NEW INSIGHTS INTO THE STATISTICAL SIGNAL MODEL AND THE PERFORMANCE BOUNDS OF ACOUSTIC ECHO CONTROL

Gerald Enzner and Peter Vary

Institute of Communication Systems and Data Processing (IND)
Aachen University (RWTH), D–52056 Aachen, Germany
Phone: +49-241-8026960 E-mail: {enzner,vary}@ind.rwth-aachen.de

ABSTRACT

The contribution of this paper is two-fold. At first, we introduce a modification of the linear statistical signal model in acoustic echo control. In contrast to the traditional approach, the acoustic echo path is characterized as a random process with statistical mean and covariance, while the echo path input is modeled as a deterministic signal. Based on the modified signal model, we then derive the linear MMSE estimator for the near-end speech components in the microphone signal. The result can be seen as a generalized Wiener filter that consists of an acoustic echo canceler and a postfilter for residual echo suppression.

The presented theory entails several fundamental advantages: a) the new signal model better matches the practical applications of acoustic echo control, b) it proves the principal coexistence of echo canceler and postfilter in hands-free communication systems, c) the generalized Wiener solution simplifies the realization of acoustic echo controllers, and d) we obtain a better insight into the performance bounds of acoustic echo control.

1. INTRODUCTION

A generic signal model of a hands-free voice communication system (e.g. hands-free telephone) is shown in Figure 1. In receiving direction, a possibly processed version $x'(i)$ of the received signal $x(i)$ at discrete time i is played back by the loudspeaker. In sending direction, the hands-free microphone captures the speech signal $s(i)$ of the near speaker as well as the room reflections of the signal $x'(i)$. The microphone signal $y(i)$ is therefore considered as an additive mixture of useful signal components $s(i)$ and echo signal components $d(i)$. If there is a considerable signal delay between both ends of the communication system (e.g. GSM), then the acoustic echo is not tolerated by the far speaker.

An *adaptive echo canceler* in parallel to the electroacoustic echo path [1, 2] has been identified as a seemingly ideal solution for acoustic echo control. In practice, however, it has been observed that there is always a residual echo signal after the echo canceler. The residual echo is often attributed to the presence of echo path non-linearities or to the echo path tail which is not considered by a constrained (i.e. FIR) echo canceler.

A *frequency-selective adaptive postfilter* in the sending path of the communication system has been proposed to reduce the residual echo [3, 4]. It has been demonstrated in [5] that the postfilter may indeed strongly attenuate the residual echo and preserve the subjective quality of the speech signal (full-duplex ability).

In this paper, the optimum echo canceler and postfilter coefficients will be derived jointly from the minimum mean-square error (MMSE) criterion. In contrast to [4], we will show that the combination of both filters is required even in the case of a linear unconstrained echo cancellation problem. In Section 2 of the paper, we will present the underlying signal model. In Section 3, the mathematical derivation of the optimum filters will be outlined. In Section 4, interesting interpretations of the solution and realization issues will be discussed. In Section 5, the resulting performance bounds in acoustic echo control will be analyzed by simulations.

This work is supported by Nokia Research Center, Tampere, Finland, and Nokia Mobile Phones, Bochum, Germany.

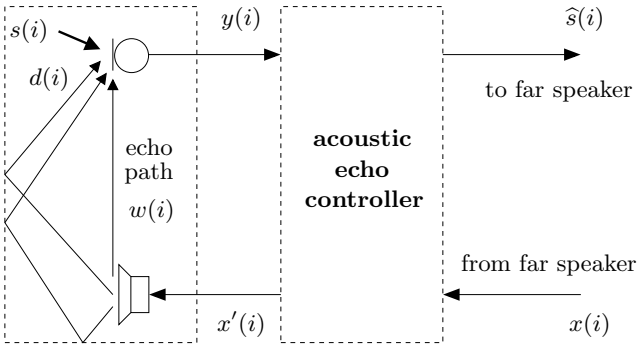


Figure 1: Acoustic front-end of a hands-free voice communication system (e.g. hands-free telephone). The output signal $\widehat{s}(i)$ of the acoustic echo controller shall approximate the useful signal $s(i)$.

