A Filter Structure for Low Delay Noise Suppression

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

Many noise suppression systems are based on (short-term) spectral-domain filtering by using an analysis-synthesis filter-bank. This filtering technique causes a relatively high signal delay as a filter-bank for the analysis and synthesis part is needed. In this contribution, a filter structure with a significantly lower signal delay is presented. The proposed low delay filter achieves a higher amount of noise reduction than an analysis-synthesis filter-bank with the same signal delay.

1. Introduction

Today, noise reduction systems are often integrated into speech processing and transmission systems, such as mobile phones, hands-free telephony, tele-conferencing systems, or hearing aids. For such systems, a low signal delay is an important necessity for a convenient and pleasant speech communication.

The enhancement of noisy speech is usually performed by means of spectral weighting employing a (DFT) *analysis-synthesis filter-bank* (AS FB). This method causes a relatively high signal delay as an analysis FB and synthesis FB is needed. In [1], an approach to reduce the signal delay of a noise reduction system based on an AS FB is presented.

A different approach to decrease the signal delay due to filtering is to employ the recently proposed *filter-bank equalizer* (FBE) [2], [3]. This filter(-bank) has a lower signal delay than the corresponding AS FB for the same number of frequency bands and prototype filter length. The noise reduction (noise attenuation and speech quality) achieved by the FBE and the corresponding AS FB is similar [4].

A modification of the FBE is presented in this contribution. The proposed *low delay filter* (LDF) allows to reduce the signal delay in a simple and flexible manner.

2. Low Delay Filter

The principle of the low delay filter (LDF) applied for noise reduction is shown in Fig. 1.

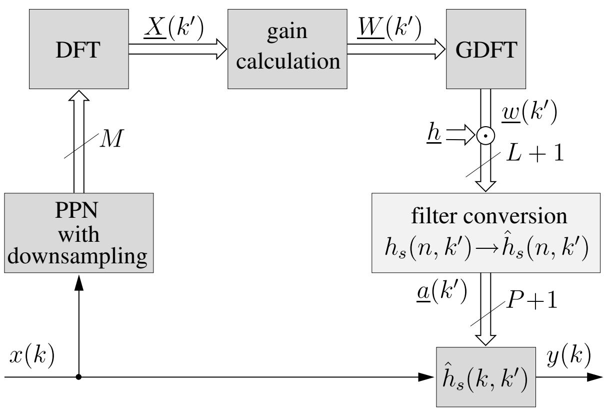

Figure 1: Low delay filter (LDF) for adaptive filtering. (Vectors are marked by the underline and the 'circle-dot' denotes an element-wise vector multiplication.)

The LDF is a modification of the uniform filter-bank equalizer (FBE) [2], [3]. The block 'filter conversion' is inserted to further reduce the signal delay without requiring a parameter adjustment of the gain calculation.

The M spectral coefficients $\underline{X}(k')$ are calculated at a reduced rate $k' = k/r$ by means of a DFT polyphase network (PPN) analysis FB (cf. [5]). This allows the prototype lowpass filter with impulse response $\underline{h} = [\,h(0), \ldots, h(L)\,]^T$ to be longer than the DFT size M to increase frequency selectivity. The calculation of the M spectral gains $\underline{W}(k') = [\,W_0(k'), \ldots, W_{M-1}(k')\,]^T$ can be done by common spectral speech estimators, e.g., [6], [7]. The obtained spectral gains with $0 \leq W_i(k') \leq 1$ are of zero phase. The generalized discrete Fourier transform (GDFT)

$$w_n(k') = \sum_{i=0}^{M-1} W_i(k')\, e^{-j\,\frac{2\pi}{M}\,(n-n_0)\,(i-i_0)} \qquad (1)$$

with $n_0 = L/2$ and $i_0 = 0$ yields $L+1$ time-domain weighting factors $w_n(k')$, due to the implicit periodicity of the (G)DFT, with linear phase. (This can also be achieved by a DFT of the gains $W_i(k')$ followed by a cyclic shift of $w_n(k')$ by $L/2$ samples.) The filter coefficients

$$h_s(n, k') = h(n)\, w_n(k') \;\; ; \;\; n = 0, \ldots, L \qquad (2)$$

constitute the time-domain filter of the FBE.
The signal delay is further reduced by approximation of the time-domain filter of Eq. (2) by a filter with lower filter degree P and impulse response $\hat{h}_s(n, k')$. The time-varying coefficients $\underline{a}(k')$ of this filter are calculated by the additional block 'filter conversion' in Fig. 1. In the following, two different filter approximations will be proposed to realize the low delay filter (LDF).

