ON THE PROBLEM OF ACOUSTIC ECHO CONTROL IN CELLULAR NETWORKS

Gerald Enzner, Hauke Krüger, and Peter Vary

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

ABSTRACT

If acoustic echo cancelers are implemented in cellular networks, e.g., in base stations, the echo attenuation is severely degraded by the nonlinearity and/or unpredictability of the effective echo path. In this contribution, we consider a network model (or rather echo path model) which takes the statistical behavior of speech encoders into account if contained in the echo path. Using this model, we show that the optimum filter structure of network acoustic echo controllers can be derived from the theory of combined acoustic echo and noise control for hands-free telephones.

1. INTRODUCTION

In mobile communication systems with significant transmission delay, e.g. GSM and UMTS, a hands-free voice interface imposes the need for acoustic echo control (AEC). In the classical setup, the AEC is located in the mobile phone to be as close as possible to the acoustic interface. This strategy has the advantage that the electroacoustic coupling between sending and receiving direction of the system obeys to a nearly linear filter model which facilitates the reduction of acoustic echo by linear adaptive filters [1]. Unfortunately, the low-complexity constraint restricts the choice of adaptive algorithms and limits the AEC performance.

Alternatively, the AEC could be realized in the cellular network. Here, the resources required for adaptive filtering (processing power and memory) are no longer a limiting factor of the AEC performance. Furthermore, network operators are enabled to control the acoustic echo of hands-free telephones (in addition to line echo control). However, if the network AEC is based on the classical system identification approach, the achievable echo attenuation will suffer from the unpredictable behavior of the *effective echo path* from the network to the mobile and back to the network. The unpredictability of the echo path can be caused, among other reasons, by lossy speech encoders or transmission errors.

In the literature, several sub-optimal filter structures for network AECs were treated. In [2], a weighted spectral subtraction was suggested to attenuate the echo, but an echo canceler was not utilized at all. In [3],[4], an acoustic echo canceler combined with a nonlinear processor (e.g. in the form of a center clipper) is recommended. In [5], the combination of acoustic echo cancellation and decorrelation of the residual echo was proposed.

In contrast to the previous work, we consider a general two-filter structure (comprising acoustic echo canceler and statistical postfilter) to improve AEC performance in the network. The remainder of the paper is organized as follows: After a brief analysis of network architectures in Section 2, we develop a statistical network model suitable for acoustic echo control in Section 3. On the basis of this model, Section 4 derives the optimum filters of an AEC unit that is located in the cellular network. Section 5 presents simulation results for the case of speech transmission in GSM.

2. CELLULAR NETWORK ARCHITECTURES

2.1. Tandem-Free Operation (TFO) Network

In the TFO network as shown by Figure 1, the same speech codec has to be available in both mobiles. In the case without AEC and assuming idealized channel conditions, the network serves as a lossless bridge between both sides of the communication system. An AEC in the coded domain maps the coded input signal to a coded and echo compensated output signal. The design of this procedure can be seen as a joint optimization of coding elements and acoustic echo control features (e.g. linear prediction, quantization, and adaptive filtering). An analytical treatment of this issue will become highly nonlinear. Moreover, since the standardization of ideal TFO networks is still in progress [6], we do not further consider AEC design in the coded domain in this paper.

2.2. Transcoding Network

We turn to the prevalent situation in cellular networks. As shown in Figure 2, mobile A and B apply different speech codecs and the network performs the translation by decoding to a PCM signal and re-encoding. Compared to TFO, this strategy involves a loss in signal quality, but in most cases there is no other known way to realize the conversion between different codec formats.

From the viewpoint of acoustic echo control, the advantage is that reconstructed waveforms are available in the network. Typically, we have two independent devices AEC-A and AEC-B which are responsible for the attenuation of the echo of mobile A and B, respectively. Unfortunately, the presence of two speech encoders in the effective echo path (one in sending and one in receiving direction of the AEC) deteriorates the performance of the system identification approach significantly. As speech encoders are usually based on time-variant filtering and quantization, they are often referred to as *nonlinearities* in the echo path [3, 4].

