METHOD FOR TEMPORAL INTERPOLATION OF SHORT-TERM SPECTRA AND ITS APPLICATION TO ADAPTIVE SYSTEM IDENTIFICATION

Mohamed Krini

Nuance Communications Aachen GmbH
Acoustic Speech Enhancement Research
Söflinger Str. 100, Ulm, Germany
mohamed.krini@nuance.com

Gerhard Schmidt

Christian-Albrechts-Universität zu Kiel
Digital Signal Processing and System Theory
Kaiserstr. 2, Kiel, Germany
gus@tf.uni-kiel.de

ABSTRACT

In this paper an extension of analysis filterbanks utilized in adaptive system identification schemes is presented. The extension operates as a postprocessing stage following conventional polyphase filterbanks or short-term Fourier transforms. The idea of the new method is to exploit the redundancy of succeeding short-term spectra for computing interpolated temporal supporting points. For its efficient implementation some approximations are performed – it can be shown that the postprocessing stage can be easily realized based on the weighted sum of subband signals. The new method allows for significant increase of the frameshift (subsampling rate), leading to a reduction of the computational complexity while keeping the convergence speed and the steady-state performance. Alternatively, the frameshift can be kept unchanged – in this case an improved steady-state convergence can be achieved. Real-time measurements performed with systems for acoustic echo cancellation have shown that significant improvements in terms of echo reduction can be achieved while increasing the amount of required memory only marginally.

Index Terms— Adaptive filters, echo cancellation, system identification, filterbanks, aliasing

1. INTRODUCTION

In a variety of applications such as hands-free telephony or speech dialogue systems, the desired speech signal is often corrupted by echoes (multipath propagation from a loudspeaker to a microphone) and by background noise. For reducing the undesired signal components, speech enhancement algorithm schemes like echo cancellation and noise reduction are applied. In the majority of cases these algorithms are realized in the subband or frequency domain in order to reduce the computational complexity.

An echo cancellation unit within a speech processing system tries to estimate the impulse response of a loudspeaker-enclosure-microphone (LEM) system. For estimating the echo components in the subband domain, the microphone as well as the reference signal are first segmented into overlapping blocks of appropriate size (segments are often overlapped by $50 - 75$ %) and a window function is applied. Afterwards, each block is transformed into the short-term frequency (subband) domain. The resulting reference subband signals are convolved with an unknown LEM subband impulse response to estimate the subband echo signals [5]. These estimated signals are subtracted from the microphone subband signal to determine the error signals for the filter update. In addition, Wiener-type filtering can be utilized to reduce remaining echo components (if the estimated LEM impulse response has estimation errors) as well as stationary background noise. Finally, the enhanced subband signals are converted back into the time domain using an inverse DFT and appropriate windowing. After adding overlapping blocks the broadband signal is determined. Further details about subband signal processing can be found, e.g., in [2].

For adaptive system identification in the subband domain – as it is used for echo cancellation – the analysis filterbanks are required to generate only a very low amount of aliasing. However, when increasing the subsampling rate (frameshift), on the one hand side the computational complexity can be reduced but on the other hand aliasing components within subband signals are increased, leading to low echo reduction. For practical purposes a compromise between performance and computational complexity has to be found. For larger subsampling rates the required amount of echo attenuation (about 30 dB [4]) can not be achieved anymore. Therefore, a new method is proposed in the following which is able to enhance the convergence behavior for larger subsampling rates. The contribution is organized as follows: First, a brief overview about conventional filterbanks applied for echo cancellation is given. Afterwards, the new method of time-frequency interpolation is derived as well as its efficient realization is explained. The section after that details the application of the proposed method for echo cancellation schemes. The paper concludes with some simulation results and a summary.

2. TYPES OF FILTERBANKS APPLIED FOR ECHO CANCELLATION

To reduce aliasing effects or to enhance the frequency selectivity of DFT-modulated filterbanks a so called polyphase filterbank can be utilized [8]. With a polyphase filterbank structure a subsampling rate close to the DFT order can be chosen, depending on the used length of the prototype low-pass filter. That kind of filterbank scheme is able to reduce the computational complexity due to its large subsampling rates, however, it also increases the delay significantly. Unfortunately, such a high delay is not suitable for applications such as hands-free telephony. In [1], adaptive subband filtering with critical subsampling has been investigated. In order to explicitly cancel the aliasing components the application of adaptive cross filters between the subbands are proposed. Experiments for echo cancellation have shown that using cross filters is leading to a slightly reduced convergence speed as well as to an increased computational load. The application of filterbanks based on allpass filters as an alternative to FIR-based filterbanks results in high sidelobe attenuation [7]. It has been found that these filterbanks are computational efficient. The use of polyphase IIR filterbanks has the disadvantage of non-linear phase distortion and appearance of narrowband high energy aliasing terms