2. NEW SIGNAL MODEL AND OPTIMALITY CRITERION FOR ACOUSTIC ECHO CONTROL

We assume that the impulse response $w(i)$ entirely models the electroacoustic coupling between the loudspeaker signal $x'(i)$ and the microphone signal $y(i)$. If sufficiently good transducers are used, the linear echo path is widely accepted as a realistic model for the acoustic environment of hands-free telephones. To provide transparent sound at least to the near speaker (telephone owner), we set the loudspeaker signal equal to the received signal, i.e., $x'(i) = x(i)$. The microphone signal $y(i)$ then reads

$$\begin{aligned} y(i) &= s(i) + d(i) \\ &= s(i) + w(i) * x(i) \\ &= s(i) + \sum_p w(p)x(i-p)\ . \end{aligned} \tag{1}$$

In the traditional theory of acoustic echo control, the speech signal $s(i)$ and the received signal $x(i)$ are both modeled as independent *random processes*, while the echo path

$w(i)$ is treated as an unknown *deterministic* parameter. For these model assumptions, the MMSE optimization of an echo canceler leads to the well-known Wiener solution, i.e., the echo canceler ideally copies the true echo path in order to compensate the acoustic echo in the microphone signal.

Here, we propose an alternative signal model which better reflects the practical applications of acoustic echo control. As the speech signal $s(i)$ is not observable alone, it is still modeled as a stationary *random process* with zero mean and autocorrelation $\varphi_{ss}(n) = \mathcal{E}\{s(i)s(i+n)\}$. The echo signal $d(i)$, however, is now considered as the linear convolution of a *measurable (i.e. deterministic)* loudspeaker signal $x(i)$ and the unknown echo path coefficients $w(i)$. Due to the uncertainty about the acoustic echo path, the coefficients $w(i)$ are now modeled as an independent *random process* with non-constant statistical expectation $w_o(i)$ and covariance $\varphi_{w_r w_r}(n)$:

$$w_o(i) = \mathcal{E}\{w(i)\} \tag{2}$$

$$w_r(i) = w(i) - w_o(i) \tag{3}$$

$$\varphi_{w_r w_r}(n) = \mathcal{E}\{w_r(i)w_r(i+n)\}\ . \tag{4}$$

If the mean $w_o(i)$ exists, it represents a systematic (i.e. deterministic) component of the statistical echo path at lag i, while the stationary sequence $w_r(i)$ is a purely random (i.e. zero-mean) component. The significance of the proposed signal model will be further demonstrated in the course of this paper.

The full-duplex operation of the hands-free telephone in Figure 1 requires the strong attenuation of the echo signal $d(i)$ by the acoustic echo controller, but ideally, the echo attenuation is subject to the undistorted reproduction of the useful signal $s(i)$ at the system output $\widehat{s}(i)$. Mathematically, this conflict can be expressed as a statistical optimization problem which aims, for example, at the MMSE between $s(i)$ and $\widehat{s}(i)$:

$$\epsilon^2 = \mathcal{E}\big\{\big(s(i) - \widehat{s}(i)\big)^2\big\} \ \rightarrow\ \min\ . \tag{5}$$

To facilitate the computation of the system output $\widehat{s}(i)$ according to the MMSE criterion (without making further assumptions on the statistics of the involved signals), we formulate the echo cancellation problem as a general unconstrained linear filtering problem, where the output signal $\widehat{s}(i)$ is obtained as a linear combination of the input signals $x(i)$ and $y(i)$:

$$\widehat{s}(i) = w_2'(i) * y(i) + w_1'(i) * x(i) \tag{6a}$$

$$= w_2(i) * \big[y(i) - w_1(i) * x(i)\big] \tag{6b}$$

$$= w_2(i) * e(i)\ . \tag{6c}$$

Mathematically, the filter structures in (6a) and (6b) are equivalent as they can be uniquely transformed into each other. In principle, we can either optimize the linear filters $w_1'(i)$ and $w_2'(i)$, or alternatively $w_1(i)$ and $w_2(i)$. It turns out, however, that the solution for the filter structure in (6b) is simpler and more intuitive. To be in line with the common literature on acoustic echo control, we will refer to $w_1(i)$ as the echo canceler and to $w_2(i)$ as the postfilter for residual echo suppression. The abbreviation $e(i) = y(i) - w_1(i) * x(i)$ obviously takes the meaning of the error signal after echo cancellation.

The formulation as an unconstrained (i.e. IIR) filtering problem was chosen to emphasize that the echo cancellation problem is not under-modeled by a restricted adaptive filter length here. That implies that the echo canceler $w_1(i)$ entirely covers the span of the linear echo path $w(i)$. As $w_1(i) = w(i)$ clearly eliminates the echo, it will be interesting to clarify the role of the postfilter $w_2(i)$ in this seemingly simple constellation.

