Reducing the Complexity or the Delay of Adaptive Subband Filtering

Patrick Hannon[1], *Mohamed Krini*[1], *Gerhard Schmidt*[2], *and Arthur Wolf*[1]

SVOX Deutschland GmbH[1]
Magirus-Deutz-Str. 16, 89077 Ulm, Germany
web: www.svox.com

Christian-Albrechts-Universität zu Kiel[2]
Kaiserstr. 2, 24143 Kiel, Germany
web: www.dss.tf.uni-Kiel.de

Abstract

In this contribution a method to reduce the computational complexity of adaptive subband filtering is presented. Alternatively, the approach can be utilized to keep the complexity as it is in a conventional scheme, but to reduce the delay introduced by non-critical subsampled filterbanks (meaning to shorten the length of the filterbanks' prototype filters). This is achieved by using different subsampling rates for the reference channels on one hand and for the desired and the output channels on the other hand. Experiments and real-time measurements performed with systems for acoustic echo cancellation show either a reduction of complexity of about 30 percent or a delay reduction of about 50 percent (depending on the optimization objective). At the same time the steady-state error (remaining error power due to aliasing after convergence) can be kept at the same level.

1 Introduction

In a variety of applications adaptive subband filtering is utilized – on the one hand due to its low computational complexity (compared to equivalent broadband structures) and on the other hand due to its straightforward algorithmic realizations (e.g. SNR-dependent weighting of the subband signals for noise suppression purposes). Enhancing speech signals that have been disturbed by echoes (multipath propagation from a loudspeaker to a microphone), by stationary background noise as present in automotive environments, and by competing (undesired) speakers is a common application of adaptive subband systems.

Such systems usually consist of an echo cancellation unit that tries to estimate the impulse response of so-called *loudspeaker-enclosure-microphone* (LEM) systems. If the estimation procedure performs well, one can convolve the loudspeaker's signal with the estimated impulse responses to obtain estimated echo signals. These signals are subtracted from microphone signals to *decouple* the electro-acoustic transducers [5]. If more than one microphone is used, the individual sensor signals can be combined (adaptively) in order to suppress undesired recording directions while keeping signals from a predefined direction as undisturbed as possible. In addition, Wiener-type filtering can be applied to reduce stationary background noise as well as remaining echoes. The basic structure of such a system is depicted in Fig. 1. Further details about such systems can be found in [3].

The individual algorithmic components described above have different requirements on the analysis and synthesis filterbanks. Since speech signals are involved, short-term stationarity can be assumed, resulting in frameshifts (temporal distance between two subband sample instants) up to 15 ms on one hand. On the other hand several recommendations or regulations from ETSI, ITU, or VDA limit the tolerable overall system delay to 30 – 50 ms. The complexity of most processing units is inversely proportional to the frameshift. Thus, the larger this value can be chosen, the lower are the demands on the utilized hardware and the cheaper will be the price of resulting products.

For several algorithmic units (e.g. noise and residual echo suppression) it is only necessary that the combined analysis-synthesis system has no or low aliasing components. However, whenever a system identification is involved – as is the case for echo cancellation – the aliasing properties

