



HAL
open science

Uniform and Warped Low Delay Filter-Banks for Speech Enhancement

Heinrich W. Löllmann, Peter Vary

► **To cite this version:**

Heinrich W. Löllmann, Peter Vary. Uniform and Warped Low Delay Filter-Banks for Speech Enhancement. *Speech Communication*, 2007, 49 (7-8), pp.574. 10.1016/j.specom.2007.04.009 . hal-00499185

HAL Id: hal-00499185

<https://hal.science/hal-00499185v1>

Submitted on 9 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Uniform and Warped Low Delay Filter-Banks for Speech Enhancement

Heinrich W. Löllmann, Peter Vary

PII: S0167-6393(07)00063-5

DOI: [10.1016/j.specom.2007.04.009](https://doi.org/10.1016/j.specom.2007.04.009)

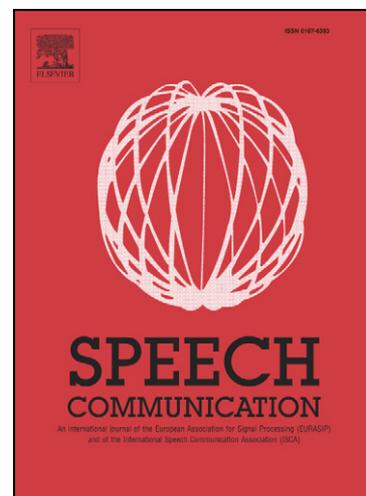
Reference: SPECOM 1630

To appear in: *Speech Communication*

Received Date: 2 July 2006

Revised Date: 25 February 2007

Accepted Date: 3 April 2007



Please cite this article as: Löllmann, H.W., Vary, P., Uniform and Warped Low Delay Filter-Banks for Speech Enhancement, *Speech Communication* (2007), doi: [10.1016/j.specom.2007.04.009](https://doi.org/10.1016/j.specom.2007.04.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Uniform and Warped Low Delay Filter-Banks for Speech Enhancement

Heinrich W. Löllmann^{a,*} Peter Vary^a

^a*Institute of Communication Systems and Data Processing,
RWTH Aachen University, D-52056 Aachen, Germany*

Abstract

A versatile filter-bank concept for adaptive subband filtering is proposed, which achieves a significantly lower algorithmic signal delay than commonly used analysis-synthesis filter-banks. It is derived as an efficient implementation of the filter-bank summation method and performs time-domain filtering with coefficients adapted in the uniform or non-uniform frequency-domain. The frequency warped version of the proposed filter-bank has a lower computational complexity than the usual warped analysis-synthesis filter-bank for most parameter configurations. The application to speech enhancement shows that the same quality of the enhanced speech can be achieved but with lower signal delay. For systems with tight signal delay requirements, modifications of the new filter-bank design are discussed to further decrease its signal delay by approximating the original time-domain filter by an FIR or IIR filter of lower degree. This approach can achieve a very low signal delay and reduced computational complexity with almost no loss for the perceived speech quality.

Key words: filter-bank equalizer, frequency warping, time-varying filters, low delay, speech enhancement

1 Introduction

The wide-spread use of mobile communication devices and the availability of digital hearing aids has lead to an increased demand for powerful speech

* Corresponding author.

Email addresses: loellmann@ind.rwth-aachen.de (Heinrich W. Löllmann), vary@ind.rwth-aachen.de (Peter Vary).

URL: www.ind.rwth-aachen.de (Peter Vary).

enhancement systems. They aim to ease the speech communication in environments where ambient noise impairs the intelligibility and fidelity of the transmitted speech signal. Most algorithms for (single-channel) noise reduction are realized by short-term spectral weighting, e.g., [1,2,3,4]. Besides the algorithm for the computation of the spectral gain factors, the choice of the spectral transform or filter-bank, respectively, influences significantly the performance of the speech enhancement system in terms of speech quality, computational complexity and signal delay. Especially the system delay is often an important design criterion. An example are speech processing systems for hearing aids where a high system delay can cause 'coloration effects', e.g., [5]. These distortions occur when the processed speech signal interacts with the original speech signal which reaches the cochlea with minimal delay via bone conduction or through the hearing aid vent.

Due to its low computational complexity, the uniform DFT analysis-synthesis filter-bank (AS FB) is a common choice, which comprises the weighted overlap-add method as a variant thereof [6]. The frequency resolution of the uniform DFT filter-bank, however, is not well adapted to that of the human auditory system, which can be modeled by the Bark frequency bands [7]. Therefore, some authors have proposed the use of non-uniform AS FBs for noise reduction to obtain an improved (subjective) quality for the enhanced speech [8,9,10,11]. For such applications, the allpass transformed DFT filter-bank [12,13,14,15] is of special interest, because this frequency warped filter-bank achieves a very good approximation of the Bark scale [16]. This property is not only of interest for perceptual based speech enhancement, but also for speech and audio processing applications in general. Thereby, the adapted non-uniform frequency resolution of a warped filter-bank allows to use a lower number of frequency channels than the uniform filter-bank.

A drawback of allpass transformed filter-banks is their increased computational complexity since the delay elements are replaced by allpass filters. This also leads to an altered phase response, which can become audible for subband filters with high group delays. The compensation of this effect increases signal delay and algorithmic complexity. Moreover, a lower downsampling rate is required in comparison to the uniform filter-bank to avoid noticeable aliasing distortions, cf. [8,9].

In this paper, an alternative design is proposed to tackle these problems. A versatile and efficient filter-bank concept with uniform and warped frequency resolution is developed to perform adaptive subband filtering with significantly lower signal delay than the conventional AS FB. For this purpose, the recently proposed concept of the filter-bank equalizer (FBE) [17,18] is revised. The FBE performs time-domain filtering with coefficient update in the frequency-domain. A similar approach has been proposed independently in [5] for dynamic range compression in hearing aids. For acoustic echo cancel-

lation and active noise control applications, a related time-domain filtering approach can be found in [19] where the coefficients are adapted in the uniform frequency-domain.

An important novel aspect of this contribution is that the proposed filter-bank design takes explicitly into account the effects of time-varying coefficients. It is shown how the structure of the time-domain filter influences the overall algorithmic complexity, the signal delay, and the perceptual quality of the filtered (speech) signal. It is revealed how the compensation of phase modifications due to the allpass transformation is achieved with significantly lower effort than for the conventional warped AS FB. Moreover, a modification of the FBE is proposed to achieve a very low signal delay. Thereby, the original time-domain filter is approximated by a filter of lower degree with negligible loss for the perceived speech quality. An application of the devised filter-bank structures is speech enhancement with tight signal delay requirements.

The paper is organized as follows: In Section 2, different filter-banks for adaptive subband filtering are reviewed and our alternative design approach is motivated. In Section 3, the time-varying FBE with uniform frequency resolution is derived and its efficient implementation is investigated. In Section 4, the allpass transformed FBE is devised. In Section 5, a modification of the uniform and allpass transformed FBE is discussed to further decrease the signal delay. In Section 6, the developed filter-banks are applied to noise reduction and the achieved performance is analyzed. The paper concludes with Section 7.

2 Filter-Banks for Subband Filtering

A common technique for adaptive subband filtering is to use an *analysis-synthesis filter-bank* (AS FB). The frequently used uniform DFT AS FB has complex modulated subband filters

$$h_i(n) = h(n) e^{-j \frac{2\pi}{M} i n} \quad (1)$$

$$g_i(n) = g(n) e^{-j \frac{2\pi}{M} i (n+1)} \quad (2)$$

$$i = 0, 1, \dots, M-1 \quad ; \quad n = 0, 1, \dots, L$$

with $h(n)$ and $g(n)$ denoting the impulse responses of the prototype filters for the analysis and synthesis filter-bank, respectively, where $L+1 \geq M$. As illustrated in Fig. 1, such a filter-bank can be efficiently implemented by means of a polyphase network (PPN) with the discrete Fourier transform (DFT) computed by the fast Fourier transform (FFT), cf. [20,21]. A non-uniform time-frequency resolution can be obtained by an *allpass transformation* where the delay elements are replaced by allpass filters of first order [12,15]. Most

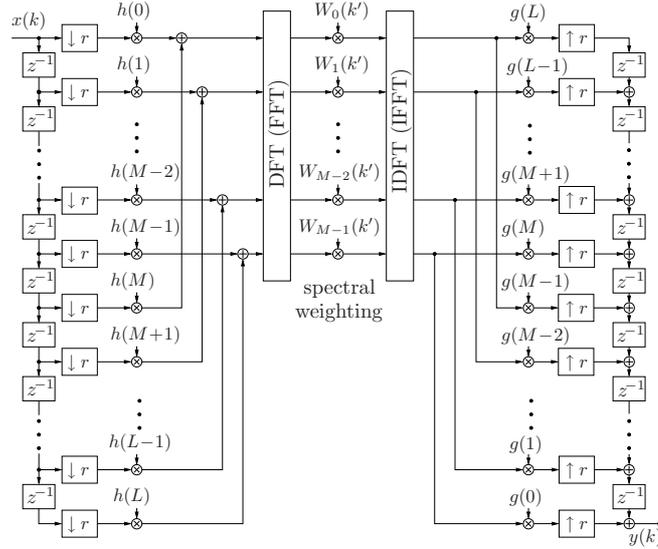


Figure 1. Polyphase network (PPN) implementation of a DFT analysis-synthesis filter-bank (AS FB) for a prototype filter length of $L + 1 = 2M$. The allpass transformed (APT) AS FB is obtained by replacing the delay elements with allpass filters.

arithmetic operations are performed at a decimated sampling rate, indicated by the sample index k' .