3. A GENERALIZED WIENER SOLUTION FOR ACOUSTIC ECHO CONTROL

To derive the desired MMSE solution for acoustic echo control, we substitute the general filter structure (6b) into (5) and compute the partial derivatives of the mean-square error ϵ^2 with respect to the coefficients $w_1(n)$ and $w_2(n)$. The following expressions for the derivatives are obtained using only the previously made assumptions of a deterministic signal $x(i)$ and a random variable $s(i)$ with zero mean:

$$\frac{\partial \epsilon^2}{\partial w_1(n)} = -2\big[w_2(i) * \mathcal{E}\{y(i)\} - w_2(i) * w_1(i) * x(i)\big] \cdot \big[w_2(i-n) * x(i-n)\big] \tag{7}$$

$$\frac{\partial \epsilon^2}{\partial w_2(n)} = -2\big[w_2(i) * \mathcal{E}\{y(i)\} - w_2(i) * w_1(i) * x(i)\big] \cdot \big[w_1(i-n) * x(i-n)\big] + 2\,\mathcal{E}\Big\{\big[w_2(i) * e(i) - s(i)\big] \cdot y(i-n)\Big\}\ . \tag{8}$$

The reason for the remaining time index i after the evaluation of the statistical expectation lies in the deterministic nature of $x(i)$. In the next step, we use the linear statistical signal model in (1) to find that $\mathcal{E}\{y(i)\} = \mathcal{E}\{w(i)\} * x(i)$. Now it can be easily seen that $\partial\epsilon^2/\partial w_1(n) = 0$ if $w_1(i) = \mathcal{E}\{w(i)\}$. Therefore, the optimum echo canceler in the time-domain is given by

$$w_1(i) = w_o(i)\ . \tag{9}$$

This result is in line with the commonly known Wiener solution for a purely deterministic echo path model. For the convenience in our later analysis, we also define the frequency response $W_1(\Omega) = \mathcal{F}\{w_1(n)\}$ of the optimum filter:

$$W_1(\Omega) = \mathcal{F}\{\mathcal{E}\{w(i)\}\} = \mathcal{E}\{W(\Omega)\}\ . \tag{10}$$

Here, $W(\Omega) = \mathcal{F}\{w(i)\}$ is the complex frequency response of the echo path and we define its statistical expectation as $W_o(\Omega) = \mathcal{E}\{W(\Omega)\}$.

We now consider Equation (8) to find $w_2(i)$ that satisfies $\partial\epsilon^2/\partial w_2(n) = 0$. The first part of (8) is very similar to (7) and vanishes for the optimum filter $w_1(i) = w_o(i)$. The second part of (8) can be simplified if we again use the signal model in (1) and further assume the statistical independence of $s(i)$ and $w_r(i)$. We obtain the following equation (consisting of deterministic and statistical terms) for the optimum filter $w_2(n)$:

$$w_2(n) * \big[\varphi_{ss}(n) + x(-n) * x(n) * \varphi_{w_r w_r}(n)\big] = \varphi_{ss}(n)\ . \tag{11}$$

In order to compute the optimum postfilter explicitly, we apply the Fourier transform to (11) and solve for the frequency response $W_2(\Omega) = \mathcal{F}\{w_2(n)\}$:

$$W_2(\Omega) = \frac{\Phi_{ss}(\Omega)}{\Phi_{ss}(\Omega) + |X(\Omega)|^2 \cdot \Phi_{w_r w_r}(\Omega)}\ . \tag{12}$$

In this result, $\Phi_{ss}(\Omega) = \mathcal{F}\{\varphi_{ss}(n)\}$ denotes the power spectral density (PSD) of the speech signal $s(i)$. With the previous definitions in (3) and (4), the symbol $\Phi_{w_r w_r}(\Omega) = \mathcal{F}\{\varphi_{w_r w_r}(n)\}$ defines the frequency-dependent system distance between the echo path $w(i)$ and the optimum echo canceler $w_1(i) = w_o(i)$. For brevity, we will call $\Phi_{w_r w_r}(\Omega)$ the echo path covariance in the frequency-domain and we shall keep in mind that it is a direct measure for the uncertainty about the acoustic echo path.

