SYSTEM IDENTIFICATION FOR MULTI-CHANNEL LISTENING-ROOM COMPENSATION USING AN ACOUSTIC ECHO CANCELLER

*Stefan Goetze, †Markus Kallinger, ‡Alfred Mertins, and *Karl-Dirk Kammeyer

*University of Bremen
Dept. of Communications Eng.
28334 Bremen, Germany

†University of Oldenburg
Signal Processing Group
26111 Oldenburg, Germany

‡University of Lübeck
Institute for Signal Processing
23538 Lübeck, Germany

ABSTRACT

Modern hands-free telecommunication devices jointly apply several subsystems, e.g. for noise reduction (NR), acoustic echo cancellation (AEC) and listening-room compensation (LRC). In this contribution the combination of an equalizer for listening room compensation and an acoustic echo canceller is analyzed. Inverse filtering of room impulse responses (RIRs) is a challenging task since they are, in general, mixed phase systems having hundreds of zeros inside and outside near the unit circle in the z-domain. Furthermore, a reliable estimate of the RIR which shall be inverted is important. Since RIRs are time-variant due to possible changes of the acoustic environment, they have to be identified adaptively. If an AEC (or any other adaptive method) is used to identify the time variant room impulse responses the estimate's distance to the real RIRs may be too high for a satisfying equalization, especially in periods of initial convergence of the AEC or after RIR changes. Therefore, we propose to estimate the convergence state of the AEC and to incorporate this knowledge into the equalizer design.

***Index Terms*—** Listening Room Compensation (LRC), Acoustic Equalization, Acoustic Echo Cancellation (AEC), System Identification, Least-Squares Equalizer (LS-EQ)

1. INTRODUCTION

High quality hands-free video-conferencing systems employ several subsystems in combination. Ambient noise has to be suppressed and acoustic echoes due to the acoustic coupling between microphones and loudspeakers have to be canceled out. Furthermore, blind dereverberation and listening-room compensation shall increase speech intelligibility for the far-end and the near-end listener, respectively. This contribution will focus on the combination of an acoustic echo canceller (AEC) with a subsystem for listening-room compensation (LRC) and the influences of imperfect system identification by the AEC on the equalization.

The acoustic coupling between loudspeakers and microphones leads to echoes in the microphone paths that are transmitted back to the far-end user. An AEC delivers estimates of the acoustic echoes which can be subtracted from the microphone signals. For the case of a single loudspeaker this is done by identifying the room impulse responses (RIRs).

The second problem of hands-free devices investigated in this contribution is the reverberation of the far-end speech signal which is caused by reflections in the near-end room and is increased by the spatial distance between the near-end user of a teleconferencing system and its loudspeakers. This leads to a loss of speech intelligibility of the far speech signal radiated by the loudspeakers. For this purpose a LRC filter is placed in the signal path in front of the loudspeaker to reduce the influence of the RIRs at the microphone positions. Equalizers for LRC need information about the RIRs which can be obtained from the AEC filters. However, the deviation of the RIR estimates from the true RIRs may be too large for a good room equalization. In this paper we propose a method to avoid signal distortions due to unsatisfactory RIR estimates by using the knowledge about the AEC convergence state for the equalizer (EQ) design. Furthermore, spatial robustness is increased by the proposed method which is an important property since the system user may be located at some spatial distance from the microphones.

In the remainder of this contribution we will briefly review the concepts of LRC in Section 2. In Section 3 we discuss the capability of an AEC to correctly identify the RIR and propose a method to incorporate the knowledge of the convergence state of the AEC into the EQ design. By this, severe distortions of the speech signal can be avoided in case of an insufficient AEC estimate. Simulation results concerning the influence of an imperfect system estimate by the AEC on the equalizer are presented in Section 4, and Section 5 concludes the paper.

Notation: Vectors and matrices are printed in boldface while scalars are printed in italic. k is the discrete time index. The superscripts ^{T}, *, and ^{H} denote the transposition, the complex conjugation and the Hermitian transposition, respectively. The operator $*$ denotes the convolution of two sequences, $\mathrm{E}\{\cdot\}$ the expectation, and the operator convmtx$\{\mathbf{h}, L_c\}$ generates a convolution matrix of size $(L_c + L_h - 1) \times L_c$.

