Room Impulse Response Reshaping by p-Norm Optimization based on Estimates of Room Impulse Responses

Jan Ole Jungmann[1], Stefan Goetze[2], Alfred Mertins[1]

[1] *University of Lübeck, Institute for Signal Processing, D-23538 Lübeck, Email: {jungmann, mertins}@isip.uni-luebeck.de*

[2] *Fraunhofer Institute for Digital Media Technology, D-26129 Oldenburg, Email: s.goetze@idmt.fraunhofer.de*

Abstract

Hands-free telecommunication raised several real-world problems, such as corruption of the desired signal by additive noise, acoustic echoes, and reverberation. This paper addresses the mutual impacts of the subsystems for Acoustic Echo Cancellation (AEC) and Listening Room Compensation (LRC) based on p-norm optimization. In acoustic systems for LRC the equalizer is placed in front of the loudspeaker. An estimate of the room impulse response (RIR) is necessary for the equalizer to compensate for the influence of the RIR at the position of the reference microphone where the human user is assumed to be located. Since the RIR is identified by the acoustic echo canceller anyway, its estimate can be used to design the equalizer. The quality of dereverberation in dependance of the degree of system identification will be investigated in this contribution. Furthermore, the influence of the equalizer on the AEC is analyzed.

RIR Reshaping by p-Norm Optimization

In [3] the *well-known* least-squares based design rule to build an equalizer that renders certain properties on the global impulse response has been generalized by introducing a p-norm based objective function. Two weighting windows for the desired and the unwanted part of the global impulse response are defined. The optimization problem reads

$$\min_{\mathbf{h}} : f(\mathbf{h}) = \log\left(\frac{\|\mathbf{g}_u\|_{p_u}}{\|\mathbf{g}_d\|_{p_d}}\right), \qquad (1)$$

with $\mathbf{h}$ being the equalizer one aims to design, $\mathbf{g}_d = \mathrm{diag}\{\mathbf{w}_d\}\,\mathbf{C}\mathbf{h}$ being the desired part of the global impulse response of length L_g ($\mathbf{g}_u$ accordingly), $\mathbf{C}$ being the convolution matrix made up of the RIR $c(n)$, $\|\cdot\|_p$ denoting the p-norm of a vector, and $\mathrm{diag}\{\cdot\}$ transforming a vector into a diagonal matrix. The optimal solution is approximated by applying a gradient-descent procedure.

Design of the Weighting Windows

Mertins et. al proposed to exploit psychoacoustic findings to reduce the *audible* echoes and introduced weighting functions that capture the compromise temporal masking limit of the human auditory system [3]. The definitions for the windows read as follows:

$$\mathbf{w}_d = [\underbrace{0,0,\ldots,0}_{N_1},\underbrace{1,1,\ldots,1}_{N_2},\underbrace{0,0,\ldots,0}_{N_3}]^T \qquad (2)$$

and

$$\mathbf{w}_u = [\underbrace{0,0,\ldots,0}_{N_1+N_2},\underbrace{\mathbf{w}_0^T}_{N_3}]^T \qquad (3)$$

where $N_1 = t_0 f_s$, $N_2 = 0.004\mathrm{s} f_s$, and $N_3 = L_g - N_1 - N_2$ with f_s being the sampling frequency and t_0 being the time taken by the direct sound, respectively. The window $\mathbf{w}_0$, with its reciprocal being the compromise masking limit according to [1], is defined as

$$w_0(n) = 10^{\frac{3}{\log(N_0/(N_1+N_2))}\log\left(\frac{n}{N_1+N_2}\right)+0.5} \qquad (4)$$

with $N_0 = (0.2\mathrm{s} + t_0)\, f_s$ and time index n ranging from $N_1 + N_2 + 1$ to $L_g - 1$.

Spectral Distortions

To attenuate an additional coloration of the loudspeaker signal, we propose to weaken potential spectral distortions introduced by the equalizer by applying a *short* linear prediction error filter that is specifically designed for the equalizer.

System Identification

Since the AEC is an adaptive filter which identifies the room impulse response, its estimate can be used for the equalizer design. Figure 1 depicts the general setup of a combined LRC/AEC scenario.

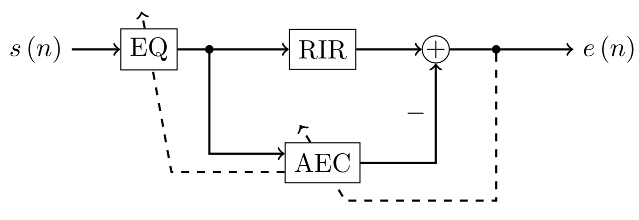

Figure 1: Combined system with LRC (EQ) and acoustic echo canceller (AEC); RIR denotes the listening room, containing the speaker and the microphone.

The RIR can be split up into one part $\hat{\mathbf{c}}(n)$ that is correctly identified by the AEC and an estimation error $\tilde{\mathbf{c}}(n)$, due to underestimating the order of the RIR and insufficient convergence of the AEC:

$$\mathbf{c}(n) = \hat{\mathbf{c}}(n) + \tilde{\mathbf{c}}(n)\,. \qquad (5)$$

Simulation Results

The RIR was simulated having a reverberation time of $\tau_{60} = 200$ ms and a length of $L_c = 2000$ taps. The filter orders of the AEC and the equalizer $\mathbf{h}(k)$ were set to 1800 and 2000, respectively. As input signals Gaussian white noise and a recorded speech signal (female speaker) were used. The equalizer was redesigned by optimizing the p-norm based objective function (equation (1)) every 1000 updates of the AEC with the weighting windows defined in (2) and (3); further we set $p_d = 10$ and $p_u = 20$. The length of the prediction error filter has been set to $L_p = 50$ taps.