In our derivation, echo canceler and postfilter have been deduced jointly from the MMSE criterion and, therefore, we will refer to the combination of (10) and (12) as the *generalized Wiener solution* for acoustic echo control. A block diagram of the optimal filter structure immediately follows from Equation (6b) and is shown in Figure 2(a).

For the sake of completeness, we have to mention that the previous assumption of stationarity is certainly not realistic in the application of acoustic echo control. Therefore, in practice, the optimum filters have to be updated on the basis of short-term stationary signal frames. The implementation of the digital filters can be realized approximately by a fast convolution in the DFT domain, but also by a direct convolution in the time-domain.

4. DISCUSSION OF THE GENERALIZED WIENER SOLUTION AND SPECIAL CASES

We will distinguish three cases which can be relevant for practical applications, differing in the degree of generality and in the way they utilize *a priori* information in form of the mean and the covariance of the acoustic echo path.

a) General Statistical Case

We have shown that the optimum solution for acoustic echo control is generally given by the following two optimum filters in the frequency-domain (noting that the two filters separately take the mean and the covariance of the echo path into account):

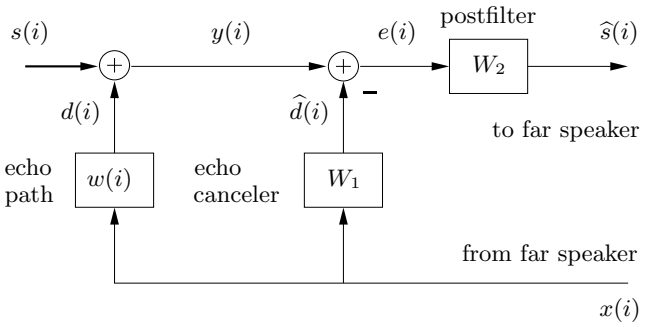
$$W_1(\Omega) = \mathcal{E}\{W(\Omega)\} \tag{13}$$

$$W_2(\Omega) = \frac{\Phi_{ss}(\Omega)}{\Phi_{ss}(\Omega) + |X(\Omega)|^2 \cdot \Phi_{w_r w_r}(\Omega)} \,. \tag{14}$$

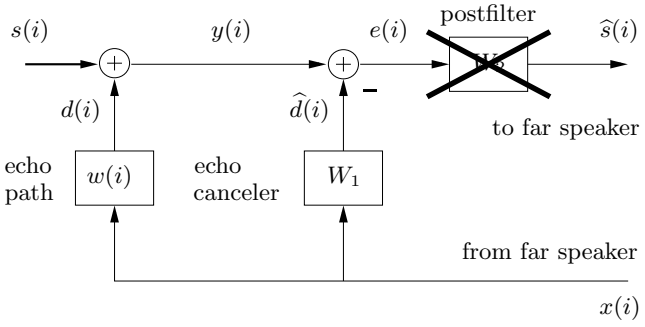
It turns out in practice that this general two-filter solution is very much suitable for acoustic echo control. In most applications, it is indeed possible to determine a (time-varying) systematic component $W_o(\Omega) = \mathcal{E}\{W(\Omega)\}$ of the echo path using adaptive filters and the system identification approach [1, 2]. Nevertheless, an uncertainty $\Phi_{w_r w_r}(\Omega)$ about the echo path always remains and, therefore, the postfilter is an indispensable component of advanced hands-free communication systems.

In contrast to the traditional theory of echo canceler and postfilter [1, 4], the generalized Wiener solution simplifies the design and realization of acoustic echo controllers, e.g.:

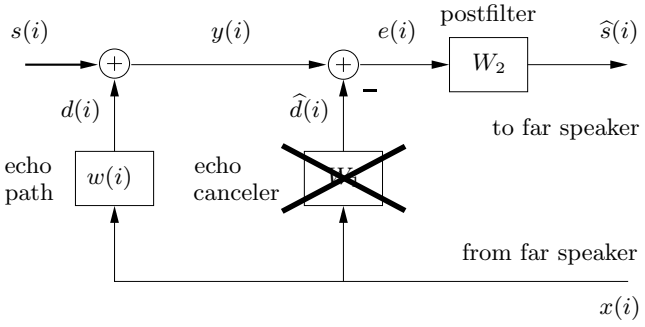
- The postfilter in (14) requires the instantaneous power spectrum (periodogram) $|X(\Omega)|^2$ of the received signal $x(i)$. For non-stationary speech signals, this *periodogram* can be determined much easier than the time-varying PSD of $x(i)$, which appears in the traditional approach.
- Another advantage over the traditional approach is that the optimum filters in (13) and (14) are characterized by *purely statistical parameters* of the echo path. Hence, we may utilize the extensive literature on statistical parameter estimation in order to adjust the filter coefficients.



