

Energy Efficiency of Network-Based Acoustic Echo Control in Mobile Radio

Matthias Pawig, Peter Vary

Institute of Communication Systems and Data Processing, RWTH Aachen University, 52056 Aachen
Email: {pawig,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

In this contribution we investigate the energy efficiency of network-based acoustic echo control (AEC). Usually, AEC is implemented in the mobile terminal. Due to the limited resources of the mobile terminal in terms of the battery, computational capacity and memory, the complexity of the AEC algorithm has to be limited. In this paper we will investigate the idea to move the AEC processing from the mobile terminal to a network-based processing unit which allows to use more sophisticated algorithms. As this might require the use of improved speech codecs with higher bit rate, we will evaluate the trade-off between increased power for the radio transmission with additional bit rate and the reduced power for signal processing at the mobile terminal. This implies to take into consideration the attenuation of the radio channel according to, e.g., the Okumura-Hata path loss model.

1 Introduction

Due to the small dimensions of mobile telephones and thus a non-negligible acoustic echo path from loudspeaker to microphones, especially in the hands-free mode, an acoustic echo severely reduces the perceived quality of the terminal call for the far end user. Users expect reliable acoustic echo attenuation with full duplex capability, which is a particularly challenging signal processing task.

Usually a linear echo path is assumed and an adaptive echo canceller is applied within the mobile terminal prior to transmission over the network, e.g., [1–4]. Sophisticated control mechanisms are available to handle double-talk and to achieve fast and robust adaptation of the echo canceller coefficients in time-varying and noisy environments [5–7].

In practice, state-of-the-art algorithms for small terminal equipment have to take the constraints of limited computational complexity and low memory consumption into account. In order to ease these constraints and to reduce the power and thus energy consumption at the mobile terminal, some of the signal processing complexity can be moved to a network-based unit (NBU), e.g., at the entry point into the provider network like the transcoding unit in GSM (TRAU) or at a management unit for audio conferencing. Additionally, this approach gives control over the speech transmission quality in the hands of the network operator. In previous work [8], we have shown that a relocation of acoustic noise reduction in such a system is possible without significant changes to the algorithm.

Unfortunately, realization of the AEC in the NBU is not straightforward, since the achievable echo attenuation will suffer from the behavior of the *effective echo path* from the network to the mobile and back to the network. This path includes lossy speech encoders and radio transmission with error correction. Eventually, error concealment has to be applied in case of residual bit errors. This limits the effectiveness of network based linear adaptive echo control. Several structures for network AEC were treated in, e.g., [9–11].

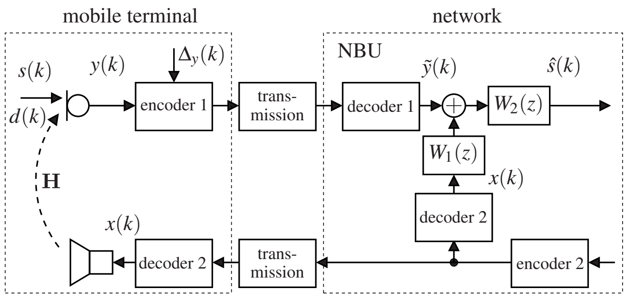

Figure 1: System overview.

In this paper, we employ the structure introduced in [11]. This approach introduces an additional postfilter, which in the proposed system is included in the postfilter of the frequency domain Kalman filter algorithm in [7].

We show by simulation that the performance of network-based AEC significantly improves by using higher bit rate codecs, which reduces the non-linear distortions of the effective echo path. However, such an increase in bit rate also increases the required signal power for transmission. We will evaluate and compare the power budget of a system with a) signal processing at the mobile terminal and b) signal processing at a network-based unit after transmission with higher bit rate taking the radio path attenuation into account according to the Okumura-Hata path loss model [12]. The resulting equations for this trade-off can be used to analyze the effectiveness of any kind of network-based processing.

This paper is structured as follows: In Section 2, we will shortly describe the employed network-based acoustic echo control system and demonstrate the increased performance when using a higher bit rate transmission. In Section 3 the framework for evaluating the overall system power budget is introduced and an exemplary calculation for a system with network based processing is shown. Section 4 will give some conclusions and an outlook to further investigations.