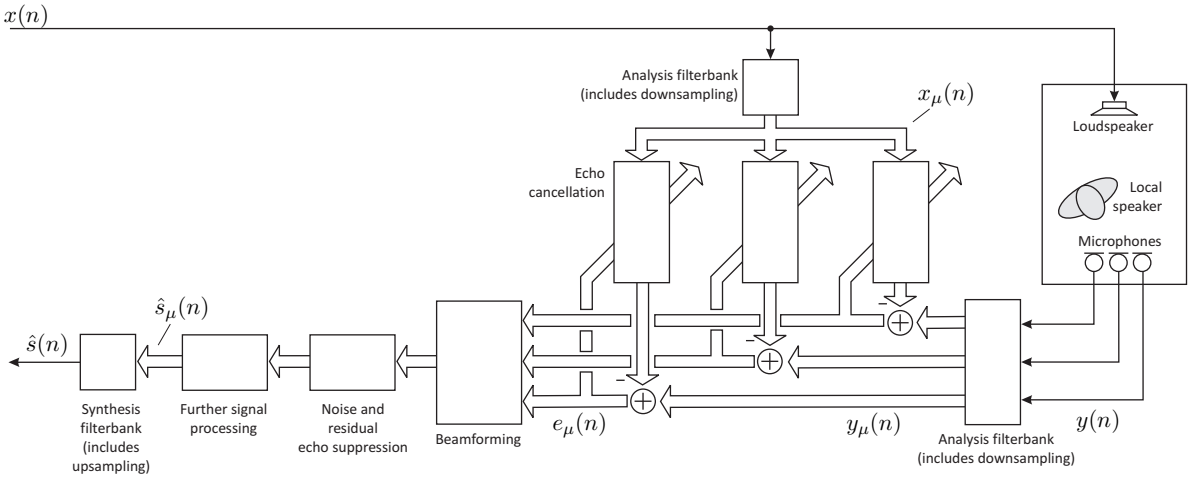

Figure 1 - Basic structure of a speech enhancement system operating in subbands.

of the analysis filterbank alone (without synthesis) are important. In the remainder of this contribution, we will use a hands-free speech enhancement system as an example for showing the limits of current setups and the benefit of a small modification within the analysis filterbank (presented in Sec. 3). However, the modification is not restricted to improve only speech enhancement systems. It can be used in any adaptive subband processing scheme where system identification of a broadband system is a major aim.

2 Subband Schemes

A variety of filterbank types can be used for adaptive subband systems. For systems that focus on speech enhancement often subband decomposition schemes that are based on the short-term Fourier transform are used. We will describe these systems first. Even though the achievable computational complexity reduction with this type of frequency decomposition is very impressive, an extension to polyphase filterbanks allows for even larger complexity reductions. However, this has to be "paid for" with an increased overall system delay. This "dilemma" is described at the end of this section.

2.1 Overlap-add Based Schemes

The process of decomposing a block of input samples (here denoted for the microphone signal $y(n)$) into a block of subsampled subband signals can be denoted as

$$y_\mu(n) = \sum_{m=0}^{N-1} y(nR-m)\, h_m\, e^{-j\frac{2\pi}{N}\mu m}. \tag{1}$$

The index μ denotes here the (frequency) index of the resulting subband signal, N is the DFT order, and R represents the subsampling rate (or equivalently the frameshift in broadband samples). The coefficients h_m denote the window function that is applied. On one hand the outputs of this decomposition can be interpreted as supporting points of the short-term spectrum computed around the broadband time index $nR-N/2$ or on the other hand as bandpass filtered subsampled (time-domain) signals. We will use the latter interpretation in the following.

After several signal processing stages (see, e.g., Fig 1) the enhanced subband signals can be converted back to a broadband signal by computing first an inverse DFT:

$$\hat{s}_{\text{pre}}(n,m) = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{\mu=0}^{N-1} \hat{s}_\mu(n)\, e^{j\frac{2\pi}{N}\mu m}, & \text{if } 0 \le m < N, \\ 0, & \text{else.} \end{cases} \tag{2}$$

The resulting signal blocks are weighted with a window function w_m. After adding overlapping blocks the output signal is obtained by

$$\hat{s}(nR+m) = \sum_{k=-\infty}^{\infty} \hat{s}_{\mathrm{pre}}(n-k, m-kR)\, w_{m-kR}. \tag{3}$$

Here it is assumed that the analysis and synthesis window functions have values different from zero only between 0 and $N-1$. If one restricts the processing in between the analysis and synthesis filterbank to a constant weighting $\hat{s}_\mu(n) = const. \cdot y_\mu(n)$, perfect reconstruction, meaning $\hat{s}(n-N) = y(n)$, can easily be guaranteed. A variety of solutions for pairs of h_m and w_m exist. Given an analysis window function h_m one possible synthesis window function (the one with minimum Euclidean norm) can be computed as

$$\boldsymbol{w} = [\boldsymbol{H}^{\mathrm{T}}\, \boldsymbol{H}]^{-1} \boldsymbol{H}^{\mathrm{T}}\, \boldsymbol{c}\,. \tag{4}$$

The matrix $\boldsymbol{H}$ contains shifted parts of the analysis window function and is defined as

$$\boldsymbol{H} = \begin{bmatrix} h_0 & 0 & \dots & 0 & h_R & 0 & \dots & 0 & \dots \\ 0 & h_1 & \dots & 0 & 0 & h_{R+1} & \dots & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \dots \\ 0 & 0 & \dots & h_{R-1} & 0 & 0 & \dots & h_{2R-1} & \dots \end{bmatrix}. \tag{5}$$

The resulting vector $\boldsymbol{w} = [w_0,\, w_1,\, \dots,\, w_{N-1}]^{\mathrm{T}}$ describes all synthesis filter coefficients between 0 and $N-1$. The vector $\boldsymbol{c}$ consists of R ones: $\boldsymbol{c} = [1,\, 1,\, \dots,\, 1]^{\mathrm{T}}$.

