A WARPED LOW DELAY FILTER FOR SPEECH ENHANCEMENT

Heinrich W. Löllmann and Peter Vary

Institute of Communication Systems and Data Processing (IND)
RWTH Aachen University, D-52056 Aachen, Germany
{loellmann|vary}@ind.rwth-aachen.de

ABSTRACT

Warped analysis-synthesis filter-banks with Bark-scaled frequency bands are used for speech enhancement systems to improve the subjective speech quality. In this contribution, an alternative warped filter(-bank) structure is proposed which has a significantly lower signal delay and algorithmic complexity. The warped moving-average low delay filter allows to decrease the signal delay in a simple and flexible manner. The warped auto-regressive low delay filter has minimum phase property and can achieve a delay of only a few samples. The application to speech enhancement shows that a similar subjective quality for the enhanced speech can be achieved as by means of a warped analysis-synthesis filter-bank.

1. INTRODUCTION

Frequency warped filter-banks obtained by allpass transformation [1],[2] are able to approximate the Bark frequency scale with great accuracy [3]. This property to mimic the frequency resolution of the human auditory system is exploited by speech and audio processing applications. An example are speech enhancement systems where non-uniform (warped) analysis-synthesis filter-banks (AS FBs) are used to achieve an improved (subjective) speech quality, e.g., [4]. However, an allpass transformed filter-bank has a higher computational complexity and signal delay in comparison to the corresponding uniform filter-bank, which can only be partly compensated by using a smaller number of frequency channels. This makes it difficult to employ warped AS FBs for applications where system delay and computational complexity are strictly limited, such as noise reduction systems for mobile communication devices or hearing-aids.

A warped filter-bank with a significantly lower signal delay and algorithmic complexity than for the corresponding warped AS FB is proposed in [5],[6], termed as filter-bank equalizer (FBE). For dynamic-range compression in hearing-aids, a similar approach has been presented independently in [7].

In this contribution, a modification of the FBE concept is proposed to further decrease its signal delay and algorithmic complexity with almost no loss for the perceived subjective quality of the enhanced speech. Thus, the devised low delay filter (LDF) is of interest for speech enhancement systems with demanding requirements for the permitted system delay.

In Section 2.1, the concept of the uniform LDF is introduced first. In Section 2.2, the moving-average (MA) LDF is discussed; and the auto-regressive (AR) LDF is treated in Section 2.3. The more general warped LDF is proposed in Section 3. A comparison of warped AS FB and warped LDF is given in Section 4. The paper concludes with Section 5.

2. UNIFORM LOW DELAY FILTER

2.1. Concept

The *filter-bank equalizer* (FBE) [5],[6] performs time-domain filtering with coefficients adapted in the uniform or non-uniform frequency-domain. If applied to speech enhancement, the FBE achieves a very similar objective and subjective quality for the enhanced speech as the corresponding[1] *analysis-synthesis filter-bank* (AS FB) but with a significantly lower algorithmic signal delay and lower computational complexity [8].

A further reduction of the signal delay and algorithmic complexity can be achieved by approximating the time-domain filter of the FBE by a filter of lower degree. This modification of the FBE concept, termed as *low delay filter* (LDF), is illustrated in Fig. 1. The uniform LDF [9] is regarded first, before introducing the more general warped LDF in Section 3. The M spectral coefficients (sub-band signals) $X_i(k')$ are calculated at intervals of r samples by means of a DFT[2] analysis filter-bank