at the filter boundaries. In [6], a delayless adaptive filtering for echo cancellation has been proposed which is also able to reduce aliasing distortions compared to conventional subband adaptive filters. In this case the filter weights are adjusted in the subband domain and transformed afterwards to an equivalent time-domain filter. Since the filtering operation is performed in the time domain, the computational complexity is remarkably increased. Recently, a method to improve the steady-state convergence has been reported in [3] where the FFT of the reference signal is more often computed compared to all other FFTs/IFFTs used within a hands-free system.

In contrast to the state-of-the-art schemes that mainly try to design or optimize appropriate low-pass prototype filters for filterbanks, we propose in this contribution to apply a time-frequency interpolation method as a postprocessing stage after a conventional frequency analysis. As it can be observed later on, a significant reduction of the aliasing terms can be achieved without inserting any additional delay in the signal path and with only few operations by means of multiplications and additions.

3. TEMPORAL INTERPOLATION OF SPECTRA

For the derivation of the new temporal-frequency interpolation method, first the input signal is segmented into overlapping blocks of length N according to:

$$\boldsymbol{y}(nR) \quad = \quad \big[y(nR),\, ...,\, y(nR-N+1)\big]^{\mathrm{T}}, \tag{1}$$

where the parameter R corresponds to the used subsampling rate (frameshift) and the element n denotes the frame index. The subsampled input vector is windowed with a window function (e.g. Hann window), $h_k \in [0, N-1]$, and transformed into the frequency or subband domain using a DFT or filterbank:

$$Y\left(e^{j\Omega_\mu}, n\right) \quad = \quad \sum_{k=0}^{N-1} y(nR-k)\, h_k\, e^{-j\Omega_\mu k}\,. \tag{2}$$

Note that Eq. (2) can also be interpreted as subsampled output signals of an analysis filterbank. The used frequency supporting points Ω_μ are equidistantly distributed over the normalized frequency range: $\Omega_\mu = 2\pi\mu/N$, $\mu \in \{0, \ldots, N-1\}$. For the sake of simplicity, the vector containing all subband signals can be rewritten in a matrix-vector notation as:

$$\boldsymbol{Y}\big(e^{j\Omega}, n\big) \quad = \quad \boldsymbol{D}\,\boldsymbol{H}\,\boldsymbol{y}(nR)\,, \tag{3}$$

where the quantity $\boldsymbol{D}$ specifies a DFT matrix of order N and $\boldsymbol{H}$ characterizes a diagonal matrix consisting of the window function coefficients:

$$\boldsymbol{H} \;=\; \mathrm{diag}\big\{\boldsymbol{h}\big\} \;=\; \begin{bmatrix} h_0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & h_{N-1} \end{bmatrix}. \tag{4}$$

The principle idea of the interpolation method is to exploit the correlation of successive input signal blocks for interpolating an additional signal frame in between of the originally generated frames. The proposed interpolation is performed in the frequency or subband domain. Here the interpolated subband signals are computed by weighted addition of the current and a number of previous input short-term spectra according to:

$$\boldsymbol{Y}'\big(e^{j\Omega}, n\big) \quad = \quad \boldsymbol{S} \begin{bmatrix} \boldsymbol{Y}\big(e^{j\Omega}, n\big) \\ \vdots \\ \boldsymbol{Y}\big(e^{j\Omega}, n-M+1\big) \end{bmatrix}, \tag{5}$$

with $\boldsymbol{S}$ describing the interpolation matrix and M being the amount of used input spectra. Thus, the interpolated subband signals $\boldsymbol{Y}'(e^{j\Omega}, n)$ corresponds exactly to that signal block which would be computed with an analysis filterbank at a reduced rate.