A paraunitary AS FB has a system delay of L samples, e.g., [20]. A lower delay of down to $M - 1$ samples can be achieved by a biorthogonal cosine modulated AS FB [22]. An alternative is to employ the lifting scheme for the design of uniform and allpass transformed low delay filter-banks [23,24]. However, these designs do not consider *time-varying* spectral gain factors, where aliasing components are not canceled in the same way as for $W_i(k') = 1 \forall i$. Moreover, the warped DFT AS FB proposed in [24] has still a very high signal delay ($d_0 \gg M$) and a high computational complexity.

A lower signal delay and aliasing-free signal reconstruction can be achieved by using the *filter-bank summation method* (FBSM) for adaptive subband filtering as illustrated by Fig. 2. The FBSM can be derived from the filter-bank interpretation of the short-time DFT, e.g., [25]. It has a lower signal delay than the corresponding¹ conventional AS FB with downsampling, but exhibits a significantly higher algorithmic complexity as no downsampling of the subband signals $x_i(k)$ can be performed. In the following, the FBSM is investigated in detail to derive a low delay filter-bank without the computational complexity of the FBSM.

¹ FBSM and AS FB use the same type of analysis filter-bank having identical values for L and M . The same filter-bank can often not be taken due to different design constraints for signal reconstruction.

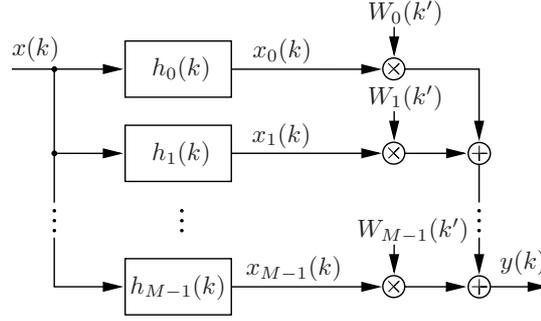


Figure 2. Filter-bank summation method (FBSM) with time-varying spectral gain factors.

3 Time-Varying Filter-Bank Equalizer

3.1 Concept

For the FBSM of Fig. 2, the real input signal $x(k)$ is split into $i = 0, 1, \dots, M-1$ subband signals $x_i(k)$ by means of M bandpass filters. The adaptation of the time-varying gain factors $W_i(k')$ can be done by the same algorithms as for the AS FB based on the subband signals $x_i(k)$. This adaptation is performed at intervals of r sample instants with k' defined by²

$$k' = \lfloor k/r \rfloor r ; r = 1, 2, \dots \quad (3)$$

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

The (fixed) impulse response $h_i(n)$ of the i -th bandpass filter shall be given by modulation of a prototype lowpass filter with real impulse response $h(n)$ of length $L + 1$ according to

$$h_i(n) = \begin{cases} h(n) \Phi(i, n) & ; i = 0, 1, \dots, M-1 \\ & n = 0, 1, \dots, L \\ 0 & ; \text{else.} \end{cases} \quad (4)$$

The general modulation sequence $\Phi(i, n)$ can be interpreted as transformation kernel of the filter-bank and has periodicity

$$\Phi(i, n + m M) = \Phi(i, n) \rho(m) ; m \in \mathbb{Z} \quad (5)$$

where mostly $\rho(m) = 1$. \mathbb{Z} denotes the set of all integer numbers.

² This definition is more suitable for the following treatment than the common convention $k' r = k$.

For speech and audio processing, the generalized discrete Fourier transform (GDFT) with evenly-stacked frequency channels is of special interest, cf. [25]. The corresponding modulation sequence reads

$$\Phi_{\text{GDFT}}(i, n) = \exp\left\{-j \frac{2\pi}{M} i(n - n_0)\right\} \quad (6)$$

$$i = 0, 1, \dots, M - 1 ; \quad n, n_0 \in \mathbb{Z}.$$

The common DFT is included as special case for $n_0 = 0$.

The input-output relation for the FBSM of Fig. 2 is given by

$$y(k) = \sum_{i=0}^{M-1} W_i(k') \sum_{n=0}^L x(k - n) h_i(n). \quad (7)$$

Insertion of Eq. (4) leads to

$$y(k) = \sum_{n=0}^L x(k - n) h(n) \sum_{i=0}^{M-1} W_i(k') \Phi(i, n). \quad (8)$$

The second summation is the spectral transform of the gain factors $W_i(k')$ which yields the time-domain weighting factors

$$w_n(k') = \sum_{i=0}^{M-1} W_i(k') \Phi(i, n) ; \quad n = 0, 1, \dots, L \quad (9)$$

$$\doteq T\{W_i(k')\}. \quad (10)$$

These $L + 1$ weighting factors have the periodicity

$$w_{n+mM}(k') = w_n(k') \quad (11)$$

for the GDFT of Eq. (6). The input-output relation finally reads

$$y(k) = \sum_{n=0}^L x(k - n) h(n) w_n(k') \quad (12)$$

$$= \sum_{n=0}^L x(k - n) h_s(n, k'). \quad (13)$$

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

$$h_s(n, k') \doteq h(n) w_n(k') \quad (14)$$

are the product of the impulse response $h(n)$ of the prototype lowpass filter and the time-varying weighting factors $w_n(k')$ adapted in the short-term spectral-domain. This efficient implementation of the FBSM (which resembles

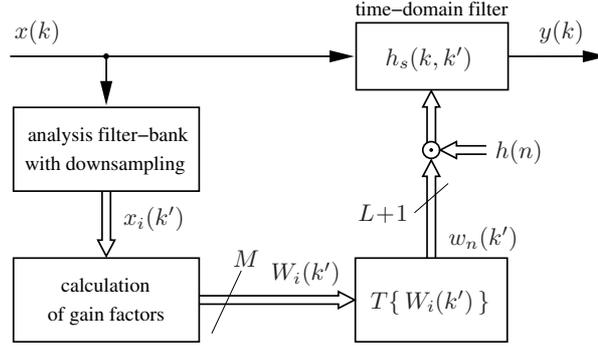


Figure 3. Filter-bank equalizer (FBE) for adaptive subband filtering.

a filter-bank used as equalizer) is termed as *filter-bank equalizer (FBE)* [17,18]. A sketch of this time-varying filter-bank structure is given in Fig. 3. An advantage is that the choice of the downsampling rate r for the subband signals $x_i(k')$ is not governed by restrictions for aliasing-free signal reconstruction as for the AS FB.

For DFT based noise reduction, the spectral gain factors have the properties

$$\epsilon \leq W_i(k') \leq 1 \ ; \ W_i(k') \in \mathbb{R} \ ; \ 0 \leq \epsilon < 1 \quad (15a)$$

$$W_i(k') = W_{M-i}(k') \ ; \ i = 0, 1, \dots, M-1, \quad (15b)$$

where \mathbb{R} marks the set of all real numbers. The (implicit) limitation by ϵ is beneficial to avoid 'musical noise', cf. [26]. The DFT of the gain factors yields time-domain weighting factors $w_n(k')$ corresponding to a (non-causal) zero-phase FIR filter. A causal linear-phase FIR filter with

$$h_s(n, k') = h_s(L-n, k') \ ; \ n = 0, 1, \dots, L \quad (16)$$

is obtained for the generalized DFT of Eq. (6) with $n_0 = L/2$ (L even), and requires a linear-phase FIR prototype filter with symmetry

$$h(n) = h(L-n) \ ; \ n = 0, 1, \dots, L. \quad (17)$$

The GDFT of the gain factors $W_i(k')$ can be accomplished by the FFT with a subsequent cyclic shift of the obtained time-domain weighting factors by n_0 samples. The gain factors of Eq. (15) are calculated by the magnitude of the subband signals $|x_i(k')|$ such that the DFT analysis filter-bank instead of the GDFT filter-bank can be taken for the FBE in Fig. 3.

3.2 Prototype Filter Design

A filter-bank achieves perfect (signal) reconstruction, if the output sequence is given by the input sequence or its delayed version

$$y(k) = x(k - d_0) ; d_0 \geq 0 \quad (18)$$

where $W_i(k') = 1$. This is achieved for the uniform FBE, if the general modulation sequence of Eq. (4) has the property

$$\sum_{i=0}^{M-1} \Phi(i, n) = \begin{cases} c \neq 0 & ; n = n_0 \\ 0 & ; n \neq n_0 \end{cases} \quad (19)$$

$$n, n_0 \in \{0, 1, \dots, M-1\},$$

and if a generalized M -th band filter with impulse response

$$h(n) = \begin{cases} \frac{1}{c\rho(m_c)} & ; n = n_0 + m_c M \\ & \rho(m_c) \neq 0 ; m_c \in \mathbb{Z} \\ 0 & ; n = n_0 + m M \\ & m \in \mathbb{Z} \setminus \{m_c\} \\ \text{arbitrary} & ; \text{else} \end{cases} \quad (20)$$

is used as prototype lowpass filter [18]. Eq. (18) is then fulfilled with $d_0 = n_0 + m_c M$. The condition of Eq. (19) applies for the GDFT as well as the Hadamard or Walsh transform, cf. [27], but not for the discrete cosine transform (DCT) in general. A relation between DFT FBE and DCT FBE is established in [17] for gain factors with property of Eq. (15).