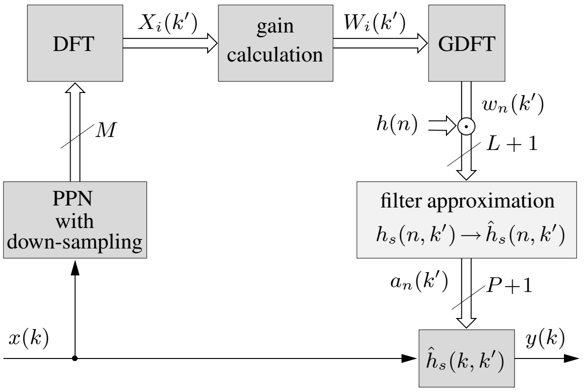

Figure 1: *Low delay filter (LDF) for adaptive noise reduction.*

$$X_i(k') = \sum_{n=0}^{L} x(k'-n)\, h(n)\, e^{-j\,\frac{2\pi}{M}\,i\,n} \qquad (1)$$
$$i = 0, 1, \ldots, M-1$$

[1] FBE and AS FB use (almost) the same analysis filter-bank.

[2] Other spectral transforms are discussed in [6].

This work was supported by GN ReSound, Eindhoven.

with $h(n)$ denoting the real impulse response of the prototype lowpass filter of length $L+1 \geq M$. This analysis filter-bank can be efficiently realized by means of a polyphase network (PPN) with down-sampling [10],[6] with the DFT computed by the Fast Fourier Transform (FFT), e.g., [11]. The spectral gains $W_i(k')$ can be calculated by any spectral speech estimator for noise suppression, e.g., [12]. The obtained real spectral gains with $0 \leq W_i(k') \leq 1$ are of zero phase. The (evenly-stacked) generalized discrete Fourier transform (GDFT) of the spectral gains $W_i(k')$ yields $L+1$ time-domain weighting factors

$$w_n(k') = \sum_{i=0}^{M-1} W_i(k')\, e^{-j\,\frac{2\pi}{M}\,i\,(n-n_0)} \qquad (2)$$
$$n = 0, 1, \ldots, L$$

where the variable n_0 ensures coefficients with non-zero phase[3]. For example, the choice $n_0 = L/2$ (L even) yields weighting factors with linear phase property, that is, $w_n(k') = w_{L-n}(k')$. The GDFT of Eq. (2) can be efficiently realized by a FFT of the gains W_i followed by a cyclic shift of the time-domain coefficients by n_0 samples. The time-varying FIR filter coefficients

$$h_s(n,k') = h(n)\, w_n(k') \quad ; \quad n = 0, 1, \ldots, L \qquad (3)$$

constitute the time-domain filter of the filter-bank equalizer [6]. The signal delay is now further reduced by approximating the (FIR) filter of Eq. (3) and degree L by a filter of lower degree P and impulse response $\hat{h}_s(n,k')$, cf. Fig. 1. By this, the signal delay is reduced without requiring an adjustment of the spectral gain calculation since the transform size M is not changed. The efficient realization of the LDF by means of an FIR and IIR filter approximation is discussed in the sequel.

2.2. Moving-Average Low Delay Filter

The time-domain filter of Eq. (3) can be approximated by an FIR filter of degree $P < L$ following a technique very similar to FIR filter design by windowing, e.g., [11]. The impulse response[4] $h_s(n)$ of Eq. (3) is truncated by a window sequence of length $P+1$ according to

$$\hat{h}_s(n) = a_n = h_s(n+n_c)\,\mathrm{win}_P(n) \quad ; \quad n = 0, 1, \ldots, P \quad (4)$$

with the general window sequence given by

$$\mathrm{win}_P(n) \begin{cases} \neq 0 & ; \; 0 \leq n \leq P \\ = 0 & ; \; \text{else}\,. \end{cases} \qquad (5)$$

The window sequence and value for n_c can be chosen, e.g., to obtain an FIR filter with linear phase response. This approximation of the original filter by an FIR filter is termed as *moving-average low delay filter* (MA LDF). The (MA) low delay filter comprises the overall system according to Fig. 1, and the term MA filter only refers to the actual FIR time-domain filter with impulse response $\hat{h}_s(n)$.