Fig. 2 shows an example for a pair of perfect reconstructing analysis and synthesis filters. For the analysis filter a Hann window according to, e.g., [6] was chosen and the synthesis window was designed according to Eq. (4) for a frameshift $R = 82$. As we will see later on, with this setup a subband system can achieve echo reductions by means of cancellation of about 30 dB.

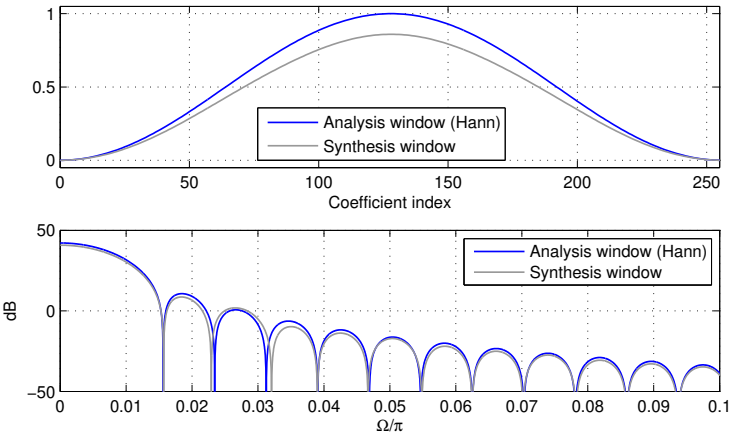

Figure 2 - Analysis and synthesis windows for an FFT size of $N = 256$, based on an Hann window. The synthesis window was designed according to Eq. (4) for a frameshift $R = 82$.

We will focus here only on so-called *overlap-add* based structures. Their counterpart – *overlap-save* based schemes – will not be covered in this contribution since a constraint (realized by additional DFTs, projections and IDFTs) for avoiding circular convolution effects is required. This would lead – for the applications treated here – to a larger computational complexity (compared to overlap-add based structures).

For most types of filterbanks (including overlap-add based structures as well as polyphase filterbanks, which will be discussed in the next section) an alternative processing view is helpful for understanding the aliasing properties [1]. Eq. (1) can be realized very efficiently by multiplying signal segments with a window function h_m and an FFT of windowed signal vector. This computes blockwise all required output signals (also called spectral bins or subbands) in one step. Alternatively, Eq. (1) represents a convolution of the input signal $y(n)$ with a N-tap FIR bandpass filter $h_m\, e^{-j2\pi\mu m/N}$ and a downsampling process (subsampling factor: R) afterwards. The latter type of realization would not be very efficient. However, it allows for straightforward aliasing analysis.

Multiplying an impulse response h_m with a complex modulation term $e^{-j2\pi\mu m/N}$ represents a shift by $\Omega = 2\pi\mu/N$ in the Fourier domain. Subsampling the bandpass filtered signal by a factor R equals summing R shifted and compressed versions of the spectrum of the bandpass filtered signal in the spectral domain [1]:

$$Y_\mu\left(e^{j\Omega}\right) = \frac{1}{R}\sum_{r=0}^{R-1} Y\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r\right)}\right)\, H\left(e^{j\left(\frac{\Omega}{R}+\frac{2\pi}{N}\mu-\frac{2\pi}{R}r\right)}\right). \tag{6}$$

