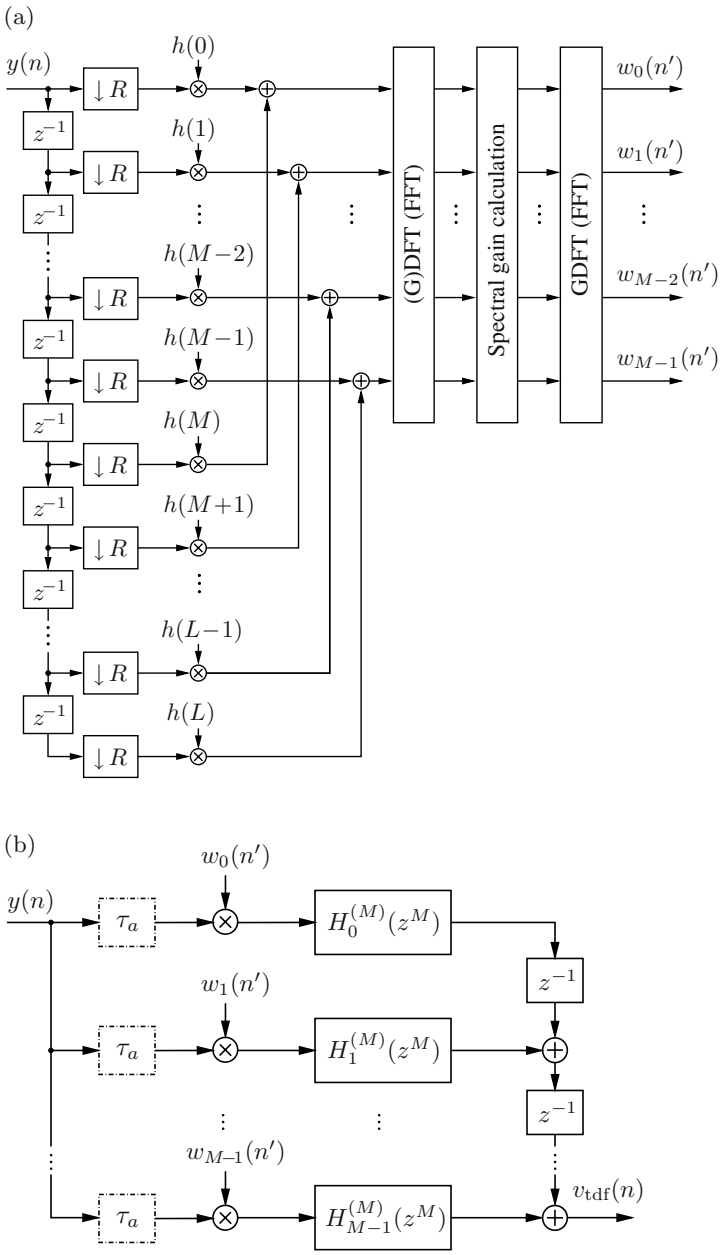


Fig. 2.15. PPN implementation of the GDFT FBE: (a) calculation of the time-domain weighting factors $w_l(n')$, (b) time-domain filter in transposed direct form.

2.3.6 The Non-Uniform Filter-Bank Equalizer

The allpass transformation can be applied to the FBE to achieve a non-uniform time-frequency resolution. The following treatment will show that the obtained allpass transformed FBE has some distinctive advantages in comparison to the corresponding allpass transformed AS FB[11] treated in Sec. 2.2.4.

2.3.6.1 Concept

The application of the allpass transformation to the bandpass filter of Eq. 2.29 yields the warped frequency responses

$$H_i(z = e^{j\,\varphi_a(\Omega)}) = \sum_{l=0}^{L} h(l)\,\Phi(i,l)\,e^{-j\,l\,\varphi_a(\Omega)} \tag{2.67}$$

$$= \widetilde{H}_i(e^{j\,\Omega})\,; \quad i \in \{\,0, 1, \ldots, M-1\,\}\,. \tag{2.68}$$

As before, the tilde-notation is used to mark quantities changed by the allpass transformation. The uniform FBE can be seen as a special case for $a = 0$ where $\varphi_a(\Omega) = \Omega$. The effect of this frequency warping on the frequency characteristic of the subband filters has been discussed already in Sec. 2.2.4. Fig. 2.16 provides a block diagram of the allpass transformed FBE. This warped FBE can be implemented efficiently by the PPN structures derived in Sec. 2.3.5 where the delay elements are substituted by allpass filters.

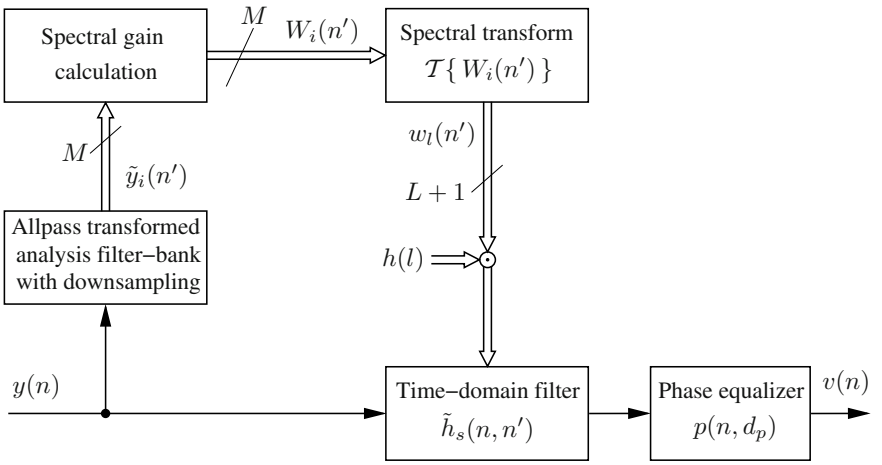


Fig. 2.16. Allpass transformed FBE for adaptive subband filtering.

[11] The corresponding AS FB uses the same type of analysis filter-bank as the FBE with identical values for L and M. An identical analysis filter-bank can not always be used due to different design constraints for signal reconstruction.

2.3.6.2 Compensation of Phase Distortions

The uniform FBE with perfect reconstruction fulfills Eq. 2.2, which can be expressed in the frequency-domain by the relation

$$V(e^{j\,\Omega}) = Y(e^{j\,\Omega}) \cdot e^{-j\,d_0\,\Omega}\,. \tag{2.69}$$

If the allpass transformation is applied, this relation turns into

$$\widetilde{V}(e^{j\,\Omega}) = Y(e^{j\,\Omega}) \cdot e^{-j\,d_0\,\varphi_a(\Omega)}\,. \tag{2.70}$$

Thus, the input signal $y(n)$ is filtered by an allpass chain of length d_0. Perfect reconstruction is achieved for a linear phase characteristic. This can be accomplished with an arbitrarily small error by means of a (fixed) phase equalizer connected to the output of the FBE as indicated in Fig. 2.16 [44, 48]. The design of the phase equalizer is similar to that for the AS FB structure II of Fig. 2.7(b). However, a phase equalizer of significantly lower degree is needed due to the lower signal delay d_0 in comparison to the AS FB.