2.3. Auto-Regressive Low Delay Filter

A significantly lower signal delay than for the MA filter can be achieved by a recursive minimum phase filter. Here, an allpole filter or auto-regressive (AR) filter, respectively, is considered. This approximation neglects the phase response of the original filter which, however, is tolerable (for noise reduction applications) due to the insensitivity of the human ear towards phase modifications, cf. [13]. The $P+1$ coefficients a_n of the AR filter

$$\hat{H}_s(z) = H_{\mathrm{AR}}(z) = \frac{a_0}{1 - \sum\limits_{n=1}^{P} a_n\, z^{-n}} \qquad (6)$$

are determined by the filter coefficients $h_s(n)$ of Eq. (3) with methods taken from parametric spectrum analysis, e.g., [11]. A relation between the coefficients $h_s(n)$ and a_n can be established by the Yule-Walker equations

$$\begin{bmatrix} \varphi(1) \\ \vdots \\ \varphi(P) \end{bmatrix} = \begin{bmatrix} \varphi(0) & \ldots & \varphi(1-P) \\ \vdots & \ddots & \vdots \\ \varphi(P-1) & \ldots & \varphi(0) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} \qquad (7)$$

with

$$\varphi(\lambda) = \sum_{n=0}^{L-|\lambda|} h_s(n)\, h_s(n+\lambda) \quad ; \quad 0 \leq |\lambda| \leq P \qquad (8)$$

$$a_0 = \sqrt{\varphi(0) - \sum_{n=1}^{P} a_n\, \varphi(-n)}\,. \qquad (9)$$

The used auto-correlation method to calculate $\varphi(n)$ ensures a symmetric Toeplitz structure for the auto-correlation matrix in Eq. (7). This allows to solve the Yule-Walker equations efficiently by means of the Levinson-Durbin recursion, e.g., [11]. The obtained AR filter is always stable and of minimum phase as the auto-correlation matrix is positive-definite. This IIR filter approximation yields the *auto-regressive low delay filter* (AR LDF) in analogy to the terminology of the previous section.

A general IIR filter (ARMA filter) approximation is also possible, but this approach is much more complex and prone to numerical inaccuracies, cf. [11].

3. WARPED LOW DELAY FILTER

A low delay filter with non-uniform frequency resolution can be obtained by digital frequency warping using an allpass transformation [1],[2]. This transformation is achieved by substituting all delay elements of the discrete filters by allpass filters $z^{-1} \rightarrow H_A(z)$. A (causal) real allpass filter of first order is used here. Its frequency response reads

$$H_A(e^{j\Omega}) = \frac{e^{-j\Omega} - \alpha}{1 - \alpha\, e^{-j\Omega}} = e^{-j\,\varphi_\alpha(\Omega)} \qquad (10)$$
$$\alpha \in \mathbb{R} \quad ; \quad |\alpha| < 1$$

$$\varphi_\alpha(\Omega) = -\Omega + 2\,\arctan\left(\frac{\sin\Omega}{\cos\Omega - \alpha}\right)\,. \qquad (11)$$

The warped LDF is obtained directly by applying the allpass transformation to the analysis filter-bank and the time-domain

[3] In principle, the GDFT of Eq. (2) has to be used for the analysis filter-bank of Eq. (1) as well [6]. However, the gain calculation for noise suppression is based on the magnitude $|X_i(k')|$ such that a common DFT analysis filter-bank can also be taken.

[4] The time-dependency of the filter coefficients on k' is omitted for the sake of simplicity.

filter. The frequency response of the warped filter is given by

$$\widetilde{H}_s(e^{j\,\Omega}) = H_s(e^{j\,\varphi_\alpha(\Omega)}) \tag{12}$$

where the tilde-notation is used here to mark quantities altered by allpass transformation. Thus, the allpass transformation leads to a frequency warping $\Omega \rightarrow \varphi_\alpha(\Omega)$. For a positive value of α, a higher frequency resolution is obtained for the lower frequency bands and vice versa.