The aim of the anti-aliasing bandpass filter is to suppress most of the shifted spectra ($r \neq r_0(\mu)$, the meaning of $r_0(\mu)$ will be explained a few lines below) as much as possible and to keep the original spectrum for one shift $r = r_0(\mu)$. Which of the shifted spectra is the desired one and which are the undesired ones is determined by the passband of the involved bandpass filter. Assuming h_m is designed to be a lowpass filter, the passband center frequencies of the shifted filter spectra can be determined as $2\pi\mu/N - 2\pi r/R$. Resolving this for the parameter r that achieves a center frequency closest to $\Omega = 0$ leads to

$$r_0(\mu) = \text{round}\left\{\frac{\mu\, R}{N}\right\}. \tag{7}$$

Thus, the spectrum of the subsampled subband signal $y_\mu(n)$ can be divided into a *desired* spectral component

$$Y_{\mu,\text{des}}\left(e^{j\Omega}\right) = \frac{1}{R}\, Y\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r_0(\mu)\right)}\right)\, H\left(e^{j\left(\frac{\Omega}{R}+\frac{2\pi}{N}\mu-\frac{2\pi}{R}r_0(\mu)\right)}\right) \tag{8}$$

and undesired *aliasing* components

$$Y_{\mu,\text{al}}(e^{j\Omega}) = \frac{1}{R}\sum_{r=0,r\neq r_0(\mu)}^{R-1} Y\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r\right)}\right)\, H\left(e^{j\left(\frac{\Omega}{R}+\frac{2\pi}{N}\mu-\frac{2\pi}{R}r\right)}\right). \tag{9}$$

The larger the subsampling rate R is chosen the more aliasing terms are summed in the undesired components. Additionally, the most critical aliasing components (the ones resulting from $r \equiv \text{mod}(r_0(\mu) - 1, R)$ as well as $r \equiv \text{mod}(r_0(\mu) + 1, R)$ start to become larger and larger since parts of the transition or even pass band of the bandpass filter are also included in these components if R is chosen too large. Fig. 3 depicts the aliasing components for three different setups ($R = 64$, $R = 96$, and $R = 128$). Note that the individual diagrams have been compressed such that a transformation of the x-axis into a frequency scale in Hertz would lead to an equal Hertz-per-length value for all diagrams.

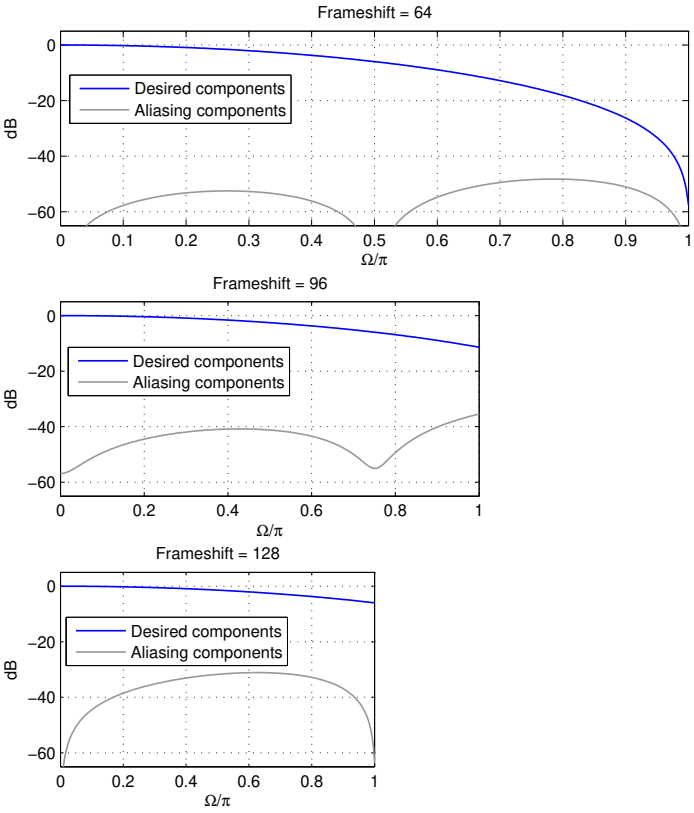

Figure 3 - Aliasing properties (for subband $\mu = 0$) for an FFT size of $N = 256$, a Hann window for the analysis, and three different subsampling rates ($R = 64$, 96, 128).

The decomposition of the reference signal $x(n)$ is achieved in an equivalent manner. If a system identification (echo cancellation) according to Fig. 4 should be performed, the dependence on the aliasing properties within the subband domain becomes visible. While each of the shifted components of the microphone subband signal have been modified individually by the system to be identified, $G(e^{j\Omega})$, only one modification per subband can be performed by means of an adaptive filter $G_\mu(e^{j\Omega})$.