2. LISTENING ROOM COMPENSATION

The common setup for listening room compensation is depicted in Fig. 1. The equalization filter $\mathbf{c}_{\mathrm{EQ}}$ precedes the RIRs $\mathbf{H}$. Its goal is to reduce the reverberation introduced by the RIRs. Since a RIR is a mixed-phase system, in general, only its minimum-phase component can be inverted by a causal IIR filter [1]. Finite-length equalizers minimize the error between the overall system of $\mathbf{c}_{\mathrm{EQ}}$ convolved with $\mathbf{h}$ and a desired target system $\mathbf{d}$ [2], as depicted in Fig. 1 and described in Section 2.1. Thus, the goal of LRC is to minimize differences between the signals $y_i[k]$, which a human listener at the position of the reference microphones perceives, and the original unreverberated signal $s_f[k]$.

2.1. Multi-Channel Least-Squares Equalization

An equalization scheme which shall approximate the overall system of the concatenation of RIRs and EQ filter to a desired system $\mathbf{d}$ is depicted in Fig. 1. The identification of the RIRs is done by the AEC filters $\mathbf{C}_{\mathrm{AEC}}$.

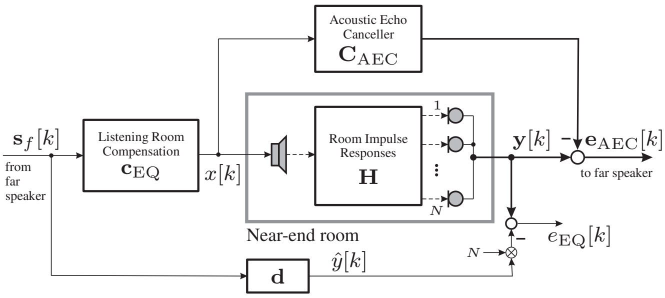

Fig. 1. Multi-channel setup for listening room compensation and AEC.

The minimization of $\mathrm{E}\left\{||e_{\mathrm{EQ}}[k]||^2\right\}$ for the error signal $e_{\mathrm{EQ}}[k] = \mathbf{s}_f^T[k]\mathbf{H}\mathbf{c}_{\mathrm{EQ}} - \mathbf{s}_f^T[k]\mathbf{d}$ leads to the well known least squares equalizer

$$\mathbf{c}_{\mathrm{EQ}} = \mathbf{H}^{+}\mathbf{d} \tag{1}$$

with

$$\mathbf{c}_{\mathrm{EQ}} = \left[\, c_{\mathrm{EQ},0},\, c_{\mathrm{EQ},1},\, \ldots,\, c_{\mathrm{EQ},L_{c,\mathrm{EQ}}-1}\,\right]^T \tag{2}$$

$$\mathbf{H} = \left[\mathbf{H}_1^T,\, \mathbf{H}_2^T,\, \ldots,\, \mathbf{H}_N^T\right]^T \tag{3}$$

$$\mathbf{H}_i = \text{convmtx}\left\{[h_{i,0}, h_{i,1}, \ldots, h_{i,L_h}]^T,\, L_{c,\mathrm{EQ}}\right\} \tag{4}$$

$$\mathbf{s}_f[k] = \left[\mathbf{s}_{f,1\mathrm{ch}}^T[k],\, \ldots,\, \mathbf{s}_{f,1\mathrm{ch}}^T[k]\right]^T_{N(L_{c,\mathrm{EQ}}+L_h-1)\times 1} \tag{5}$$

$$\mathbf{s}_{f,1\mathrm{ch}}[k] = \left[\, s_f[k],\, \ldots,\, s_f[k - L_h - L_{c,\mathrm{EQ}} + 2]\,\right]^T \tag{6}$$

$$\mathbf{d} = \left[\mathbf{d}_1^T,\, \mathbf{d}_2^T,\, \ldots,\, \mathbf{d}_N^T\right]^T \tag{7}$$

$$\mathbf{d}_i = [\underbrace{0,\, \ldots,\, 0}_{k_0},\, d_0,\, d_1,\, \ldots,\, d_{L_d-1},\, \underbrace{0,\, \ldots,\, 0}_{L_h+L_{c,\mathrm{EQ}}-1-L_d-k_0}\,]^T. \tag{8}$$