A phase equalizer can be applied to the output signal of the warped MA filter to obtain approximately a (generalized) linear phase response, cf. [6]. The phase equalizer can be omitted for small filter degrees P as the ear does not perceive the phase modifications due to the allpass transformation in this case.

The direct implementation of the warped allpole filter is not possible as the allpass transformation yields delay-less feedback loops. An efficient approach to eliminate them has been proposed by Steiglitz [14]. The warped AR filter is now given by

$$\widetilde{H}_{\text{AR}}(z) = \frac{a_0\,\tilde{a}_0}{1 - \tilde{a}_0\,\frac{(1-\alpha^2)\,z^{-1}}{1-\alpha\,z^{-1}}\,\sum_{n=1}^{P} \tilde{a}_n\,H_A(z)^{n-1}} \tag{13}$$

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

$$\tilde{a}_P = a_P \tag{14a}$$

$$\tilde{a}_n = a_n + \alpha\,\tilde{a}_{n+1} \quad ; \quad n = P-1,\ldots,1 \tag{14b}$$

$$\tilde{a}_0 = (1 + \tilde{a}_1\,\alpha)^{-1}\,. \tag{14c}$$

It can be shown that the warped AR filter keeps the minimum phase property for $|\alpha| < 1$ and, thus, remains stable.

The algorithmic complexity for the warped AR LDF is listed in Table 1. The variable $\mathcal{M}_{\text{div}}$ marks the number of multiplications needed for a division operation, and $\mathcal{M}_{\text{sqrt}}$ represents the number of multiplications needed for a square-root operation, whose values dependent on the used numeric procedure. Accordingly, the variables $\mathcal{A}_{\text{div}}$ and $\mathcal{A}_{\text{sqrt}}$ mark the additions needed for a division and square-root operation, respectively. (A value of 15 will be taken for each of these variables later in Section 4.)

	computation of $\tilde{h}_s(n, k')$
multiplications	$\frac{1}{r}(2M\log_2 M + 2\,L{+}2) + 2\,L$
additions	$\frac{1}{r}(3M\log_2 M{+}L{+}1{-}M){+}2\,L$
delay elements	$L{+}2\,M$
	computation of a_n
multiplications	$\frac{1}{r}\left((P{+}1)\,(L{+}4) + P\,(\mathcal{M}_{\text{div}} + \mathcal{M}_{\text{sqrt}})\right)$
additions	$\frac{1}{r}\left((P{+}1)\,(L{+}2) + P\,(\mathcal{A}_{\text{div}} + \mathcal{A}_{\text{sqrt}})\right)$
memory	$(3\,P)$
	computation of $\tilde{a}_n$ and actual filtering
multiplications	$\frac{1}{r}\,(P{+}\mathcal{M}_{\text{div}}) + 3\,P{+}1$
additions	$\frac{1}{r}\,(P{+}\mathcal{A}_{\text{div}}) + 3\,P$
delay elements	$P+1$

Table 1: *Algorithmic complexity in terms of required average number of real multiplications and real additions per sample instant, and number of delay elements (memory) for a warped AR low delay filter.*

The real allpass filter of first order can be realized with 2 real multiplications, 2 real additions and one delay element. The regarded (G)DFT can be computed in-place by the radix-2 FFT algorithm, cf. [11]. Thereby, the FFT of a real sequence of size M can be computed by a complex FFT of size $M/2$ with approximately half the algorithmic complexity.

The algorithmic complexity for the warped MA LDF can be derived from Table 1 as well with the difference that the calculation of the MA filter coefficients a_n according to Eq. (4) requires only $1/r(P+1)$ multiplications for a non-rectangular window. However, the degree P of the (warped) AR filter is usually chosen to be lower than for the MA filter such that both LDFs have a comparable computational complexity (see Section 4).

The switching of the time-domain filter coefficients $a_n(k')$ during operation can lead to perceptually annoying artifacts (e.g., 'click sounds') which can be avoided by an appropriate smoothing over time.