The described phase compensation is also effective for *time-varying* filter coefficients if the symmetry of Eq. 2.48 holds. For the warped FBE with direct form filter, the (frozen-time) frequency response reads

$$\widetilde{H}_s(e^{j\,\Omega}, n') = \sum_{l=0}^{L} h_s(l, n')\, e^{-j\,l\,\varphi_a(\Omega)}\,. \tag{2.71}$$

If the real filter coefficients have the symmetry of Eq. 2.51, it can be shown that the transfer function of Eq. 2.71 can be expressed by

$$\widetilde{H}_s(e^{j\,\Omega}, n') = e^{-j\,\frac{L}{2}\,\varphi_a(\Omega)} \begin{cases} \sum\limits_{l=0}^{\frac{L}{2}} 2\,\mathcal{A}(\Omega, l, L, n') - h_s\left(\frac{L}{2}, n'\right) & ;\ L \text{ even} \\ \sum\limits_{l=0}^{\frac{L-1}{2}} 2\,\mathcal{A}(\Omega, l, L, n') & ;\ L \text{ odd} \end{cases} \tag{2.72a}$$

$$\mathcal{A}(\Omega, l, L, n') = h_s(l, n') \cos\left(\left[\frac{L}{2} - l\right] \varphi_a(\Omega) \right)\,. \tag{2.72b}$$

The non-linear phase term $\varphi_a(\Omega)\, L/2$ can be compensated by a phase equalizer which has to fulfill Eq. 2.23 with $d_P = L/2$ [48]. The expressions to the right of the curly brace are real and cause only phase shifts of $\pm\pi$. (This can also be shown for the more general case of a complex prototype filter with linear phase response, cf. [44].) Thus, the system has a *generalized linear phase response* despite the time-varying coefficients, if a sufficient phase compensation is performed. A system with a generalized linear phase response has a constant group delay, if the discontinuities that result from the addition of constant phase shifts due to the real function are neglected [59]. For

the transposed direct form filter, Eq. 2.72 is approximately fulfilled in case of moderately time-varying coefficients.

For speech and audio processing, a phase equalizer might be omitted for smaller phase distortions ($d_P < 20$) as the human ear is relatively insensitive towards minor phase distortions.

2.3.7 Comparison between FBE and AS FB

A comparison with the uniform and warped AS FB of Sec. 2.2 shows that the concept of the FBE exhibits the following benefits:

- The FBE is not affected by aliasing distortions for the reconstructed signal. Hence, the difficult problem of achieving a sufficient aliasing cancellation, especially for time-varying spectral gain factors, does not occur. One consequence is that the FBE does not require a prototype filter with high degree and/or a low subsampling rate R to reduce aliasing distortions. Moreover, the prototype filter design for the FBE is less complex than for the AS FB.
- The allpass transformation of the FBE does not cause aliasing distortions for the reconstructed signal. Thus, only the phase modifications due to the frequency warping need to be compensated. Therefore, near-perfect reconstruction can be achieved with an arbitrarily small error and much lower efforts than for the AS FB.
- The signal delay d_0 for the FBE is significantly lower than for the corresponding AS FB. The uniform AS FB with linear-phase prototype filter has a signal delay of $d_0 = L$. In contrast, the uniform FBE with linear-phase prototype filter has a signal delay of $d_0 = L/2$, which can be further reduced by a non-linear phase filter. Accordingly, the design objective of Eq. 2.23 for the phase equalizer applies with $d_p = L$ for the warped AS FB (structure II) and with $d_P = L/2$ for the warped FBE. Therefore, a phase equalizer with a significantly lower filter degree can be used for the FBE which results a lower signal delay.
- The warped FBE can achieve an almost linear overall phase characteristic even for time-varying coefficients, which can be beneficial for multi-channel processing.

A drawback of the FBE is its higher algorithmic complexity for some configurations, which is exposed in the following.

2.3.8 Algorithmic Complexity

Tab. 2.1 contrasts the algorithmic complexity of the developed uniform and warped FBE to that of the corresponding uniform and warped AS FB for the same values L, M, R, and N_p. The real DFT of size M can be computed in-place by a radix-2 FFT, e.g., [63]. The FFT of a real sequence of size M can

Table 2.1. Algorithmic complexity for different realizations of a polyphase network (PPN) DFT filter-bank with real valued prototype filter of length $L+1 = l_M\, M$.

	2 real FFTs	Remaining operations	Additional operations due to allpass transformation
AS FB			
Multiplications	$2\,\frac{M}{R}\,\log_2 M$	$\frac{1}{R}(2\,L+2+M)$	$4\,L+N_p+1$
Summations	$3\,\frac{M}{R}\,\log_2 M$	$\frac{1}{R}(L-M+1)+L$	$4\,L+N_p$
Delay elements	$2\,M$	$2\,L$	N_p
Direct form FBE			
Multiplications	$2\,\frac{M}{R}\,\log_2 M$	$L+1+M$	$2\,L+N_p+1$
Summations	$3\,\frac{M}{R}\,\log_2 M$	L	$2\,L+N_p$
Delay elements	$2\,M$	L	N_p
Transposed direct form FBE			
Multiplications	$2\,\frac{M}{R}\,\log_2 M$	$L+1+M+\frac{L}{R}$	$4\,L+N_p+1$
Summations	$3\,\frac{M}{R}\,\log_2 M$	$L+\frac{1}{R}(L+1-M)$	$4\,L+N_p$
Delay elements	$2\,M$	$2\,L$	N_p

be calculated by a complex FFT of size $M/2$, which requires approximately half the algorithmic complexity as a complex M-point FFT [62]. The GDFT can be computed by the FFT with similar complexity as for the DFT.

The last column contains the additional operations and delay elements due to the allpass transformation. The implementation of an allpass filter according to Fig. 2.4 is considered. This requires two real multiplications, two real summations, and one delay element for a real allpass coefficient $a = \alpha$. A LS phase equalizer of degree N_P is applied to compensate phase distortions.

It should be noted that allpass transformed filter-banks are usually operated with a smaller number of channels M than uniform filter-banks. As reasoned before, a higher subsampling rate R and a lower phase equalizer degree N_p are needed for the warped FBE in comparison to the warped AS FB. Therefore, the warped FBE has a lower algorithmic complexity than the corresponding warped AS FB for most parameter configurations. Contrariwise, the uniform AS FB has a lower complexity than the uniform FBE. A design example for these filter-banks is given later in Sec. 2.5.

2.4 Further Measures for Signal Delay Reduction

Even though the FBE causes only about half the algorithmic signal delay than the corresponding AS FB, a further reduced delay might be required for applications with very demanding system delay constraints. One example are

the initially mentioned hearing aid devices. For such cases, a modification of the FBE concept will be discussed, which allows a further lowering of the signal delay and computational complexity with almost no loss for the perceived quality of the enhanced speech.

2.4.1 Concept

One approach to reduce the signal delay of a filter-bank is to reduce the transform size M to allow for a lower prototype filter degree and to adjust the gain calculation to the altered time-frequency resolution (smoothing factors etc.), e.g., [28]. For the FBE, a further reduction of the signal delay can also be accomplished by approximating the original time-domain filter by a filter of lower degree. This approach offers a greater flexibility in the choice of the time-domain filter and requires no adjustment of the spectral gain calculation. The principle is illustrated in Fig. 2.17. In contrast to the FBE of Fig. 2.11, an additional module for the filter approximation is inserted, which evaluates the new $L_D + 1$ filter coefficients $a_l(n')$ from the $L + 1$ original FIR filter coefficients $h_s(l, n')$.

