# Concept for Combined Reduction of Acoustic Echo and Coding Noise in Digital Cellular Networks

*Gerald Enzner, Hauke Krüger, and Peter Vary*

Institut für Nachrichtengeräte und Datenverarbeitung, RWTH Aachen, D-52056 Aachen

{enzner,krueger,vary}@ind.rwth-aachen.de

## Abstract

Acoustic echo cancelers in base stations of cellular networks suffer from the unpredictability of the effective echo path. A main reason of this unpredictability is the quantization in speech encoders. In order to resolve this problem, we suggest a design of network acoustic echo controllers on the basis of a statistical echo path model. In contrast to earlier work, we consider an optimization criterion which aims at the combined reduction of acoustic echo and coding noise.

## 1.   Introduction

The acoustic echo control (AEC) components of cellular networks are usually implemented in the mobile phone. On the one hand, this has the advantage that the AEC can rely on a nearly linear echo path model which facilitates echo reduction by linear adaptive filtering, e.g., [1], [2]. On the other hand, the low-complexity constraint in the mobile phone often limits the achievable AEC performance.

If the AEC is realized in the fixed part of the cellular network (literature references in the next section), then the resources required for adaptive filtering, i.e., 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. Unfortunately, such a *network AEC* suffers from the unpredictability of the *effective echo path* from the network to the mobile and back to the network.

In contrast to a previous paper on this issue [3], we will consider a different optimization criterion for network AEC design. Instead of the *coded and transmitted near-end speech*, we will now choose the *clean speech signal* as the target signal of an MMSE optimization of the network AEC. This modification results in a more intuitive formulation of the optimality

and it allows for a simplified derivation of the optimum solution.

In addition to that, the new optimization criterion leads to a naturally combined reduction of the aforementioned echo *and* the coding noise (quantization noise) due to digital speech transmission. However, since the effect of echo in a communication is clearly more disturbing than the (perceptually masked) coding noise, our signal processing solution mainly focuses on the echo reduction problem in the cellular network.

## 2.   Previous Work

In literature, several sub-optimal filter structures for network AEC were treated. In [4], a weighted spectral subtraction was suggested in order to attenuate the echo, but an echo canceler was not utilized at all. In [5],[6], an acoustic echo canceler combined with a nonlinear post-processor (e.g. in the form of a center clipper) is recommended. In [7], a combination of acoustic echo cancellation and residual echo decorrelation was described, while in [8], the same authors proposed echo cancellation combined with postfiltering for residual echo suppression. In [9] and [10], the postfilter was modified according to psychoacoustically (perceptually) motivated design criteria.

In [3], we optimized a general two-filter structure (comprising acoustic echo canceler and statistical postfilter) according to the MMSE criterion. This optimization was performed on the basis of a new statistical network model for acoustic echo control, i.e., the effective echo path was modeled as an unpredictability (rather than a deterministic nonlinearity, e.g., [11]). This approach has been justified by the fact that our major source of echo path nonlinearity is the quantization in the speech encoder. It was demonstrated by simulations in a GSM environment that the resulting AEC unit delivers an acceptable output signal quality in single talk and double talk situations.

For the sake of completeness, it should be mentioned that the individual problem of coding noise reduction (by postfiltering in receivers) has a quite long tradition in literature, e.g., [12], [13], [14], [15] and references therein. In the following considerations, the coding noise reduction is a desirable byproduct of the design of the network AEC.

## 3. System Model for AEC Design

Of course, the optimal design of a network AEC depends on the network architecture. Different network types were analyzed in [3] and a suitable transcoding architecture as shown in Figure 1 was identified. Two independent devices AEC-A and AEC-B are responsible for the attenuation of echo from mobile A and B, respectively. Note that the downlink speech encoders are "out of sight" of the AEC in order to avoid a downlink echo path nonlinearity. Clearly, that will require an additional decoder in each of the AECs.

Assuming a lossless and delayless radio channel, it was further shown in [3] that the left part of Figure 1 – being relevant for the design of AEC-A – can be described by the compact system model in Figure 2. The microphone signal $y(i)$ contains near-end speech $s(i)$ and acoustic echo $d(i)$. The acoustic echo path is given by the transfer function $W(z)$ and has the decoded signal $x_Q(i)$ received from the far speaker as input. Furthermore, Enc A and Dec A were

replaced by a linear predictive model-encoder and decoder: $A_y(z)$ and $A_y^{-1}(z)$ denote analysis and synthesis filtering, respectively, and the quantization of the residual $y_r(i)$ inbetween introduces a statistically independent additive white noise distortion $\Delta_y(i)$. To take level-adaptive quantization into account, we assume that the quantization noise power $\sigma_\Delta^2$ follows the (short-term) power $\sigma_{y_r}^2$ of the residual, i.e.,