The spectrum of the desired component within the microphone signal (after the subband decomposition) can be denoted as:

$$D_\mu\left(e^{j\Omega}\right) = \frac{1}{R}\sum_{r=0}^{R-1} X\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r\right)}\right)\, G\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r\right)}\right)\, H\left(e^{j\left(\frac{\Omega}{R}+\frac{2\pi}{N}\mu-\frac{2\pi}{R}r\right)}\right). \tag{10}$$

For this signal a (linear) convolution takes place before the subsampling operation. The spectrum of the estimated subband signal can be denoted as

$$\hat{D}_\mu\left(e^{j\Omega}\right) = \hat{G}_\mu\left(e^{j\Omega}\right)\, \frac{1}{R}\sum_{r=0}^{R-1} X\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r\right)}\right)\, H\left(e^{j\left(\frac{\Omega}{R}+\frac{2\pi}{N}\mu-\frac{2\pi}{R}r\right)}\right). \tag{11}$$

Note that now we have the downsampling operation first and the convolution with the estimated subband impulse response afterwards (see Fig. 4). Usually an optimum cancellation, meaning $\hat{D}_\mu\left(e^{j\Omega}\right) = D_\mu\left(e^{j\Omega}\right)$, cannot be achieved. However, keeping the power of the difference of both signals as small as possible can be achieved by optimizing the following three criteria in parallel:

$$\left|H\left(e^{j\Omega}\right)\right| \;\rightarrow\; 0, \text{ for } \frac{\pi}{R} < \Omega < 2\pi - \frac{\pi}{R}, \tag{12}$$

$$\left|H\left(e^{j\Omega}\right)\right| \;\rightarrow\; 1, \text{ for } 0 \le \Omega < \frac{\pi}{N} \;\wedge\; 2\pi - \frac{\pi}{N} < \Omega \le 2\pi, \tag{13}$$

$$\hat{G}_\mu\left(e^{j\Omega}\right) \;\rightarrow\; G\left(e^{j\left(\frac{\Omega}{R}-\frac{2\pi}{R}r_0(\mu)\right)}\right). \tag{14}$$

Requirements (12) and (13) can be achieved by appropriate design of the weighting function h_m. For fulfilling requirement (14) an appropriate adaptation algorithm as well as low aliasing components according to Eq. (9) are necessary. To show the influence of these components a system identification was performed. The impulse response that should be identified was measured in a car at a sample rate of $f_\mathrm{s} = 11025\,\mathrm{Hz}$. In order to be able to neglect all other performance restricting factors no background noise was involved in this simulation, a large delay was introduced to be able to model also non-causal parts of the subband impulse responses, and sufficient coefficients were spent for modeling the entire tail of the impulse response.

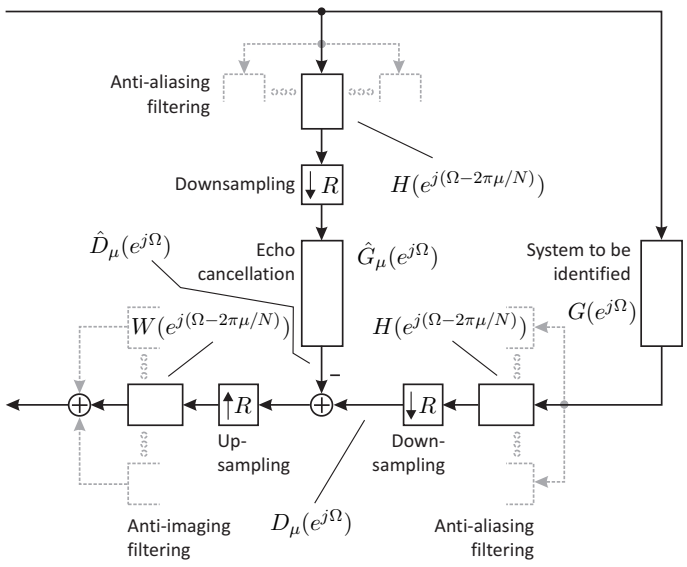

Figure 4 - Basic setup of a system identification performed using an adaptive subband structure. Only those components necessary for understanding the system identification problem caused by aliasing are depicted (no noise components, etc.).