In the following, we investigate an FIR and IIR filter approximation for the uniform FBE first, before the results are extended to the more general case of allpass transformed filters.

2.4.2 Approximation by a Moving-Average Filter

The time-domain filter of the FBE can be approximated by an FIR filter of lower degree $L_D < L$ following a technique very similar to FIR filter design by

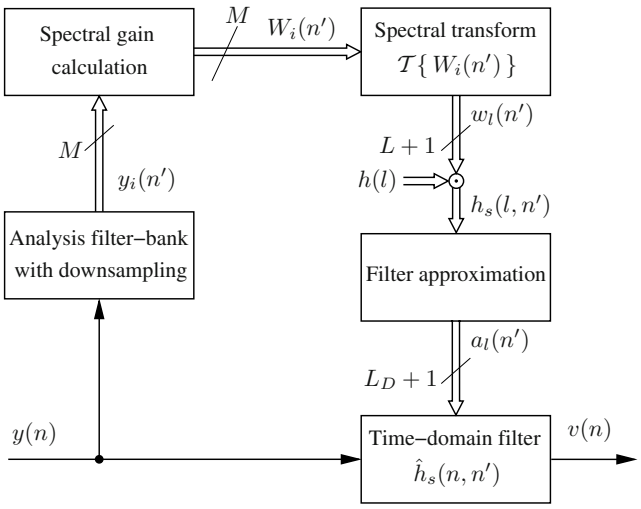


Fig. 2.17. Modification of the FBE to achieve a further reduced signal delay.

windowing, e.g., [63]. The impulse response $h_s(l, n')$ of Eq. 2.38 is truncated by a window sequence of length $L_D + 1$. This yields the FIR filter coefficients

$$a_l(n') = \hat{h}_s(l, n') = h_s(l + l_c, n') \, \mathrm{win}_{L_D}(l) \, ; \quad l \in \{ 0, 1, \ldots, L_D \} \tag{2.73}$$

where the window sequence is defined by Eq. 2.44. The value for l_c determines the part of the impulse response to be truncated, e.g., to maintain the symmetry of Eq. 2.51 for linear-phase filters. The truncation by a window results in a smoothed frequency response which is influenced by the choice of the window sequence, cf. [59].

This modified FBE based on an FIR filter approximation is named as *moving-average low delay filter* (MA LDF) [46]. The term 'low delay filter' (LDF) refers to the overall system according to Fig. 2.17, while the term 'MA filter' denotes the actual time-domain filter.

2.4.3 Approximation by an Auto-Regressive Filter

Instead of a (linear-phase) FIR filter, a *minimum-phase* IIR filter is now considered for the filter approximation. A filter can always be decomposed into an allpass filter and a minimum-phase filter, e.g., [59]. The group delay of the minimum-phase filter is lower than or equal to the group delay of the original filter for all frequencies. The approximation of a mixed-phase filter by a minimum-phase filter leads to a different phase response. This effect, however, is mostly tolerable for speech and audio processing as the human ear is relatively insensitive towards phase modifications, cf. [80, 84]. An allpole filter or auto-regressive (AR) filter is used here, because the calculation of its coefficients demands a lower computational complexity than for a general minimum-phase IIR filter (with zeros outside the origin), cf. [63]. The filter to be approximated contains no sharp zeros in its spectrum, if the threshold ϵ in Eq. 2.47 is greater than zero, which further supports the idea of an AR filter approximation.

The transfer function of the considered AR filter of degree L_D is given by

$$\hat{H}_s(z, n') = H_{\mathrm{AR}}(z, n') = \frac{a_0(n')}{1 - \sum_{l=1}^{L_D} a_l(n') \, z^{-l}} \, . \tag{2.74}$$

The AR filter coefficients can be determined by the Yule-Walker equations, e.g., [63]

$$\begin{bmatrix} \varphi_{\bar{h}\bar{h}}(1) \\ \vdots \\ \varphi_{\bar{h}\bar{h}}(L_D) \end{bmatrix} = \begin{bmatrix} \varphi_{\bar{h}\bar{h}}(0) & \ldots & \varphi_{\bar{h}\bar{h}}(1 - L_D) \\ \vdots & \ddots & \vdots \\ \varphi_{\bar{h}\bar{h}}(L_D - 1) & \ldots & \varphi_{\bar{h}\bar{h}}(0) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_{L_D} \end{bmatrix} \tag{2.75}$$

where the dependence on n' is omitted for the sake of simplicity. The $L_D + 1$ auto-correlation coefficients $\varphi_{\bar{h}\bar{h}}(\lambda)$ are computed by the rule[12]

$$\varphi_{\bar{h}\bar{h}}(\lambda) = \sum_{l=0}^{L-|\lambda|} \bar{h}(l)\,\bar{h}(l+\lambda)\,; \quad 0 \le |\lambda| \le L_D \tag{2.76a}$$

$$\bar{h}(l) = h_s(l)\,\mathrm{win}_L(l)\,; \quad l \in \{\,0, 1, \ldots, L\,\}\,. \tag{2.76b}$$

The scaling factor a_0 in Eq. 2.74 is given by

$$a_0 = \sqrt{\varphi_{\bar{h}\bar{h}}(0) - \sum_{l=1}^{L_D} a_l\,\varphi_{\bar{h}\bar{h}}(l)} \tag{2.77}$$

and ensures that the AR filter and the original filter have both the same amplification. The calculation of the auto-correlation coefficients according to Eq. 2.76 guarantees a symmetric Toeplitz structure for the auto-correlation matrix of Eq. 2.75. This allows to solve the Yule-Walker equations efficiently by means of the Levinson-Durbin recursion. The auto-correlation matrix is positive-definite so that the obtained AR filter has minimum-phase property, which implies a stable filter, cf. [63].

The devised modification of the FBE is named as *auto-regressive low delay filter* (AR LDF) in analogy to the terminology of the previous section [46].

2.4.4 Algorithmic Complexity

The algorithmic complexity for the low delay filter concept – in terms of computational complexity and memory consumption – is listed in Tab. 2.2. The complexity for the calculation of the original filter coefficients $h_s(l, n')$ has been discussed in Sec. 2.3.8. The variable $\mathcal{M}_{\mathrm{div}}$ marks the number of multiplications needed for a division operation, and $\mathcal{M}_{\mathrm{sqrt}}$ represents the number of multiplications needed for a square-root operation. Accordingly, the variables $\mathcal{A}_{\mathrm{div}}$ and $\mathcal{A}_{\mathrm{sqrt}}$ denote the additions needed for a division and square-root operation, respectively. Their values depend on the numeric procedure and accuracy used to execute these operations. (An equivalent of 15 operations will be assigned to these variables for the later complexity assessment in Sec. 2.5.). A rectangular window is assumed for Eq. 2.76b.

Most of the computational complexity for the AR filter conversion is required to compute the $L_D + 1$ auto-correlation coefficients according to Eq. 2.76. A lower computational complexity can be achieved by calculating Eq. 2.76 by means of the fast convolution or the Rader algorithm [65] with savings dependent on L_D and L.

[12] An alternative to this 'auto-correlation method' is the use of the 'covariance method'. However, this results in a more complex procedure for the calculation of the AR filter coefficients, cf. [81].

Table 2.2. Algorithmic complexity for the MA and AR low delay filter (LDF).