3.1. Computation of the Interpolation Matrix

For computing the interpolation matrix $\boldsymbol{S}$, we first define an interpolated short-term spectrum by:

$$\boldsymbol{Y}'\big(e^{j\Omega}, n\big) \quad = \quad \boldsymbol{D}\,\boldsymbol{H}'\,\boldsymbol{y}'(nR)\,, \tag{6}$$

where $\boldsymbol{y}'(nR)$ characterizes an extended input signal frame containing the last $N+M'$ samples of the input signal: $\boldsymbol{y}'(nR) = [y(nR),\, ...,\, y(nR-N-M'+1)]^{\mathrm{T}}$, with $M' = (M-1)\,R$. Furthermore, an extended matrix of the filter coefficients is specified:

$$\boldsymbol{H}' \quad = \quad \Big[\boldsymbol{0}^{(N\times R/2)} \quad \boldsymbol{H} \quad \boldsymbol{0}^{(N\times M'-R/2)}\Big]. \tag{7}$$

The aim of $\boldsymbol{H}'$ is to add $N \times R/2$ zeros before the original diagonal window matrix and $N \times (M' - R/2)$ after. For the derivation it is assumed that the chosen subsampling R is even-valued. Reformulating Eq. (6) by using the interpolation matrix $\boldsymbol{S}$, the following expression is obtained:

$$\boldsymbol{Y}'\big(e^{j\Omega}, n\big) \quad = \quad \boldsymbol{S}\,\widetilde{\boldsymbol{D}}\,\widetilde{\boldsymbol{H}}\,\boldsymbol{y}'(nR)\,, \tag{8}$$

where $\widetilde{\boldsymbol{D}}$ describes an extended transformation matrix (block-diagonal DFT matrix of size $MN \times MN$) defined by

$$\widetilde{\boldsymbol{D}} \quad = \quad \begin{bmatrix} \boldsymbol{D} & \cdots & \boldsymbol{0}^{(N\times N)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}^{(N\times N)} & \cdots & \boldsymbol{D} \end{bmatrix}. \tag{9}$$

A second extended window matrix $\widetilde{\boldsymbol{H}}$ with a dimension of $MN \times (N+M')$ is computed according to:

$$\widetilde{\boldsymbol{H}} \quad = \quad \Big[\boldsymbol{H}_0^{\mathrm{T}},\, \boldsymbol{H}_1^{\mathrm{T}},\, ...,\, \boldsymbol{H}_{M-1}^{\mathrm{T}}\Big]^{\mathrm{T}}. \tag{10}$$

The first element matrix $\boldsymbol{H}_0$ adds $N \times M'$ zero values after the diagonal window matrix, whereas the remaining matrices $\boldsymbol{H}_1, \boldsymbol{H}_2$, etc., represent cyclic shifts of $\boldsymbol{H}_0$. This means that equal row indices of adjacent submatrices are rotated by R elements. Thus, the first and the last element matrices are defined according to:

$$\boldsymbol{H}_0 \;=\; \Big[\boldsymbol{H} \;\; \boldsymbol{0}^{(N\times M')}\Big] \;\; \text{and} \;\; \boldsymbol{H}_{M-1} \;=\; \Big[\boldsymbol{0}^{(N\times M')} \;\; \boldsymbol{H}\Big]. \tag{11}$$

Using the definitions of Eqs. (6) and (8) result in several solutions for the interpolation matrix $\boldsymbol{S}$ that depend in general on the input signal vector. A solution that is independent of the input signal can be obtained as:

$$\boldsymbol{S} \quad = \quad \boldsymbol{D}\,\boldsymbol{H}'\,\widetilde{\boldsymbol{H}}^{\dagger}\,\widetilde{\boldsymbol{D}}^{\dagger},$$

where $\widetilde{\boldsymbol{H}}^{\dagger}$ and $\widetilde{\boldsymbol{D}}^{\dagger}$ characterize the Moore-Penrose inverse which is defined as

$$\boldsymbol{A}^{\dagger} \quad = \quad \Big[\mathrm{adj}\big\{\boldsymbol{A}\big\}\,\boldsymbol{A}\Big]^{-1}\,\mathrm{adj}\big\{\boldsymbol{A}\big\}\,.$$

The abbreviation adj{...} is denoting the adjoint of the matrix.

