THE TIGHT RELATION BETWEEN ACOUSTIC ECHO CANCELLATION AND RESIDUAL ECHO SUPPRESSION BY POSTFILTERING*

Gerald Enzner[1], *Rainer Martin*[2], *Peter Vary*[1]

[1]*Institute of Communication Systems and Data Processing*
ind, *Aachen University of Technology, D–52056 Aachen*
Phone: +49-241-80-26960, *E-mail:* {enzner,vary}@ind.rwth-aachen.de

[2]*Institute of Communications Technology*
Technical University of Braunschweig, Schleinitzstrasse 22, D–38106 Braunschweig
Phone: +49-531-391-2485, *E-mail:* r.martin@tu-bs.de

Abstract: In the acoustic environment of mobile hands-free telephones we have to expect low signal-to-noise ratios and considerable acoustic feedback at the local microphone. Adaptive filters are typically used for feedback cancellation. However, there is often residual echo due to insufficient performance of the echo canceler. It has been shown in [1, 2, 3] that the postfilter for combined residual echo and noise suppression improves the feedback attenuation in the duplex connection.

In this paper, we will clarify differences and similarities between echo cancellation and postfiltering: Echo cancellation in principle performs *in–phase* feedback attenuation, whereas postfiltering performs suppression of residual echo *irrespective of the signal phase*. Interestingly, echo cancellation and postfiltering for residual echo suppression *rely on the same control parameter*, the residual echo power. The assessment of this parameter is indeed the essential problem in echo control systems. In that respect, echo cancellation and postfiltering are reduced to the same estimation problem. We will demonstrate this fact in the example of a frequency–domain implementation.

1 Introduction

Figure 1 shows our model of the acoustic scenario of the hands–free telephone application together with our setup of echo and noise reduction filters. The local microphone signal at the sampling time index i,

$$y(i) = s(i) + n(i) + d(i) \ , \tag{1}$$

is additively composed of clean near speech $s(i)$, local background noise $n(i)$, and acoustic echo $d(i)$.

The linear echo canceler W yields a possibly inaccurate estimate $\widehat{d}(i)$ of the acoustic echo $d(i)$. Insufficient performance of the echo canceler can be caused variously:

*This work is supported by Nokia Research Center (NRC), Tampere, Finland, and Nokia Mobile Phones (NMP), Bochum, Germany.

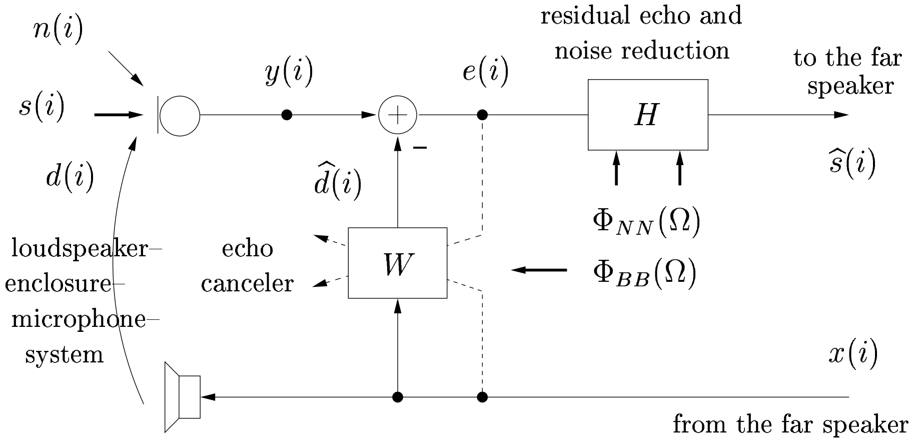

Figure 1 - Combined acoustic echo and noise reduction for (mobile) hands-free telephony.

- After enclosure dislocations, any type of an acoustic echo canceler needs a certain period of time to reconverge to the steady state of the system. The residual echo related to this transient phase of the echo canceler can be quite large.
- Even in the steady state of the system, we typically observe a small component of residual echo. This component is due to the tradeoff between tracking performance and steady state accuracy of acoustic echo cancelers.
- In many real–world applications, the loudspeaker–enclosure–microphone (LEM) system cannot be entirely covered by the echo canceler (e.g. due to complexity constraints). The number of model tap weights of the echo canceler falls short of the actual number of taps of the LEM system. In this case, the residual echo originates from the "late" part of the LEM system response. We will not specifically treat this problem in the present paper.