2.3. Transcoding Network with Reduced Nonlinearity

The aforementioned echo path nonlinearities can be at least partly avoided by the modified network in Figure 3. Here, the nonlinear encoders A and B in the network have been moved “out of sight” of AEC-A and AEC-B. This modification is possible if we assume that no signal processing (e.g. filtering, loss control) is carried out in receiving direction of the AECs. Clearly, this will require an additional speech decoder in each of the AECs. We further note that it would not provide an additional benefit to move the network decoders “out of sight” of the AECs, since the echo path nonlinearity (or rather unpredictability) is merely due to the encoders.

The latter network will serve as the basis for our AEC design in Sections 3 and 4. As the network structure is still symmetric, the network AECs can be realized independently for mobile A and B.

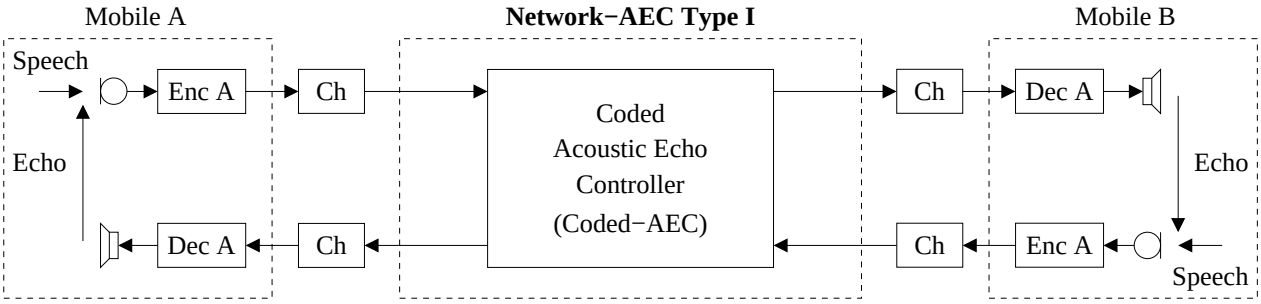

Fig. 1. TFO network with AEC in the coded domain. Legend: Enc = Encoder, Dec = Decoder, Ch = Channel.

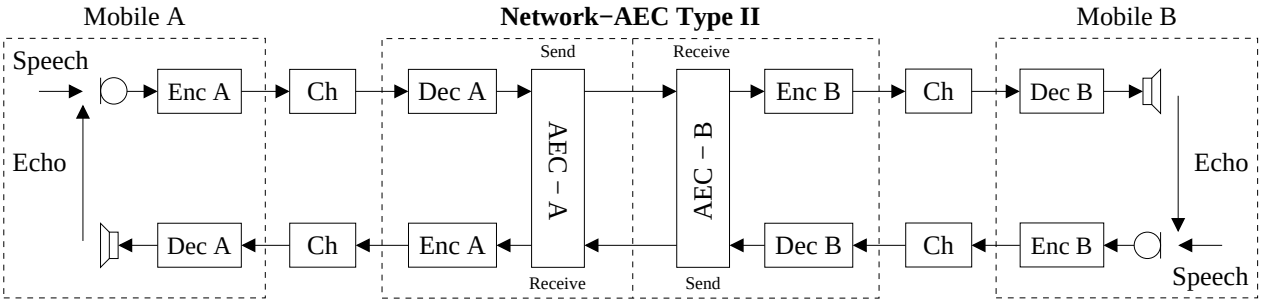

Fig. 2. Transcoding network with AEC in the waveform domain. Echo path nonlinearity in sending and receiving direction of the AEC.

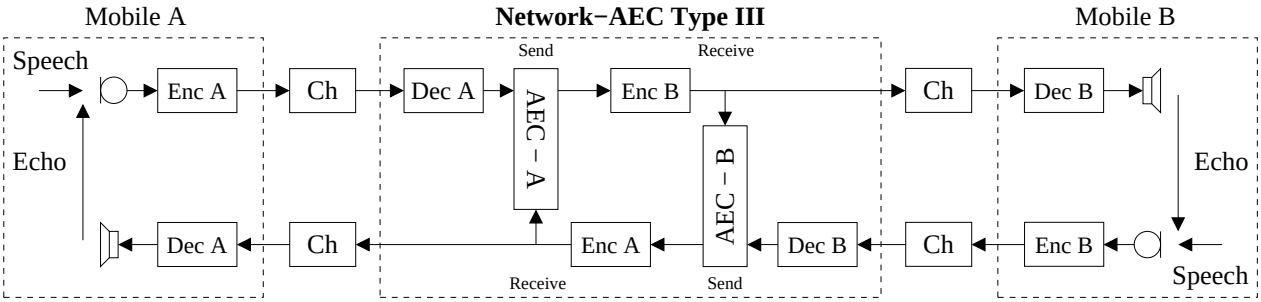

Fig. 3. Alternative transcoding network with echo path nonlinearity only in sending direction of the AEC.