For the considered GDFT of Eq. (6), $c = M$ and $\rho(m) = 1$, and a suitable FIR prototype lowpass filter to fulfill Eq. (20) is given by

$$h(n) = \frac{1}{M} \frac{\sin\left(\frac{2\pi}{M}(n - d_0)\right)}{\frac{2\pi}{M}(n - d_0)} \text{win}_L(n). \quad (21)$$

The general window sequence of length $L + 1$ is defined by

$$\text{win}_L(n) = \begin{cases} \text{arbitrary} & ; 0 \leq n \leq L \\ 0 & ; \text{else.} \end{cases} \quad (22)$$

A common window sequence is

$$\text{win}_L(n, \beta) = \begin{cases} \beta + (\beta - 1) \cos\left(\frac{2\pi}{L} n\right) & ; 0 \leq n \leq L \\ & 0.5 \leq \beta \leq 1 \\ 0 & ; \text{else.} \end{cases} \quad (23)$$

The rectangular window ($\beta = 1$), the Hann window ($\beta = 0.5$) and the Hamming window ($\beta = 0.54$) are included as special cases, cf. [28].

For this prototype filter and the considered GDFT, a *linear-phase* FIR time-domain filter according to Eq. (16) is obtained for

$$d_0 = n_0 = L/2 \quad (24)$$

with L being even.

3.3 Implementation for Different Filter Structures

The choice of the filter structure plays an important role for time-varying filters as well as for filter implementations with finite precision arithmetic, e.g., [21]. Fig. 4 shows the realization of a filter with finite impulse response (FIR) by means of the direct form (df) and the transposed direct form (tdf).

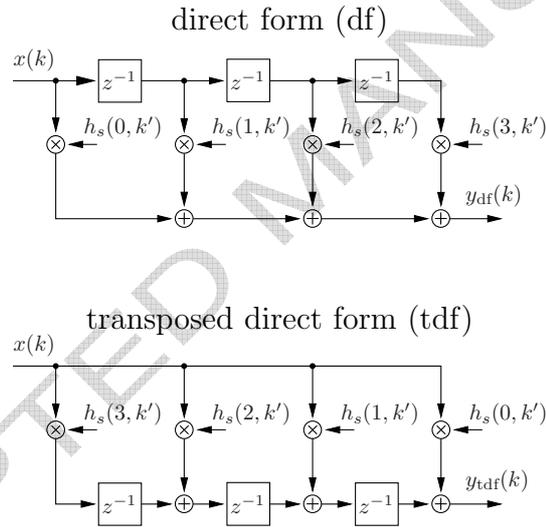


Figure 4. Direct filter forms for a time-varying FIR filter of degree $L = 3$.

The input-output relation for this two filter forms are given by

$$y_{df}(k) = \sum_{n=0}^L x(k-n) h_s(n, k') \quad (25)$$

$$y_{tdf}(k) = \sum_{n=0}^L x(k-n) h_s(n, k' - n). \quad (26)$$

Here, only the direct forms are considered since they do not require a conversion of the time-varying filter coefficients such as the cascade form or the

parallel form, cf. [21]. Obviously, the derived FBE according to Eq. (13) uses a time-domain filter in the *direct form*.

For the *transposed direct form*, Eq. (14) is inserted into Eq. (26) which leads to

$$\begin{aligned}
 y_{\text{tdf}}(k) &= \sum_{n=0}^L x(k-n) h(n) w_n(k'-n) \\
 &= \sum_{n=0}^L x(k-n) h(n) \sum_{i=0}^{M-1} W_i(k'-n) \Phi(i, n) \\
 &= \sum_{i=0}^{M-1} \sum_{n=0}^L x(k-n) W_i(k'-n) h_i(n) \quad (27)
 \end{aligned}$$

due to Eqs. (9) and (4). The last equation corresponds to the FBSM of Fig. 2 but with the gain factors $W_i(k')$ applied *before* the bandpass filters. Here, the gains are adapted based on the subband signals $x_i(k')$ which means that they must be calculated 'in advance'. The obtained filter-bank structure is illustrated in Fig. 5. For many real-time applications, a time-mismatch of τ_c

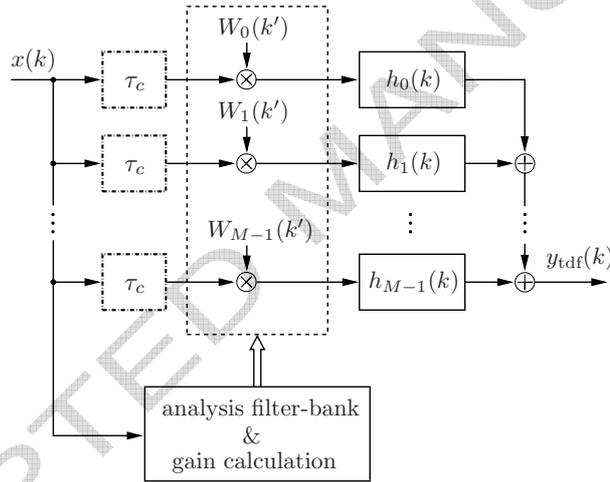


Figure 5. Filter-bank summation method (FBSM) corresponding to a filter-bank equalizer (FBE) with time-domain filter in transposed direct form. The dash-dotted boxes mark optional delay elements needed to account for the signal delay of the analysis filter-bank.

samples is usually less problematic than an additional signal delay.

Switching the coefficients of a digital filter during operation can lead to 'filter-ringing' effects, which might be perceived by perceptually annoying artifacts. The application to speech enhancement revealed that the FBE with transposed direct form time-domain filter provides a better perceived speech quality than the direct form filter (see Fig. 4). This can be explained by comparing the equivalent FBSMs of Fig. 2 and Fig. 5: If the transposed direct form is used for the time-domain filter of the FBE, the transients caused by the switching

gain factors are effectively smoothed by bandpass filters in contrast to the direct form implementation.

3.4 Polyphase Network Implementation

An efficient polyphase network (PPN) implementation of the FBE shall be developed. This eases the utilization of long prototype filters with $L > M$ to improve frequency selectivity and facilitates the use of recursive prototype filters.

Time-varying filters can be represented by their frozen-time transfer function or frozen-time impulse response, respectively, [29]. This description is not exact but sufficient for our treatment of different filter forms. The direct form filter of Eq. (14) at sample instant k' is represented by

$$H_s(z, k') = \sum_{n=0}^L w_n(k') h(n) z^{-n}. \quad (28)$$

The (type 1) polyphase components for the prototype filter are given in the z -domain by [20]

$$H_\lambda^{(M)}(z^M) = \sum_{m=-\infty}^{\infty} h(\lambda + m M) z^{-mM} \quad (29)$$

$$0 \leq \lambda \leq M - 1.$$

The transfer function³ of Eq. (28) can then be written

$$H_s(z, k') = \sum_{\lambda=0}^{M-1} w_\lambda(k') \sum_{m=0}^{l_M-1} h(\lambda + m M) z^{-(\lambda+mM)}$$

$$= \sum_{\lambda=0}^{M-1} w_\lambda(k') H_\lambda^{(M)}(z^M) z^{-\lambda} \quad (30)$$

due to Eq. (11) and with $l_M = \lceil \frac{L+1}{M} \rceil$.

The subband signals $x_i(k)$ of the FBSM in Fig. 2 are represented in the z -domain by

$$X_i(z) = X(z) \sum_{n=0}^L h(n) z^{-n} \Phi(i, n) \quad (31)$$

$$i = 0, 1, \dots, M - 1.$$

³ For the sake of brevity, the terms transfer function and impulse response shall refer to the frozen-time representation or the conventional one dependent on the context.

Applying Eq. (29) and Eq. (5) with $\rho(m) = 1$ leads to

$$X_i(z) = X(z) \sum_{\lambda=0}^{M-1} \sum_{m=0}^{l_{M-1}-1} h(\lambda + m M) \times z^{-(\lambda+m M)} \Phi(i, \lambda) \quad (32)$$

$$= X(z) \sum_{\lambda=0}^{M-1} z^{-\lambda} H_{\lambda}^{(M)}(z^M) \Phi(i, \lambda). \quad (33)$$

The developed PPN FBE according to Eq. (30) and Eq. (33) is illustrated in Fig. 6.