The lengths of the RIRs, the LRC filter and the desired systems are denoted by L_h, $L_{c,\mathrm{EQ}}$ and L_d, respectively. Although in practical environments a RIR is of infinite length it can be truncated after L_h samples since it is sufficiently decayed. In (1) $\mathbf{H}^{+}$ is the Moore-Penrose pseudoinverse of the channel convolution matrix $\mathbf{H}$ and $\mathbf{d}$ is a vector of size $N(L_{c,\mathrm{EQ}} + L_h - 1) \times 1$ containing the desired systems for each microphone channel which can be chosen to be delayed unit impulses, band- or highpasses.

2.2. System Identification

Since the LRC filter has to be placed in front of the acoustic environment, an estimate of the RIRs is needed as we can see from equation (1). Even for adaptive LMS-like algorithms for LRC that minimize the output error of the LRC system, an estimate of the RIRs is needed for the input of the so-called filtered-X LMS [3, 4]. An AEC which will be briefly described in Section 3 delivers estimates of the room impulse responses (RIRs) which can be used by the equalizer. Another method to access the RIRs would be, for example, ongoing measurement by maximum length sequences (MLS) [5]. However, this would be a protracted process because averaging over time is necessary and it would result in an audible perturbation for the near-end speaker.

3. SYSTEM IDENTIFICATION BY AN ACOUSTIC ECHO CANCELLER

In hands-free communications the desired signal of the near-end speaker is superimposed by acoustic echoes due to the coupling between loudspeaker and microphones. AEC filters for single-microphone setups calculate estimates for the echoes $\psi[k]$ in the microphone paths by identifying the RIRs as depicted in Fig. 2 for a single-channel system. The echo estimate $\hat{\psi}[k]$ is then subtracted from the microphone signal $y[k] = s_n[k] + \psi[k]$. Note that for the following discussions an inactive near speaker $s_n[k] = 0$ is assumed, and thus $y[k] = \psi[k]$. To detect periods of an inactive near speaker the voice activity detection (VAD) of [9] is used. Since the AEC filter is, in general, shorter than the RIR and due to the so-called *tail effect* [6] the identification of the RIRs may be biased even after complete convergence. Thus a RIR $\mathbf{h}[k]$ can be split up in one part $\hat{\mathbf{h}}[k]$ which is correctly identified by the AEC and an estimation error $\tilde{\mathbf{h}}[k]$.

$$\mathbf{h}[k] = \hat{\mathbf{h}}[k] + \tilde{\mathbf{h}}[k] \tag{9}$$

$$= \begin{bmatrix} \mathbf{c}_{\mathrm{AEC}}[k] \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{h}}[k] \tag{10}$$

with

$$\mathbf{h}[k] = [h_0[k]\,,\, h_1[k]\,,\, \ldots,\, h_{L_h-1}[k]]^T \tag{11}$$

$$\hat{\mathbf{h}}[k] = \left[c_{\mathrm{AEC},0}[k],\, c_{\mathrm{AEC},1}[k],\, \ldots,\, c_{\mathrm{AEC},L_{c,\mathrm{AEC}}-1}[k],\; 0,\, \ldots,\, 0\right]^T \tag{12}$$

$$\tilde{\mathbf{h}}[k] = \left[\tilde{h}_0[k]\,,\, \tilde{h}_1[k]\,,\, \ldots,\, \tilde{h}_{L_h-1}[k]\right]^T \tag{13}$$

If the RIR is estimated by an AEC the estimation error $\tilde{\mathbf{h}}[k]$ is also known as the AEC system misalignment. Fig. 2 shows the single-channel system of LRC filter $\mathbf{c}_{\mathrm{EQ}}[k]$ and AEC $\mathbf{c}_{\mathrm{AEC}}[k]$ and the decomposition of the RIR $\mathbf{h}[k]$ into the part modeled by the AEC and the system misalignment vector $\tilde{\mathbf{h}}[k]$.