2 Network-Based AEC

An overview of the proposed system with network-based processing is shown in Fig.1. At the microphone, the near end speech signal $s(k)$ and the echo signal $d(k)$ are captured as a mixture $y(k) = s(k) + d(k)$. We propose transmitting the mixture $y(k)$ using the uplink speech codec 1 and performing the signal processing on the transmitted microphone signal $\tilde{y}(k)$. Since moderate transmission errors do not lead to a significant reduction of AEC performance, in this paper error-free transmission is assumed. In the down-link path from the network to the mobile terminal the loudspeaker signal $x(k)$ can be provided as input to the AEC using the downlink speech decoder 2. Uplink speech codec 1 and speech codec 2 might be different.

The echo control structure used in this paper was proposed in [11], where the transmission is modeled by a combination of linear prediction analysis and synthesis filters $A_{x,y}(z)$ and $A^{-1}_{x,y}(z)$ in addition to a quantizer which introduces the speech encoder distortion $\Delta_y(k)$. Although this model does not fully represent all of the influences of low bit rate speech encoders, it can be used as a first approximation of the statistical behavior of the transmission in the effective echo path. The AEC system is composed of an adaptive filter W_1 and a postfilter W_2 as in [7].

The microphone input to the AEC can be written as

$$\tilde{y}(k) = s(k) + s_\Delta(k) + d(k) + d_\Delta(k), \quad (1)$$

where the signal $s_\Delta(k)$ and $d_\Delta(k)$ denote the specific distortions caused by the near end speech signal $s(k)$ and the echo signal $d(k)$, respectively. The task of echo control is to remove the reconstructed echo signal $\tilde{d}(k) = d(k) + d_\Delta(k)$. Unfortunately, the term $d_\Delta(k)$ is not correlated with the far end speaker signal $x(k)$ and thus cannot be compensated by an adaptive filter alone.

The AEC postfilter $W_2(\Omega)$ is taken from [7] as

$$W_2(l,i) = \frac{\Phi_{SS}(l,i)}{\Phi_{w_r,w_r}(l,i)\cdot|X(l,i)|^2 + \Phi_{SS}(l,i)}, \quad (2)$$

where l is the FFT frequency index, i is the frame index. $\Phi_{w_r,w_r}(l,i)$ represents the state of convergence and $\Phi_{SS}(l,i)$ is an estimate of the near speech power spectral density (PSD). The estimation is detailed in the reference as a result of the generalized Wiener solution.

In order to consider the effect of the effective echo quantization noise $d_\Delta(k)$, it can be treated as an additive noise term in a modified postfilter according to

$$W_{2,\mathrm{mod}}(l,i) = \frac{\Phi_{SS}(l,i)}{\Phi_{w_r,w_r}(l,i)\cdot|X(l,i)|^2 + \Phi_{SS}(l,i) + K\Phi_{\hat{D}\hat{D}}(l,i)},$$

where $\Phi_{\hat{D}\hat{D}}(l,i)$ is the PSD estimate of the echo which is obtained by the adaptive filter W_1. The factor K is based on the assumption in [11] that the quantization noise power σ_Δ follows the power of the residual $\sigma^2_{y_r}$

$$K \approx \frac{\sigma^2_\Delta}{\sigma^2_{y_r}}. \quad (3)$$

This is based on the approximation of the effective quantization noise PSD

$$\Phi_{YY,\Delta} \approx K\Phi_{YY} = K(\Phi_{SS} + \Phi_{DD}) \quad (4)$$

at the decoder output. The PSD of the effective quantization noise component $d_\Delta(k)$ is thus estimated by $K\cdot\Phi_{DD}$. In practice, the factor K was determined once in advance for each speech codec by measuring the average SNR of the specific speech codec for a range of speech signals and empirically fine-tuning the parameter to achieve postfiltering performance without speech distortions.

Simulation results are shown in Fig. 2 for the single talk case with encoding and decoding by the AMR 12.2 kbit/s codec for both postfilters $W_2(l,i)$ and $W_{2,\mathrm{mod}}(l,i)$. The input signal was taken from the NTT database. Since the usually employed ERLE measure is not obtainable due to the non-linear effective echo path, the employed quality measure is echo reduction (ER), which is measured as

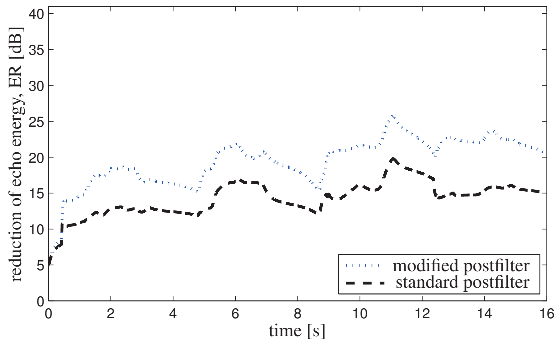

Figure 2: Single talk AEC results after transmission with the AMR 12.2 kbit/s codec with and without the postfilter modification.