AEC Convergence

The convergence of the AEC is influenced by the additional coloration introduced by the equalizer.

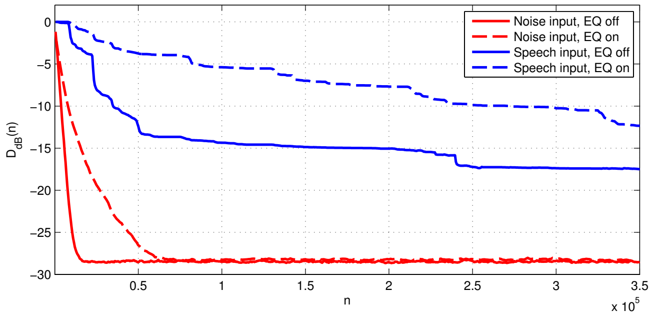

Figure 2: Relative System Misalignment $D_{dB}(n)$.

Figure 2 shows the relative system misalignment

$$D_{dB}(n) = 10 \cdot \log_{10} \frac{\|\mathbf{c}(n) - \hat{\mathbf{c}}(n)\|^2}{\|\mathbf{c}(n)\|^2} \quad (6)$$

with the quadratic vector norm $\|\mathbf{c}(n)\|^2 = \mathbf{c}(n)^T \mathbf{c}(n)$ for the two input signals $s(n)$ (noise or speech) and for the cases of active and inactive equalizer. In the case of noise input, the use of the equalizer only results in a slightly lower converge rate and slightly decreased overall system identification. By choosing a speech signal as an input signal the use of the equalizer results in a decrease of both the convergence rate and the system identification performance.

Influence of the AEC on the EQ

For evaluation of the LRC subsystem we use a revised version of the reverberation quantization (RQ) measure [2]. The proposed measure captures the audible reverberation by integrating the impulse response's energy that exceeds the compromise temporal masking limit on a logarithmic scale:

$$\mathrm{RQ_r} = \sum\nolimits_{n=N_0}^{L_g-1} g_{\mathrm{EM}}(n) \quad (7)$$

with

$$g_{\mathrm{EM}}(n) = \begin{cases} 20 \cdot \log_{10}(|g(n)| \cdot w_u(n)) & , |g(n)| > \frac{1}{w_u(n)} \\ 0 & , \text{otherwise} \end{cases} \quad (8)$$

and N_0 being the discrete time index that is 4 ms later than the direct impulse of $g(n)$.

If the RIR is completely reshaped, then no time coefficient exceeds the temporal masking limit and $\mathrm{RQ_r} = 0$.

Figure 3 shows the $\mathrm{RQ_r}$ value in dependance of the system misalignment of the AEC for the noise and the speech input signal. It should be mentioned that the x-axis is flipped so that high values for $\mathrm{D_{dB}}$, which indicate *bad* convergence, are left and smaller values indicating *good* convergence are right. The dashed horizontal line at $\mathrm{RQ_r} = 5218$ indicates the unequalized RIR. When the equalizer has been designed with perfect knowledge of the RIR one reaches $\mathrm{RQ_r} = 0.96$.

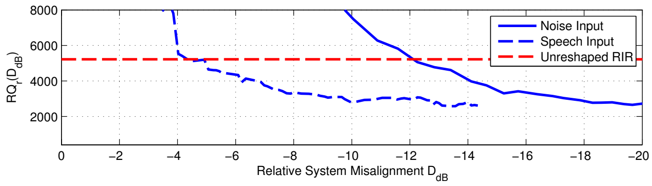

Figure 3: $\mathrm{RQ_r}$ value of the equalized system depending on the degree of system identification measured by $\mathrm{D_{dB}}$.

Generally, it can be seen that a high system misalignment results in bad dereverberation performance. By choosing speech as the input signal, dereverberation can be achieved when the misalignment of the system estimate is below -5.06 dB. With the noise input signal, the misalignment must be lower than -12.1 dB to reduce the audible reverberations.

Conclusions

In this contribution we analyzed the mutual influences of the video-conferencing subsystems *listening-room compensation* by utilizing the p-norm optimization and *acoustic echo cancellation.* The quality of the system identification was shown in dependance of the additional coloration of the loudspeaker signal introduced by the equalizer. Furthermore the performance of the audible echo reduction was analyzed in dependance of the degree of system identification.

References

[1] L. D. Fielder. Analysis of traditional and reverberationreducing methods for room equalization. *J. Audio Eng. Soc.*, 51:3–26, 2003.

[2] T. Mei and A. Mertins. On the robustness of room impulse response reshaping. In *Proc. International Workshop on Acoustic Echo and Noise control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[3] A. Mertins, T. Mei, and M. Kallinger. Room impulse response shortening/reshaping with infinity- and p-norm optimization. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(2):249–259, 2010.