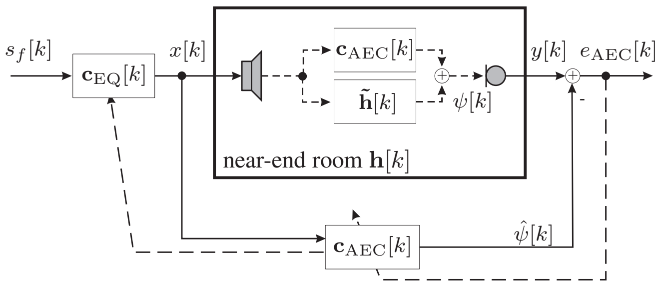

Fig. 2. Combined system with least squares equalizer and acoustic echo canceller. The RIR can be split into a part modeled by the AEC $\mathbf{c}_{\mathrm{AEC}}[k]$ and the system misalignment $\tilde{\mathbf{h}}[k]$.

The AEC filter is updated by minimizing its error signal $\mathrm{E}\left\{e_{\mathrm{AEC}}^2[k]\right\}$ by a gradient algorithm (e.g. the Partitioned Frequency Block LMS (PFBLMS) [7]). Thus, especially in periods of initial convergence or after RIR changes the system identification is insufficient and an EQ designed on its basis will introduce severe signal distortions. For a known misalignment vector $\tilde{\mathbf{h}}[k]$ at a fixed time instance k the EQ error signal can be modified:

$$e_{\mathrm{EQ}}[k] = \mathbf{s}^T[k](\hat{\mathbf{H}} + \tilde{\mathbf{H}})\mathbf{c}_{\mathrm{EQ}} - \mathbf{s}^T[k]\mathbf{d}. \tag{14}$$

Here the convolution matrices $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ are built similar to (3) and (4) by replacing $\mathbf{H}$ and $\mathbf{H}_i$ by $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{H}}_i$ or $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}_i$, respectively. Minimization of $\mathrm{E}\{e^2_{\mathrm{EQ}}[k]\}$ according to (14) leads to

$$\mathbf{c}_{\mathrm{EQ}} = \left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \tilde{\mathbf{H}}^T\tilde{\mathbf{H}} + \hat{\mathbf{H}}^T\tilde{\mathbf{H}} + \tilde{\mathbf{H}}^T\hat{\mathbf{H}}\right)^{-1}(\hat{\mathbf{H}}+\tilde{\mathbf{H}})^T\mathbf{d}. \quad (15)$$

With the simplifying assumption of $\tilde{\mathbf{h}}$ and $\mathbf{h}$ being uncorrelated $\mathrm{E}\{\tilde{\mathbf{H}}^T\hat{\mathbf{H}}\} = \mathbf{0}$ and a zero-mean system misalignment vector $\mathrm{E}\{\tilde{\mathbf{h}}\} = \mathbf{0}$ the EQ filter reduces to

$$\mathbf{c}_{\mathrm{EQ}} = \left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^T\mathbf{d}. \quad (16)$$

The system misalignment vector $\tilde{\mathbf{h}}[k]$ is unknown for real-world environments and difficult to estimate on its full length. However, different algorithms exist for estimating the norm of the system misalignment vector $\mathrm{E}\{||\tilde{\mathbf{h}}[k]||^2\}$, often also called coupling factor [8], because it describes the coupling between loudspeaker and AEC error signal $e_{\mathrm{AEC}}[k]$ if no disturbances are present. A prominent method to estimate the norm $\mathrm{E}\{||\tilde{\mathbf{h}}[k]||^2\}$ is to introduce an artificial delay of $L_\Delta \approx 20$ to 40 samples directly after the microphone and extrapolating the system misalignment of the AEC filter at those coefficients to the full length of the filter. For a more detailed description see [8]. In this contribution the estimate of the norm of the system misalignment is based on the ratio of the power of the AEC error signal $e^2_{\mathrm{AEC}}[k]$ and the power of the loudspeaker signal $x^2[k]$ which is updated in periods of an inactive near speaker ($s_n[k] = 0$)