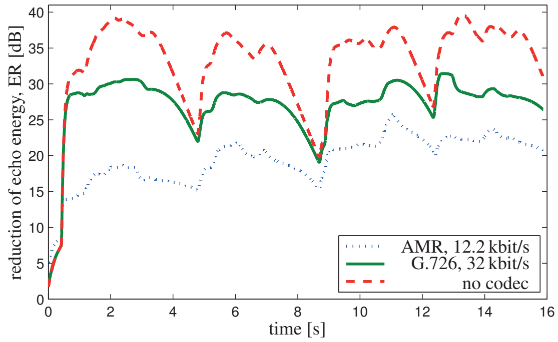

Figure 3: Single talk AEC results for different codecs.

the ratio of the short-term microphone input power to the short-term output energy after postfiltering during far end single talk according to

$$\mathrm{ER(k)} = 10\log_{10}\frac{\sigma^2_{\tilde{y}}(k)}{\sigma^2_{\hat{s}}(k)}. \quad (5)$$

For the far end single talk case, this represents the reduction of echo energy, however since the background noise is not affected by the AEC, the ER is bound by the background noise energy, which was 35 dB below the echo energy at the microphone in these simulations. It can be observed that the modified postfilter increases the reduction of echo by several dB.

In the next step, the potential for improving the quality of the AEC in the network by using a better codec 1 was analyzed. In Fig. 3 the performance of the AEC is shown examplarily for 3 different representations of $y(k)$: no coding, AMR codec at 12.2 kbit/s, and G.726 codec with a data rate of 32 kbit/s. It can be observed that the performance of AEC improves significantly when using the G.726 compared to AMR due to a reduction of non-linearity in the effective echo path. Thus, it can be concluded that network-based AEC benefits strongly from increased transmission bit rate.

3 Energy Efficiency

Besides quality aspects, the reduction of signal processing power in the mobile terminal is the motivation for network-based processing. Therefore, the overall power consumption of the system has to be taken into account. An increase in transmitted bit rate will directly lead to an increase in required transmission power.

3.1 Path Loss Calculation

In order to calculate the relation between the consumed signal processing power and the required transmission power, absolute values for these powers have to be estimated. Since the transmission power and thus the signal-to-noise-ratio of the radio transmission depends on the distance between transmitter and receiver, this distance has to be taken into account. In this paper, the calculation of transmission powers employs the Okumura-Hata path loss model [12]. This model was used to determine a starting point of the required power for a GSM transmission by calculating the attenuation on the radio channel according to

$$\frac{L_0}{\mathrm{dB}} = 69.55 + 26.16 \log_{10}\left(\frac{f_0}{\mathrm{MHz}}\right) - 13.82 \cdot \log_{10}\left(\frac{h_T}{\mathrm{m}}\right) + \left(44.9 - 6.55 \cdot \log_{10}\left(\frac{h_T}{\mathrm{m}}\right)\right) \cdot \log_{10}\left(\frac{d}{\mathrm{km}}\right) - a\left(\frac{h_R}{\mathrm{m}}\right), \tag{6}$$

where f_0 is the transmission frequency, h_T is the height of the cell tower, h_R is the height of the mobile terminal, and d is the distance between cell tower and mobile terminal. For the so called “typical urban” scenario, the parameter a is defined by

$$a\left(\frac{h_R}{\mathrm{m}}\right) = \left(1.1 \cdot \log_{10}\left(\frac{f_0}{\mathrm{MHz}}\right) - 0.7\right)\frac{h_R}{m} \tag{7}$$

$$- \left(1.56 \cdot \log_{10}\left(\frac{f_0}{\mathrm{MHz}}\right) - 0.8\right). \tag{8}$$

Using the determined path loss L_0, the required transmission power P_0 can be calculated as

$$\frac{P_0}{\mathrm{dBm}} = \frac{P_{RX}}{\mathrm{dBm}} + \frac{L_0}{\mathrm{dB}} - \frac{G_A}{\mathrm{dB}} + \frac{L}{\mathrm{dB}}, \tag{9}$$

i.e, $P_{RX} = P_0 \cdot 2^{-\gamma}$ with required power P_{RX} at the receiver (receiver sensitivity), antenna gain G_A and an additional loss L which covers the fading margin and implementation losses (e.g. RX antenna cable at the base station).