2.1 Moving-Average Low Delay Filter

The original filter (of the FBE) can be approximated by an FIR filter of degree $P < L$ employing a technique very similar to FIR filter design by windowing, e.g., [5]. The impulse response $h_s(n)$ of Eq. (2) is truncated by a window sequence of length $P{+}1$ according to[1]

$$\hat{h}_s(n) = a_n = h_s(n)\, \mathrm{win}_P(n - n_c) \qquad (3)$$

$$\text{with} \quad \mathrm{win}_P(n) \begin{cases} \neq 0 & ; \; 0 \leq n \leq P \\ = 0 & ; \; \text{else}\,. \end{cases} \qquad (4)$$

[1]The time-dependency of the filter coefficients on k' will be dropped in the sequel for the sake of simplicity.

The window sequence and the value n_c can be chosen, e.g., to obtain a linear phase filter. The approximation of the original filter by a moving-average (MA) filter results in the MA LDF. To ease the treatment, the term (MA) LDF shall refer to the overall system according to Fig. 1, and the term (MA) filter includes only the actual filter with impulse response $\hat{h}_s(n)$.

2.2 Auto-Regressive Low Delay Filter

A lower signal delay than for the MA filter can be achieved by a minimum phase filter. Here, an allpole filter or auto-regressive (AR) filter, respectively, will be regarded since the phase response must not be approximated for noise reduction applications.
The $P+1$ coefficients a_n of the AR filter

$$\widehat{H}_s(z) = \frac{a_0}{1 - \sum\limits_{n=1}^{P} a_n\, z^{-n}} \qquad (5)$$

are determined by the filter coefficients $h_s(n)$ of Eq. (2) by methods taken from parametric spectrum analysis, e.g., [5]. The relation between the coefficients $h_s(n)$ and a_n is given by the Yule-Walker equations

$$\begin{bmatrix} \varphi(1) \\ \vdots \\ \varphi(P) \end{bmatrix} = \begin{bmatrix} \varphi(0) & \ldots & \varphi(1{-}P) \\ \vdots & \ddots & \vdots \\ \varphi(P{-}1) & \ldots & \varphi(0) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} \qquad (6)$$

with

$$\varphi(\lambda) = \sum_{n=0}^{L-|\lambda|} h_s(n)\, h_s(n+\lambda) \;\; ; \;\; 0 \leq |\lambda| \leq P \qquad (7)$$

$$a_0 = \sqrt{\varphi(0) - \sum_{n=1}^{P} a_n\, \varphi(-n)}\,. \qquad (8)$$

The used auto-correlation method to calculate $\varphi(n)$ ensures a symmetric Toeplitz structure for the auto-correlation matrix in Eq. (6), which allows to solve the Yule-Walker equations efficiently by means of the Levinson-Durbin recursion, e.g., [5]. The obtained AR filter is always stable and of minimum phase since the auto-correlation matrix is positive-definite.

3. Algorithmic Complexity & Signal Delay

The algorithmic complexity for the two presented LDFs is listed in Table I.

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

Table I: Algorithmic complexity in terms of required average number of real multiplications and real additions per input sample, and number of delay elements (memory) for a MA LDF and AR LDF of filter degree P.

The DFT and GDFT can be calculated efficiently by means of the FFT, cf. [5], [4]. The algorithmic complexity for the calculation of the filter coefficients $h_s(n,k')$ and the actual time-domain filtering is equal for both LDFs.
The variable $\mathcal{M}_{\text{div}}$ refers to 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. These values depend on the numeric procedure and accuracy used to perform these operations. Accordingly, the variables $\mathcal{A}_{\text{div}}$ and $\mathcal{A}_{\text{sqrt}}$ denote the additions needed for a division and square-root operation, respectively. (The complexity for the gain calculation in Fig. 1 is not considered in Table I.)
The AR filter degree can be chosen significantly lower than the MA filter degree for a similar amount of noise reduction, in terms of noise attenuation and speech quality (cf. Sec. 4), such that both approaches require a comparable computational complexity.
The realization of discrete filters can be done by different filter forms, e.g., [5]. The implementation of the MA / AR filter by the transposed direct form II has been found suitable to avoid filter-ringing effects. These effects can occur for time-varying filter coefficients and become audible by disturbing artifacts.
The (algorithmic) signal delay of the filter can be determined by

$$d_0 = \arg\max_{\lambda\in\mathbb{Z}} \{\varphi_{xy}(\lambda)\} \qquad (9)$$

with $\varphi_{xy}(\lambda)$ denoting the cross-correlation sequence between the input sequence $x(k)$ and the output sequence $y(k)$ of the filter[2]. The signal delay of a MA filter with linear phase and degree P amounts to $P/2$ samples. A maximal signal delay of only a few samples has been observed for the AR filter regarded in Sec. 4. The AS FB requires non critically down-sampling $(r < M)$, i.e., an overlap between consecutive signal frames, to avoid aliasing effects. This restriction is not given for the LDF because of the time-domain filtering.