$$\mathrm{E}\left\{||\tilde{\mathbf{h}}[k]||^2\right\} = \alpha_g \mathrm{E}\left\{||\tilde{\mathbf{h}}[k-1]||^2\right\} + (1-\alpha_g)\frac{\overline{e^2_{\mathrm{AEC}}}[k]}{\overline{x^2}[k]} \quad (17)$$

with recursively smoothed squared amplitudes

$$\overline{e^2_{\mathrm{AEC}}}[k] = \alpha_e \overline{e^2_{\mathrm{AEC}}}[k-1] + (1-\alpha_e)e^2_{\mathrm{AEC}}[k] \quad (18)$$
$$\overline{x^2}[k] = \alpha_x \overline{x^2}[k-1] + (1-\alpha_x)x^2[k]. \quad (19)$$

A voice activity detector (VAD) is implemented based on the normalized cross correlation approach by [9]. This VAD is needed anyway by the AEC to stop the adaptation in presence of an active near speaker and thus does not lead to an increased computational load. With the assumption of a white system misalignment (16) can be approximated by

$$\mathbf{c}_{\mathrm{EQ}} = \left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \mathrm{E}\left\{||\tilde{\mathbf{h}}||^2\right\}\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^T\mathbf{d}, \quad (20)$$

only depending on accessible variables such as the convolution matrix $\hat{\mathbf{H}}$ built from the AEC coefficients, the norm of the AEC misalignment vector $\mathrm{E}\{||\tilde{\mathbf{h}}||^2\}$ given by (17) and the desired response $\mathbf{d}$. In the following section the EQ design rules given by (15), (16), and (20) are evaluated by means of their performance to reduce reverberation introduced by the RIR.

4. SIMULATION RESULTS

The filter orders of the AEC and the EQ were $L_{c,\mathrm{AEC}}$=2048 and $L_{c,\mathrm{EQ}}$=1024, respectively. For the AEC a PFBLMS algorithm [7] was implemented. The RIR was simulated [10] having a length of L_h=4096 for different room reverberation times of $\tau_{60} = 200$ms to $\tau_{60} = 900$ms. We chose $\mathbf{d}$ containing 40^{th} order finite impulse response (FIR) highpasses with band limits at 200Hz at a sampling frequency of $f_s = 8000$Hz. The delay introduced by the equalizer was fixed to $k_0 = 170$ samples.

The equalizers are evaluated by means of the signal-to-reverberation-ratio-enhancement (SRRE)

$$\mathrm{SRRE} = \mathrm{SRR}_{\mathrm{out}} - \mathrm{SRR}_{\mathrm{bypass}} \quad (21)$$

that is defined here similar to the common signal-to-noise ratio enhancement (SNRE) which is widely used for evaluating noise reduction algorithms. In (21) $\mathrm{SRR}_{\mathrm{out}}$ is the signal-to-reverberation-ratio (SRR) after processing and $\mathrm{SRR}_{\mathrm{in}}$ is the SRR for an equalizer switched to bypass which means $\mathbf{c}_{\mathrm{EQ}} = \mathbf{d}$. By this, the delay in the target signal path is taken into account. With the definition of the microphone signal for an EQ switched to bypass $y_b[k] = s_f[k] * d[k] * h[k]$ we can calculate $\mathrm{SRR}_{\mathrm{out}}$ and $\mathrm{SRR}_{\mathrm{bypass}}$ by

$$\mathrm{SRR}_{\mathrm{out}} = \frac{L}{K}\sum_{\ell=0}^{K/L} 10\log_{10}\frac{\sum_{k=0}^{L-1}\hat{y}[\ell L+k]^2}{\sum_{k=0}^{L-1}(\hat{y}[\ell L+k]-y[\ell L+k])^2}$$

$$\mathrm{SRR}_{\mathrm{bypass}} = \frac{L}{K}\sum_{\ell=0}^{K/L} 10\log_{10}\frac{\sum_{k=0}^{L-1}\hat{y}[\ell L+k]^2}{\sum_{k=0}^{L-1}(\hat{y}[\ell L+k]-y_b[\ell L+k])^2}$$

with K, $L = 128$, and ℓ being the signal length, the block length and the block index, respectively. As depicted in Fig. 1, $\hat{y}[k]$ is the desired signal at the output of the target system $\mathbf{d}$ and $y[k]$ is the microphone signal.