3.2 Transmission of Additional Bit Rate

In order to transmit additional bit rate to support network-based processing, the uplink transmission to the network has to be changed. Generally, this means using a different combination of speech coding and channel coding. These modifications will need to increase the transmission power.

In a first approach, we use a speech codec 1 which achieves a better quality at the expense of a higher bit rate R_1 in comparison to the original codec (which is the same as codec 2) with bit rate R_0. The relative bit rate increase is defined as

$$r_b = \frac{R_1}{R_0}. \tag{10}$$

If we assume that for error protection a channel code with the same protection rate is used, the increased transmission power can be quantified as

$$P_1 = r_b \cdot P_0. \tag{11}$$

The reasoning for the proportional power increase is that the receiver requires a certain minimum of energy E_b per received bit

$$E_b = \frac{P_0}{R_0} = \frac{P_1}{R_1}. \tag{12}$$

3.3 Signal Processing Power

In order to compare the overall system power required for a system with network based processing to a system with mobile terminal-based processing, the power consumption of the signal processor has to be taken into account. If less instructions per second are required by moving complexity to the network, the processor, e.g., can be utilized at a lower clock frequency, directly reducing the processing power. The calculations in this paper assume a linear relation between *million instructions per second* (MIPS) and power consumption in mW. Thus, the power saved by removing the signal processing complexity ΔI can be expressed as

$$\Delta P_{DSP} = \Delta I \cdot e_{DSP}, \tag{13}$$

where e_{DSP} denotes the efficiency of the processor in $\frac{mW}{\mathrm{MIPS}}$.

3.4 System Comparisons

The framework of different power calculations stated in the previous subsections enables a comparison of different systems with and without network-based processing for a specific system setup. In order to estimate the validity of moving acoustic signal processing to the network-based unit in order to conserve power at the mobile terminal, we can now evaluate the overall power budget.

In the following calculations, a reduction of 35 MIPS of signal processing complexity is assumed. Table 1 shows a breakdown of the savings for the different components. System A uses mobile phone signal processing and speech transmission in the uplink (codec 1 in Fig. 1) with the AMR codec. System B uses network based processing and speech transmission in the uplink with the G.726 codec (codec 1), motivated by the results of Section 2. Thus, the bitrate factor between the systems is $r_B = 2.6$. In the downlink (codec 2) the AMR codec is used.

For the transmission power calculations using the Oku–mura-Hata model, the following settings have been chosen which represent a typical GSM base station transceiver: receiver sensitivity $P_{RX} = -113\,\mathrm{dBm}$, antenna gain $G_A = 17.43\,\mathrm{dB}$, additional losses $L = 6\,\mathrm{dB}$, radio frequency $f_0 = 900\,\mathrm{MHz}$, mobile terminal height $h_T = 1.7\,\mathrm{m}$, and cell tower height $h_R = 60\,\mathrm{m}$.

Since the efficiency of DSPs strongly depends on the employed technology, the calculations have been performed for a sensible range of efficiencies from $e_{DSP} = 0.05\,\frac{mW}{\mathrm{MIPS}}$ to $e_{DSP} = 1\,\frac{mW}{\mathrm{MIPS}}$. The reference transmission for the comparison was the transmission of the AMR bit rate of

	System A	System B	ΔI
speech encoder	20 MIPS	5 MIPS	15 MIPS
AEC	10 MIPS	0 MIPS	10 MIPS
Noise Reduction	10 MIPS	0 MIPS	10 MIPS
Total savings			35 MIPS

Table 1: Signal processing complexity at the mobile terminal.

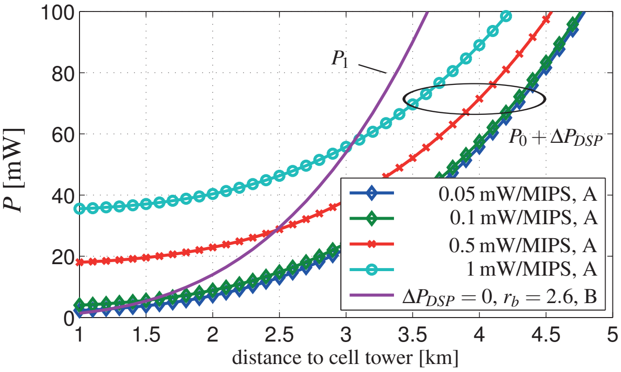

Figure 4: Power P_0 for transmission and ΔP_{DSP} for signal processing (System A) for different DSP technologies compared to transmission power P_1 for system B. $P_{RX} = -113$ dbm.