We assume that the time–varying impulse response $g(m,i)$ of the residual echo system can be modeled at time i by a causal FIR filter with N taps for $0 \leq m \leq N-1$. The residual echo $b(i) = d(i) - \widehat{d}(i)$ is then given by the linear convolution

$$b(i) = \sum_{m=0}^{N} g(m,i)x(i-m) \ . \tag{2}$$

The residual echo $b(i)$ and the background noise $n(i)$ can be suppressed by the postfilter H with input signal

$$e(i) = s(i) + n(i) + b(i) \ . \tag{3}$$

Due to the noise reduction part, the postfilter is typically implemented in the Discrete Fourier Transform (DFT) domain, for example by a Wiener filter relying on estimates of the background noise and residual echo power spectral density (PSD), $\Phi_{NN}(\Omega)$ and $\Phi_{BB}(\Omega)$, respectively.

The concept of joint control of acoustic echo cancellation and postfiltering (for residual echo suppression) was introduced in [4, 5] within the framework of sub–band echo cancellation. In this concept, the echo canceler step–size and the postfilter coefficients are uniquely related to each other. The objective of the present paper is to direct the attention of system designers to this important relation. We apply this concept to the efficient DFT based implementation of acoustic echo control. In the latter case, the link between echo cancellation and postfiltering can be established through the residual echo PSD $\Phi_{BB}(\Omega)$. It is the control parameter that postfiltering [1] and step–size control for acoustic echo cancellation [6] have in common [7]. Unfortunately, the residual echo is not a measurable signal and, thus, the residual echo PSD $\Phi_{BB}(\Omega)$ must be estimated from the available signals. Algorithms have been developed for this purpose in [7].

We chose the following organization of the paper: In Section 2 we will recall the frequency-domain adaptive filter (FDAF) [8] utilized for echo cancellation. Associated to that is an optimum step–size for the FDAF to guarantee the robustness of the echo canceler against the observation noise $s(i) + n(i)$ [6]. In Section 3, we will present the idea of a DFT based postfilter and how it is concatenated efficiently with the FDAF. Eventually, in Section 4, we will discuss the tight relation and the interaction between the echo canceler and the postfilter, given the residual echo PSD $\Phi_{BB}(\Omega)$.

2 Frequency–Domain Adaptive Filter W

Firstly, we summarize the frequency–domain adaptive filtering (FDAF) algorithm [8, 9] for a DFT length of $M = 2N$ corresponding to an effective length N of the echo canceler. Secondly, we recall an optimum step–size control for the FDAF [6] essentially relying on the residual echo PSD $\Phi_{BB}(\Omega)$.

2.1 Filtering and Update Equations

The DFT spectrum $E(\Omega_\ell, kR)$ at frame index $k \in \mathbb{Z}$ is obtained from the windowed time domain signal $e(i)$

$$\begin{aligned} E(\Omega_\ell, kR) &\doteq DFT\{e(kR - M + R + i)w_e(i)\} \\ &\doteq \sum_{i=0}^{M-1} e(kR - M + R + i)w_e(i)\mathrm{e}^{-j\Omega_\ell i} \end{aligned} \tag{4}$$

with frame shift $R \leq M/2 = N$ and the normalized discrete frequency index $\Omega_\ell = 2\pi\ell/M$ for $\ell = 0, 1, \ldots, M-1$. The rectangular window function applied to the signal $e(i)$ is defined as

$$w_e(i) = \begin{cases} 1 & \text{for } M/2 \leq i \leq M-1 \\ 0 & \text{otherwise .} \end{cases} \tag{5}$$

The same notation holds for the DFT coefficients $X(\Omega_\ell, kR)$ corresponding to the excitation signal $x(i)$ when we use the extended window function

$$w_x(i) = w_e(i) + w_e(i + M/2) = \begin{cases} 1 & \text{for } 0 \leq i \leq M{-}1 \\ 0 & \text{otherwise .} \end{cases} \tag{6}$$