4. COMPARISON OF ANALYSIS-SYNTHESIS FILTER-BANK AND LOW DELAY FILTER

The discussed filter(-bank) designs have been employed for noise reduction. The regarded warped M-channel DFT AS FB employs an analysis and synthesis prototype filter of degree $L+1 = M = 64$. A relatively low down-sampling factor of $r = M/8$ is needed to avoid aliasing effects due to the non-uniform frequency bands. (A higher value for r can be permitted at the expense of a longer prototype filter with $L \gg M$, cf. [4].)

The warped AS FB is compared with a warped MA LDF ($L = 63$, $M = 64$, $P = 32$) and a warped AR LDF ($L = 63$, $M = 64$, $P = 12$). A higher down-sampling factor of $r = M/2$ than for the AS FB is taken as aliasing effects are negligible due to the time-domain filtering.

An allpass coefficient of $\alpha = 0.4$ is chosen for the frequency warping which yields a good approximation of the Bark scale for the regarded sampling frequency of 8 kHz [3]. FIR phase equalizers with 141 taps and 45 taps have been employed for the warped AS FB and the warped MA LDF, respectively, to compensate phase distortions due to the allpass transformation, cf. [6],[4]. A phase equalizer is not needed for the warped AR filter.

The spectral gains $W_i(k')$ are determined by the Wiener rule (MMSE estimator). The required a priori SNR is calculated by the decision-directed approach with noise PSD estimation based on minimum statistics, see [12]. The gains are adapted at intervals of $M/2 = 32$ samples in all cases to ease the comparison of the filter structures. Noise of a moving tank and car noise from the NOISEX-92 database are added to a male and female speech sequence at a signal-to-noise ratio (SNR) of 0 dB and 15 dB, respectively. In the simulation, speech and noise can be filtered separately with coefficients adapted for the noisy speech $x(k) = s(k) + n(k)$, such that the output sequence reads $y(k) = \hat{s}(k) = \bar{s}(k) + \bar{n}(k)$. With these separate sequences, the segmental speech SNR ($\text{SNR}_{\text{seg}}^{\text{speech}}$) and the segmental noise (power) attenuation (NA) can be calculated (e.g., Chap. 4 in [12]). These two time-domain measures account for the trade-off between speech distortions and noise power reduction. Because of their strong correlation, the algorithmic signal delay due to the filtering κ_0 is determined by means of the cross-correlation sequence $\varphi_{s\bar{s}}(\lambda)$ between the clean speech $s(k)$ and the filtered speech $\bar{s}(k)$ according to

$$\kappa_0 = \arg \max_{\lambda \in \mathbb{Z}} \{\varphi_{s\bar{s}}(\lambda)\} \; . \tag{15}$$

A perceptual evaluation of the speech quality of the enhanced speech $y(k) = \hat{s}(k)$ is performed by the PESQ measure [15].

The obtained results and properties of the three filter(-bank) structures are listed in Table 2. The LDFs achieve a significantly lower algorithmic signal delay and computational complexity in comparison to the AS FB with almost no loss for the perceived *subjective* speech quality as indicated by the PESQ measures. This complies with informal listening tests where the speech quality was rated similar for all three filter structures. The AR LDF is able to achieve a very low signal delay at the price of a decreased *objective* speech quality (lower segmental speech SNR) since the phase is neglected by the AR filter approximation. However, this does apparently not lead to a diminished subjective speech quality.

instrumental measures for speech enhancement

	0 dB			15 dB		
	$\mathrm{SNR}_{\mathrm{seg}}^{\mathrm{speech}}$ [dB]	NA [dB]	PESQ	$\mathrm{SNR}_{\mathrm{seg}}^{\mathrm{speech}}$ [dB]	NA [dB]	PESQ
AS FB	6.78	11.86	1.60	18.18	8.80	2.72
MA LDF	6.26	11.74	1.59	17.25	8.80	2.72
AR LDF	4.74	11.69	1.60	10.54	8.73	2.73

signal delay and algorithmic complexity

	delay κ_0 [samples]	real multiplications	real additions	delay elements
AS FB	141	605	518	396
MA LDF	45	225	285	269
AR LDF	$0-2$	238	236	236