Fig. 3 compares the equalizers according to eqns. (1), (15), and (20) by means of the SRRE for different room reverberation times τ_{60} from 200ms (left) to 900 ms (right) against the AEC convergence state expressed by means of its normalized system misalignment $D_{dB} = 10\log_{10}||\tilde{\mathbf{h}}||^2/||\mathbf{h}||^2$. Low values for D_{dB} are reached for a well converged AEC that delivers reliable RIR estimates. $D_{dB} = 0$ dB indicates initial convergence or a change of the RIR, e.g. To avoid the non-uniqueness problem for the RIR identification [6] we restrict the number of loudspeakers to $M = 1$. Simulation results for $N = 1$ (single-channel case) and $N = 4$ microphones are shown in Fig. 3 in the upper and lower part, respectively. For the multi-channel case (subplots d-f) the SRRE is averaged over all channels. The microphones were arranged in a line array with an inter-microphone distance of 5cm.

The solid (blue) lines show the EQ performance for the EQ design based on the RIR estimate delivered by the AEC only, which means that the least squares EQ (1) is applied by taking the AEC filter coefficients as a direct estimate for the RIR. As we can see, a direct and straightforward implementation of (1) by applying an AEC for system identification may not lead to sufficient improvement or even to a deterioration of the SRR for a high system misalignment which will be the case most of the time for high room reverberation times.

The horizontal dash-dotted (green) lines indicate the performance of an EQ designed according to (15) with *a-priori* knowledge of the full system misalignment vector $\tilde{\mathbf{h}}[k]$. Thus, they can be interpreted as upper limits for the improvement that can be achieved by the EQ for given RIRs and a given EQ order. It should be mentioned that the maximum achievable SRR improvement depends on the room reverberation time and the absolute positions of sources and microphones. Numerous positions have been simulated and Fig. 3 shows some representative results. It can be seen that the maximum possible SRR enhancement decreases for higher room reverberation times due to the higher energy in the reverberation tail of the RIRs and also for the multi-channel case compared to the single-channel

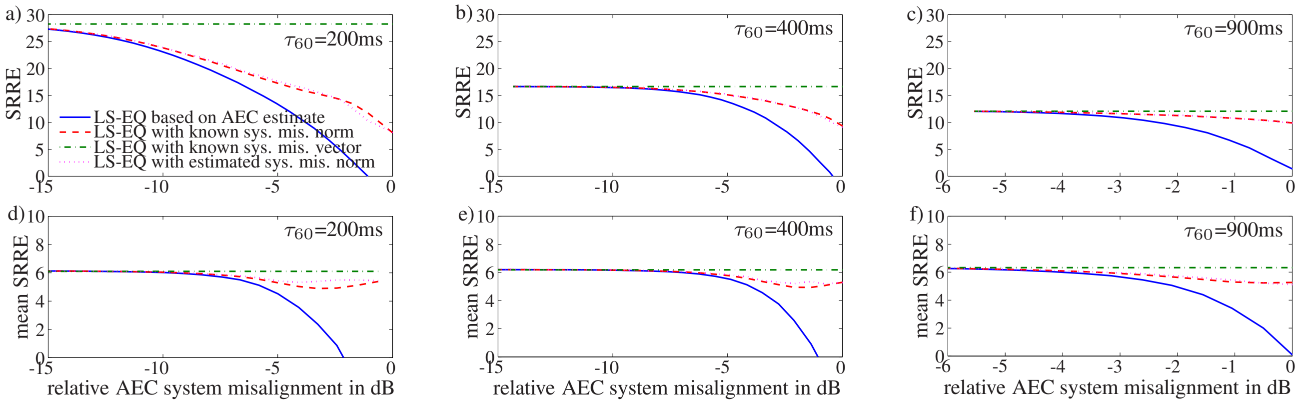

Fig. 3. Comparison of EQ designs according to eqns. (1), (15), and (20) by means of the SSRE depending on the AEC system misalignment for $\tau_{60} = 100$ms ... 900ms. Upper plots (a-c) show single-channel system and lower plots (d-f) show 4-microphone system.

case. The latter is due to the fact that the equalization is done by one filter for all four RIRs and thus a mean equalization is achieved [3]. This leads to a loss of SRR enhancement but to an increased spatial robustness which is very important in a hands-free scenario, since the user will not be located exactly at the positions of the microphones.

