SYSTEM IDENTIFICATION FOR LISTENING-ROOM COMPENSATION BY MEANS OF ACOUSTIC ECHO CANCELLATION AND ACOUSTIC ECHO SUPPRESSION FILTERS

Feifei Xiong, Jens-E. Appell and Stefan Goetze

Fraunhofer Institute for Digital Media Technology (IDMT), Project group
Hearing-, Speech- and Audio-Technology (HSA), 26129, Oldenburg, Germany
Email: {feifei.xiong,jens.appell,s.goetze}@idmt.fraunhofer.de

ABSTRACT

Subsystems for dereverberation and acoustic echo cancellation (AEC) / acoustic echo suppression (AES) are important components in high-quality hands-free telecommunication systems. This contribution describes and analyzes a combined system for dereverberation and AEC/AES. The system identification inherently achieved by the AEC/AES system is used for the design of the room impulse response (RIR) equalization filter, i.e. the listening-room compensation (LRC) system. We use complex RIR smoothing and decoupled filtered-X least-mean-squares (dFxLMS) gradient algorithm for LRC and a combined AEC/AES system for the system identification necessary for the LRC filter design. The performance of the combined system and the mutual influences of LRC and AEC/AES are analyzed.

Index Terms— Listening-Room Compensation, Dereverberation, Acoustic Echo Cancellation, Complex Smoothing, Decoupled Filtered-X LMS, System Identification

1. INTRODUCTION

Hands-free telecommunication systems usually comprise several subsystems to tackle different disturbances like acoustic echoes, ambient noise and reverberation [1]. Fig. 1 shows a block diagram of a hands-free system which contains an AEC filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and an AES post-filter $\boldsymbol{p}[k]$ to reduce acoustic echoes as well as an LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ to compensate for the reverberation caused by the RIR $\boldsymbol{h}[k]$ at the near-end listener's position.

Due to numerous reflections at the room boundaries, reverberant signals sound distant and echoic, leading to a significant loss in speech intelligibility [2]. One common approach to reduce reverberation is to equalize the RIR by pre-filtering the loudspeaker signal. This dereverberation approach is known as LRC. The knowledge of the RIR is required to design the equalizer $\boldsymbol{c}_{\mathrm{EQ}}[k]$ which can be obtained by the AEC/AES subsystem. AEC filters typically are based on non-blind system identification to subtract an estimate of the echo $\hat{\psi}[k]$ from the microphone output $y[k]$ [1]. Although AES filters are *only* based on a reliable estimate of the echo power spectral density (PSD) and system identification is just *one* way to obtain this PSD we will focus on AES systems based on system identification filters since system identification is needed for the LRC subsystem.

This work was partially supported by the project AAL-2009-2-049 "Adaptable Ambient Living Assistant" (ALIAS) co-funded by the European Commission (EC) and the Federal Ministry of Education and Research (BMBF) in the Ambient Assisted Living (AAL) program and the EU-FP7 project S4EeB (grant agreement no. 284628).

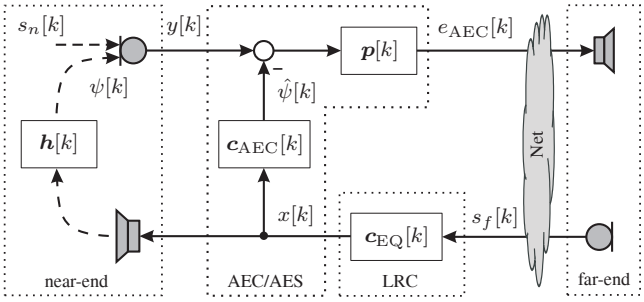

Fig. 1. Block diagram of a hands-free system with subsystems for LRC and AEC/AES.