These demands lead to the following setup: FFT size $N = 256$, Number of coefficients for the subband adaptive filters $N_\mathrm{ec} = 16$, a Hann window for the analysis weighting function and a synthesis window function according to Eq. (4).

In Fig. 5 nine adaptation runs performed with white noise as excitation signal are depicted. While adaptation with a frameshift of $R = 64$ leads to a good steady state error reduction (about 40 dB) the performance degrades more and more with increasing frameshift. However, the computational complexity decreases by about 50 percent when changing the frameshift from 64 to 128. For that reason it is desired to optimize the convergence performance for larger frameshifts.

2.2 Polyphase Filterbanks

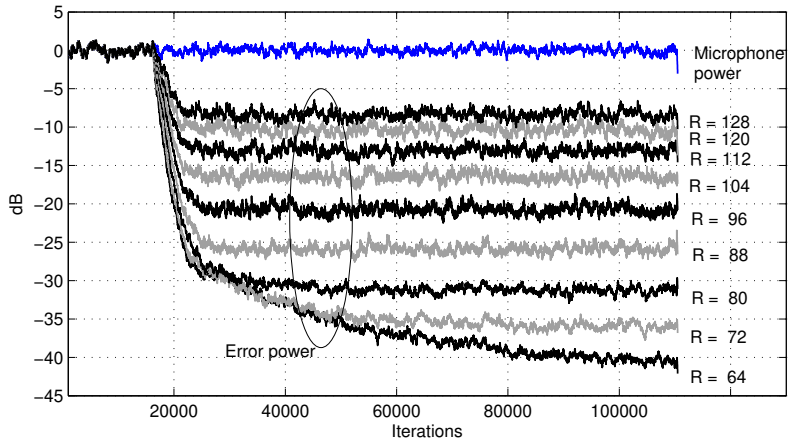

Figure 5 - Several convergence runs for frameshifts ranging from 64 to 128. All synthesis filters were designed such that perfect reconstruction is achieved. All remaining conditions were equal for each simulation run.

Overlap-add structures can be extended to so-called *non-critically subsampled polyphase filterbanks* [1]. This means that the length N_{lp} of the window functions h_m and w_m (also called prototype lowpass filters) is allowed to be larger than the amount of subbands (determined by the DFT size N). This leads to much better (lower) aliasing components without increasing the computational complexity significantly. Depending on the length N_{lp} of the filters h_m and w_m, frameshifts of size R close to N can be selected. Typical design procedures (e.g. [7]) can achieve a frameshift $R \approx 3/4\,N$. However, this comes with filter orders of about $N_{\mathrm{lp}} = 6 \ldots 8\,N$.

Nearly all considerations made until now are still valid if polyphase filterbanks instead of overlap-save based structures are utilized except Eqs. (1) and (3). The necessary polyphase extensions for the analysis part lead to [2]

$$y_\mu(n) = \sum_{m=0}^{N-1} e^{-j\frac{2\pi}{N}\mu m} \sum_{p=0}^{N_{\mathrm{lp}}/N-1} y(nR - pN - m)\, h_{m+pN} \tag{15}$$

and for the synthesis part to

$$\hat{s}(nR+m) = \sum_{k=-\infty}^{\infty} \hat{s}_{\mathrm{pre}}(n-k, m-kR) \sum_{p=0}^{N_{\mathrm{lp}}/N-1} w_{m-kR-pN}. \tag{16}$$

2.3 The Complexity-Delay Dilemma

Extending an overlap-add structure to a polyphase one allows for selecting a much larger frameshift, which comes with a decrease in computational complexity. Extending for example the setup depicted in Fig. 2 ($N = 256$, $R = 82$) can lead to a system with nearly the same convergence properties but allowing for frameshifts up to $R = 160$ (the computational complexity is nearly halved). In this case the lowpass filter h_m needs a length of about $N_{\mathrm{lp}} \approx 5N$, meaning that also the delay of the polyphase structure is 5 times larger than that of the overlap-add structure. In several applications, including hands-free telephone, such a large delay is not tolerable any more (e.g. because of ITU or ETSI recommendations).