The dashed (red) curves show the EQ performance if only the norm of the system misalignment is known *a-priori* and the dotted (magenta) line if it is estimated by (17). As we can see from Fig. 3 the use of $||\tilde{\mathbf{h}}[k]||^2$ leads to significant improvements compared to the use of the RIR estimates given by the AEC only and that it is a good approximation of the use of the misalignment vector especially for higher room reverberation times and for a multi-channel scenario. The use of (17) as an estimate for $\mathrm{E}\{||\tilde{\mathbf{h}}[k]||^2\}$ for the proposed EQ design (20) leads to good approximations.

Since the user of a hands-free system will not be located directly at the position of the microphones an example for an EQ design with spatial mismatch is shown in Fig. 4.

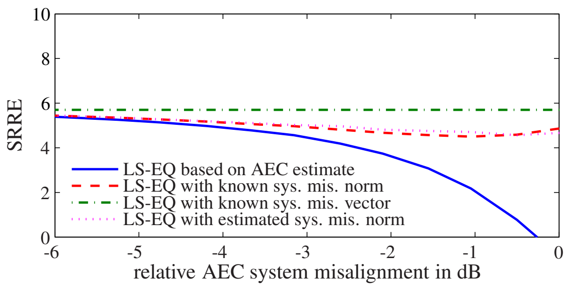

Fig. 4. Signal-to-Reverberation enhancement for spatial displacement. Distance between user of the system and of mic array: 20cm

The distance between the user and the microphone array, for which the RIR identification is done, is 20cm. As it can be seen from Fig. 4 a spatial displacement of course degrades the performance of the system compared to Fig. 3f) but an SRR gain of about 5dB is still possible, which is quite a good result considering the findings in [11].

5. CONCLUSION

In this contribution the impact of an imperfect system identification on the performance of an equalizer for listening room compensation (LRC) was investigated for a combined system of LRC and acoustic echo cancellation (AEC). The performance of equalization with respect to the signal-to-reverberation ratio enhancement (SRRE) was analyzed depending on the degree of system identification. If an AEC is utilized for estimating an unknown RIR, estimation errors severely degrade the LRC performance if no additional measures are taken especially in periods of initial convergence or at RIR changes. An enhanced scheme for LRC was proposed which incorporates the AEC system misalignment for the EQ filter design. It was shown that good results can be achieved even if the unknown system misalignment vector is approximated by an estimate of its norm only.

6. REFERENCES

[1] S. T. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *Journal of the Acoustical Society of America (JASA)*, vol. 66, pp. 165–169, Jul. 1979.

[2] J. N. Mourjopoulos, "Digital Equalization of Room Acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, Nov. 1994.

[3] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

[4] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, 1985.

[5] J. Borish and J. B. Angell, "An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise," *Journal of the Audio Engineering Society*, vol. 31, pp. 478–488, Jul./Aug. 1983.

[6] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, Mar 1998.

[7] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, pp. 14–34, January 1992.

[8] A. Mader, H. Puder, and G. Schmidt, "Step-Size Control for Acousatic Echo Cancellation Filters – an Overview," *Elsevier Signal Processing*, vol. 80, no. 9, pp. 1697–1719, September 2000.

[9] T. Gänsler and J. Benesty, "A Frequency Domain Double-Talk Detector based on a Normalized Cross Correlation Vector," *EURASIP Signal Processing*, vol. 81, no. 2001, pp. 1783–1787, 2001.

[10] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[11] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an Acoustic Reverberant Environment: Robustness Results," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.