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

The AEC unit has to be designed such that the output signal $\widehat{s}(i)$ approximates an echo-free transmission of the near-end speech $s(i)$.
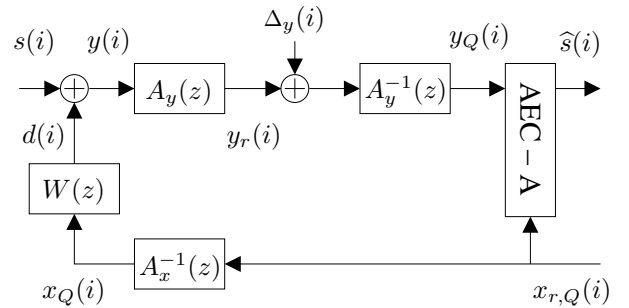


Fig. 2: Linear predictive system model for network AEC design.

## 4. The Optimization Criterion

Depending on the conversation mode (near-end single talk, far-end single talk, or double talk), the quantization noise $\Delta_y(i)$ will appear after decoding in the form of coding noise, *fictitious echo*, or both. Independent of the specific situation, we write the AEC input $y_Q(i)$ as

$$y_Q(i) = s(i) + d(i) + q(i) \ , \tag{2}$$

where $q(i)$ is the quantization noise $\Delta_y(i)$ processed by the decoder $A_y^{-1}(z)$. Due to the independence of $\Delta_y(i)$, the distortion $q(i)$ cannot be removed by an acoustic echo canceler (even if it is perceived as echo). In order to achieve a suppression of the distortion $q(i)$ at the output of the AEC, we formulate the optimality of the AEC by the following minimum mean-square error (MMSE) criterion:

$$\mathcal{E}\big\{\big(s(i) - \widehat{s}(i)\big)^2\big\} \to \min \ . \tag{3}$$
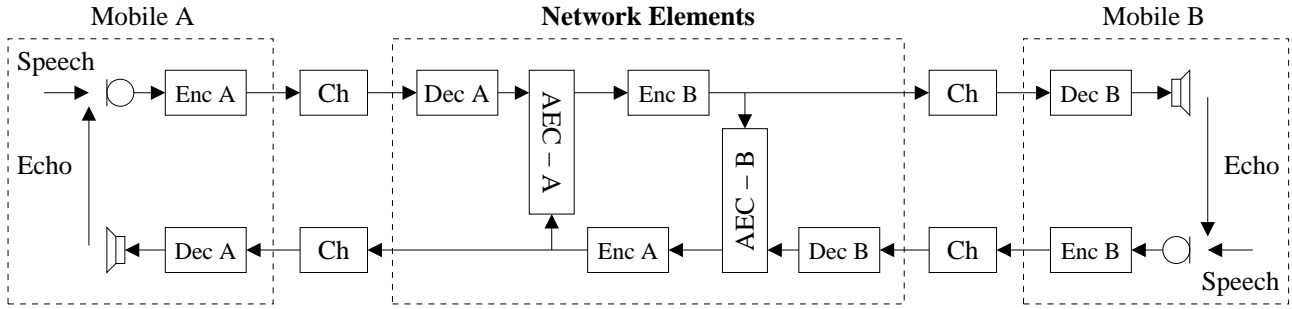
Fig. 1: Transcoding network with uplink echo path nonlinearities (unpredictabilities). The cellular network may utilize different codecs A and B to transmit over the air interface of mobile A and B. *Legend:* Enc = Speech Encoder, Dec = Speech Decoder, Ch = Radio Channel.

## 5. Derivation of Optimum AEC Filters

The solution to (3) is now approached in the frequency domain. 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., $W(\Omega) = W(z)\mid_{z=e^{j\Omega}}$.

### 5.1 Statistical Analysis of the System Model

According to Figure 2, the AEC input $Y_Q(\Omega)$ can be expressed in terms of the microphone signal $Y(\Omega)$ or, alternatively, in terms of the desired speech signal $S(\Omega)$:

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

Comparing this relation to (2), it turns out that the spectrum $Q(\Omega) = \Delta_y(\Omega)A_y^{-1}(\Omega)$ corresponds to the processed quantization noise $q(i)$ at the AEC input. The power spectral density (PSD) of $Q(\Omega)$ is thus given by:

$$\Phi_{qq}(\Omega) = \frac{\sigma_\Delta^2}{|A_y(\Omega)|^2} \tag{5a}$$

$$= \frac{\sigma_\Delta^2}{\sigma_{y_r}^2}\Phi_{yy}(\Omega) \tag{5b}$$

$$= K\Phi_{yy}(\Omega) \ . \tag{5c}$$

Here, $\Phi_{yy}(\Omega) = \sigma_{y_r}^2/|A_y(\Omega)|^2$ is the PSD of the microphone signal $y(i)$, assuming that the sum $y(i) = s(i) + d(i)$ can be described with sufficient accuracy as an autoregressive process. The last equality is obtained invoking (1).