(a) Generalized Wiener solution for a statistical echo path model with non-zero mean and non-zero covariance.



(b) Classical Wiener solution for a deterministic echo path.



(c) The Wiener solution for an echo path with zero mean.

Figure 2: Generalized Wiener solution and two special cases.

b) Deterministic Case

If there is no uncertainty about the echo path, i.e., $\mathcal{E}\{w(i)\} = w(i)$ and $\varphi_{w_r w_r}(n) = 0$, the general solution degenerates to:

$$W_1(\Omega) = W(\Omega) \tag{15}$$

$$W_2(\Omega) = 1 \,. \tag{16}$$

The echo canceler is now an ideal copy of the echo path and the echo canceled error signal $e(i)$ passes the postfilter unprocessed, as shown by Figure 2(b). This result is commonly known as the Wiener solution for acoustic echo cancellation and is indeed optimal for a deterministic echo path $w(i)$.

In practice, the deterministic echo path model requires very sophisticated control mechanisms to let the echo canceler coefficients follow the true echo path with sufficient accuracy. In time-varying and noisy acoustic environments, this strategy may not deliver sufficient echo attenuation.

c) Zero-Mean Case
If no systematic information is available about the echo path, i.e. $\mathcal{E}\{w(i)\} = 0$ and $\varphi_{w_r w_r}(n) = \varphi_{ww}(n) = \mathcal{E}\{w(i)w(i+n)\}$, we obtain the following optimum filters:

$$W_1(\Omega) = 0 \quad (17)$$

$$W_2(\Omega) = \frac{\Phi_{ss}(\Omega)}{\Phi_{ss}(\Omega) + |X(\Omega)|^2 \cdot \Phi_{ww}(\Omega)}, \quad (18)$$

using the definition $\Phi_{ww}(\Omega) = \mathcal{F}\{\varphi_{ww}(n)\}$. As shown by Figure 2(c), the echo canceler does not have an influence on the output signal, i.e., the statistical postfilter degenerates to an MMSE equalizer which performs the suppression of the echo signal alone.

An MMSE equalizer alone is not recommended for acoustic echo control. Experiments have shown that the speech signal at the system output may sound distorted, while the background noise can be strongly modulated by the presence of the received signal.

5. PERFORMANCE BOUNDS

As the quality of acoustic echo controllers can be distinguished mainly by their duplex ability, we consider a hard *double talk* situation with an input signal-to-echo ratio of $\mathrm{SER}_y = \sigma_s^2/\sigma_d^2 = 0\,\mathrm{dB}$. We have chosen a car acoustic environment, where the length of the echo path is about 500 taps at 8 kHz sampling frequency. The background noise level at the hands-free microphone has been adjusted such that the signal-to-noise ratio of the near speaker is 10 dB. The echo attenuation that can be achieved by echo canceler and postfilter is measured in terms of the echo return loss enhancement (ERLE) [1, 4]. To analyze the selectivity of the filters, we consider the resulting $\mathrm{SER}_{\hat{s}} = \sigma_s^2/\sigma_{s-\hat{s}}^2$ at the system output.

The performance of the generalized Wiener solution is investigated for different qualities of *a priori* knowledge about the echo path, i.e., we consider different echo path uncertainties $\Phi_{w_r w_r}(\Omega) = \beta \cdot \Phi_{ww}(\Omega)$. The factor $0 < \beta < 1$ controls the smooth transition between the purely deterministic case and the zero-mean case. Given the echo path covariance $\Phi_{w_r w_r}(\Omega)$, the speech PSD $\Phi_{ss}(\Omega)$ could be estimated by spectral subtraction [1], but here it is known *a priori* to find performance bounds.