In this contribution, two approaches will be discussed to design the LRC filter, i.e. complex smoothing (CS) of the RIR [3] and the decoupled filtered-X least-mean-squares (dFxLMS) gradient approach [4]. Complex smoothing only removes the minimum-phase component since, in general, no causal and stable system exists for direct inversion of mixed-phase systems, such as RIRs [5]. The dFxLMS gradient algorithm adaptively converges towards the well-known least-squares (LS) solution for LRC, however at much lower computational complexity. Furthermore, two different approaches are applied to update the AEC systems that cancel the acoustic echo and at the same time provide an RIR estimate for the LRC filter. Here, the PFBLMS [6] is used for the AEC $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and spectral subtraction for the AES filter $\boldsymbol{p}[k]$. While the AEC provides the more accurate estimate, the AES filter converges faster.

Notation: Vectors are printed in boldface while scalars are printed in italic. k, n and ℓ are the discrete time, frequency and block index, respectively. All frequency domain variables are printed in sans-serif typeface, e.g. $\mathsf{x}[\ell]$. The superscript ^{T} denotes the transposition and the symbols $*$, $\oslash$ represent the convolution and element-by-element division of two vectors. $\mathcal{F}(\cdot, L)$ and $\mathcal{F}^{-1}(\cdot, L)$ are the discrete Fourier transform (DFT) and the inverse DFT (IDFT) of size L.

The reminder of this paper is organized as follows: Section 2 introduces the LRC approaches based on CS and the dFxLMS algorithm. The AEC/AES system is introduced in Section 3. Simulation results are shown in Section 4 for the possible combinations and Section 5 concludes this paper.

2. LISTENING-ROOM COMPENSATION

Although dereverberation has been research topic for several decades now [5, 2, 3, 4], it is still challenging topic mainly due to the non-minimum-phase property of RIRs which does not allow for a direct stable and causal inversion [5]. Therefore, we will evaluate two methods in the following that circumvent this problem, i.e. complex impulse response smoothing [3] and gradient adaptive algorithms for least-squares approximation [4]. In Fig. 2, left lower part shows the approach based on the fractional-octave CS of the measured RIR extended by iterative homomorphic method to overcome the magnitude distortion [7], and right lower part visualizes the dFxLMS algorithm [4] to compute the equalizer coefficients. Please refer to the given references for a more detailed discussion of the two algorithms.

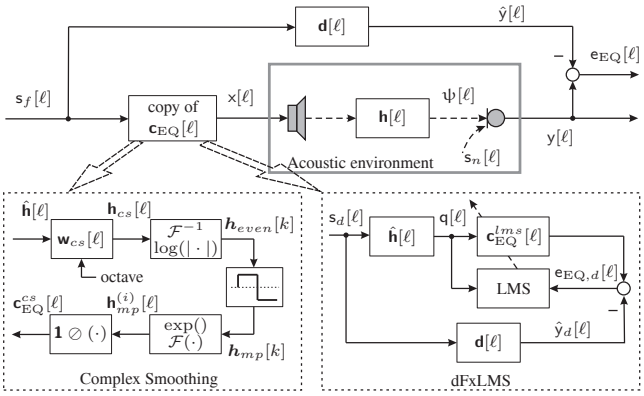

Fig. 2. Block diagram of the LRC system with two different approaches to obtain LRC filter coefficients $\mathbf{c}_{\mathrm{EQ}}[\ell]$, one is based on the CS of the RIR (left lower part), and the other on the dFxLMS gradient algorithm (right lower part).

Algorithm 1 Equalizer based on complex smoothing [3] and iterative homomorphic method [7].

$\mathbf{h}_{cs}[\ell] = \hat{\mathbf{h}}[\ell] * \mathbf{w}_{cs}[\ell]$

$$\mathsf{w}_{cs}[\ell] = \begin{cases} \frac{b-(b-1)\cos(\pi\ell/\mathsf{m}[\ell])}{2b(\mathsf{m}[\ell]+1)-1} & \ell = 0,\ldots,\mathsf{m}[\ell] \\ \frac{b-(b-1)\cos(\pi(\ell-L)/\mathsf{m}[\ell])}{2b(\mathsf{m}[\ell]+1)-1} & \ell = L-\mathsf{m}[\ell],\ldots,L-1 \\ 0 & \text{else} \end{cases}$$