3. DERIVATION OF A SYSTEM MODEL FOR ACOUSTIC ECHO CONTROL

3.1. Model Encoder and Decoder

An analytical treatment of network AEC is facilitated by the linear predictive transmission model in Figure 4. The encoder consists of an analysis filter $A_y(z)$ and a scalar quantization Q of the decorrelated residual $y_r(i)$. The corresponding decoder $A_y^{-1}(z)$ uses the quantized signal $y_{r,Q}(i)$ to determine the output $y_Q(i) \approx y(i)$.

The quantization introduces a distortion $\Delta_y(i)$ which can be modeled (for high rate) as a statistically independent additive white noise as shown by Figure 5. Let us further assume that the quantization noise power $\sigma_\Delta^2 = \mathcal{E}\{\Delta_y^2(i)\}$ follows the (short-term) power $\sigma_{y_r}^2 = \mathcal{E}\{y_r^2(i)\}$ of the residual $y_r(i)$, i.e.,

$$\sigma_\Delta^2 = K \cdot \sigma_{y_r}^2 \ . \tag{1}$$

This model of speech transmission shall *not* be considered as an exact reproduction of the functionality of standardized low bit-rate speech encoders for cellular networks (e.g. *ETSI GSM 06.60* enhanced full-rate, *GSM 06.20* half-rate, *GSM 06.90* adaptive multirate, *ITU-T G.729*). However, we assume that the model serves at least as a first approximation to describe the statistical behavior of speech codecs in the effective echo path.

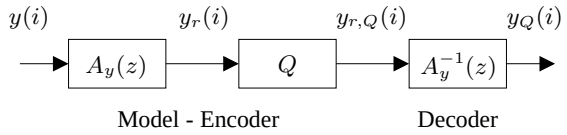

Fig. 4. Model of linear predictive speech transmission.

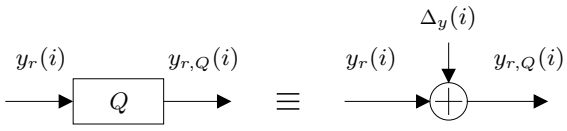

Fig. 5. Additive white noise model of the quantizer Q.

3.2. System Model for Acoustic Echo Control

We now consider that part of the transcoding network in Figure 3 which is relevant for the design of AEC-A. In Figure 6, encoder A and decoder A have been replaced by the model-encoder and decoder as introduced in the previous section. The microphone signal $y(i)$ contains near-end speech $s(i)$ and acoustic echo $d(i)$. The acoustic echo path is described by the transfer function $W(z)$ and has the decoded signal $x_Q(i)$ received from the far speaker as input. The output signal $\widehat{s}(i)$ of the AEC unit shall approximate an echo-free transmission of the near-end speech $s(i)$.

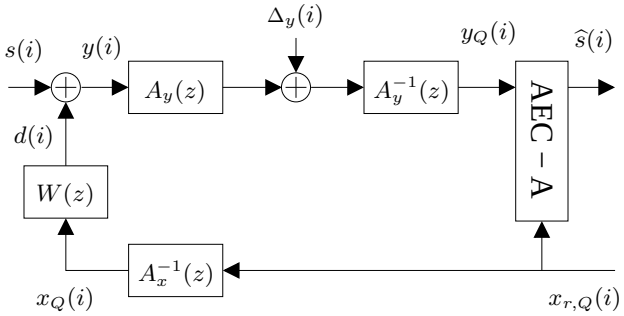

Fig. 6. Simplified system model for acoustic echo control.

The system model in Figure 6 assumes a distortionless and delayless transmission channel. In practice, the channel delay has to be compensated by the AEC unit. Any kind of linear distortion by the transmission can be considered as part of the *effective echo path* from the network to the mobile and back to the network. However, if the channel is responsible for nonlinear distortions (e.g. bit errors or packet loss), then this requires a specific treatment by the AEC. A more precise analysis of the impact of the transmission channel is beyond the scope of this paper.