Calculation of $h_s(l, n')$ and MA/AR filtering	
Multiplications	$\frac{1}{R}(2M \log_2 M + 2L + 2) + L_D + 1$
Summations	$\frac{1}{R}(3M \log_2 M + L + 1 - M) + L_D$
Delay elements	$L + 2M + L_D$
Calculation of MA filter coefficients $a_l(n')$	
Multiplications	$\frac{1}{R}(L_D + 1)$
Summations	0
Registers	0
Calculation of AR filter coefficients $a_l(n')$	
Multiplications	$\frac{1}{R}\left((L_D + 1)(L + 4) + L_D(\mathcal{M}_{\text{div}} + \mathcal{M}_{\text{sqrt}})\right)$
Summations	$\frac{1}{R}\left((L_D + 1)(L + 2) + L_D(\mathcal{A}_{\text{div}} + \mathcal{A}_{\text{sqrt}})\right)$
Registers	$3L_D$

The MA filter conversion needs no multiplications, if a rectangular window is used for Eq. 2.73. However, the AR filter degree is usually chosen significantly lower than the MA filter degree so that both approaches have a similar overall algorithmic complexity as exemplified later in Sec. 2.5.

2.4.5 Warped Filter Approximation

The discussed filter approximations can also be applied to the allpass transformed FBE [47]. In the process, the delay elements of the analysis filter-bank and the time-domain filter are replaced by allpass filters according to Eq. 2.15. For the obtained warped MA LDF, a phase equalizer can be applied to obtain an approximately linear phase response according to Sec. 2.3.6.2.

The direct realization of the warped AR filter is not possible since the allpass transformation leads to delayless feedback loops. Different approaches have been proposed to solve this problem for a real allpass transformation ($a = \alpha$) [29, 30, 74]. Here, the algorithm of Steiglitz [74] is preferred due to its low computational efforts for time-varying filters. The modified transfer function of the allpass transformed AR filter is given by

$$\widetilde{H}_{\text{AR}}(z) = \frac{a_0\,\tilde{a}_0}{1 - \tilde{a}_0\,\dfrac{(1-\alpha^2)\,z^{-1}}{1-\alpha\,z^{-1}}\,\displaystyle\sum_{l=1}^{L_D}\tilde{a}_l\,H_A(z)^{l-1}} \tag{2.78}$$

with coefficients $\tilde{a}_l$ calculated by the recursion

$$\tilde{a}_{L_D} = a_{L_D} \tag{2.79a}$$

$$\tilde{a}_l = a_l - \alpha\,\tilde{a}_{l+1}\,; \quad l = L_D - 1, \ldots, 1 \tag{2.79b}$$

$$\tilde{a}_0 = \frac{1}{1 + \tilde{a}_1\,\alpha}\,. \tag{2.79c}$$

The computation of the new filter coefficients $\tilde{a}_l(n')$ needs only L_D multiplications, L_D summations, and one division at intervals of R sample instants. The warped AR filter according to Eq. 2.78 requires $3L_D + 2$ real multiplications and $3L_D$ real summations per sample instant as well as $L_D + 1$ delay elements.

It can be proven that the minimum-phase property of the AR filter is always maintained for a real allpass transformation but not for a complex allpass transformation in general. This is an important result as it guarantees stability for the warped AR filter. The use of a fixed phase equalizer (as for the warped MA LDF) is neither feasible nor required.

The cross-fading approach of Eq. 2.61 can not be applied to the coefficients of an IIR filter. Instead, we use a second filter with the previous coefficients to achieve a smooth transition by a cross-fading of both output signals. This general approach can be expressed by

$$\bar{H}_{\mathrm{g}}(z, n) = \big(1 - c_f(n)\big) \cdot H_{\mathrm{g}}(z, n' - R) + c_f(n) \cdot H_{\mathrm{g}}(z, n') \tag{2.80a}$$

$$c_f(n) = \frac{n - n'}{R}\,. \tag{2.80b}$$

A second filter is required for this smoothing technique which can be applied to arbitrary filters and does not cause an additional signal delay.

2.5 Application to Noise Reduction

In this section, the treated filter-bank designs are employed for noise reduction to compare the achieved performance with regard to speech quality, computational complexity, and signal delay.

2.5.1 System Configurations

The filtering of the noisy speech is done by the DFT AS FB, the GDFT FBE, and the MA/AR LDF. The uniform and allpass transformed version of these filter-banks are used each.[13] A real allpass coefficient of $a = 0.4$ is considered, which yields a good approximation of the Bark scale for a sampling frequency

[13] The low delay filter of Sec. 2.4 can be seen as a filter-bank system as it is derived from the FBE or FBSM, respectively.

of 8 kHz [73]. In all cases, a transform size of $M = 64$ and a linear-phase FIR prototype filter of degree $L = 64$ are used.[14]

A square-root Hann window derived from Eq. 2.45 is employed as common prototype filter for the DFT AS FB. The uniform AS FB uses a subsampling rate of $R = 32$ and the warped AS FB a rate of $R = 8$. A higher subsampling rate R can be used for the warped AS FB as well. However, this increases the signal delay significantly since subband filters with a higher stopband attenuation are needed to achieve a sufficient aliasing cancellation. A LS phase equalizer with a filter degree of $N_p = 141$ is applied to the filter-bank output according to Sec. 2.2.4.2.

The GDFT FBE is implemented in the transposed direct form. The MA LDF possesses a filter degree of $L_D = 48$. The LS FIR phase equalizer with filter degree $N_p = 80$ and $N_p = 56$ is applied to the warped FBE and the warped MA LDF, respectively. The considered AR LDF has a filter degree of $L_D = 16$. The cross-fading technique is used to avoid filter-ringing artifacts. A subsampling rate of $R = 64$ is used for the analysis filter-banks of FBE and LDF.

The spectral gain factors are computed by the super-Gaussian joint MAP estimator [50]. This joint spectral amplitude and phase estimator is derived by the more accurate assumption that the real and imaginary parts of the speech DFT coefficients are rather Laplace distributed (considered here) or Gamma distributed than Gaussian distributed. The needed *a priori* SNR is determined by the decision-directed approach [20] with a fixed smoothing parameter of 0.9. The short-term noise power spectral density is estimated by minimum statistics [54]. Speech presence uncertainty is taken into account by applying soft-gains [52]. Independent of the subsampling rate R of the filter-bank, the spectral gain factors are always adapted at intervals of 64 sample instants and no individual parameter tuning is performed to ease the comparison.

2.5.2 Instrumental Quality Measures

The used audio signals of 8 kHz sampling frequency are taken from the noisy speech corpus NOIZEUS presented in [32]. A total of 20 sentences spoken by male and female speakers is used, each disturbed by five different, instationary noise sequences (airport, babble, car, station, and street noise) with signal-to-noise ratios (SNRs) between 0 dB and 15 dB.