3.2. Approximated Interpolation

Once the general solution for the interpolation matrix $\boldsymbol{S}$ is formulated, we can try to simplify and approximate the matrix. In Fig. 1 the log-magnitudes of the elements of the interpolation matrix are shown for $M = 2$ and $N = 256$. From this result, one can observe that the matrix $\boldsymbol{S}$ contains only few coefficients being significantly different from zero. This results from the diagonal structure of the matrix $\boldsymbol{H}$, the sparseness of the extended window matrix $\widetilde{\boldsymbol{H}}$ and the

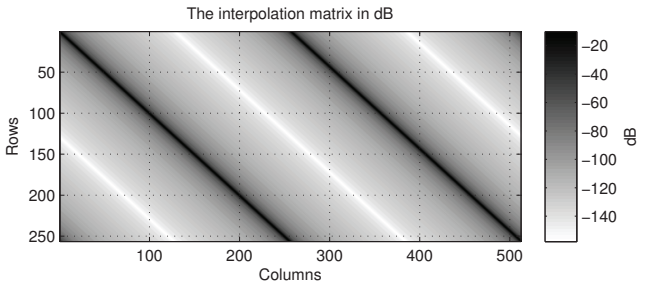

Fig. 1. Magnitude of the elements of the interpolation matrix $\boldsymbol{S}$ in dB with $N = 256$ and $M = 2$.

orthogonal eigenfunctions included in the transformation matrices. Thus, the computation of the temporally interpolated spectrum can be approximated very efficiently as described in the following. Since $\boldsymbol{S}$ is a sparse matrix, the interpolation can be realized as a postprocessing stage after an analysis filterbank using a weighted sum of subband signals. The weighting coefficients for the i-th subband can be easily extracted from the interpolation matrix according to:

$$g_p^{(i,m)} = S_{i,L_i+mN+p}\,. \tag{12}$$

The parameter i in Eq. (12) specifies the row and the quantity $L_i + mN + p$ the column of the interpolation matrix, with $m \in [0,\, M-1]$ and $p \in \big[0,\, K_i - L_i\big]$. The interpolated spectrum for the i-th subband is then determined by:

$$Y'\big(e^{j\Omega_i}, n\big) \;=\; \sum_{m=0}^{M-1} \sum_{k=L_i}^{K_i} g_{k-L_i}^{(i,m)}\, Y\big(e^{j\Omega_k}, n-m\big)\,. \tag{13}$$

Experiments have shown that it is sufficient to use only 5 to 10 complex multiplications and additions for computing one interpolated subband signal. The filter order of $\boldsymbol{g}^{(i,m)}$ for the i-th subband is defined by the difference $K_i - L_i$ with

$$L_i \;=\; \max\left\{0, i - \left\lfloor \frac{P}{2} \right\rfloor\right\} \text{ and} \tag{14}$$

$$K_i \;=\; \min\left\{i + \left\lceil \frac{P}{2} \right\rceil - 1, N-1\right\}\,, \tag{15}$$

with P being the maximal filter order used for the interpolation. Fig. 2 shows the principle realization of an analysis filterbank with time-frequency interpolation as a postprocessor by means of weighted sum of subband signals for $M = 2$.

4. APPLICATION TO ECHO CANCELLATION AND EXPERIMENTAL RESULTS

The subsampling unit within the reference path produces the major part of the convergence problem for echo cancellation. To overcome

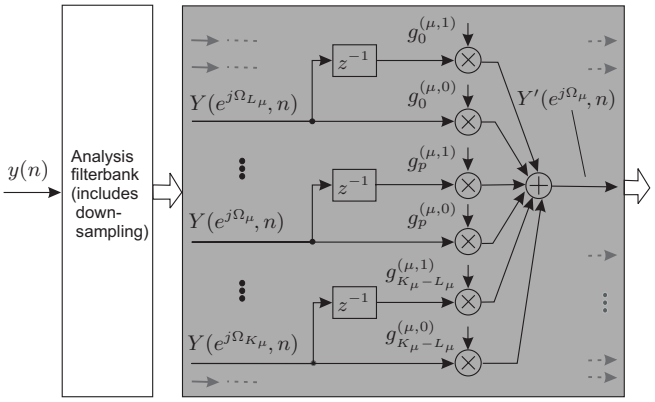

Fig. 2. Analysis filterbank with time-frequency interpolation as a postprocessor by means of weighted sum of subband signals.

a low steady-state convergence at large subsampling rates, the new interpolation method can be implemented. Fig. 3 depicts the proposed structure for subband echo cancellation with additional time-frequency interpolation applied only in the reference channel. First we suggest to apply the same subsampling rate R for the reference and the microphone path. The resulting reference subband signals $Y(e^{j\Omega_\mu}, n)$ – after decomposition of $y(n)$ using an analysis filterbank (anti-aliasing filtering and downsampling) – are subsequently fed to a time-frequency interpolation unit that includes temporally interpolating the time series of the short-term spectra. The original reference subband signals as well as the output of the time-frequency