$\boldsymbol{h}_{even}[k] = \mathcal{F}^{-1}(\log(|\mathbf{h}_{cs}[\ell]|), L)$

$$h_{mp}[k] = \begin{cases} h_{even}[k] & k = 0,\ L/2 \\ 2h_{even}[k] & k = 1,\ldots,L/2-1 \\ 0 & k = L/2+1,\ldots,L-1 \end{cases}$$

for $i = 1 : O^{cs}$ **do**
  $\boldsymbol{h}^{(i)}_{mp}[k] = \boldsymbol{h}_{mp}[k]/2^i$
  $\mathbf{h}^{(i)}_{mp}[\ell] = \exp(\mathcal{F}(\boldsymbol{h}^{(i)}_{mp}[k], L))$
  $\mathbf{c}^{(i)}_{mp}[\ell] = \mathbf{1} \oslash \mathbf{h}^{(i)}_{mp}[\ell]$
end for
$\mathbf{c}^{cs}_{\mathrm{EQ}}[\ell] = \prod_{i=1}^{O^{cs}} \mathbf{c}^{(i)}_{mp}[\ell]$

It was shown e.g. in [7] that using the smoothed RIR to design the inverse filter for the equalization leads to perceptually good results. Algorithm 1 summarizes this approach to design the equalizer $\mathbf{c}^{cs}_{\mathrm{EQ}}[\ell]$. In general, the CS function $\mathbf{w}_{cs}[\ell]$ has a low-pass characteristic with sufficient stop-band attenuation to avoid additional artifacts. $\mathbf{m}[\ell]$ is defined as the smoothing index corresponding to the fractional octave scaling, and L denotes the DFT size (here equal to the equalizer length L_{EQ}). In sequence, a reduced-complexity smoothed RIR can be obtained for the iterative homomorphic technique [7] to calculate the equalizer coefficients based on its minimum-phase component. On the other hand, the phase distortion may occur because the direct inversion of the all-pass component is not possible in practice.

Gradient algorithms that converge to the least-squares solution generally are computationally much more efficient than a direct inversion [4]. By using a modified error signal $\mathsf{e}_{\mathrm{EQ},d}[\ell]$ as shown in Fig. 2 compared to the conventional filtered-X least-mean-square (FxLMS) algorithm [8] which is based on the error signal $\mathsf{e}_{\mathrm{EQ}}[\ell]$, the update branch of the dFxLMS can be designed without any coupling to the microphone output $\mathsf{y}[\ell]$. Instead of depending on the signal statistics of the input signal $\mathsf{s}_f[\ell]$, the update path is driven by an independent excitation $\mathsf{s}_d[\ell]$ which can be optimized to achieve higher convergence speed [4]. Here, a Gaussian white excitation is used. Furthermore, the so-called overclocking by a factor O^{lms} allows for further increasing the convergence speed by calculating O^{lms} filter updates for each block of input samples. Algorithm 2 summarizes this approach. Note that $\hat{\mathbf{h}}[\ell]$ is estimated by an AEC filter as described in Section 3 since otherwise, the knowledge of the RIR is usually not available in real-world scenarios.

Algorithm 2 Decoupled filtered-X LMS algorithm [4].

for $i = 1 : O^{lms}$ **do**
  $\mathsf{q}[\ell + i - 1] = \mathbf{s}_d^T[\ell + i - 1]\,\hat{\mathbf{h}}[\ell]$
  $\hat{\mathsf{y}}_d[\ell + i - 1] = \mathbf{s}_d^T[\ell + i - 1]\,\mathbf{d}[\ell]$
  $\mathsf{e}_{\mathrm{EQ},d}[\ell + i - 1] = \mathbf{q}^T[\ell]\,\mathbf{c}^{lms}_{\mathrm{EQ}}[\ell + i - 1] - \hat{\mathsf{y}}_d[\ell + i - 1]$
  $\mathbf{c}^{lms}_{\mathrm{EQ}}[\ell+i] = \mathbf{c}^{lms}_{\mathrm{EQ}}[\ell+i-1] + \mu_{\mathrm{EQ}}\,\boldsymbol{\Phi}^{-1}_{\mathbf{qq}}[\ell]\,\mathbf{q}[\ell]\,\mathsf{e}_{\mathrm{EQ},d}[\ell+i-1]$
end for
Copy updated EQ coefficients $\mathbf{c}^{lms}_{\mathrm{EQ}}[\ell + i]$ to upper branch