3 Complexity or Delay Reduction

To overcome the complexity-delay dilemma without reducing quality in terms of speed of convergence or steady-state performance too much, a simple modification can be applied. The aliasing terms [Eq. (9)] that cause the major part of the convergence problem are generated by the subsampling unit within the reference path. Thus, one can split the downsampling unit into two stages. First, a reduction by a factor $R_1 < R$ is performed, then the convolution is performed, and afterwards a second downsampling stage (reduction factor $R_2 = R/R_1$) is applied as depicted in Fig. 6.

This modification widens the transition range for the analysis filter h_m and allows for better approximation of requirements (12) and (13). Even if the convolution seems to be performed at a higher rate now, the result of this processing stage is really required only every R_2 samples. Thus, the convolution operates in practice at the same rate as all other components. The amount of samples and filter coefficients involved in the convolution and the adaptation, however, is now larger than in the basic scheme: R_2 as many as before. Also the decomposition of the reference signal has to be computed R_2 times as often as in the original setup. However, all other components (adaptation control, beamforming, noise suppression, residual echo suppression, etc.) are still operating at the main subsampled rate, which can now be chosen much lower than before.

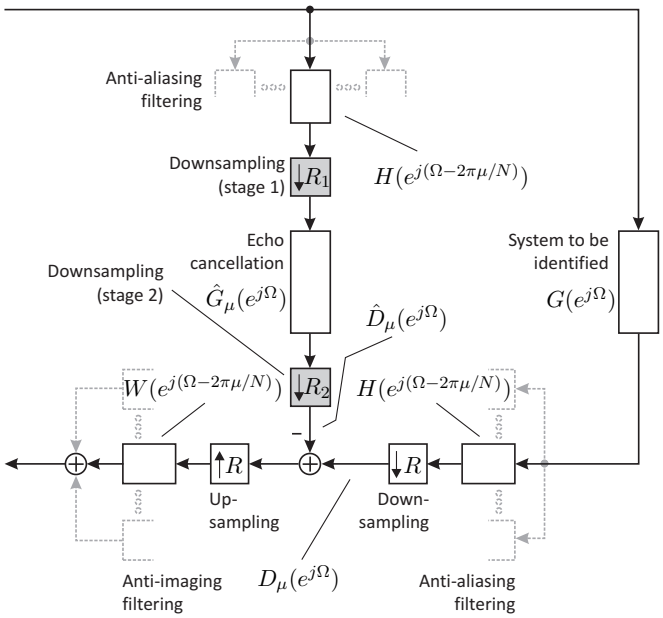

Figure 6 - Modified setup.

3.1 Results

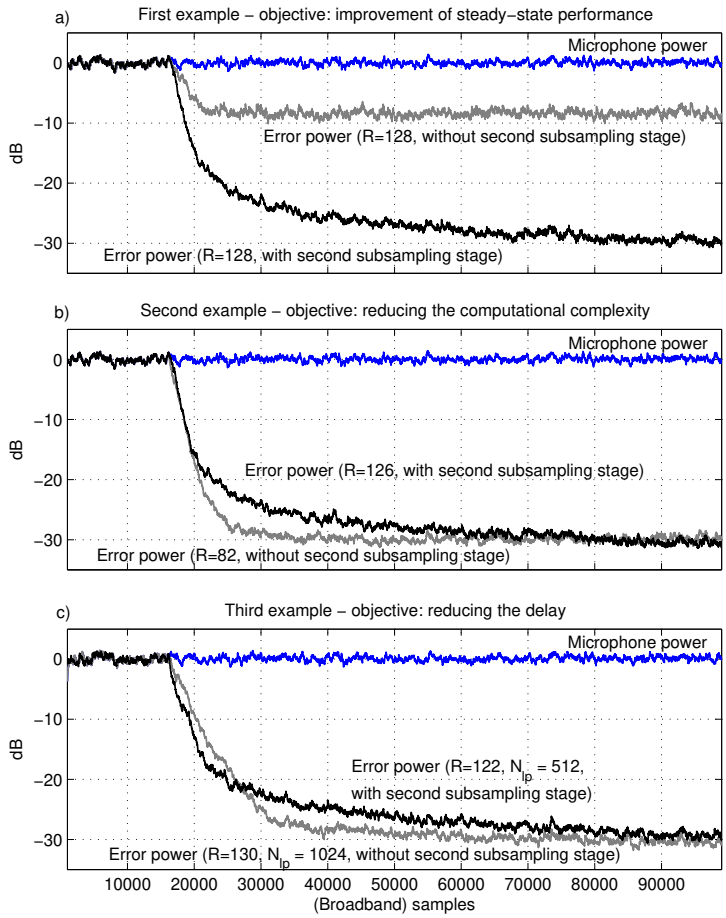

Figure 7 - Three examples that should show the benefits and potentials of the new scheme.