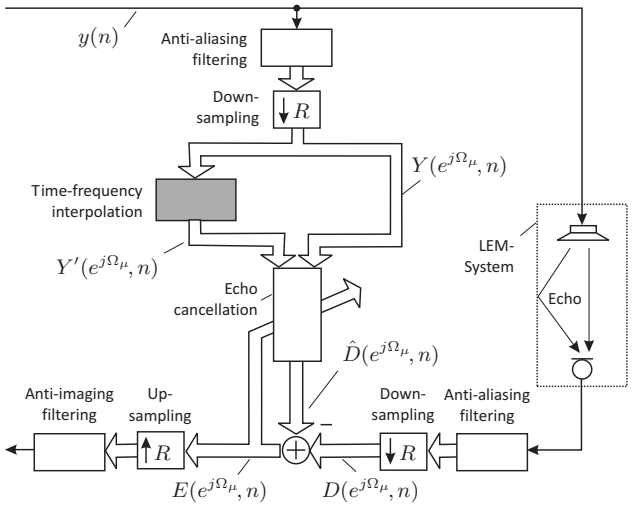

Fig. 3. Proposed system for subband echo cancellation with additional time-frequency interpolation in the reference path.

interpolation $Y'(e^{j\Omega_\mu}, n)$ are fed to the echo cancellation for estimating the subband echo signals. The use of both the reference subband signals as well as its interpolated version reduces the unwanted effects of aliasing. The subband echo signals are estimated by a convolution of the input subband signals with the estimated LEM subband impulse response according to:

$$\begin{aligned}\hat{D}(e^{j\Omega_\mu}, n) \;=\; &\sum_{i=0}^{V-1} W_{i,\mu}(n)\, Y(e^{j\Omega_\mu}, n-i)\;... \\ &+ \sum_{i=0}^{V-1} W'_{i,\mu}(n)\, Y'(e^{j\Omega_\mu}, n-i)\,.\end{aligned} \tag{16}$$

$W'_{i,\mu}(n)$ and $W_{i,\mu}(n)$, $i \in [0, V-1]$, are the subband filter coefficients for the interpolated and non-interpolated part, respectively. V denotes the number of filter coefficients. It has to be noticed that the convolution is still only operating at the original subsampling rate R. The estimated echoes are subtracted from the microphone subband signals $D(e^{j\Omega_\mu}, n)$ to determine the error $E(e^{j\Omega_\mu}, n)$ for the filter update. For adaptation of the filter coefficients a typically gradient-based optimization procedure (e.g. the NLMS algorithm) can be utilized. The amount of samples and filter coefficients involved in the convolution and the adaptation, however, is now larger than in a basic scheme: twice as many as before. All other components used within a complete handsfree-system like the adaptation control of the echo canceller, residual echo suppression, and beamforming are still operating at the original subsampled rate R. Real-time experiments have shown that with the new interpolation method a much higher subsampling rate can be used and thus a significant reduction of the computational complexity can be achieved (even if twice as many filter coefficients are required for echo cancellation).

To show the performance and the accurateness of the proposed method a simulation example in terms of steady-state convergence is shown in Fig. 4. For this measurement, white noise as reference excitation has been used with a Hann-windowed 256-FFT – local speech and noise are not considered in this simulation. First graph from the top shows the normalized power of the microphone, whereas the second, third, and fourth curves depict the power of the error signal (steady-state performance) for different subsampling rates. Using a subsampling rate of $R = 64$, which corresponds to a blockoverlap of 75 %, a high echo attenuation can be measured. However, for many applications even higher complexity reductions are necessary. Increasing the subsampling rate from 64 to 128 (50 %-blockoverlap) the computational complexity can be reduced by half while the performance is strongly degraded due to increased aliasing terms – only about 9 dB echo attenuation can be obtained (see second graph). Third signal shows the performance of the proposed method with $M = 2$, $P = 10$ and the same higher subsampling rate of $R = 128$. A significant improvement of about 22 dB in terms of echo reduction can be achieved. The performance of a 75 % overlap can not be achieved but it has to be noticed that in a real application the performance is in the majority of cases limited to about 30 dB (e.g. due to the background noise). The performance of the echo cancellation in terms of measured maximum convergence (after the filter coefficients are converged) for different blockoverlap values without and with the new time-frequency interpolation method is shown in Fig. 5. For this experiment the same test setup as before has been utilized.