### 5.2 Optimum Filtering

Using the linear relation between the acoustic echo $D(\Omega)$ and the AEC input signal $X_{r,Q}(\Omega)$, cf. Figure 2, we can easily rewrite (4) as

$$Y_Q(\Omega) = S(\Omega) + \widetilde{W}(\Omega)X_{r,Q}(\Omega) + Q(\Omega) \ , \ (6)$$

where the abbreviation $\widetilde{W}(\Omega) = W(\Omega)A_x^{-1}(\Omega)$ stands for the serial concatenation of the decoder $A_x^{-1}(\Omega)$ in the mobile, i.e., Dec A in Figure 1, and the acoustic echo path $W(\Omega)$.

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 [16]. The solution to the optimization problem therefore comprises an acoustic echo canceler $W_1(\Omega)$ and a statistical postfilter $W_2(\Omega)$ as shown in Figure 3, i.e., the spectrum $\widehat{S}(\Omega)$ of the AEC output $\widehat{s}(i)$ can be expressed by the following formula:

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

Based on the analytic approach in [16] and using the "noise" PSD in (5c), the optimum filters $W_1(\Omega)$ and $W_2(\Omega)$ of a network AEC in the frequency domain can be jointly determined as:

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

$$W_2(\Omega) = \frac{\Phi_{ss}(\Omega)}{\Phi_{ss}(\Omega) + K\Phi_{yy}(\Omega)} \ . \tag{9}$$

It should be noted that the echo attenuation by adaptive echo cancelers generally depends on the degree of unpredictability of the effective
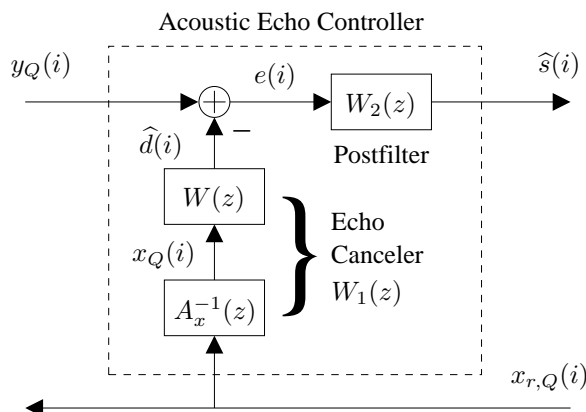
Fig. 3: Network AEC according to MMSE.

echo path. Assuming a time-invariant acoustic impulse response $W(z)$, then the quantization SNR $K^{-1}$ is a direct measure for this unpredictability. For example, if $K^{-1} = 10$, a maximum echo attenuation of about $10\,\text{dB}$ will be attained by the echo canceler. In this case, and if the input echo level before echo cancellation is not larger than the near-end speech level, a postfilter can successfully suppress the residual echo and preserve the near-end speech. These properties have been demonstrated by simulation results, e.g., [3], [17], [18].

## References

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

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

[3] Gerald Enzner and Peter Vary, "On the problem of acoustic echo control in cellular networks," in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control*, September 2005.

[4] 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.

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

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

[7] 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.

[8] H. Gnaba, P. Scalart, M. Turki-Hadj Alouane, and M. Jaidane-Saidane, "Combined acoustic echo canceller for the GSM network," in *Proc. of Euro. Signal Processing Conf. (EUSIPCO)*, September 2002.

[9] Xiaojian Lu and Benoit Champagne, "A centralized acoustic echo canceller exploiting masking properties of the human ear," in *Proc. of IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, April 2003.

[10] H. Gnaba-Daassi, P. Scalart, and C. Marro, "A centralized acoustic echo canceller based on perceptual properties," in *Proc. of Euro. Signal Processing Conf. (EUSIPCO)*, September 2005.

[11] Alexander Stenger, Lutz Trautmann, and Rudolf Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive volterra filters," in *Proc. of IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Phoenix (Arizona), March 1999, pp. 877–880.

[12] N.S. Jayant, "ADPCM coding of speech with backward-adaptive algorithms for noise feedback and postfiltering," in *Proc. of IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, April 1987.

[13] Juin-Hwey Chen and Allen Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, pp. 59–71, January 1995.

[14] Azhar Mustapha and Suat Yeldener, "An adaptive post-filtering technique based on a least squares approach," in *IEEE Workshop on Speech Coding*, Haikko Manor, Porvoo (Finland), June 1999.

[15] Wei Chen, Peter Kabal, and Turaj Z. Shabestary, "Perceptual postfilter estimation for low bit rate speech coders using Gaussian mixture models," in *Proc. of INTERSPEECH*, Lisbon (Portugal), September 2005.

[16] 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.

[17] 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.

[18] Gerald Enzner and Peter Vary, "New insights into the statistical signal model and the performance bounds of acoustic echo control," in *Proc. of Euro. Signal Processing Conf. (EUSIPCO)*, Antalya (Turkey), September 2005.