In the following we will show three examples that should demonstrate the potential and the benefits of the two-stage downsampling unit in the reference channel. The second downsampling factor will always be $R_2 = 2$. For all simulation the NLMS algorithm has been used. We will start with a basic system using an FFT size $N = 256$ and a Hann window for the analysis decomposition. Speech enhancement components such as noise or residual echo suppression would perform well with a frameshift of $R = 128$. The performance of the echo cancellation in terms of maximum convergence, however, would be very low as we can see in Fig. 7a) (as well as in Fig. 5). Applying the new scheme leads to a significant improvement – also depicted in Fig. 7a). This improvement of about 20 dB more echo reduction comes with the cost of one additional FFT as well as increased filter orders for the echo cancellation filters. However, for an entire hands-free speech enhancement system this is about a 2 to 5 % increase of complexity.

The second example compares the echo performance of two setups – see Fig. 7b). First a conventional scheme with a frameshift of $R = 82$. Increasing the frameshift to $R = 126$ without any modification would cause severe problems for the echo reduction performance. Applying the new method leads to a reduction of about 30 % in terms of overall system complexity while inserting only small performance degradations in terms of speed of convergence and steady-state performance.

Finally, we compare two polyphase filterbank structures designed according to [7] – see Fig. 7c). If a frameshift of $R = 130$ should be achieved, prototype lowpass filters of length $N_{\mathrm{lp}} = 4N = 1024$ are required. At a sampling rate of $f_{\mathrm{s}} = 11025\,\mathrm{Hz}$, this would lead to a delay of about 93 ms. With the new scheme prototype lowpass filters of length $N_{\mathrm{lp}} = 2N = 512$ and a frameshift of $R = 122$ achieve nearly the same performance. Again the overall complexity is increased by about 10 to 15 % – the delay, however, can be reduced by 50 % and is now only 46 ms.

4 Conclusions and Outlook

In this contribution a small modification by means of performing the downsampling in two stages was presented for the subband decomposition of the reference signal in adaptive subband systems. With this add-on it is possible to increase the main downsampling factor of such systems, leading to a reduction in computational complexity of all signal processing components except those for echo cancellation. Even if the complexity of echo cancellation units can not be reduced that clearly, the filters can operate now at subsample rates that would not lead to sufficient convergence without the modification. Alternatively, the two-stage reference downsampling procedure can be utilized to keep the complexity as it is in a polyphase filterbank based scheme, but to reduce the delay introduced by non-critical subsampled filterbanks.

The method has been implemented in a variety of commercial products for hands-free speech enhancement systems [4]. It has helped to save quite a bit of computational load while keeping the quality degradation that comes along with it to a degree not noticeable by listeners. Further improvements might also be achieved if the window function, respectively the prototype lowpass filter, of the synthesis filterbank is designed such that it helps to suppress those components (in terms of frequency regions) that cause the largest remaining distortions.

References

[1] R. E. Crochiere, L. R. Rabiner, *Multirate Digital Signal Processing*, Upper Saddle River, NJ, USA: Printice-Hall, 1983.

[2] H. G. Göckler, A. Groth, *Multiratensysteme*, Wilburgstetten, Germany: Schlembach, 2004 (in German).

[3] E. Hänsler, G. Schmidt, *Acoustic Echo and Noise Control*, Hoboken, NJ, USA: Wiley, 2004.

[4] Harman/Becker Automotive Systems, "Low Complexity Echo Compensation", EPO patent application, EP000001936939A1, 2006.

[5] W. Kellermann, *Analysis and Design of Multirate Systems for Cancellation of Acoustical Echoes*, Proc. ICASSP, pp. 2570-2573, 1988.

[6] H. L. Van Trees, *Optimum Array Processing*, New York, NY, USA: Wiley, 2002.

[7] G. Wackersreuther, "On the Design of Filters for Ideal QMF and Polyphase Filter Banks", *AEÜ*, vol. 39, no. 2, pp. 123 - 13, 1985.