The quality of the enhanced speech is evaluated by informal listening tests and instrumental quality measures. (An overview of this topic is provided by Chap. 9.) A common time-domain measure for the quality of the enhanced speech $v(n) = \hat{s}(n)$ is given by the *segmental SNR*

[14] A lower number of frequency channels can be used for warped filter-banks whereas a value of $M = 256$ is often preferred for speech enhancement using the uniform DFT filter-bank (at 8 kHz sampling frequency). However, such different configurations are not considered to ease the comparison of the filter-banks.

$$\mathrm{SNR}_{\mathrm{seg}}/\,\mathrm{dB} = \frac{1}{\mathcal{C}(\mathbb{F}_s)} \sum_{m \in \mathbb{F}_s} 10\,\log_{10} \left(\frac{\sum_{\mu=0}^{K-1} s^2(m\,K + \mu - \tau_0)}{\sum_{\mu=0}^{K-1} \left(\hat{s}(m\,K + \mu) - s(m\,K + \mu - \tau_0)\right)^2} \right) . \quad (2.81)$$

The calculation comprises only frames with speech activity ($m \in \mathbb{F}_s$) whose total number is denoted by $\mathcal{C}(\mathbb{F}_s)$.

In a simulation, the clean speech $s(n)$ and the additive background noise $b(n)$ can be filtered separately with coefficients adapted for the disturbed speech $y(n) = s(n) + b(n)$. This provides the filtered speech $\bar{s}(n)$ and filtered noise $\bar{b}(n)$ separately, where

$$v(n) = \hat{s}(n) = \bar{s}(n) + \bar{b}(n)\,. \quad (2.82)$$

The algorithmic *signal delay* of non-linear phase systems is determined here by the maximum of the cross-correlation between the clean speech $s(n)$ and the processed speech $\bar{s}(n)$ (due to their strong correlation) according to

$$\tau_0 = \arg\, \max_{\lambda \in \mathbb{Z}} \left\{ \varphi_{s\bar{s}}(\lambda) \right\} . \quad (2.83)$$

The achieved *segmental noise attenuation* is calculated by the expression

$$\mathrm{NA}_{\mathrm{seg}}/\,\mathrm{dB} = \frac{1}{\mathcal{C}(\mathbb{F})} \sum_{m \in \mathbb{F}} 10 \log_{10} \left(\frac{\sum_{\mu=0}^{K-1} b^2(m\,K + \mu - \tau_0)}{\sum_{\mu=0}^{K-1} \bar{b}^2(m\,K + \mu)} \right) \quad (2.84)$$

where $\mathbb{F}$ marks the set of all frame indices including speech pauses, and $\mathcal{C}(\mathbb{F})$ denotes the total number of frames.

A frequency-domain measure for the speech quality is provided by the mean *cepstral distance* (CD) between the clean speech $s(n)$ and the processed speech $\bar{s}(n)$, cf. [64]. For all instrumental measures, a frame size of $K = 256$ is used, and 40 cepstral coefficients are considered for the CD measure.

2.5.3 Simulation Results for the Uniform Filter-Banks

The instrumental speech quality obtained with the different uniform filter-banks is plotted in Fig. 2.18. The uniform FBE achieves the same (or even better) objective speech quality as the uniform AS FB. Tab. 2.3 reveals that the FBE possesses a slightly higher algorithmic complexity but achieves a significantly lower signal delay. The MA and AR LDF achieve a further reduction of the signal delay and algorithmic complexity. Contrary to the MA LDF, the enhancement by the AR LDF leads to a significantly decreased *objective* speech quality. The AR filter approximation causes phase modifications

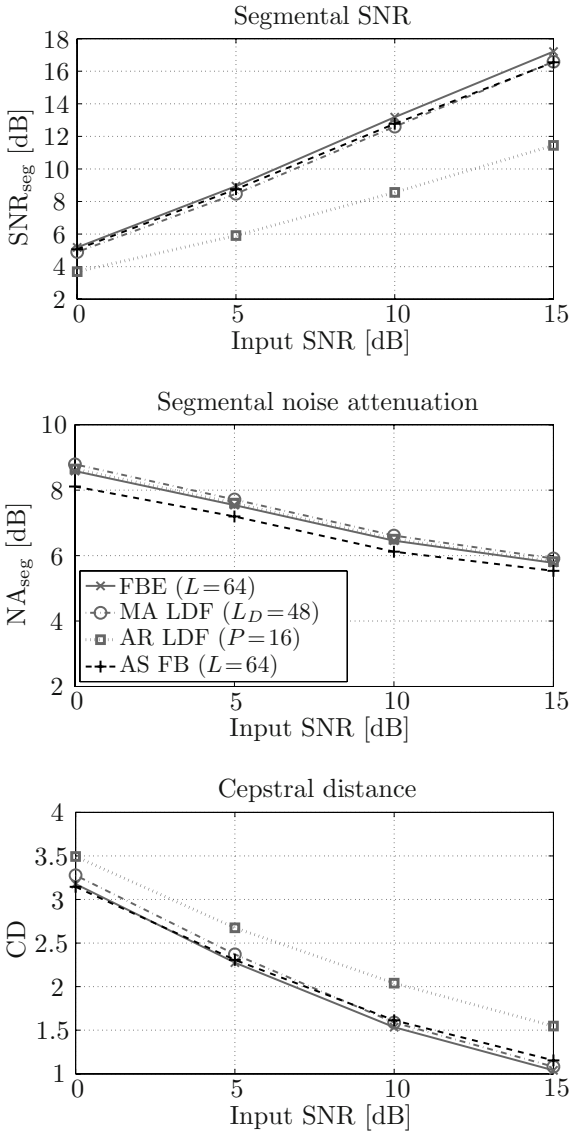


Fig. 2.18. Objective quality measures obtained with the uniform filter-banks.

which have a very detrimental effect on the segmental SNR measure. (Such an effect can also be observed for warped filter-banks with an imperfect phase compensation.) However, informal listening tests have revealed only negligible differences for the perceived *subjective* speech quality. Therefore, a perceptual evaluation of the speech quality (PESQ) according to [33] has been conducted in addition. This PESQ measure ranges from −0.5 (bad quality) to 4.5

Table 2.3. Measured signal delay and average algorithmic complexity per sample for the uniform filter-banks ($M = L = 64$).

Uniform filter-bank	Signal delay [samples]	Summations (real)	Multiplications (real)	Delay elements
AS FB	64	101	31	256
FBE	32	83	142	256
MA LDF	24	67	64	240
AR LDF	0-2	75	74	272

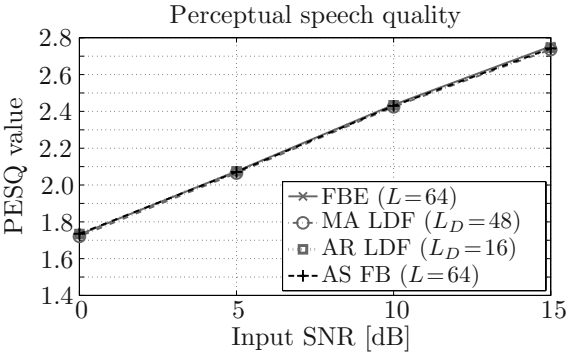


Fig. 2.19. Perceptual evaluation of the speech quality for the enhanced speech $\hat{s}(k)$ achieved with the uniform filter-banks.

(excellent quality). The PESQ measure is mainly used for the assessment of speech codecs, but also employed as a perceptual quality measure for speech enhancement systems, e.g., [4]. The measured PESQ values in Fig. 2.19 show that all four filter-banks achieve an almost identical perceptual speech quality. The PESQ measure is no all-embracing quantity for the subjective speech quality, but it complies well with the results of our informal listening tests. Thus, the low delay filter concept is suitable to achieve a further reduced signal delay in a flexible and simple manner with negligible loss for the perceived (subjective) speech quality.