3. ACOUSTIC ECHO CANCELLATION

Reduction of acoustic echoes can be achieved by two common methods, i.e. by adaptive filters to estimate and subtract the echoes (AEC) and by short-term spectral suppression filters (AES). We use the PF-BLMS algorithm [6] based on the well-known NLMS [8] and the Wiener post-filter (PF) using residual echo PSD estimate based on system identification [9, 10], not only because they are efficient for echo cancellation, but also since RIR identification is needed for the LRC system.

Fig. 3 shows a combined system with subsystems for AEC and AES. Using switches S_1 and S_2, the system output signal $\mathsf{e}_{\mathrm{AEC}}[\ell]$ can be decided to choose the error signal of the PFBLMS algorithm $\mathsf{e}^{lms}_{\mathrm{AEC}}[\ell]$ or of the AES approach $\mathsf{e}^{psd}_{\mathrm{AEC}}[\ell]$. The respective subsystem update is switched off if the output is not used, resulting in the different versions of the RIR estimation $\hat{\mathbf{h}}[\ell]$ as well (also see (1)). A partitioned calculation is applied to be able to use a short DFT length [6]. In addition, a double-talk detector (DTD) is applied to stop filter adaptation in periods of an active near-end speaker $\mathsf{s}_n[\ell]$ [11]. The step-size μ_{AEC} in Algorithm 3 is determined by DTD to prevent divergence of AEC/AES.

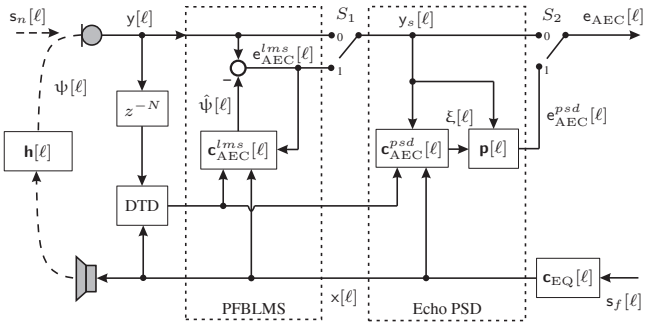

Fig. 3. Block diagram of the AEC system with two different approaches, one is the PFBLMS algorithm (front part), the other is based on the echo PSD estimate with the Wiener PF (latter part).

Algorithm 3 Partitioned frequency block LMS algorithm.

1: $\hat{\mathbf{y}}[\ell] = \mathbf{x}^T[\ell]\,\mathbf{c}^{lms}_{\mathrm{AEC}}[\ell]$
2: $\mathbf{e}^{lms}_{\mathrm{AEC}}[\ell] = \mathbf{y}[\ell] - \hat{\mathbf{y}}[\ell]$
3: $\mathbf{c}^{lms}_{\mathrm{AEC}}[\ell+1] = \mathbf{c}^{lms}_{\mathrm{AEC}}[\ell] + \mu_{\mathrm{AEC}}[\ell]\,\mathbf{x}[\ell]\mathbf{e}^{lms}_{\mathrm{AEC}}[\ell] \oslash (\mathbf{\Phi}_{\mathbf{xx}}[\ell] + \delta\mathbf{I})$

A reliable estimate of the echo PSD is crucial for a good AES performance. The Wiener PF $\mathbf{p}[\ell]$ suppresses the echoes by subtracting the (residual) echo PSD, as summarized in Algorithm 4 and based on [9, 10]. The first order recursive smoothing factor ν is usually set very close to 1, but also should depend on the DTD to prevent divergence, i.e. $\nu = 1$ in case the near-end speaker is active.