The constrained FDAF algorithm updates the frequency–domain adaptive weights $W(\Omega_\ell, kR)$ according to

$$W(\Omega_\ell, (k+1)R) = W(\Omega_\ell, kR) + DFT\Big\{q(i) \cdot IDFT\Big\{\mu \frac{X^*(\Omega_\ell, kR)E(\Omega_\ell, kR)}{\Phi_{XX}(\Omega_\ell, kR)}\Big\}\Big\} \tag{7}$$

using the projection (i.e. constraining) window $q(i)$

$$q(i) = w_e(i + M/2) = \begin{cases} 1 & \text{for } 0 \leq i \leq M/2 - 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

and the DFT spectrum $E(\Omega_\ell, kR)$ corresponding to the error signal $(M/2 \leq i \leq M-1)$

$$e(kR - M + R + i) = y(kR - M + R + i) - IDFT\{X(\Omega_\ell, kR)W(\Omega_\ell, kR)\}\ . \tag{9}$$

In Equation (7), the normalization PSD $\Phi_{XX}(\Omega_\ell, kR)$ is usually approximated by first order recursive smoothing of $\frac{1}{2}|X(\Omega_\ell, kR)|^2$ with $0 < \lambda < 1$ adjusted to the short–time stationarity of the excitation. The factor $\frac{1}{2}$ accounts for the extended window (6) that is used for the spectral analysis of the excitation signal (in contrast to other signals):

$$\Phi_{XX}(\Omega_\ell, kR) = \lambda\Phi_{XX}(\Omega_\ell, (k-1)R) + (1-\lambda)\frac{1}{2}|X(\Omega_\ell, kR)|^2 \tag{10}$$

The parameter μ in (7) denotes the non–negative step-size factor which must be chosen to ensure the robustness of the FDAF against the observation noise $n(i) + s(i)$. This issue will be treated in Section 2.2.

The FDAF algorithm is illustrated in Figure 2 as a memoryless black-box module for frame index k. At the inputs, the excitation signal $x(i)$ and the microphone signal $y(i)$ are required. The set of estimated LEM system coefficients $W(\Omega_\ell, kR)$ is the one that was predicted in the previous iteration. At the outputs, the echo compensated error signal $e(i)$ is observed together with its DFT spectrum $E(\Omega_\ell, kR)$ and the estimated LEM system coefficients $W(\Omega_\ell, (k+1)R)$. We now turn to the computation of the optimum step–size parameter μ.

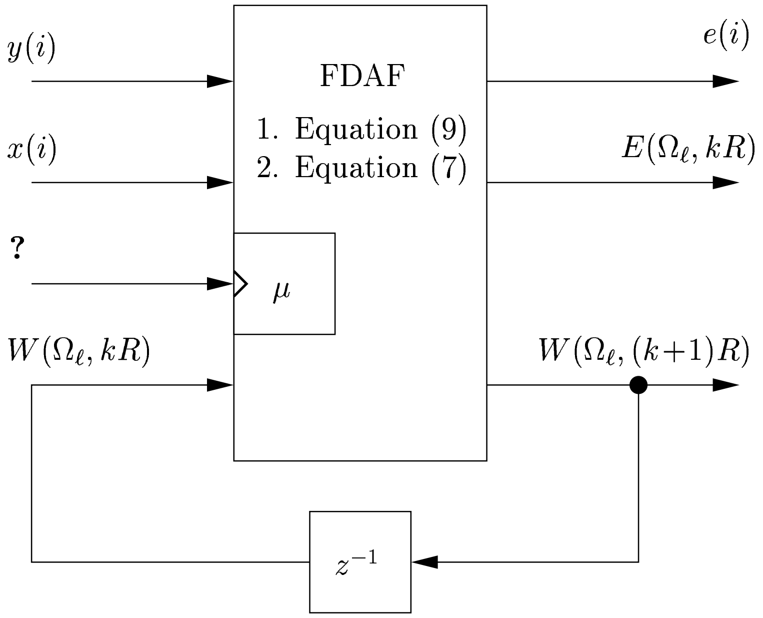

Figure 2 - Black-box module of the FDAF with input and output interfaces.

2.2 Optimum Step–Size

Consider the sampled frequency response $G(\Omega_\ell, kR) = DFT\{g(i, kR)\}$ corresponding to the time–varying residual echo system $g(m, i)$ as defined in the introduction. We will call the magnitude squared frequency response $|G(\Omega_\ell, kR)|^2$ of the residual echo system the convergence state or the misalignment of the echo canceler. $|G(\Omega_\ell, kR)|^2$ could also be understood as a frequency dependent echo return loss (ERL) jointly provided by the acoustical transfer and the echo canceler.