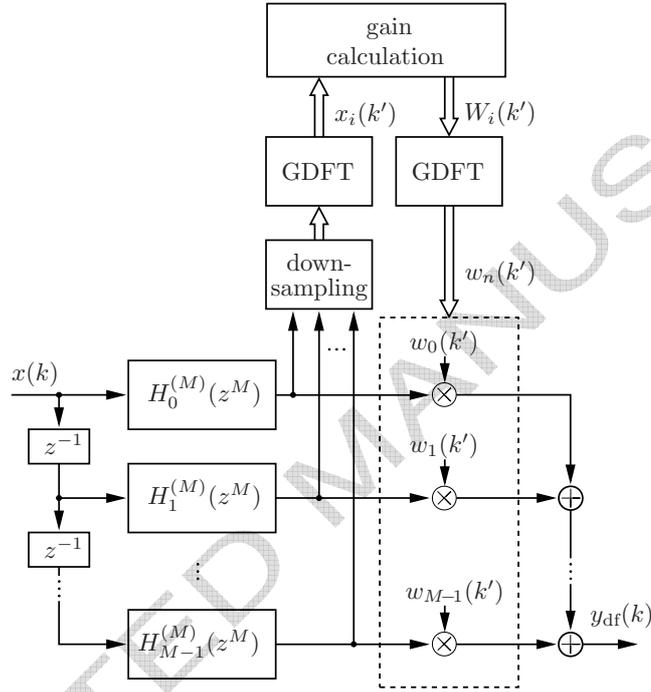


Figure 6. Polyphase network filter-bank equalizer (PPN FBE) for the direct form filter.

A filter in transposed direct form is obtained from the direct form by transposition of its signal flow graph [21]: All signal directions as well as the input and output are reversed. The nodes become summations and vice versa. The obtained PPN implementation for the transposed direct form is shown in Fig. 7. The weighting factors $w_n(k')$ according to Eq. (9) are calculated by a separate network as shown in Fig. 6, but with the difference that the downsampling can be performed directly after the delay elements as for the PPN analysis filter-bank in Fig. 1. In contrast to the direct form FBE, a second PPN is needed for the FBE in transposed direct form.

The developed PPN implementations of the FBE allows to use an FIR or IIR prototype lowpass filter. A polyphase decomposition for FIR filters and

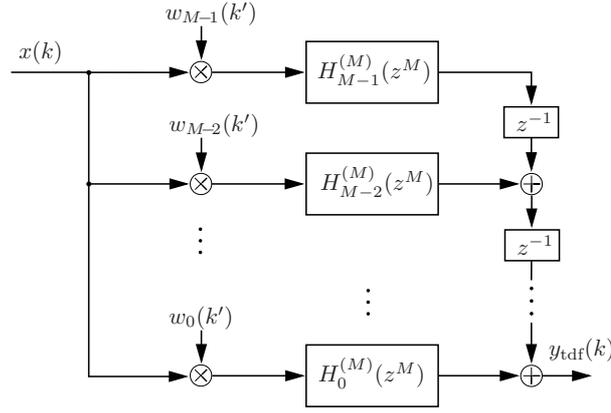


Figure 7. Polyphase network (PPN) implementation for the time-domain filter of the FBE in transposed direct form.

IIR filters has been proposed in [30] and [31], respectively. Thus, the PPN implementation enables a practical realization of Eq. (12) for L being infinite, that is, a *recursive* prototype filter. The design of IIR M -th band filters is proposed in [32] (cf. Section 3.2).

4 Non-Uniform Filter-Bank Equalizer

4.1 Concept

A well-known approach to obtain a digital filter-bank with non-uniform frequency bands is the frequency warping technique based on an allpass transformation [12,13,14,15]. This transformation is achieved by substituting all delay elements of the discrete (subband) filters by allpass filters

$$z^{-1} \rightarrow H_A(z). \quad (34)$$

A (causal) allpass filter of first order with transfer function

$$H_A(z) = \frac{z^{-1} - a^*}{1 - a z^{-1}} ; \quad |a| < |z| \quad (35a)$$

$$|a| < 1 ; \quad a = \alpha e^{j\gamma\Omega} \in \mathbb{C} ; \quad \alpha, \gamma \in \mathbb{R} \quad (35b)$$

and frequency response

$$H_A(z = e^{j\Omega}) = \frac{e^{-j\Omega} - a^*}{1 - a e^{-j\Omega}} = e^{-j\varphi_a(\Omega)} \quad (36a)$$

$$\varphi_a(\Omega) = 2 \arctan \left(\frac{\sin \Omega - \alpha \sin \gamma}{\cos \Omega - \alpha \cos \gamma} \right) - \Omega \quad (36b)$$

is used for the allpass transformation, where the asterisk denotes the conjugate complex. The allpass transformation of the bandpass filters of Eq. (4) yields the warped frequency responses

$$H_i(z = e^{j\varphi_a(\Omega)}) = \sum_{n=0}^L h(n) \Phi(i, n) e^{-jn\varphi_a(\Omega)} \quad (37)$$

$$\doteq \tilde{H}_i(e^{j\Omega}) ; \quad i = 0, 1, \dots, M-1. \quad (38)$$

The frequency warping $\Omega \rightarrow \varphi_a(\Omega)$ due to the allpass transformation is solely determined by the allpass pole a respective $\varphi_a(\Omega)$ according to Eq. (36). This effect is exemplified in Fig. 8. For a real coefficient $a = \alpha$ with $\alpha > 0$, a higher

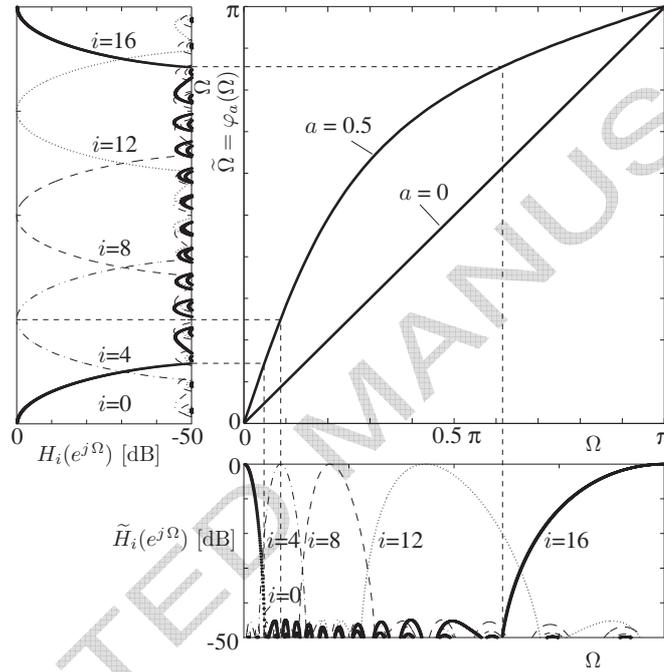


Figure 8. Allpass transformation of bandpass filters for $M = 16$.

frequency resolution is achieved for the lower frequency bands and vice versa. Thereby, the adjustment of the frequency resolution is accomplished without requiring a renewed prototype filter design. The uniform FBE is included as special case for $a = 0$ where $H_A(z) = z^{-1}$.

4.2 Warped Phase Characteristics

The allpass transformation of a filter-bank leads to an altered (warped) phase response, which might become audible for prototype filters with high group delays. Hence, measures to compensate this effect are a key element in the

design of warped⁴ filter banks, even though a perfect compensation is usually not required for speech and audio processing applications.

For the warped FBE, the (frozen-time) frequency response of the direct form time-domain filter reads

$$\widetilde{H}_s(e^{j\Omega}, k') = \sum_{n=0}^L h_s(n, k') e^{-jn\varphi_a(\Omega)}. \quad (39)$$

For the transposed direct form, this representation provides a suitable approximation for the following treatment. If the real filter coefficient have the symmetry of Eq. (16), the transfer function of Eq. (39) can be written

$$\widetilde{H}_s(e^{j\Omega}, k') = e^{-j\frac{L}{2}\varphi_a(\Omega)} \times \begin{cases} \sum_{n=0}^{\frac{L}{2}} 2\mathcal{A}(\Omega, n, L, k') - h_s\left(\frac{L}{2}, k'\right) & \text{for } L \text{ even} \\ \sum_{n=0}^{\frac{L-1}{2}} 2\mathcal{A}(\Omega, n, L, k') & \text{for } L \text{ odd} \end{cases} \quad (40a)$$

$$\mathcal{A}(\Omega, n, L, k') = h_s(n, k') \cos\left(\left[\frac{L}{2} - n\right]\varphi_a(\Omega)\right). \quad (40b)$$

The expressions to the right of the curly brace are real. Thus, the time-varying uniform FBE ($a = 0$) has a *generalized linear-phase* as defined in [21]. For the warped FBE, an almost (generalized) linear-phase response can be obtained by filtering its output sequence with a *phase equalizer*. The frequency response of the (fixed) phase equalizer should fulfill the general requirement

$$e^{-jd_p\varphi_a(\Omega)} P_{\text{gen}}(e^{j\Omega}) \stackrel{!}{=} e^{-j\tau_p\Omega} \quad (41)$$

$$\tau_p \geq 0 \quad ; \quad \tau_p \in \mathbb{R}.$$

For the warped FBE according to Eq. (40), $d_p = L/2$. An 'ideal' phase equalizer is obviously given by the inverse transfer function of an allpass filter chain of length d_p :

$$P_{\text{ideal}}(z, d_p) = H_A(z)^{-d_p} \quad ; \quad |z| < \frac{1}{|a|}. \quad (42)$$

The impulse response is anti-causal, i.e., $p_{\text{ideal}}(k, d_p) = 0$ for $k \geq 0$. One solution is to approximate the desired phase equalizer by a causal FIR filter

⁴ Here, the more general concept of frequency warping will always refer to an allpass transformation of first order such that both terms are used interchangeably.

with impulse response

$$p_{\text{FIR}}(k, d_p) = \begin{cases} p_{\text{ideal}}(k - L_p, d_p) & ; k = 0, \dots, L_p \\ 0 & ; \text{else.} \end{cases} \quad (43)$$

The design and properties of phase equalizers for warped filter-banks are treated in more detail in [33]. A phase equalizer might be omitted, if d_p is lower than 20 due to the insensitivity of the human auditory system towards minor phase distortions.