3.3. An Optimization Criterion

In order to define an optimization criterion for the AEC, we have to consider that the decoder $A_y^{-1}(z)$ creates an additional *fictitious echo* $d_\Delta(i)$ from the independent quantization noise $\Delta_y(i)$. We may write the AEC input $y_Q(i)$ as:

$$y_Q(i) = s(i) + s_\Delta(i) + d(i) + d_\Delta(i)\ . \tag{2}$$

The reconstructed speech signal $\widetilde{s}(i) = s(i) + s_\Delta(i)$ appears at the AEC input together with the reconstructed echo signal $\widetilde{d}(i) = d(i) + d_\Delta(i)$. Unfortunately, the disturbance $\widetilde{d}(i)$ cannot be removed entirely by an acoustic echo canceler since the fictitious echo $d_\Delta(i)$ is not correlated with the far speaker signal $x_{r,Q}(i)$. To achieve a suppression of the fictitious echo in the output signal of the AEC, we formulate the optimality of the AEC by the following minimum mean-square error (MMSE) criterion:

$$\mathcal{E}\big\{\big(\widetilde{s}(i) - \widehat{s}(i)\big)^2\big\} \ \rightarrow\ \min\ . \tag{3}$$

In the next section, we will define $s_\Delta(i)$ and $d_\Delta(i)$ more precisely and derive optimum filters for the AEC.

4. OPTIMUM FILTERS FOR THE AEC UNIT

The analysis is carried out in the frequency-domain. The involved signals are written as, e.g., $Y(\Omega) = \mathcal{F}\{y(i)\}$, where $\mathcal{F}$ denotes the Fourier transform, while transfer functions are written as, e.g., $W(\Omega) = W(z)\,|_{z=e^{j\Omega}}$.

4.1. Statistical Analysis of the System Model

According to Figure 6, the AEC input $Y_Q(\Omega)$ can be expressed in terms of the microphone signal $Y(\Omega)$:

$$\begin{aligned} Y_Q(\Omega) &= \big(Y(\Omega)A_y(\Omega) + \Delta_y(\Omega)\big)A_y^{-1}(\Omega) \\ &= S(\Omega) + D(\Omega) + \Delta_y(\Omega)A_y^{-1}(\Omega)\ . \end{aligned} \tag{4}$$

Comparing this relation to (2), it turns out that the spectrum $Y_\Delta(\Omega) = \Delta_y(\Omega)A_y^{-1}(\Omega)$ corresponds to the effective quantization noise $s_\Delta(i) + d_\Delta(i)$ at the AEC input. The power spectral density (PSD) of $Y_\Delta(\Omega)$ is given by

$$\begin{aligned} \Phi_{YY,\Delta}(\Omega) &= \frac{\sigma_\Delta^2}{|A_y(\Omega)|^2} \\ &\approx \frac{\sigma_\Delta^2}{\sigma_{y_r}^2}\Phi_{yy}(\Omega) \\ &= K\big(\Phi_{ss}(\Omega) + \Phi_{dd}(\Omega)\big)\ . \end{aligned} \tag{5}$$

Here, $\Phi_{yy}(\Omega) \approx \sigma_{y_r}^2/|A_y(\Omega)|^2$ is the PSD of the microphone signal $y(i) = s(i) + d(i)$, where the approximation reflects that the sum of two speech signals (autoregressive processes) cannot be described exactly as an autoregressive process. The last equality has been obtained invoking (1). $\Phi_{ss}(\Omega)$ and $\Phi_{dd}(\Omega)$ are the PSDs of independent near speech and acoustic echo, respectively.

Based on (5), we can now associate the PSDs $K\Phi_{ss}(\Omega)$ and $K\Phi_{dd}(\Omega)$ with the previously defined components $s_\Delta(i)$ and $d_\Delta(i)$ of the effective quantization noise. This can be justified by the assumption that the fictitious echo $d_\Delta(i)$ should have the same spectral shape as the acoustic echo $d(i)$. We recall, however, that in contrast to the acoustic echo, the fictitious echo is independent of the far speaker signal $x_{r,Q}(i)$.