2.5.4 Simulation Results for the Warped Filter-Banks

The curves for the objective speech quality obtained by means of the different warped filter-banks are plotted in Fig. 2.20. The measured PESQ values are not plotted again since they are as close together as in Fig. 2.19 but all about 0.25 PESQ units higher. Thus, the warped filter-banks achieve an improved instrumental speech quality in comparison to the corresponding uniform filter-banks. These results comply with our informal listening tests where the speech enhanced by the warped filter-banks was rated to be superior.

The measured signal delay and algorithmic complexity of the used allpass transformed filter-banks are listed in Tab. 2.4. It shows the increase of the

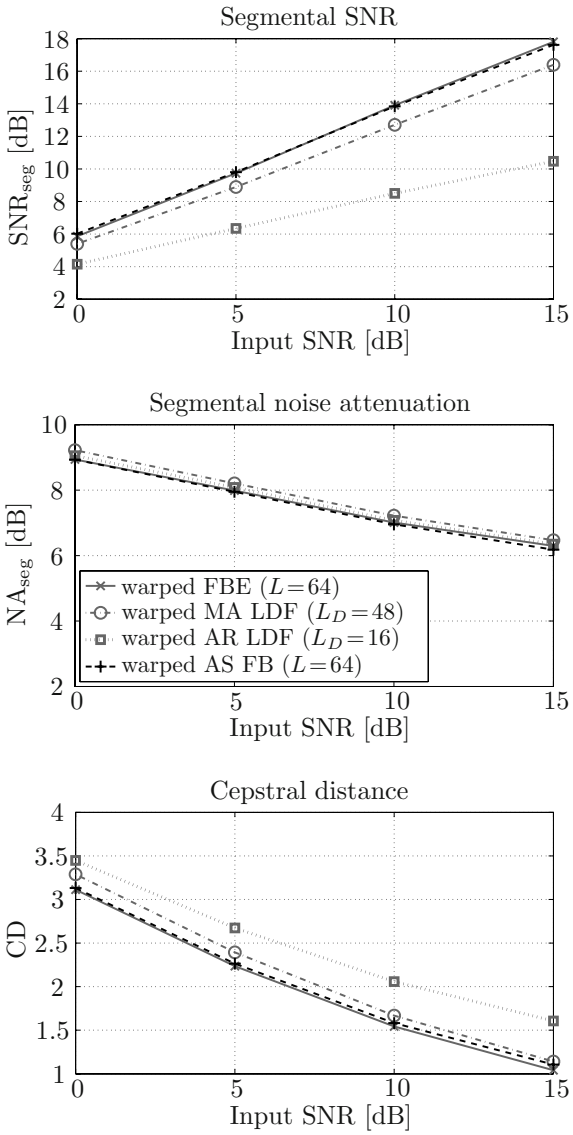


Fig. 2.20. Objective quality measures obtained by the allpass transformed filterbanks.

signal delay and algorithmic complexity due to the allpass transformation if the *same* values for M and L are taken (see also Sec. 2.3.8). The warped FBE causes a significantly lower signal delay and possesses a lower algorithmic complexity than the corresponding warped AS FB, but achieves the same objective and subjective speech quality.

Table 2.4. Measured signal delay and average algorithmic complexity per sample for the allpass transformed filter-banks ($M = L = 64$).

Warped filter-bank	Signal delay [samples]	Summations (real)	Multiplications (real)	Delay elements
AS FB	141	605	518	397
FBE	80	418	478	336
MA LDF	56	347	335	296
AR LDF	0-2	268	268	274

As for the uniform filter-banks, a further reduction of the signal delay and algorithmic complexity can be achieved by the low delay filter approximation with no loss for the subjective speech quality. The AR LDF is a minimum-phase system and causes a very low signal delay of only a few samples.

2.6 Conclusions

Filter-bank systems used for speech and audio processing have to fulfill several, partly conflicting requirements. A low signal delay and low algorithmic complexity are important for many applications such as mobile communication devices or digital hearing aids. A non-uniform, Bark-scaled frequency resolution is desirable to achieve a high speech and audio quality with a small number of frequency bands.

In this chapter, we have investigated different design approaches for such filter-banks. The main focus lies on allpass transformed filter-bank systems. These frequency warped filter-banks are a generalization of the uniform filter-bank. They are attractive for speech and audio processing due to their ability to mimic the Bark frequency bands of human hearing with great accuracy. However, the use of an allpass transformed analysis-synthesis filter-bank (AS FB) leads to a high signal delay as well as a high algorithmic complexity. This is attributed to the fact that synthesis subband filters of high degree are needed to compensate the aliasing and phase distortions caused by the allpass transformation of the analysis filter-bank.

These problems are addressed by the alternative concept of the filter-bank equalizer (FBE). It is derived as an efficient realization of the filter-bank summation method (FBSM) and performs time-domain filtering with coefficients adapted in the frequency-domain. Perfect signal reconstruction is achieved for a broad class of transformations with significantly lower efforts than for the common AS FB. The reconstructed signal is (inherently) aliasing-free so that a prototype filter with a high filter degree to achieve a high stopband attenuation is not essential. It is shown how the FBE can be efficiently implemented by means of a polyphase network (PPN) structure. The explicit consideration of the time-varying coefficients in the derivation has revealed the influence of the filter structure on system delay, computational complexity, and signal

quality. This has shown, among others, how the transposed direct form implementation achieves a stronger smoothing effect for time-varying coefficients in comparison to the direct form implementation, which is beneficial to avoid artifacts for the processed signal.

The presented allpass transformed FBE achieves a near-perfect, aliasing-free signal reconstruction with significantly lower efforts than allpass transformed AS FBs. The uniform FBE has a higher algorithmic complexity than the corresponding uniform AS FB for most parameter configurations (L, M, R), while the opposite applies for the allpass transformed FBE in comparison to the allpass transformed AS FB. The uniform and warped FBE achieve a significantly lower algorithmic signal delay than the corresponding AS FBs. A nearly linear phase response can be maintained even for time-varying coefficients, which can be exploited, e.g., for binaural signal processing in hearing aids.

The proposed filter-bank design provides a versatile concept for applications such as low delay speech enhancement. The uniform and warped FBE achieve the same (or even better) objective and subjective quality for the enhanced speech as comparable AS FBs, but with a significantly lower signal delay. The frequency warping can be utilized either to achieve an improved speech quality or to use a lower number of frequency channels.

The concept of the low delay filter (LDF) is a modification of the FBE to achieve a further reduction of signal delay and algorithmic complexity with almost no compromise on the perceived (subjective) speech quality. In this process, the time-domain filter of the FBE is approximated by a moving-average (MA) filter or auto-regressive (AR) filter of lower degree. The use of the uniform and warped MA filter allows to maintain a constant (near-linear) phase characteristic, which is beneficial, e.g., for multi-channel processing. The uniform and warped AR filter are minimum-phase systems and can achieve an algorithmic signal delay of only a few sample instants.

The use for noise reduction has been considered primarily here, but the presented low delay filter-bank concepts are also suitable for other speech and audio processing algorithms.