An FIR phase equalizer designed for $d_p = L$ is used for the warped DFT AS FB treated in [8,9]. The warped DFT AS FB proposed in [24] uses a synthesis filter-bank consisting of L phase equalizers designed for $d_p = 1, \dots, L$. However, both designs do not consider time-varying gains. Moreover, the warped FBE requires a phase equalizer with a significantly lower filter degree than the corresponding allpass transformed AS FBs.

4.3 Algorithmic Complexity

Table 1 compares the algorithmic complexity of the derived uniform and warped PPN FBE to that of the corresponding uniform and warped PPN AS FB of Fig. 1, excluding the complexity for the computation of the spectral gain factors. The implementation of an allpass filter with real pole $a = \alpha$ requires, for instance, two multiplications, two summations, and one delay element. The real DFT of size M can be computed in-place by the radix-2 FFT. Thereby, the FFT of a real sequence of size M requires approximately half the algorithmic complexity as a complex M -point FFT [34]. The GDFT can be calculated with similar complexity as the DFT. A design example is given later in Section 6.

5 Further Measures for Delay Reduction

5.1 Concept

Even though the FBE achieves about half the algorithmic signal delay than the corresponding AS FB, a further reduction might be needed for applications with very demanding and strict system delay requirements. A prominent example are speech enhancement systems for hearing aids. For such cases, a modification of the filter-bank equalizer concept is discussed, which allows a

	2 real FFTs	remaining operations	operations for APT
analysis-synthesis filter-bank (AS FB)			
mult.	$2M/r \log_2 M$	$(2L + 2 + M)/r$	$4L$
add.	$3M/r \log_2 M$	$(L - M + 1)/r + L$	$4L$
mem.	$2M$	$2L$	-
direct form filter-bank equalizer (df FBE)			
mult.	$2M/r \log_2 M$	$L + 1 + M$	$2L$
add.	$3M/r \log_2 M$	L	$2L$
mem.	$2M$	L	-
transposed direct form filter-bank equalizer (tdf FBE)			
mult.	$2M/r \log_2 M$	$(1 + 1/r)(L + 1 + M)$	$4L$
add.	$3M/r \log_2 M$	$(1 + 1/r)L$	$4L$
mem.	$2M$	$2L$	-

Table 1

Algorithmic complexity in terms of average number of real multiplications and real additions per sample instant, and number of delay elements (memory) for different realizations of a polyphase network (PPN) DFT filter-bank. A real prototype filter of length $L + 1 = m M$ is assumed. The last column contains the additional operations due to the allpass transformation (APT) of the respective filter-bank (without phase equalizer).

further reduction of the signal delay with almost no loss for the perceived quality of the enhanced speech.

The signal delay of a filter-bank can be reduced by taking a lower DFT size and adjusting the calculation of the spectral gain factors to the changed time-frequency resolution, cf. [35]. Here, a further reduction of the signal delay can also be accomplished by approximating the time-domain filter of the FBE by a filter of lower degree $P < L$. This offers a greater flexibility in the choice of the time-domain filter and requires no adjustment of the gain calculation. Different filter approximations for the uniform FBE are investigated first, before extending the results to allpass transformed filters.

5.2 Approximation by a Moving-Average (MA) Filter

The time-domain filter of the FBE can be approximated by an FIR filter of degree $P < L$ following a technique very similar to FIR filter design by windowing, e.g., [28]. The impulse response $h_s(n, k')$ of Eq. (14) is truncated by a window sequence of length $P + 1$ according to

$$\hat{h}_s(n, k') = h_s(n + n_c, k') \text{win}_P(n) \quad ; \quad n = 0, 1, \dots, P. \quad (44)$$

The constant value for n_c determines the part of the impulse response to be truncated, e.g., to maintain linear-phase property. This modified FBE based on an FIR filter approximation will be termed as *moving-average low delay filter* (MA LDF). The low delay filter comprises the overall system including the adaptation of the gain factors, and the term MA filter refers only to the time-domain filter with coefficients $\hat{h}_s(n, k')$.

The MA LDF achieves a reduced signal delay in a very simple and flexible manner and a constant (linear) phase characteristics of the time-domain filter can be preserved.

5.3 Approximation by an Auto-Regressive (AR) Filter

The approximation of the original time-domain filter of the FBE by an auto-regressive (AR) filter is of special interest. This allpole filter provides a very low signal delay due to its minimum phase property, and the calculation of its coefficients demands a lower computational complexity than for a general (minimum phase) IIR filter, e.g., [28]. The modification of the phase response is tolerable for speech processing as the human ear is relatively insensitive towards phase modifications, cf. [14]. The filter to be approximated contains no sharp zeros in its spectrum if Eq. (15) applies, which further supports the AR filter approximation. In the following, the dependence of the filter on k' is omitted for the sake of simplicity.

The transfer function of an AR filter of degree P is given by

$$\hat{H}_s(z) = H_{\text{AR}}(z) = \frac{a_0}{1 - \sum_{n=1}^P a_n z^{-n}}. \quad (45)$$

Methods to determine the $P + 1$ AR filter coefficients a_n are known from parametric spectrum analysis, e.g., [28]. The AR filter coefficients are determined

by the Yule-Walker equations

$$\begin{bmatrix} \varphi_{\bar{h}\bar{h}}(1) \\ \vdots \\ \varphi_{\bar{h}\bar{h}}(P) \end{bmatrix} = \begin{bmatrix} \varphi_{\bar{h}\bar{h}}(0) & \dots & \varphi_{\bar{h}\bar{h}}(1-P) \\ \vdots & \ddots & \vdots \\ \varphi_{\bar{h}\bar{h}}(P-1) & \dots & \varphi_{\bar{h}\bar{h}}(0) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix}. \quad (46)$$

The auto-correlation coefficients $\varphi_{\bar{h}\bar{h}}$ are calculated here from the original impulse response $h_s(n)$ by means of the weighted auto-correlation method

$$\varphi_{\bar{h}\bar{h}}(\lambda) = \sum_{n=0}^{L-|\lambda|} \bar{h}(n) \bar{h}(n+\lambda) \quad ; \quad 0 \leq |\lambda| \leq P \quad (47a)$$

$$\text{with} \quad \bar{h}(n) = h_s(n) \text{win}_L(n) \quad ; \quad n = 0, 1, \dots, L. \quad (47b)$$

The scaling factor a_0 in Eq. (45) is given by

$$a_0 = \sqrt{\varphi_{\bar{h}\bar{h}}(0) - \sum_{n=1}^P a_n \varphi_{\bar{h}\bar{h}}(-n)} \quad (48)$$

and ensures that the AR filter and the original filter have both the same amplification. The auto-correlation method of Eq. (47) leads to a symmetric Toeplitz structure for the auto-correlation matrix in Eq. (46). This allows to solve the Yule-Walker equations efficiently by means of the Levinson-Durbin recursion. The obtained AR filter has minimum phase property and is therefore always stable, cf. [28]. This modification of the FBE is termed as *auto-regressive low delay filter* (AR LDF) in analogy to the terminology of the previous section.

5.4 Algorithmic Complexity

The algorithmic complexity for the presented low delay filters (LDFs) is listed in Table 2.

The real GDFT calculated by the FFT has been considered as transformation kernel (cf. Section 4.3). The variable \mathcal{M}_{div} marks the number of multiplications needed for a division operation, and $\mathcal{M}_{\text{sqr}}t$ represents the number of multiplications needed for a square-root operation. Their values depend on the numeric procedure and accuracy used to perform these operations. Accordingly, the variables \mathcal{A}_{div} and $\mathcal{A}_{\text{sqr}}t$ denote the additions needed for a division and square-root operation, respectively. (An equivalent of 15 operations will be assigned to these variables for the complexity assessment in Section 6.) No windowing is assumed for Eq. (47b).