4.2. Optimum Filtering

Based on (2) and using the linear relation between the acoustic echo $D(\Omega)$ and the received signal $X_{r,Q}(\Omega)$, i.e. $D(\Omega) = W(\Omega)A_x^{-1}(\Omega)X_{r,Q}(\Omega)$, we rewrite the AEC input $Y_Q(\Omega)$ in the frequency-domain as:

$$Y_Q(\Omega) = \widetilde{S}(\Omega) + W(\Omega)A_x^{-1}(\Omega)X_{r,Q}(\Omega) + D_\Delta(\Omega)\ . \tag{6}$$

The spectrum $\widetilde{S}(\Omega)$ corresponds to the reconstructed speech $\widetilde{s}(i)$ for which we can determine the PSD $\Phi_{\widetilde{s}\widetilde{s}}(\Omega) = (1+K)\Phi_{ss}(\Omega)$. The product $W(\Omega)A_x^{-1}(\Omega)$ stands for the serial concatenation of the decoder $A_x^{-1}(\Omega)$ and the acoustic echo path $W(\Omega)$. The spectrum $D_\Delta(\Omega)$ corresponds to the fictitious echo $d_\Delta(i)$ that has been associated with the PSD $K\Phi_{dd}(\Omega)$.

From (6), we observe that the optimization according to (3) can be seen as a form of combined acoustic echo and noise control as it was treated in [7]. The solution comprises an acoustic echo canceler $W_1(\Omega)$ and a statistical postfilter $W_2(\Omega)$ as shown in Figure 7, i.e., the spectrum $\widehat{S}(\Omega)$ of the AEC output $\widehat{s}(i)$ can be expressed by the following formula:

$$\widehat{S}(\Omega) = \big(Y_Q(\Omega) - W_1(\Omega)X_{r,Q}(\Omega)\big)W_2(\Omega)\ . \tag{7}$$

Based on the approach in [7], the optimum filters $W_1(\Omega)$ and $W_2(\Omega)$ in the frequency-domain can be determined as:

$$W_1(\Omega) = W(\Omega)A_x^{-1}(\Omega) \tag{8}$$

$$W_2(\Omega) = \frac{\Phi_{\widetilde{s}\widetilde{s}}(\Omega)}{\Phi_{\widetilde{s}\widetilde{s}}(\Omega) + K\Phi_{dd}(\Omega)}\ . \tag{9}$$

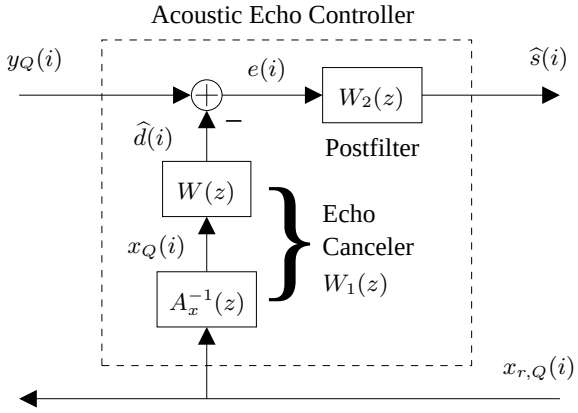

Fig. 7. Network AEC according to the MMSE criterion.

If the quantization SNR is $10\,\mathrm{dB}$, i.e., $K = 0.1$, the echo canceler $W_1(z)$ will attain a maximum attenuation of the reconstructed echo $\widetilde{d}(i)$ of about $10\,\mathrm{dB}$ (limited by the unpredictability of the fictitious echo). In the case of acoustic echo control for hands-free telephones, the echo canceler performance is similar. There, statistical postfilters $W_2(z)$ have been successfully used to reduce the residual echo after the acoustic echo canceler [8].

4.3. Implementation Issues

A realization of the AEC requires the identification of the acoustic echo path $W(z)$. Equation (6) shows that the reconstructed speech $\widetilde{S}(\Omega)$ and the ficticious echo $D_\Delta(\Omega)$ act as independent observation noises. Therefore, standard adaptive filters, e.g., NLMS, RLS, APA, or FDAF [9], will accomplish a unique system identification using the decoded signal $x_Q(i)$ as the reference input and $e(i)$, see Figure 7, as the error signal. The decoder $A_x^{-1}(z)$ which completes the echo canceler $W_1(z)$ is available in the network.

The postfilter $W_2(z)$ relies on the inverse quantization SNR K. This parameter characterizes the encoder in the mobile and is known in the network. The PSD $\Phi_{dd}(\Omega)$ of the echo signal can be calculated from the estimated echo $\widehat{d}(i) = d(i)$ that is available in the AEC. As the error signal $e(i)$ contains only reconstructed speech and ficticious echo, we have the simple relation $\Phi_{ee}(\Omega) = \Phi_{\widetilde{s}\widetilde{s}}(\Omega) + K\Phi_{dd}(\Omega)$ and thus the PSD $\Phi_{\widetilde{s}\widetilde{s}}(\Omega)$ can be obtained using the spectral subtraction technique [1].