Table 2: *Comparison of warped analysis-synthesis filter-bank (AS FB), warped moving-average low delay filter (MA LDF) and warped auto-regressive low delay filter (AR LDF) used for noise reduction. The algorithmic complexity considers the average number of real operations per sample instant excluding the complexity for the spectral gain calculation.*

Filtering with uniform frequency resolution can be regarded as special case with $\alpha = 0$ for Eq. (10) such that $H_A(z) = z^{-1}$. In this case, the LDF achieves a significantly lower signal delay than the corresponding uniform AS FB as well, where instrumental measurements and informal listening tests revealed a similar subjective speech quality for all three filter structures (hence not listed). Informal listening tests and PESQ measures judged the speech quality achieved by the warped filter(-bank) structures superior to that of their uniform counterparts ($\alpha = 0$), which complies with the findings in [4].

5. CONCLUSIONS

An alternative filter(-bank) structure to that of the uniform and warped analysis-synthesis filter-bank (AS FB) is proposed, which possesses a significantly lower signal delay and lower algorithmic complexity. The proposed MA low delay filter (LDF) allows to decrease the signal delay in a simple and flexible manner due to the employed FIR filter approximation by windowing. A near linear phase characteristic can be achieved for the warped MA LDF by employing a (fixed) phase equalizer. The devised AR LDF employs a recursive minimum phase time-domain filter and can achieve a signal delay of only a few samples.

A possible application of the proposed low delay filter are noise reduction systems for mobile communication devices or hearing-aids. The uniform and warped LDF can achieve a similar subjective quality for the enhanced speech as by means of a corresponding AS FB. Thus, the discussed filter(-bank) concept provides an efficient approach to exploit the benefits of coefficient adaptation in the uniform or warped frequency-domain while being able to fulfill demanding signal delay constraints.

6. REFERENCES

[1] C. Braccini and A. V. Oppenheim, “Unequal Bandwidth Spectral Analysis using Digital Frequency Warping,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 22, no. 4, pp. 236–244, Aug. 1974.

[2] G. Doblinger, “An Efficient Algorithm for Uniform and Nonuniform Digital Filter Banks,” in *Proc. of Intl. Symp. on Circuits and Systems (ISCAS)*, Singapore, June 1991, vol. 1, pp. 646–649.

[3] J. O. Smith and J. S. Abel, “Bark and ERB Bilinear Transforms,” *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, Nov. 1999.

[4] T. Gülzow, 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.

[5] 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.

[6] H. W. Löllmann and P. Vary, “Efficient Non-Uniform Filter-Bank Equalizer,” in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Antalya, Turkey, Sept. 2005.

[7] 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.

[8] H. W. Löllmann and P. Vary, “Generalized Filter-Bank Equalizer for Noise Reduction with Reduced Signal Delay,” in *Proc. of European Conf. on Speech Communication and Technology (Interspeech)*, Lisbon, Portugal, Sept. 2005.

[9] H. W. Löllmann and P. Vary, “Low Delay Filter for Adaptive Noise Reduction,” in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005, pp. 205–208.

[10] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[11] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, Upper Saddle River, New Jersey, 1996.

[12] J. Benesty, S. Makino, and J. Chen, Eds., *Speech Enhancement*, Springer, Berlin, Heidelberg, 2005.

[13] P. Vary, “Noise Suppression by Spectral Magnitude Estimation - Mechanism and Theoretical Limits -,” *Signal Processing, Elsevier*, vol. 8, no. 4, pp. 387–400, July 1985.

[14] K. Steiglitz, “A Note on Variable Recursive Digital Filters,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 111–112, Feb. 1980.

[15] 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.