The MA filter conversion needs no multiplications, if a rectangular window is used for Eq. (44). However, the AR filter degree is usually chosen significantly

	calculation of $h_s(n, k')$ and MA / AR filtering
mult.	$\frac{1}{r}(2M \log_2 M + 2L + 2) + P + 1$
add.	$\frac{1}{r}(3M \log_2 M + L + 1 - M) + P$
mem.	$L + 2M + P$
	filter approximation for MA LDF
mult.	$\frac{1}{r}(P + 1)$
add.	0
mem.	0
	filter approximation for AR LDF
mult.	$\frac{1}{r}((P + 1)(L + 4) + P(\mathcal{M}_{\text{div}} + \mathcal{M}_{\text{sqrt}}))$
add.	$\frac{1}{r}((P + 1)(L + 2) + P(\mathcal{A}_{\text{div}} + \mathcal{A}_{\text{sqrt}}))$
mem.	$3P$

Table 2

Algorithmic complexity in terms of average number of real multiplications and real additions per sample, and number of delay elements (memory) for the MA LDF and the AR LDF with time-domain filter of degree P .

lower than the MA filter degree such that both approaches have a similar overall algorithmic complexity as exemplified in Section 6.

5.5 Warped Low Delay Filter

The allpass transformed LDF is obtained by replacing the delay elements of the analysis filter-bank and the delay elements of the time-domain filter by allpass filters according to Eq. (34). For the warped MA LDF, a phase equalizer can be applied to obtain approximately a linear phase response as shown in Section 4.2.

The direct implementation of an allpass transformed AR filter is not possible due to delayless feedback loops. An efficient approach to eliminate them has been proposed in [36]. The modified transfer function reads

$$\tilde{H}_{\text{AR}}(z) = \frac{a_0 \tilde{a}_0}{1 - \tilde{a}_0 \frac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}} \sum_{n=1}^P \tilde{a}_n (H_A(z))^{n-1}} \quad (49)$$

with coefficients \tilde{a}_n calculated by the recursive relations

$$\tilde{a}_P = a_P \quad (50a)$$

$$\tilde{a}_n = a_n + \alpha \tilde{a}_{n+1} \quad ; \quad n = P - 1, \dots, 1 \quad (50b)$$

$$\tilde{a}_0 = (1 + \tilde{a}_1 \alpha)^{-1}. \quad (50c)$$

For the allpass transformation with real allpass coefficient $a = \alpha$, the minimum phase property is kept as proven in Appendix A. This is an important property as it guarantees stability for the warped AR filter.

Filter-ringing artifacts due to strongly varying coefficient can be avoided by interpolation in the time-domain. The used approach is a kind of 'cross-fading' and is applicable to arbitrary filters. It can be generally described by the overall transfer function

$$\bar{H}_g(z, k) = (1 - c_f(k)) H_g(z, k' - r) + c_f(k) H_g(z, k') \quad (51a)$$

$$c_f(k) = \frac{k - k'}{r}. \quad (51b)$$

A second time-domain filter with the previous filter coefficients and states is used in parallel to achieve a gradual transition from one steady state to the next one by means of the time-varying 'cross-fading' function $c_f(k)$. This smoothing does not cause an additional signal delay but requires a second time-domain filter. It should be noted that artifacts due to time-varying spectral gains must also be taken into account for AS FB designs, e.g., by non-critically downsampling etc.

6 Application to Speech Enhancement

The discussed filter-bank designs are employed for noise reduction in order to compare the achieved performance with respect to the overall quality of the enhanced speech, the algorithmic complexity and the signal delay.

6.1 Speech Enhancement Systems

The filtering of the noisy speech is done by means of the DFT AS FB according to Fig. 1, the GDFT FBE, and the MA / AR LDF. The uniform and warped versions of these filter-banks⁵ are each used. A real allpass coefficient of $a = 0.4$ is taken for the frequency warping, which provides a good approximation

⁵ The low delay filter is interpreted as a filter-bank as it is derived from the FBE.

of the Bark scale for a sampling frequency of 8 kHz [16]. A transform size of $M = 64$ and prototype filter degree of $L = 64$ is used in all cases⁶.

The analysis and synthesis prototype filter for the AS FB is a square-root Hann window obtained from Eq. (23). For the uniform AS FB, a downsampling rate of $r = 32$ is used. For the warped AS FB, a downsampling rate of $r = 8$ is needed to achieve near-perfect reconstruction for the magnitude response. (A higher downsampling rate r can be permitted at the expense of an increased signal delay, since prototype filters with higher stopband attenuation must be employed to avoid strong aliasing effects, cf. [9].) The FIR phase equalizer of Eq. (43) with filter degree $L_p = 141$ is used to achieve near-perfect signal reconstruction for the warped AS FB.

The FBE in transposed direct form is used with linear-phase prototype filter according to Eq. (21) (Hann windowing). The MA LDF in transposed direct form with filter degree $P = 32$ is taken. For the warped FBE and the warped MA LDF, the FIR phase equalizer of Eq. (43) with filter degree $L_p = 80$ and $L_p = 45$, respectively, is taken.

For the uniform and warped AR LDF with filter degree $P = 12$, the cross-fading approach according to Eq. (51) is applied to avoid filter-ringing artifacts. For FBE and LDF, a downsampling rate of $r = 64$ is taken for the analysis filter-bank.

The spectral gain factors are calculated by the super-Gaussian joint MAP estimator [3]. This joint spectral amplitude and phase estimator rests on the more accurate assumption that the real and imaginary parts of the speech DFT coefficients are rather Laplace distributed (used here) or Gamma distributed than Gaussian distributed. The required *a priori* SNR is calculated by the 'decision-directed' approach of [1] with a fixed smoothing parameter of 0.9. The needed short-term noise power spectral density is estimated by minimum statistics [37]. Speech presence uncertainty is taken into account by applying soft-gains according to [2]. Independent of the downsampling rate r of the filter-bank, the spectral gain factors $W_i(k')$ are always adapted at intervals of 64 sample instants to ease the comparison.

⁶ A lower transform size M can be used for warped filter-banks, whereas a value of $M = 256$ is often preferred for noise reduction with uniform frequency resolution and 8 kHz sampling frequency. However, individual parameter settings are not used to ease the comparison of the filter-banks.

6.2 Quality Measures

The used audio signals of 8 kHz sampling frequency are obtained from the noisy speech corpus NOIZEUS presented in [38]. A total of 20 sentences spoken by male and female speakers is taken each corrupted by four different noise sequences (airport, car, station, and street noise) with signal-to-noise ratios (SNRs) between 0 dB and 15 dB.

The quality of the enhanced speech is judged by informal listening tests and instrumental quality measures. A common time-domain measure for the overall quality of the enhanced speech $y(k) = \hat{s}(k)$ is given by the segmental (speech) SNR

$$\text{SEGSNR}/\text{dB} = \frac{1}{\mathcal{C}(\mathbb{F}_s)} \sum_{m \in \mathbb{F}_s} 10 \log_{10} \left(\frac{\sum_{\mu=0}^{N-1} s^2(mN + \mu - \kappa_0)}{\sum_{\mu=0}^{N-1} (\hat{s}(mN + \mu) - s(mN + \mu - \kappa_0))^2} \right). \quad (52)$$

The computation includes 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 undisturbed speech $s(k)$ and the input noise $n_d(k)$ can be filtered separately with coefficients adapted for the noisy speech $x(k) = s(k) + n_d(k)$. This provides the processed speech $\bar{s}(k)$ and processed noise $\bar{n}_d(k)$ separately, where $\hat{s}(k) = \bar{s}(k) + \bar{n}_d(k)$. Because of their strong correlation, the maximum of the cross-correlation between the clean and filtered speech $\varphi_{s\bar{s}}(\lambda)$ is used to determine the algorithmic signal delay

$$\kappa_0 = \arg \max_{\lambda \in \mathbb{Z}} \{\varphi_{s\bar{s}}(\lambda)\}. \quad (53)$$

The achieved noise attenuation (NA) is measured according to

$$\text{NA}/\text{dB} = 10 \log_{10} \left(\frac{1}{\mathcal{C}(\mathbb{F})} \sum_{m \in \mathbb{F}} \frac{\sum_{\mu=0}^{N-1} n_d^2(mN + \mu - \kappa_0)}{\sum_{\mu=0}^{N-1} \bar{n}_d^2(mN + \mu)} \right) \quad (54)$$

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

A frequency-domain measure for the speech quality is given by the cepstral distance (CD) between the clean speech $s(k)$ and the filtered speech $\bar{s}(k)$ as described in Appendix B. For all instrumental measures, a frame size of

$N = 256$ is used, and $N_{CD} = 40$ cepstral coefficients are considered for the CD measure.