4. Simulation Example

The described MA / AR LDF according to Fig. 1 and spectral weighting by means of a DFT AS FB (realized by the overlap-add method) have been employed for noise reduction. The soft-gain MMSE spectral estimator [6] with a decision-directed SNR estimation [7] based on minimum statistics [8] has been used to calculate the gains $W_i(k')$ at time-intervals of $r = M/2$ samples. White noise with varying amplification has been added to a male speech sequence to achieve different signal-to-noise ratios (SNRs) for the noisy input speech (16 kHz sampling frequency).
The performance of the noise reduction has been measured by the cepstral distance (CD) and the segmental SNR including only frames with speech activity, e.g., [9]. The CD is a frequency-domain measure to account for speech distortions, calculated by the clean speech and processed output speech.

[2] For speech enhancement systems, the more correlated input speech and processed speech can be taken in a simulation.

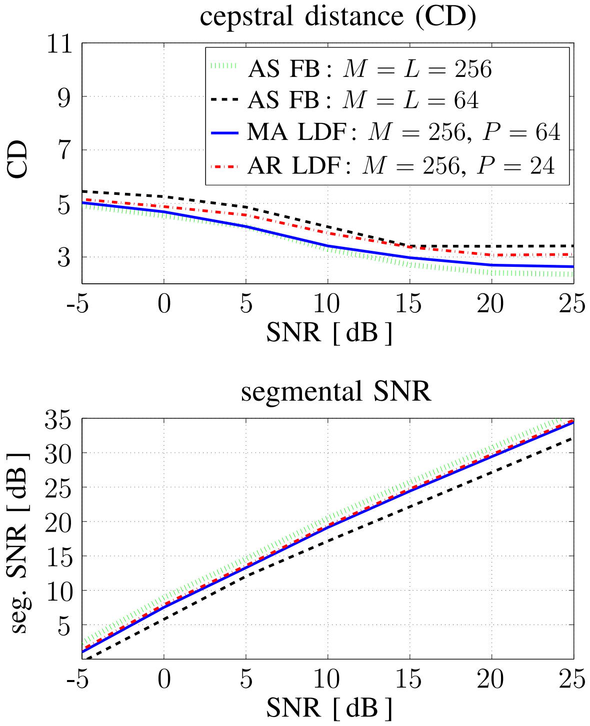

Figure 2: CD and segmental SNR achieved by noise reduction based on the LDF and the AS FB for white noise ($r = M/2$).

The simulation results are shown in Fig. 2. The use of the MA LDF and AR LDF leads to a slightly increased CD compared to the AS FB with $M = 256$ due to the filter approximation, but the LDF achieves a significantly lower signal delay: The AS FB with $M = 256$ causes an (algorithmic) signal delay of 255 samples, whereas the (linear phase) MA filter causes a delay of 32 samples. The AR filter has a varying signal delay of only up to 2 samples according to Eq. (9). Informal listening tests have revealed no significant differences for the subjective speech quality. The LDF achieves a higher segmental SNR than the AS FB with $M = 64$ and signal delay of 63 samples.

A parameter tuning, as proposed in [1], can be made to improve the noise reduction performance for this DFT size. However, this approach might be problematic in practice, if a noise reduction system is composed of different (independent) software modules for gain calculation, noise estimation, filtering etc. In this case, the LDF approach is more suitable to fulfill demanded signal delay constraints.

5. Conclusions

A low delay filter structure for adaptive noise reduction has been proposed. This approach allows to control the trade-off between signal on one hand, and achieved noise attenuation and speech quality on the other hand in a simple and flexible manner since no parameter adjustment for the spectral gain calculation is required. The MA LDF is based on an FIR filter and provides a simple method to decrease the signal delay. The AR LDF is based on a minimum phase IIR filter and can achieve a signal delay of only a few samples. The LDF approach yields a higher amount of noise reduction than an AS FB with the same delay.

References

[1] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral Subtraction Using Reduced Delay Convolution and Adaptive Averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 799–807, Nov. 2001.

[2] P. Vary, "An adaptive filter bank equalizer for speech enhancement," *Signal Processing, Elsevier, Special Issue on Applied Signal and Audio Processing*, 2006.

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

[4] 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, pp. 2105–2108.

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

[6] R. J. McAulay and M. L. Malpass, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 2, pp. 137–145, Apr. 1980.

[7] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[8] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[9] S. Gustafsson, *Enhancement of Audio Signals by Combined Acoustic Echo Cancellation and Noise Reduction*, Ph.D. thesis, IND, RWTH Aachen University, 1999.