The optimum (time and frequency dependent) step–size for the FDAF was derived in [6], such that the convergence state $|G(\Omega_\ell, (k+1)R)|^2$ becomes minimum in the mean square error (MMSE) sense, given the present $|G(\Omega_\ell, kR)|^2$:

$$\begin{aligned}\mu = \mu(\Omega_\ell, kR) &= \frac{\Phi_{BB}(\Omega_\ell, kR)}{\Phi_{EE}(\Omega_\ell, kR)} \\ &= \frac{\Phi_{BB}(\Omega_\ell, kR)}{\Phi_{SS}(\Omega_\ell, kR) + \Phi_{NN}(\Omega_\ell, kR) + \Phi_{BB}(\Omega_\ell, kR)} \\ &= \frac{|G(\Omega_\ell, kR)|^2 \Phi_{XX}(\Omega_\ell, kR)}{\Phi_{EE}(\Omega_\ell, kR)}\end{aligned} \tag{11}$$

$\Phi_{BB}(\Omega_\ell, kR)$ is the residual echo PSD at frame index k caused by the misalignment $|G(\Omega_\ell, kR)|^2$ of the weights $W(\Omega_\ell, kR)$ (with respect to the actual LEM system). $\Phi_{BB}(\Omega_\ell, kR)$ or $|G(\Omega_\ell, kR)|^2$ must be estimated from the available signals in order to implement the optimum step–size. $\Phi_{EE}(\Omega_\ell, kR)$ and $\Phi_{XX}(\Omega_\ell, kR)$ denote the PSDs of the error signal and the excitation, respectively. Those can be measured directly.

3 Postfiltering in the DFT Domain

We first discuss the conceptual background of the postfilter and then recall the linear MMSE solution (Wiener filter).

3.1 Echo Cancellation and/or Echo Suppression?

The objective of the postfilter is to suppress residual echo and background noise. In contrast to the echo canceler, the suppression of residual echo shall be performed irrespective of the signal phase. This requirement can be fulfilled much easier than the demand for in–phase cancellation of acoustic echo. It is thus evident that the postfilter for residual echo suppression still works in situations where the echo cancellation strategy fails. It must be noted, however, that postfiltering always introduces distortions to the useful signal (near speech and partly background noise). In order to limit the signal distortions to a minimum, we have to design a combined solution comprising both primary echo cancellation and postfiltering for residual echo suppression. Furthermore, there must be an interaction between both strategies (the discussion of which can be found in Section 4).

The background noise reduction for speech signals is commonly performed in the DFT domain. We adopt this way for a combined postfilter for residual echo and noise suppression. Interestingly, the DFT coefficients $E(\Omega_\ell, kR)$ corresponding to the postfilter input $e(i)$ are computed by the FDAF already (compare Figure 2). Obviously, that results in an efficient combined solution of echo cancellation (FDAF) and postfiltering in the DFT domain.

3.2 Wiener Filtering in the DFT Domain

Consider Figure 1. We want to apply a linear filter H to the echo compensated signal $e(i)$ such that $\widehat{s}(i)$ approximates $s(i) + \gamma \cdot n(i)$ in the MMSE sense, where $0 \leq \gamma \leq 1$ defines the relative level of background noise in the wanted signal.

The desired linear MMSE solution can be found approximately by spectral weighting of DFT coefficients according to

$$\widetilde{S}(\Omega_\ell, kR) = H(\Omega_\ell, kR) E(\Omega_\ell, kR) \tag{12}$$

if $H(\Omega_\ell, kR)$ is chosen on the basis of an extended Wiener rule:

$$\begin{aligned} H_\gamma(\Omega_\ell, kR) &= \frac{\Phi_{SS}(\Omega_\ell, kR) + \gamma \cdot \Phi_{NN}(\Omega_\ell, kR)}{\Phi_{SS}(\Omega_\ell, kR) + \Phi_{NN}(\Omega_\ell, kR) + \Phi_{BB}(\Omega_\ell, kR)} \\ &= \frac{\Phi_{EE}(\Omega_\ell, kR) - \Phi_{BB}(\Omega_\ell, kR) - (1-\gamma) \cdot \Phi_{NN}(\Omega_\ell, kR)}{\Phi_{EE}(\Omega_\ell, kR)} \end{aligned} \tag{13}$$

If we want to perform postfiltering for residual echo suppression only ($\gamma = 1$), we obtain the special case

$$\begin{aligned} H_1(\Omega_\ell, kR) &= \frac{\Phi_{SS}(\Omega_\ell, kR) + \Phi_{NN}(\Omega_\ell, kR)}{\Phi_{SS}(\Omega_\ell, kR) + \Phi_{NN}(\Omega_\ell, kR) + \Phi_{BB}(\Omega_\ell, kR)} \\ &= \frac{\Phi_{EE}(\Omega_\ell, kR) - \Phi_{BB}(\Omega_\ell, kR)}{\Phi_{EE}(\Omega_\ell, kR)} \end{aligned} \tag{14}$$

which will be used in Section 4 where we discuss the tight relation between echo cancellation and residual echo suppression by the postfilter.