Algorithm 4 RIR identification based on estimate of echo PSD, as well as the Wiener PF for echo suppression.

1: **for** $i = 1 : L'_{\mathrm{AEC}}$ **do**
2: $\quad \xi_i[\ell] = \mathbf{x}_i^T[\ell]\,\mathbf{c}^{psd}_{\mathrm{AEC},i}[\ell]$
3: $\quad \mathbf{c}^{psd}_{\mathrm{AEC},i}[\ell+1] = \nu\,\mathbf{c}^{psd}_{\mathrm{AEC},i}[\ell] + (1-\nu)\,\mathbf{\Phi}_{\mathbf{x}\xi,i}[\ell] \oslash \mathbf{\Phi}_{\mathbf{xx},i}[\ell]$
4: **end for**
5: $\mathbf{c}^{psd}_{\mathrm{AEC}}[\ell] = \sum_{i=1}^{L'_{\mathrm{AEC}}} \mathbf{c}^{psd}_{\mathrm{AEC},i}[\ell]$
6: $\mathbf{p}[\ell] = \mathbf{\Phi}_{\mathbf{s}_n\mathbf{s}_n}[\ell] \oslash (\mathbf{\Phi}_{\mathbf{s}_n\mathbf{s}_n}[\ell] + \mathbf{\Phi}_{\boldsymbol{\xi}\boldsymbol{\xi}}[\ell])$
7: $\quad = (\mathbf{\Phi}_{\mathbf{y}_s\mathbf{y}_s}[\ell] - \mathbf{\Phi}_{\boldsymbol{\xi}\boldsymbol{\xi}}[\ell]) \oslash \mathbf{\Phi}_{\mathbf{y}_s\mathbf{y}_s}[\ell]$

As stated before, the cooperation of these two approaches is possible since the second method (latter part in Fig. 3) is able to estimate the residual echo PSD faster, leading to higher echo cancellation as well as an improved RIR identification. In terms of aforementioned echo reduction approaches, the RIR identification for the LRC system can be formulated as

$$\hat{\mathbf{h}}[\ell] = \begin{cases} \mathbf{c}^{lms}_{\mathrm{AEC}}[\ell] & \{S_1,\, S_2\} = \{1,\, 0\} \\ \mathbf{c}^{psd}_{\mathrm{AEC}}[\ell] & \{S_1,\, S_2\} = \{0,\, 1\} \\ \mathbf{c}^{lms}_{\mathrm{AEC}}[\ell] + \mathbf{c}^{psd}_{\mathrm{AEC}}[\ell] & \{S_1,\, S_2\} = \{1,\, 1\} \end{cases} \quad (1)$$

4. SIMULATION RESULTS

An RIR of reverberation time $\tau_{60} = 500$ ms and length $L_h = 4096$, generated by the well-known image method, as depicted in Fig. 4 (a) was used for the following simulation. The LRC and AEC filter lengths and the block length are $L_{\mathrm{EQ}} = 1024$, $L_{\mathrm{AEC}} = 1024$ and $L_b = 128$, respectively, at a sampling rate of 8 kHz. We furthermore chose the parameters $b = 0.6$, $O^{cs} = 4$ and 1/3-octave filters for $\mathbf{m}[\ell]$ (CS equalizer). For the dFxLMS algorithm, the step-size and overclocking factor were chosen to $\mu_{\mathrm{EQ}} = 0.1$ and $O^{lms} = 4$.