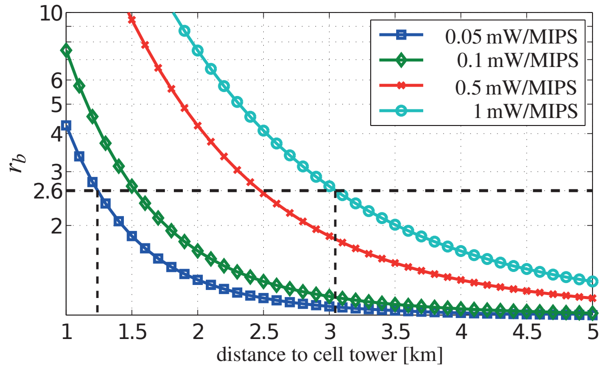

Figure 5: Allowed relative increase r_b of transmitted bit rate at same power for different DSP technologies with $P_{RX} = -113$ dbm

12.2 kbit/s in GSM with a channel frequency bandwidth of $B = 200$ kHz.

The resulting power budget can be observed in Fig. 4 for increasing distance between mobile terminal and cell tower. The figure shows the sum $P_S = P_0 + \Delta P_{DSP}$ of the powers ΔP_{DSP} for signal processing in the mobile terminal and the average transmission power P_0 for 5 different constellations. The first 4 utilize mobile terminal based signal processing (system A) with different DSP technologies and efficiencies. They are compared to system B with network-based processing. With increasing distance between mobile terminal and cell tower, the transmission power required for network-based processing rises faster due to the increased bit rate. This means that there is a cut-off point where the system with network-based processing starts to consume more power in a range of 1.25 km to 3 km depending on the DSP technology.

A different view on the trade-off between signal processing power and transmission power is given in Fig. 5. This figure shows how much additional bit rate can be transmitted when using all of the saved signal processing power to increase the bit rate. In our example whith $r_B = 2.6$, the figure shows that overall system power would be reduced within a radius of 1.25 km to 3 km from the cell tower to the mobile terminal.

4 Conclusions

In this paper, we study the feasibility to move signal processing complexity for acoustic echo and noise control from the mobile terminal to a network-based unit. As demonstrated in Section 2, acoustic echo control can be performed in the network, but benefits strongly from a transmission with higher bit rate codecs. In Section 3, we introduced methods to evaluate the relation between the required power for transmission with higher bit rates and the reduced power by moving signal processing complexity. Since the system setup contains a high number of degrees of freedom, such an evaluation has to be performed for each specific system setup and environment in practice. For the example shown in this paper, a range of operating points can be found where the power budget of the mobile terminal would be reduced by network-based signal processing. The power required for signal processing still has to be spent, but in the network is not constraint by a battery. This means that more sophisticated algorithms for combined AEC, noise reduction, and error concealment can be used to increase the overall audio quality.

The equations stated in this paper enable a quick estimation of the overall budget and can be used to motivate further development of such a relocation of complexity. These trade-off calculations can be extended to any application where processing complexity is moved to a centralized unit at the cost of additional transmission.

References

[1] G. Schmidt, “Applications of Acoustic Echo Control – An Overview,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, pp. 9–16, 2004.

[2] C. Breining, et al., “Acoustic echo control, an application of very-high-order adaptive filters,” *IEEE Signal Process. Magazine*, vol. 16, no. 4, pp. 42–69, 1999.

[3] S.L.Gay and J. Benesty, eds., *Acoustic Signal Processing for Telecommunications*. Kluwer Accademic Publishers, Dordrecht, 2000.

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

[5] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley, New York, 2004.

[6] A. Mader, H. Puder, and G. Schmidt, “Step-size Control for Acoustic Echo Cancellation Filters – An Overview,” *Signal Processing*, vol. 80, pp. 1697–1719, September 2000.

[7] G. Enzner and P. Vary, “Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones,” *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, 2006.

[8] M. Pawig and P. Vary, “Quality investigations of network based acoustic noise reduction,” in *Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*, Sept. 2011.

[9] X. Lu and B. Champagne, “Acoustic echo cancellation over a nonlinear channel,” in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control*, 2001.

[10] H. Gnaba, M. T.-H. Alouane, M. Jaidane-Saidana, 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*.

[11] G. Enzner, H. Krüger, and P. Vary, “On the problem of acoustic echo control in cellular networks,” in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sept. 2005.

[12] ITU-R, “ITU-R Rec. P.529-2: Okumura-hata propagation model in prediction methods for the terrestrial land mobile service in the VHF and UHF bands,” 1995.