6.3 Simulation Results for the Uniform Filter-Banks

The curves of the measured objective speech quality achieved with the uniform filter-banks are plotted in Fig. 9. The signal delay and algorithmic complexity

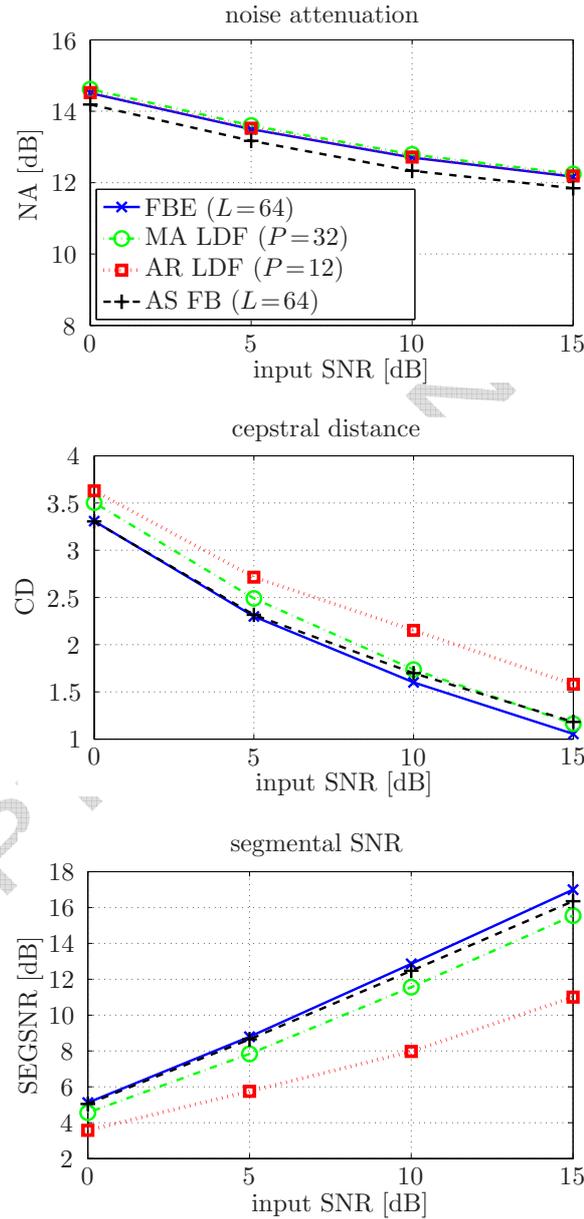


Figure 9. Objective quality measures obtained by means of the uniform filter-bank equalizer (FBE), the uniform MA and AR low delay filter (LDF), and the uniform analysis-synthesis filter-bank (AS FB).

of the filter-banks is listed in Table 3.

uniform filter-bank	signal delay [samples]	real add.	real mult.	delay elements
AS FB	64	101	31	256
FBE	32	83	144	256
MA LDF	16	51	48	225
AR LDF	0-2	62	60	216

Table 3

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

The uniform FBE achieves the same or better objective speech quality as the uniform AS FB with only a slightly higher algorithmic complexity but a significantly lower signal delay. The MA and AR LDF achieve a further reduction of the signal delay and algorithmic complexity at the price of a decreased *objective* speech quality in terms of a lower segmental SNR and higher CD. The AR filter leads to a good approximation for the magnitude response but causes phase modifications which have a very detrimental effect on the SEGSNR measure.

However, informal listening tests revealed only minor differences for the perceived *subjective* speech quality between the four systems. Therefore, a perceptual evaluation of the speech quality (PESQ) according to [39] has been performed for the enhanced speech $\hat{s}(k)$. The PESQ measure ranges from -0.5 (bad quality) to 4.5 (excellent quality), where differences of less than 0.1 PESQ units are considered to be statistically insignificant. The PESQ measure is mainly used for the assessment of speech coding systems, but also applied as a perceptual quality measure for speech enhancement systems, e.g., [4]. The measured PESQ values are shown in Fig. 10. The achieved values for the enhanced speech $\hat{s}(k)$ are almost identical for all four filter-banks. The PESQ measure can not be taken as an all-embracing measure for the subjective speech quality, cf. [40], but complies well with the impression of our informal listening tests. Thus, the proposed low delay filter can achieve a low signal delay with negligible loss for the perceived (subjective) speech quality.

6.4 Simulation Results for the Warped Filter-Banks

The curves of the measured objective speech quality achieved with the warped filter-banks are plotted in Fig. 11. The measured PESQ curves are not plotted since they were as close together as in Fig. 10 but all about 0.2 PESQ units

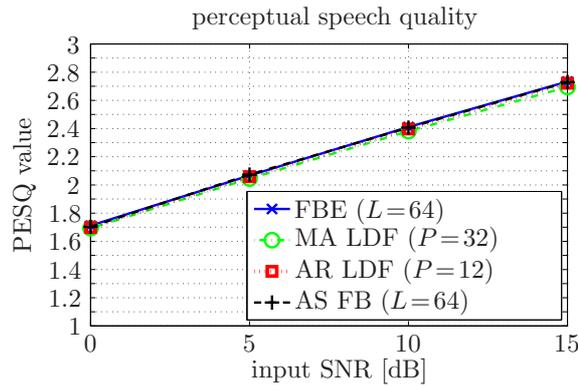


Figure 10. Perceptual evaluation of the speech quality for the enhanced speech $\hat{s}(k)$ achieved with the uniform filter-banks.

higher. The results for the instrumental measurements comply with our informal listening tests, where the overall speech quality obtained by the warped filter-banks was rated superior to that of the uniform filter-banks.

The signal delay and algorithmic complexity of the warped filter-banks is listed in Table 4. It shows the increase in algorithmic complexity and signal delay due

warped filter-bank	signal delay [samples]	real add.	real mult.	delay elements
AS FB	141	605	518	396
FBE	80	418	480	336
MA LDF	45	225	285	269
AR LDF	0-2	238	236	254

Table 4

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

to the allpass transformation. However, the warped FBE has a significantly lower signal delay and lower algorithmic complexity than the warped AS FB. As for the uniform filter-banks, a further reduction of the signal delay and algorithmic complexity can be achieved by the low delay filters with negligible loss for the perceived speech quality. The AR LDF is of minimum phase and causes an algorithmic signal delay of only a few samples.

7 Conclusions

A filter-bank design for adaptive subband filtering with low signal delay is proposed. The developed time-varying filter-bank equalizer (FBE) performs time-

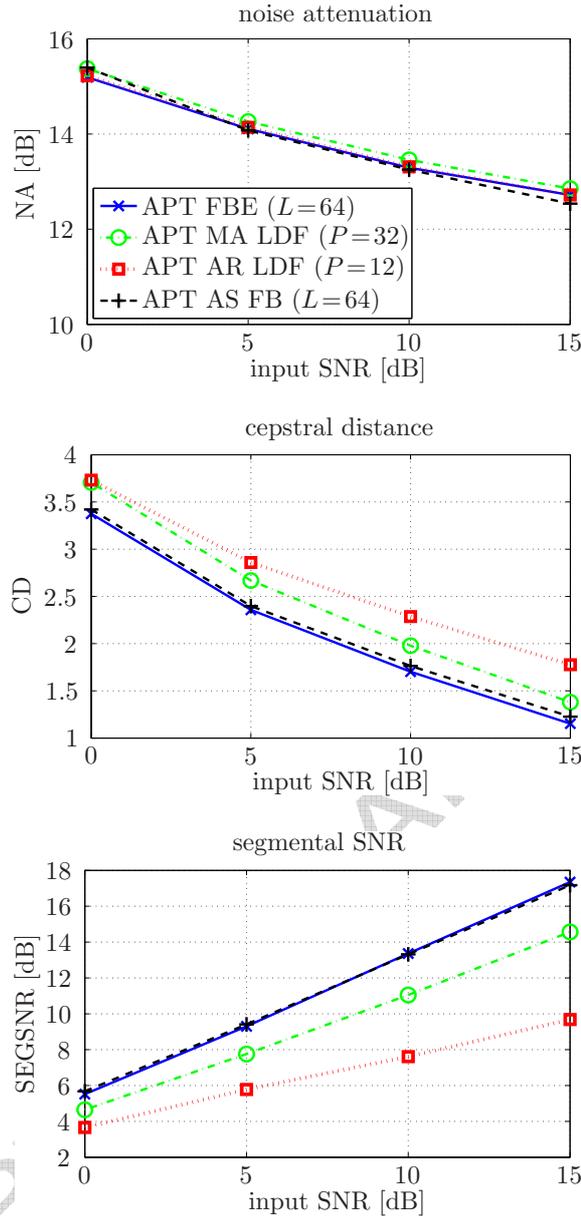


Figure 11. Objective quality measures obtained by means of the allpass transformed filter-bank equalizer (APT FBE), the APT MA and AR low delay filter (LDF), and the APT analysis-synthesis filter-bank (AS FB).

domain filtering with coefficients adapted in the frequency-domain. It is derived as an efficient realization of the filter-bank-summation method (FBSM) with explicit consideration of the time-varying coefficients. By this, the influence of the time-domain filter form can be taken into account. It is shown that the implementation of the time-domain filter in the transposed direct form yields a better perceived speech quality than the direct filter form. A non-uniform (warped) frequency resolution can be obtained by allpass transformation. The uniform and warped FBE can be efficiently implemented by means of a polyphase network (PPN), and (near-)perfect aliasing-free signal

reconstruction is achieved with significantly lower effort as with conventional analysis-synthesis filter-banks (AS FBs). The warped FBE has a lower computational complexity than the warped AS FB for most parameter configurations (L, M, r) . A phase equalizer can be employed to the warped FBE to achieve a near-linear phase response even for time-varying coefficients.

The proposed filter-bank design provides a versatile concept for low delay speech enhancement. The uniform and warped FBE achieve the same objective and subjective speech quality of the enhanced speech but with a significantly lower signal delay than conventional AS FBs. Thereby, the frequency warping can be exploited to obtain an improved speech quality or to use a lower number of frequency channels.