References

1. J. Agnew, J. M. Thornton: Just noticeable and objectionable group delays in digital hearing aids, *Journal of the American Academy of Audiology*, **11**(6), 330–336, 2000.
2. K. G. Beauchamp: *Walsh Functions and Their Applications,* London, GB: Academic Press, 1975.
3. M. G. Bellanger, G. Bonnerot, M. Coudreuse: Digital filtering by polyphase network: application to sample-rate alteration and filter banks, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-24**(2), 109–114, April 1976.
4. J. Benesty, S. Makino, J. Chen: *Speech Enhancement,* Berlin, Germany: Springer, 2005.

5. S. F. Boll: Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-27**(2), 113–120, April 1979.
6. C. Braccini, A. V. Oppenheim: Unequal bandwidth spectral analysis using digital frequency warping, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-22**(4), 236–244, August 1974.
7. C. S. Burrus, R. A. Gopinath, H. Guo: *Introduction to Wavelets and Wavelet Transforms: A Primer,* Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
8. O. Cappé: Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor, *IEEE Trans. on Speech and Audio Processing*, **2**(2), 345–349, April 1994.
9. I. Cohen: Enhancement of speech using Bark-scaled wavelet packet decomposition, *Proc. EUROSPEECH '01*, 1933–1936, Aalborg, Denmark, September 2001.
10. A. G. Constantinides: Frequency transformation for digital filters, *IEE Electronic Letters*, **3**(11), 487–489, November 1967.
11. R. E. Crochiere: A weighted overlap-add method of short-time Fourier analysis/synthesis, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-28**(10), 99–102, February 1980.
12. R. E. Crochiere, L. R. Rabiner: *Multirate Digital Signal Processing,* Upper Saddle River, NJ, USA: Prentice-Hall, 1983.
13. Z. Cvetković, J. D. Johnston: Nonuniform oversampled filter banks for audio signal processing, *IEEE Trans. on Speech and Audio Processing*, **11**(5), 393–399, September 2003.
14. R. Czarnach: Recursive processing by noncausal digital filters, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-30**(3), 363–370, June 1982.
15. I. Daubechies, W. Sweldens: Factoring Wavelet transforms into lifting steps, *Journal of Fourier Analysis and Applications*, **4**(3), 247–269, May 1998.
16. Y. Deng, V. J. Mathews, B. Farhang-Boroujeny: Low-delay nonuniform pseudo-QMF banks with application to speech enhancement, *IEEE Trans. on Signal Processing*, **55**(5), 2110–2121, May 2007.
17. G. Doblinger: An efficient algorithm for uniform and nonuniform digital filter banks, *Proc. ISCAS '91*, **1**, 646–649, Singapore, June 1991.
18. B. Dumitrescu, R. Bregović, T. Saramäki, R. Niemistö: Low-delay nonuniform oversampled filterbanks for acoustic echo control, *Proc. EUSIPCO '06*, Florence, Italy, September 2006.
19. A. Engelsberg: *Transformation-Based Systems for Single-Channel Noise Reduction in Speech Signals,* PhD thesis, Christian-Albrechts University, Ulrich Heute (ed.), Kiel, Germany: Shaker Verlag, 1998 (in German).
20. Y. Ephraim, D. Malah: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-32**(6), 1109–1121, December 1984.
21. C. Feldbauer, G. Kubin: Critically sampled frequency-warped perfect reconstruction filterbank, *Proc. ECCTD '03*, Krakow, Poland, September 2003.
22. S. Franz, S. K. Mitra, J. C. Schmidt, G. Doblinger: Warped discrete Fourier transform: a new concept in digital signal processing, *Proc. ICASSP '02*, **2**, 1205–1208, Orlando, FL, USA, May 2002.
23. E. Galijašević: *Allpass-Based Near-Perfect-Reconstruction Filter Banks,* PhD thesis, Christian-Albrechts University, Ulrich Heute (ed.), Kiel, Germany: Shaker Verlag, 2002.

24. E. Galijašević, J. Kliewer: Design of allpass-based non-uniform oversampled DFT filter banks, *Proc. ICASSP '02*, **2**, 1181–1184, Orlando, FL, USA, May 2002.
25. R. C. Gonzalez, P. Wintz: *Digital Image Processing,* London, GB: Addison-Wesley, 1977.
26. T. Gülzow, A. Engelsberg, U. Heute: Comparison of a discrete Wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement, *Signal Processing*, Elsevier, **64**(1), 5–19, January 1998.
27. T. Gülzow, T. Ludwig, U. Heute: Spectral-subtraction speech enhancement in multirate systems with and without non-uniform and adaptive bandwidths, *Signal Processing*, Elsevier, **83**(8), 1613–1631, August 2003.
28. H. Gustafsson, S. E. Nordholm, I. Claesson: Spectral subtraction using reduced delay convolution and adaptive-averaging, *IEEE Trans. on Speech and Audio Processing*, **9**(8), 799–807, November 2001.
29. A. Härmä: Implementation of recursive filters having delay free loops, *Proc. ICASSP '98*, **3**, 1261–1264, Seattle, WA, USA, May 1998.
30. A. Härmä: Implementation of frequency-warped recursive filters, *Signal Processing*, Elsevier, **80**(3), 543–548, March 2000.
31. U. Heute: Noise reduction, in E. Hänsler, G. Schmidt (eds.), *Topics in Acoustic Echo and Noise Control*, 325–384, Berlin, Germany: Springer, 2006.
32. Y. Hu, P. C. Loizou: Subjective comparison of speech enhancement algorithms, *Proc. ICASSP '06*, Tolouse, France, May 2006.
33. ITU-T Rec. P.862: Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, February 2001.
34. M. Kahrs, K. Brandenburg: *Applications of Digital Signal Processing to Audio and Acoustics,* Boston, MA, USA: Kluwer, 1998.
35. M. Kappelan: *Characteristics of Allpass Chains and their Application for Non-Equispaced Spectral Analysis and Synthesis,* PhD thesis, RWTH Aachen University, Peter Vary (ed.), Aachener Beiträge zu Digitalen Nachrichtensystemen, Aachen, Germany: Mainz Verlag, 1998 (in German).
36. M. Kappelan, B. Strauß, P. Vary: Flexible nonuniform filter banks using allpass transformation of multiple order, *Proc. EUSIPCO '96*, **3**, 1745–1748, Trieste, Italy, 1996.
37. T. Karp, A. Mertins: Lifting schemes for biorthogonal modulated filter banks, *Proc. of Intl. Conf. on Digital Signal Processing (DSP) '97*, **1**, 443–446, Santorini, Greece, July 1997.
38. T. Karp, A. Mertins, G. Schuller: Efficient biorthogonal cosine-modulated filter banks, *Signal Processing*, Elsevier, **81**(5), 997–1016, May 2001.
39. J. M. Kates, K. H. Arehart: Multichannel dynamic-range compression using digital frequency warping, *EURASIP Journal on Applied Signal Processing*, **18**, 3003-3014, 2005.
40. J. Kliewer, A. Mertins: Oversampled cosine-modulated filter banks with arbitrary system delay, *IEEE Trans. on Signal Processing*, **46**(4), 941–955, April 1998.
41. A. M. Kondoz: *Digital Speech - Coding for Low Bit Rate Communication Systems,* Chichester, UK: Wiley, 2004.
42. T.-Y. Leou, J. K. Aggarwal: Recursive implementation of LTV filters – frozen-time transfer function versus generalized transfer function, *Proc. of the IEEE*, **72**(7), 980–981, July 1984.