Accurate short–time estimates of the background noise PSD $\Phi_{NN}(\Omega_\ell, kR)$ and the residual echo PSD $\Phi_{BB}(\Omega_\ell, kR)$ are crucial for the reliability of the spectral weights $H_\gamma(\Omega_\ell, kR)$. The background noise PSD can be determined adaptively and accurately by the Minimum Statistics approach [10, 11] by which the desired noise PSD can be tracked even during speech activity. The residual echo PSD is the same as used in the step–size control of the FDAF, compare Equation (11).

Instead of Wiener filtering, we could also apply the more advanced MMSE-LSA (Log–Spectral Amplitude) estimator [12] which, however, relies in a similar way on residual echo and noise PSD estimates.

The synthesis of the output signal $\widehat{s}(i)$ of the postfilter eventually requires an inverse DFT per frame and the overlap/add method ($M - R \leq i \leq M - 1$):

$$\widehat{s}(kR - M + R + i) = \frac{2R}{M} \sum_{u=-M/2R}^{M/2R-1} IDFT\{\mathrm{e}^{-jR\pi\ell u/M} \widetilde{S}(\Omega_\ell, (k+u)R)\} \tag{15}$$

The complex exponentials are responsible for the correct recombination of signal components due different frames $k + u$. For efficiency, this kind of modulation has to be realized as a non–cyclic shift in the time–domain. Note that the synthesis of $\widehat{s}(i)$ at frame index k requires the DFT coefficients $\widetilde{S}(\Omega_\ell, (k+u)R)$ for $u = -M/2R$ to $u = M/2R - 1$. That corresponds to a signal delay of $M/2$ samples. In case we are bound to the delay constraint of $R < M/2$ samples, a truncated summation from $u = -M/2R$ to $u = 0$ must be applied.

4 Interaction of Echo Cancellation and Postfiltering

The algorithmic relation between echo cancellation and postfiltering is given by the MMSE step-size, Equation (11), and the MMSE postfilter, Equation (14). The residual echo

PSD $\Phi_{BB}(\Omega_\ell, kR)$ is the control parameter which both algorithms have in common. The estimation of $\Phi_{BB}(\Omega_\ell, kR)$ is in fact the relevant issue in both concepts. Thus, echo cancellation and postfiltering for residual echo suppression boil down to the same problem. Assume you found a reliable estimate of $\Phi_{BB}(\Omega_\ell, kR)$ to implement the optimum step–size (11) of the echo canceler, then you could easily run the postfilter (14) as well.

The relation as explained can also be expressed mathematically if we eliminate the PSD $\Phi_{BB}(\Omega_\ell, kR)$ that Equations (11) and (14) have in common. That results in the following simple statement [4, 5]:

$$\mu(\Omega_\ell, kR) + H_1(\Omega_\ell, kR) = 1 \tag{16}$$

This result is illustrated in Figure 3 where μ and H_1 according to Equations (11) and (14) are plotted versus the residual echo–to–noise power ratio (RENR) $\Phi_{BB}/(\Phi_{NN} + \Phi_{SS})$.