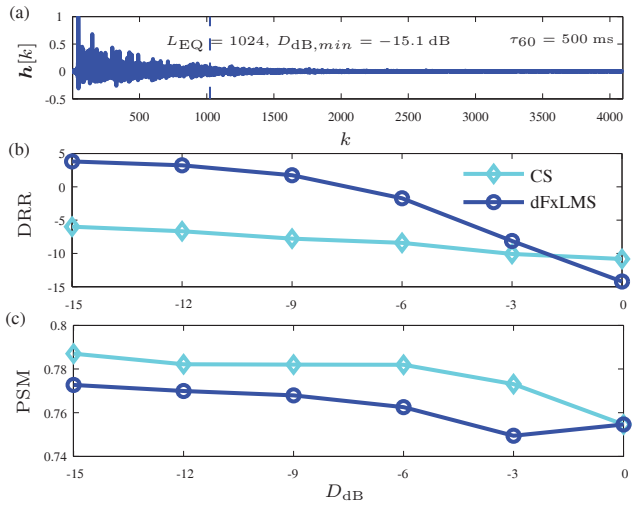

Fig. 4. (a) RIR of length $L_h = 4096$; (b) DRR performance of CS and dFxLMS approaches over system distance $D_{\mathrm{dB}} = 10\log_{10}(\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|^2/\|\boldsymbol{h}\|^2)$; (c) PSM quality measure [12] over system distance D_{dB}.

In order to grasp the first impression of the LRC filter's performance the channel-based quality measure Direct-to-Reverberation Ratio (DRR) is depicted in Fig. 4 (b) in dependence of the AEC performance in terms of system distance D_{dB}. As illustrated in Fig. 4 (b), in general, DRR is higher for better RIR identification. dFxLMS seems to perform better than CS approach specially when the RIR identification becomes more precise. This can be partially explained by the definition of the equalized system $\boldsymbol{c}_{\mathrm{EQ}} * \boldsymbol{h}$. The aim of the dFxLMS algorithm is to equalize the RIR while the CS equalizer is based on the smoothed RIR with octave effect, whose advantages may become obvious for the subjective perception. Thus, another perceptually based quality measurement is used for simulations in Fig. 4 (c), i.e. the Perceptual Similarity Measure (PSM) from PEMO-Q [12] which compares internal representations in a model of the human auditory system. PSM showed high correlations to subjective results for noise reduction and dereverberation algorithms. As seen in Fig. 4 (c), CS method behaves slightly better than the dFxLMS algorithm. Please note that the dFxLMS update speed is constrained by the convergence speed of the RIR identification in the AEC subsystem.

The loudspeaker signal $x(t)$ and the active near-end speaker's signal $s_n(t)$ are shown in Fig. 5 (a) (including periods of double-talk). Furthermore, Fig. 5 shows the system performance for the possible combination for system identification by means of AEC and/or AES filter, e.g. *CS + Combined* denotes LRC with CS method and AEC with the combined version $\{1,\, 1\}$ in (1). Signal-to-Reverberation-Ratio (SRR) and system distance D_{dB} are respectively analyzed as the performance evaluation for the LRC and AEC subsystem. With the comparison between the two approaches in the LRC system, it is clear to notice that dFxLMS performs better than CS in Fig. 5 (b) in terms of SRR. Here, the same holds as for performance in terms of DRR. A better RIR estimate enhances the

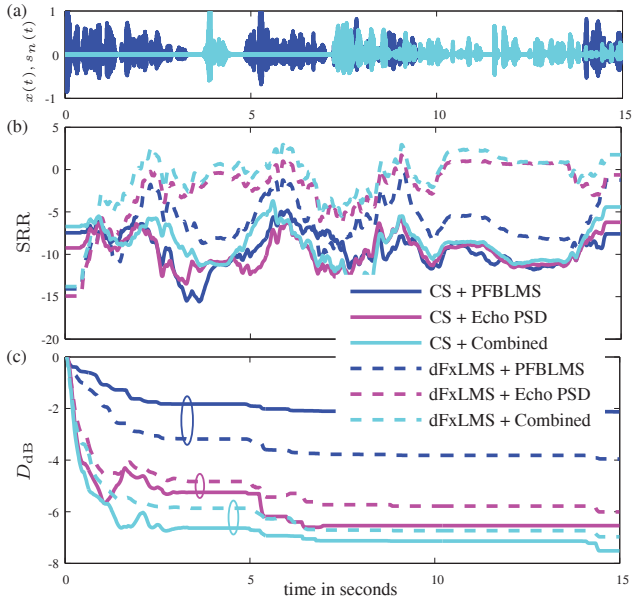

Fig. 5. Comparison of the system performance of combined system for each possible combination. (a) dark curve shows the loudspeaker signal $x(t)$, i.e. the far-end speaker's signal contributing to the echo, and light curve shows the near-end speaker's signal $s_n(t)$; (b) SRR performance of the LRC subsystem; (c) system distance D_{dB} of the AEC subsystem.