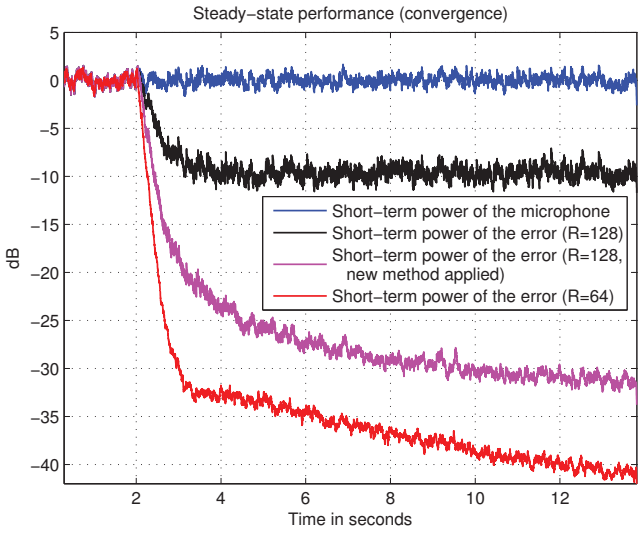

Fig. 4. Performance of echo cancellation without and with additional interpolation ($M = 2$, $N = 256$, $P = 10$).

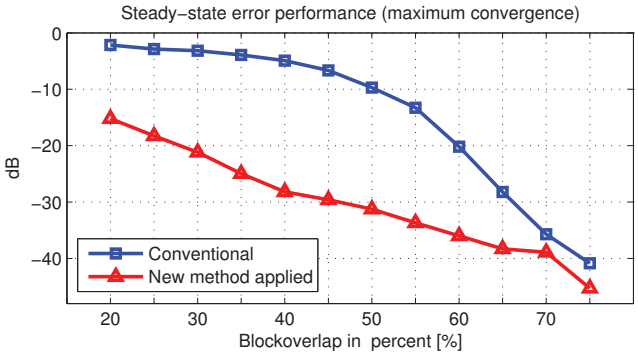

Fig. 5. Performance of the maximal steady-state error without and with time-frequency interpolation ($M = 2$, $N = 256$, $P = 10$).

To model the entire tail of the LEM impulse response, sufficient coefficients have been used. Applying the proposed scheme leads to a significant improvement for all considered blockoverlap values.

5. CONCLUSIONS

In this contribution a postprocessing scheme of analysis filterbanks applied in the reference path of adaptive system identification schemes has been presented. The proposed method is able to significantly reduce aliasing terms caused by a subsampling unit within an analysis filterbank. It has been shown that the postprocessing scheme can be realized in an efficient way based on a weighted sum of subband signals and without inserting any additional delay in the signal path. The new time-frequency interpolation method has been applied for subband echo cancellation. Experimental results have shown that the postprocessing stage allows for a significant increase of the frameshift (or subsampling rate), leading to a reduction of the computational complexity while keeping the convergence speed and the steady-state performance. In addition, an improved steady-state convergence can be achieved, if the subsampling is kept unchanged.

6. REFERENCES

[1] A. Gilloire, M. Vetterli, "Adaptive filtering in subbands with critical sampling," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 40, no. 8, pp. 1862–1875, 1992.

[2] E. Hänsler, G. Schmidt, "Acoustic echo and noise control – a practical approach," *John Wiley & Sons*, Hoboken, NJ, USA, 2004.

[3] Harman/Becker Automotive Systems, "Low complexity echo compensation," *EPO patent application*, EP 1936939 A1, 2006.

[4] ITU-T Recommendation P.340, "Transmission characteristics and speech quality parameters of hands-free terminals," 2000.

[5] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. ICASSP*, pp. 2570-2573, 1988.

[6] D. R. Morgan, J. C. Thi, "A delayless subband adaptive filter architecture," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 43, no. 8, pp. 1819–1830, 1995.

[7] P. A. Naylor, O. Tanrikulu, A. G. Constantinides, "Subband adaptive filtering for acoustic echo control using allpass polyphase IIR filterbanks," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 6, no. 2, pp. 143–155, 1998.

[8] G. Wackersreuther, "On the design of filters for ideal QMF and polyphase filter banks," *AEÜ*, vol. 39, no. 2, pp. 123–13, 1985.