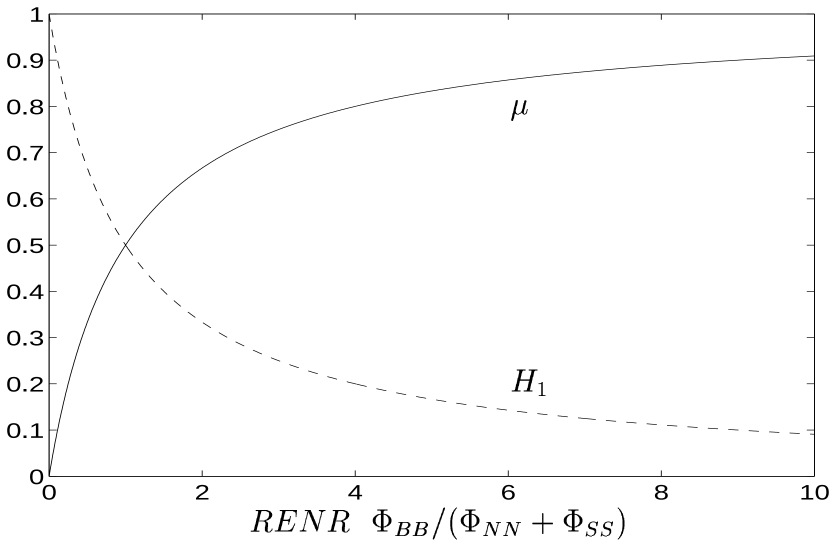

Figure 3 - Illustration of the behavior of the step–size μ and the spectral weight H_1 versus the residual echo–to–noise power ratio $\Phi_{BB}/(\Phi_{NN} + \Phi_{SS})$.

From the viewpoint of operation, echo cancellation and postfiltering are of course different. Echo cancellation performs in–phase attenuation of the acoustic feedback, whereas postfiltering suppresses the residual echo regardless of the signal phase. Another difference is that the echo canceler (which is realized as an adaptive filter) is characterized by finite convergence speed. Therefore, during the process of adaptation (when μ is large), the postfilter delivers additional attenuation in the feedback loop of the hands–free telephone (as H_1 is small). While the echo canceler converges, the residual echo PSD $\Phi_{BB}(\Omega_\ell, kR)$ decreases (thus μ becomes small) and the responsibility for echo control is gradually taken away from the postfilter (as H_1 gets larger). This interaction maintains the highest transparency with respect to the transmission of near speech and a certain level of ambient noise.

In the present paper, echo cancellation and postfiltering are both implemented in the frequency domain. The interface between the two algorithms is the frequency–domain error signal $E(\Omega_\ell, kR)$ provided by the FDAF. That means that analysis/synthesis operations (DFT/IDFT) are shared between the echo canceler and the postfilter. Obviously that results in an algorithm which is highly efficient from the viewpoint of computational complexity.

5 Conclusions

Echo cancellation with a postfilter for residual echo suppression is a very popular strategy in hands–free telephony. We have shown for the frequency–domain that the echo canceler step–size and the postfilter rely on the same control parameter, the residual echo PSD. Thus, echo cancellation and postfiltering have been reduced to the same estimation problem. Consequently, the estimation of the residual echo PSD must be considered as the key issue in the echo control problem.

In this paper, the tight relation (interaction) between echo cancellation and postfiltering was established for a DFT based solution in the MMSE sense. We have noted that the basic idea was applied to sub–band echo control as well [4, 5]. In any case, the setup in the DFT domain constitutes a very efficient solution to the echo control problem.

References

[1] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, pp. 21–32, 1998.

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

[3] S. Gustafsson, R. Martin, P. Jax, P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. Speech and Audio Processing*, July 2002.

[4] E. Hänsler, G. Schmidt, "Hands-free telephones - joint control of echo cancellation and postfiltering," *Signal Processing*, vol. 80, pp. 2295–2305, 2000.

[5] G. Schmidt, *Entwurf und Realisierung eines Multiratensystems zum Freisprechen*, Ph.D. thesis, Fortschritt–Berichte VDI Verlag, Reihe 10, Nr. 674, Düsseldorf, Germany, 2001.

[6] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Processing*, vol. 80, pp. 1733–1745, 2000.

[7] G. Enzner, R. Martin, P. Vary, "Partitioned residual echo power estimation for frequency–domain acoustic echo cancellation and postfiltering," *European Transaction on Telecommunications*, vol. 13, March 2002.

[8] E. Ferrara, "Frequency–domain adaptive filtering," *In: C.F.N. Cowan, P.M. Grant (Eds.), Adaptive Filters, Prentice–Hall, Englewood Cliffs, N.J.*, pp. 145–179, 1985.

[9] S. Haykin, "Adaptive Filter Theory," Prentice Hall, New Jersey, 1996.

[10] R. Martin, "Spectral Subtraction Based on Minimum Statistics," Proc. EUSIPCO-94, Edinburgh, pp. 1182-1185, September 12-16, 1994.

[11] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Trans. Speech and Audio Processing*, July 2001.

[12] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 443–445, April 1985.