The results of the experiment are shown in Figure 3. The $\mathrm{ERLE}_{W_1} = -10\log_{10}\beta$ on the abscissa is a direct measure for the quality of the statistical echo path parameters. The adjusted ERLE_{W_1} then leads to the improved $\mathrm{SER}_e = \sigma_s^2/\sigma_{s-e}^2 = \mathrm{ERLE}_{W_1}$ after the echo canceler. The more interesting question is related to the ERLE_{W_2} and the SER improvement that can be achieved by the postfilter. As anticipated, for high qualities of the echo path mean (i.e. high ERLE_{W_1}, nearly deterministic case) the postfilter contributes very little echo attenuation, while for a low quality of the echo path mean (i.e. $\mathrm{ERLE}_{W_1} = 0\,\mathrm{dB}$, zero-mean case) the postfilter performs the echo suppression alone. It is important to see that the $\mathrm{SER}_{\hat{s}}$ after the postfilter approaches the total $\mathrm{ERLE}_{W_{12}}$ by echo canceler and postfilter. That means that the additional echo attenuation by the postfilter is achieved at the cost of a very moderate speech distortion.

Example: In a realistic acoustic environment, a well designed echo canceler might achieve an $\mathrm{ERLE}_{W_1} = 15\,\mathrm{dB}$. Starting from $\mathrm{SER}_y = 0\,\mathrm{dB}$, we obtain an $\mathrm{SER}_e = 15\,\mathrm{dB}$ after the echo canceler. In this case, Figure 3 tells us that the ideal postfilter contributes an additional $\mathrm{ERLE}_{W_2} = 8\,\mathrm{dB}$. Therefore, a total $\mathrm{ERLE}_{W_{12}} = 23\,\mathrm{dB}$ is attained at the system output. The $\mathrm{SER}_{\hat{s}} = 22\,\mathrm{dB}$ is just 1 dB below, indicating the excellent selectivity of the postfilter.

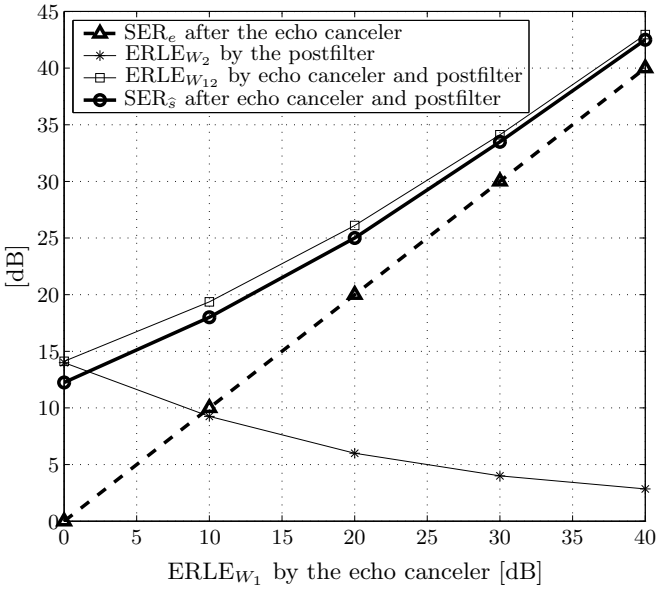


Figure 3: ERLE and output SER. Double talk, $\mathrm{SER}_y = 0\,\mathrm{dB}$.

6. CONCLUSIONS

We presented an improved signal model for acoustic echo control. In contrast to the traditional theory, the unknown echo path was characterized as a random process, while the measurable echo path input was treated as a deterministic signal. The model fits many realworld applications as it takes a systematic as well as a purely random component of the echo path into account.

Based on the new signal model, we have then optimized a general two-filter structure for acoustic echo control. The rigorous derivation of the MMSE solution has delivered new variants of the acoustic echo canceler and the postfilter for residual echo suppression. We clarified that their filter parameters can be determined easier than in the traditional framework.

The performance of the new solution was analyzed by simulations with real speech input. In a realistic acoustic environment, the optimum filters can achieve a significant echo attenuation of about 23 dB at the cost of only 1 dB speech distortion (considering a hard double talk situation).

REFERENCES

[1] Eberhard Hänsler and Gerhard Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.

[2] J. Benesty, T. Gänsler, D.R. Morgan, M.M. Sondhi, and S.L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.

[3] Rainer Martin and J. Altenhöner, "Coupled adaptive filters for acoustic echo control and noise reduction," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 1995, pp. 3043–3046.

[4] Stefan Gustafsson, Rainer Martin, and Peter Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing, Elsevier*, vol. 64, no. 1, pp. 21–32, January 1998.

[5] Gerald Enzner, Dirk Mauler, and Peter Vary, "Realtime performance of acoustic echo canceler and postfilter for residual echo suppression in the car environment," in *Proc. Deutsche Jahrestagung für Akustik (DAGA)*, March 2004.