5. SIMULATION RESULTS

We consider the simulation setup in Figure 8 to evaluate the practical relevance of the proposed theory. For the sake of simplicity, the decoders $A_x^{-1}(z)$ in the AEC and in the mobile have been merged and are represented by the decoded input signal $x_Q(i)$. We use real speech input and a measured acoustic echo path $W(z)$ with 500 coefficients at $8\,\mathrm{kHz}$ sampling frequency. The speech transmission in the effective echo path complies with the *GSM 06.60* specification. The replica $\widehat{W}(z)$ of the acoustic echo path is estimated by the adaptive algorithm in [10]. As we measured an SNR of $8\,\mathrm{dB}$ for the GSM transmission, i.e., from $y(i)$ to $y_Q(i)$, we choose $K = 0.16$ to implement the postfilter $W_2(z)$.

During remote single talk, i.e., $s(i) \approx 0$, the echo return loss enhancement (ERLE) [1] is a suitable performance measure of the AEC. We measured an ERLE of about $10\,\mathrm{dB}$ by the echo canceler and a total ERLE of about $25\,\mathrm{dB}$ by echo canceler and postfilter. We found that $45\,\mathrm{dB}$ ERLE as required in GSM can be achieved by a larger value of K or by a center clipper after the postfilter.

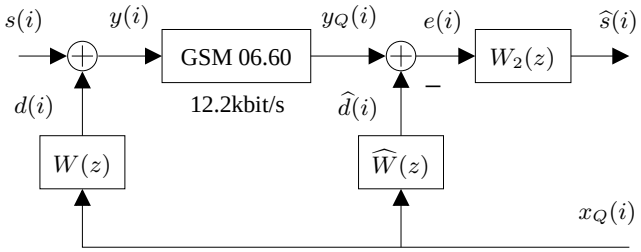

Fig. 8. Simulation setup using the GSM enhanced full-rate codec.

In the case of double talk, i.e., $\mathrm{SNR}_y = \sigma_s^2/\sigma_{y-s}^2 = 5\,\mathrm{dB}$, the GSM transmission further reduces the quality to $\mathrm{SNR}_{y_Q} = 3.5\,\mathrm{dB}$. The echo canceler increases the quality to $\mathrm{SNR}_e = 6.5\,\mathrm{dB}$ and the statistical postfilter restores an $\mathrm{SNR}_{\widetilde{s}} = 7.5\,\mathrm{dB}$ which is already close to an echo-free GSM transmission ($\mathrm{SNR}_{\widetilde{s}} = 8\,\mathrm{dB}$).

6. CONCLUSION

A linear predictive network model with open-loop quantization has been suggested for AEC design in cellular networks. On the basis of the model, we have shown that the optimum filter structure of a network AEC consists of an acoustic echo canceler and a statistical postfilter, the latter to suppress fictitious echo due to the quantization noise. Simulations confirmed the principal suitability of the statistical approach and therefore the investigation of more sophisticated network models is encouraged.

7. REFERENCES

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

[2] Xiaojian Lu and Benoit Champagne, "Acoustic echo cancellation over a nonlinear channel," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control*, September 2001.

[3] ITU-T Recommendation G.161, *Interaction Aspects of Signal Processing Network Equipment*, June 2002.

[4] ITU-T Recommendation G.168, *Digital Network Echo Cancellers*, June 2002.

[5] H. Gnaba, M. Turki-Hadj Alouane, M. Jaidane-Saidane, and P. Scalart, "Introduction of the CELP structure of the GSM coder in the acoustic echo canceller for the GSM network," in *Proc. of EUROSPEECH*, September 2003.

[6] ETSI 3GPP TS 28.062, *Inband Tandem-Free Operation (TFO) of Speech Codecs; Service Description, Stage 3*, December 2004.

[7] 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.

[8] 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. of Deutsche Jahrestagung für Akustik (DAGA)*, March 2004.

[9] Simon Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, 4th edition, 2002.

[10] Gerald Enzner and Peter Vary, "Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control*, September 2003, pp. 43–46.