43. J. Li, T. Q. Nguyen, S. Tantaratana: A simple design for near-perfect-reconstruction nonuniform filter banks, *IEEE Trans. on Signal Processing*, **45**(8), 2105–2109, August 1997.
44. H. W. Löllmann, P. Vary: Efficient non-uniform filter-bank equalizer, *Proc. EUSIPCO '05*, Antalya, Turkey, September 2005.
45. H. W. Löllmann, P. Vary: Generalized filter-bank equalizer for noise reduction with reduced signal delay, *Proc. INTERSPEECH '05*, 2105–2108, Lisbon, Portugal, September 2005.
46. H. W. Löllmann, P. Vary: Low delay filter for adaptive noise reduction, *Proc. IWAENC '05*, 205–208, Eindhoven, The Netherlands, September 2005.
47. H. W. Löllmann, P. Vary: A warped low delay filter for speech enhancement, *Proc. IWAENC '06*, Paris, France, September 2006.
48. H. W. Löllmann, P. Vary: Parametric phase equalizers for warped filter-banks, *Proc. EUSIPCO '06*, Florence, Italy, September 2006.
49. H. W. Löllmann, P. Vary: Improved design of oversampled allpass transformed DFT filter-banks with near-perfect reconstruction, *Proc. EUSIPCO '07*, Poznan, Poland, September 2007.
50. T. Lotter, P. Vary: Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model, *EURASIP Journal on Applied Signal Processing*, **7**, 1110–1126, May 2005.
51. A. Makur, S. K. Mitra: Warped discrete Fourier transform: theory and application, *IEEE Trans. on Circuits and Systems I*, **48**(9), 1086–1093, September 2001.
52. D. Malah, R. V. Cox, A. J. Accardi: Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments, *Proc. ICASSP '99*, 789–792, Phoenix, AR, USA, May 1999.
53. R. Martin, H.-G. Kang, R. V. Cox: Low delay analysis synthesis schemes for joint speech enhancement and low bit rate speech coding, *Proc. EUROSPEECH '99*, **3**, 1463–1466, Budapest, Hungary, 1999.
54. R. Martin: Noise power spectral density estimation based on optimal smoothing and minimum statistics, *IEEE Trans. on Speech and Audio Processing*, **9**(5), 504–512, July 2001.
55. R. Martin: Statistical methods for the enhancement of noisy speech, in J. Benesty, S. Makino, J. Chen (eds.), *Speech Enhancement*, 43–65, Berlin, Germany: Springer, 2005.
56. S. K. Mitra, C. D. Creusere, H. Babic: A novel implementation of perfect reconstruction QMF banks using IIR filters for infinite length signals, *Proc. ISCAS '92*, 2312–2315, San Diego, CA, USA, May 1992.
57. D. R. Morgan, J. C. Thi: A delayless subband adaptive filter architecture, *IEEE Trans. on Signal Processing*, **43**(8), 1819–1830, August 1995.
58. A. V. Oppenheim, D. Johnson, K. Steiglitz: Computation of spectra with unequal resolution using the fast Fourier transform, *Proc. of the IEEE*, **59**(2), 299–301, February 1971.
59. A. V. Oppenheim, R. W. Schafer, J. R. Buck: *Discrete-Time Signal Processing*, 2nd edition, Upper Saddle River, NJ, USA: Prentice-Hall, 1999.
60. T. W. Parks, C. S. Burrus: *Digital Filter Design,* Chichester, GB: Wiley, 1987.
61. A. Petrovsky, M. Parfieniuk, A. Borowicz: Warped DFT based perceptual noise reduction system, *Convention Paper of Audio Engineering Society*, Berlin, Germany, May 2004.

62. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: *Numerical Recipes in C,* 2nd edition, Cambridge, GB: Cambridge University Press, 1992.
63. J. G. Proakis, D. G. Manolakis: *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd edition, Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
64. S. R. Quackenbush and T. P. Barnwell III and M. A. Clements: *Objective Measures of Speech Quality*, Upper Saddle River, NJ, USA: Prentice-Hall, 1988.
65. C. M. Rader: An improved algorithm for high-speed autocorrelation with application to spectral estimation, *IEEE Trans. on Audio and Electroacoustics*, **18**(4), 439–441, December 1970.
66. K. R. Rao and P. Yip: *Discrete Cosine Transform,* New York, NY, USA: Academic Press, 1990.
67. M. Renfors, T. Saramäki: Recursive Nth-band digital filters – part I: design and properties, *IEEE Trans. on Circuits and Systems*, **34**(1), 24–39, January 1987.
68. M. Schönle, C. Beaugeant, K. Steinert, H. W. Löllmann, B. Sauert, P. Vary: Hands-free audio and its application to telecommunication terminals, *Proc. of Intl. Conf. on Audio for Mobile and Handheld Devices (AES)*, Seoul, Korea, September 2006.
69. G. D. T. Schuller and T. Karp: Modulated filter banks with arbitrary system delay: efficient implementation and the time-varying case, *IEEE Trans. on Signal Processing*, **48**(3), 737–748, March 2000.
70. H. W. Schüßler, W. Winkelnkemper: Variable digital filters, *Archiv der Elektrischen Übertragung (AEÜ)*, **24**(11), 524–525, 1970.
71. H. W. Schüßler: Implementation of variable digital filters, *Proc. EUSIPCO '80*, 123–129, Lausanne, Switzerland, September 1980.
72. B. Shankar M. R., A. Makur: Allpass delay chain-based IIR PR filterbank and its application to multiple description subband coding, *IEEE Trans. on Signal Processing*, **50**(4), 814–823, April 2002.
73. J. O. Smith, J. S. Abel: Bark and ERB bilinear transforms, *IEEE Trans. on Speech and Audio Processing*, **7**(6), 697–708, November 1999.
74. K. Steiglitz: A note on variable recursive digital filters, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **ASSP-28**(1), 111–112, February 1980.
75. M. A. Stone, B. C. J. Moore: Tolerable hearing aid delays II: estimation of limits imposed during speech production, *Ear and Hearing*, **32**(4), 325–338, 2002.
76. W. Sweldens: The lifting scheme: a custom-design construction of biorthogonal wavelets, *Applied and Computational Harmonic Analysis*, **3**(2), 186–200, 1996.
77. P. P. Vaidyanathan: *Multirate Systems and Filter Banks,* Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
78. P. Vary: On the design of digital filter banks based on a modified principle of polyphase, *AEÜ (Archive for Electronics and Communications)*, **33**, 293–300, 1979.
79. P. Vary: Digital filter banks with unequal resolution, *Short Communication Digest of European Signal Processing Conf. (EUSIPCO)*, 41–42, Lausanne, Switzerland, September 1980.
80. P. Vary: Noise suppression by spectral magnitude estimation – mechanism and theoretical limits, *Signal Processing, Elsevier*, **8**(4), 387–400, July 1985.
81. P. Vary, R. Martin: *Digital Speech Transmission: Enhancement, Coding and Error Concealment,* Chichester, GB: Wiley, 2006.
82. P. Vary: An adaptive filter-bank equalizer for speech enhancement, *Signal Processing, Elsevier*, **86**(6), 1206–1214, June 2006.

83. M. Vetterli, J. Kovačević: *Wavelets and Subband Coding,* Upper Saddle River, NJ, USA: Prentice-Hall, 1995.
84. E. Zwicker, H. Fastl: *Psychoacoustics: Facts and Models,* 2nd edition, Berlin, Germany: Springer, 1999.