The devised low delay filter (LDF) is a modification of the FBE to achieve a further reduction of the signal delay and reduced algorithmic complexity with almost no loss for the perceived subjective speech quality. The warped MA LDF can maintain a constant (near-linear) phase characteristic, which is beneficial, e.g., for speech processing with microphone arrays. The proposed uniform and warped AR LDF has minimum phase property and can achieve an algorithmic signal delay of only a few samples.

Acknowledgement

The authors are grateful for the support of GN ReSound, Eindhoven, The Netherlands.

A Preservation of Minimum Phase Property for Allpass Transformed Filters

All poles and zeros of a minimum phase filter are within the unit circle. Hence, it is sufficient to treat the allpass transformation of a causal filter with a single pole

$$F(z) = \frac{1}{1 - p_{\infty} z^{-1}} \quad (\text{A.1a})$$

$$|p_{\infty}| < 1 \ ; \ |p_{\infty}| < |z| \ ; \ p_{\infty} \in \mathbb{C} \quad (\text{A.1b})$$

which yields the warped transfer function

$$\begin{aligned}\tilde{F}(z) &= \frac{1}{1 - p_\infty \frac{z^{-1} - a^*}{1 - a z^{-1}}} \\ &= \frac{1}{1 + p_\infty a^*} \cdot \frac{1 - \tilde{p}_0 z^{-1}}{1 - \tilde{p}_\infty z^{-1}}\end{aligned}\quad (\text{A.2})$$

with zero and pole given by

$$\tilde{p}_0 = a \quad (\text{A.3a})$$

$$\tilde{p}_\infty = \frac{a + p_\infty}{1 + p_\infty a^*}. \quad (\text{A.3b})$$

The zero is always inside the unit circle due to Eq. (35b). The pole is within the unit circle if

$$|a + p_\infty| < |1 + p_\infty a^*|. \quad (\text{A.4})$$

For a real allpass pole $a = \alpha$, this equation can be converted to

$$0 < (1 - \alpha^2)(1 - |p_\infty|^2). \quad (\text{A.5})$$

Thus, a real allpass transformation according Eq. (34) yields a warped filter with minimum phase. This does not hold for a complex allpass transformation, since Eq. (A.4) does not apply for all a being complex.

B Cepstral Distance Measure

The cepstral coefficients of a speech frame m of length N are given by

$$\begin{aligned}C_s(mN + \mu) &= \text{IDFT}_N\{\ln |\text{DFT}_N\{s(mN + \mu)\}|\} \\ \mu &= 0, \dots, N - 1.\end{aligned}\quad (\text{B.1})$$

The (mean) cepstral distance (CD) between the filtered speech $\bar{s}(k)$ and original speech $s(k - \kappa_0)$ is calculated by the average

$$\text{CD} = \frac{10}{\ln(10) \cdot \mathcal{C}(\mathbb{F}_s)} \sum_{m \in \mathbb{F}_s} \text{CD}(m) \quad (\text{B.2a})$$

$$\begin{aligned}\text{CD}(m) &= \left([C_s(mN) - C_{\bar{s}}(mN)]^2 \right. \\ &\quad \left. + 2 \sum_{\mu=1}^{N_{\text{CD}}-1} [C_s(mN + \mu) - C_{\bar{s}}(mN + \mu)]^2 \right)^{\frac{1}{2}}.\end{aligned}\quad (\text{B.2b})$$

References

- [1] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.
- [2] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking Speech-Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environments," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Phoenix, USA, May 1999, pp. 789–792.
- [3] T. Lotter and P. Vary, "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1110–1126, 2005.
- [4] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*. Berlin, Heidelberg: Springer, 2005.
- [5] J. M. Kates and K. H. Arehart, "Multichannel Dynamic-Range Compression Using Digital Frequency Warping," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 3003–3014, 2005.
- [6] R. E. Crochiere, "A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 10, pp. 99–102, Feb. 1980.
- [7] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, 2nd ed. Berlin, Heidelberg: Springer, 1999.
- [8] A. Engelsberg, "Transformation-Based Systems for Single-Channel Noise Reduction in Speech Signals," Ph.D. dissertation, Christian-Albrechts University, Kiel, Germany, 1998, (in German).
- [9] T. Gölzow, A. Engelsberg, and U. Heute, "Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement," *Signal Processing, Elsevier*, vol. 64, pp. 5–19, Jan. 1998.
- [10] I. Cohen, "Enhancement of Speech Using Bark-Scaled Wavelet Packet Decomposition," in *Proc. of European Conf. on Speech Communication and Technology (EUROSPEECH)*, Aalborg, Denmark, Sept. 2001.
- [11] A. Petrovsky, M. Parfieniuk, and A. Borowicz, "Warped DFT based perceptual noise reduction system," in *Convention Paper of Audio Engineering Society*, Berlin, Germany, May 2004.
- [12] A. V. Oppenheim, D. H. Johnson, and K. Steiglitz, "Computation of Spectra with Unequal Resolution Using the Fast Fourier Transform," *Proc. of the IEEE*, vol. 59, no. 2, pp. 299–301, Feb. 1971.

- [13] C. Braccini and A. V. Oppenheim, "Unequal Bandwidth Spectral Analysis using Digital Frequency Warping," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-22, no. 4, pp. 236–244, Aug. 1974.
- [14] P. Vary, "Digital Filter Banks with Unequal Resolution," in *Short Communication Digest of European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, Sept. 1980, pp. 41–42.
- [15] G. Doblinger, "An Efficient Algorithm for Uniform and Nonuniform Digital Filter Banks," in *Proc. of Intl. Symp. on Circuits and Systems (ISCAS)*, vol. 1, Singapore, June 1991, pp. 646–649.
- [16] J. O. Smith and J. S. Abel, "Bark and ERB Bilinear Transforms," *IEEE Trans. on Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, Nov. 1999.
- [17] P. Vary, "An adaptive filter-bank equalizer for speech enhancement," *Signal Processing, Elsevier, Special Issue on Applied Signal and Audio Processing*, vol. 86, pp. 1206–1214, 2006.
- [18] H. W. Löllmann and P. Vary, "Efficient Non-Uniform Filter-Bank Equalizer," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Antalya, Turkey, Sept. 2005.
- [19] D. R. Morgen and J. C. Thi, "A Delayless Subband Adaptive Filter Architecture," *IEEE Trans. on Signal Processing*, vol. 43, no. 8, pp. 1819–1830, Aug. 1995.
- [20] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Upper Saddle River, New Jersey: Prentice-Hall, 1993.
- [21] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999.
- [22] 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*, vol. 48, no. 3, pp. 737–748, Mar. 2000.
- [23] T. Karp and A. Mertins, "Lifting Schemes for Biorthogonal Modulated Filter Banks," in *International Conference on Digital Signal Processing*, vol. 1, Santorini, Greece, June 1997, pp. 443–446.
- [24] E. Galijašević and J. Kliewer, "Design of Allpass-Based Non-Uniform Oversampled DFT Filter Banks," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, USA, May 2002, pp. 1181–1184.
- [25] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1983.
- [26] O. Cappé, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor," *IEEE Trans. on Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, Apr. 1994.

- [27] K. G. Beauchamp, *Walsh Functions And Their Applications*. London: Academic Press, 1975.
- [28] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*. Upper Saddle River, New Jersey: Prentice-Hall, 1996.
- [29] T.-Y. Leou and J. Aggarwal, "Recursive Implementation of LTV Filters - Frozen-Time Transfer Function Versus Generalized Transfer Function," *Proc. of the IEEE*, vol. 72, no. 7, July 1984.
- [30] M. G. Bellanger, G. Bonnerot, and M. Coudreuse, "Digital Filtering by Polyphase Network: Application to Sample-Rate Alteration and Filter Banks," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-24, no. 2, pp. 109–114, Apr. 1976.
- [31] P. Vary, "On the Design of Digital Filter Banks Based on a Modified Principle of Polyphase," *Archive for Electronics and Communication (AEÜ)*, vol. 33, pp. 293–300, 1979.
- [32] M. Renfors and T. Saramäki, "Recursive Nth-Band Digital Filters-Part I: Design and Properties," *IEEE Trans. on Circuits and Systems*, vol. 34, no. 1, pp. 24–37, Jan. 1987.
- [33] H. W. Löllmann and P. Vary, "Parametric Phase Equalizers for Warped Filter-Banks," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sept. 2006.
- [34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge: Cambridge University Press, 1992.
- [35] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral Subtraction Using Reduced Delay Convolution and Adaptive Averaging," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 8, pp. 799–807, Nov. 2001.
- [36] K. Steiglitz, "A Note on Variable Recursive Digital Filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 1, pp. 111–112, Feb. 1980.
- [37] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 9, no. 5, pp. 504–512, July 2001.
- [38] Y. Hu and P. C. Loizou, "Subjective Comparison of Speech Enhancement Algorithms," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, May 2006, pp. 153–156.
- [39] 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 coders," Feb. 2001.
- [40] Y. Hu and P. C. Loizou, "Evaluation of Objective Measures for Speech Enhancement," in *Proc. of Interspeech*, Philadelphia, USA, Sept. 2006.