LRC performance for both approaches. As shown in Fig. 5 (c), the method using estimated echo PSD $\{0,\ 1\}$ in (1) performs better than the PFBLMS algorithm, which is mostly due to the limited PFBLMS step-size. It is feasible to improve the initial convergence speed for PFBLMS by larger the step-size μ_{AEC} but at the higher risk of divergence, particularly for the non-stationary input. However, PFBLMS is able to provide a more precise RIR identification than the AES approach which provides faster convergence but a coarser RIR estimate. The combined version for the AEC system always shows best performance both, in terms of convergence speed and final system distance performance.

Furthermore, it is important to evaluate the mutual influences between the LRC and AEC system. Obviously, the LRC performance strongly depends on the behaviour of the AEC system due to the RIR identification $\hat{\mathbf{h}}[\ell]$. The more precise the estimated RIR is, the better final LRC performance. For the AEC subsystem, because of the pre-filtering, the input signal $\mathbf{x}[\ell]$ will be influenced by the LRC equalizer, leading to the degradation of the convergence. PFBLMS is more sensitive to this influence than the echo PSD estimate approach, as it can be clearly seen from Fig. 5 (c). Besides, CS approach seems to introduce more correlation to the input signal than dFxLMS algorithm, which results in a decreased behaviour of PFBLMS. In contrast, the echo PSD estimate method depends less on this correlation, which can be seen comparing the solid curves with the dashed curves in Fig. 5 (c).

5. CONCLUSIONS

In this contribution, a combined system for hands-free communication containing LRC and AEC subsystems has been proposed and analyzed for different possible combinations and methods employed in each subsystem. Simulation results show that dFxLMS algorithm achieved better results in terms of technical measures, but CS performs slightly better in terms of perceptually motivated measures such as PSM. An improved version for the echo cancellation and system identification is achieved by the cooperation between the conventional adaptive algorithm and spectral enhancement technique. It is found that a good choice for the combination of LRC and AEC is to combine dFxLMS algorithm for dereverberation with the combined AEC/AES system for echo cancellation and system identification.

6. REFERENCES

[1] E. Hänsler and G. Schmidt (Eds.), *Speech and Audio Processing in Adverse Environments*, Springer, 2008.

[2] P. A. Naylor and N. D. Gaubitch, "Speech Dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.

[3] P. Hatziantoniou and J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses," *Journal of the Audio Engineering Society*, vol. 48, no. 4, pp. 259–280, Apr. 2000.

[4] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "A Decoupled Filtered-X LMS Algorithm for Listening-Room Compensation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, Sept. 2008.

[5] 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, July 1979.

[6] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, January 1992.

[7] B. D. Radlović and R. A. Kennedy, "Nonminimum-Phase Equalization and its Subjective Importance in Room Acoustics," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

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

[9] S. Goetze, M. Kallinger, and K.-D. Kammeyer, "Residual Echo Power Spectral Density Estimation Based on an Optimal Smoothed Misalignment For Acoustic Echo Cancelation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005) , Eindhoven, The Netherlands*, Sept. 2005.

[10] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Enhanced Partitioned Stereo Residual Echo Estimation," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct. 2006, pp. 1326–1330.

[11] 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, Sept. 2000.

[12] R. Huber and B. Kollmeier, "PEMO-Q - A New Method for Objective Audio Quality Assessment using a Model of Auditory Perception," *IEEE Trans. on Audio, Speech and Language Processing - Special Issue on Objectiv Quality Assessment of Speech and Audio*, vol. 14, no. 6, 2006.