JOINT OPTIMIZATION OF LCMV BEAMFORMING AND ACOUSTIC ECHO CANCELLATION

W. Herbordt[†], *W. Kellermann*[†], *and S. Nakamura*[‡]

email: {herbordt, wk}@LNT.de, nakamura@slt.atr.co.jp
[†]Telecommunications Laboratory, University Erlangen-Nuremberg
Cauerstrasse 7, 91058 Erlangen, Germany
[‡]ATR Spoken Language Translation Research Laboratories
2-2-2 Hikaridai Seikacho Soraku-gun, Kyoto 619-0288, Japan

ABSTRACT

Full-duplex hands-free acoustic human/machine interfaces often require the combination of acoustic echo cancellation and speech enhancement in order to suppress acoustic echoes, local interference, and noise. In order to optimally exploit positive synergies between acoustic echo cancellation and speech enhancement, we present in this contribution a combined least-squares (LS) optimization criterion for the integration of acoustic echo cancellation and adaptive linearly-constrained minimum variance (LCMV) beamforming. Based on this optimization criterion, we derive a computationally efficient system based on the generalized sidelobe canceller (GSC), which effectively deals with scenarioes with time-varying acoustic echo paths and simultaneous presence of double-talk of acoustic echoes, local interference, and desired speakers.

1. INTRODUCTION

For audio signal acquisition in hands-free human/machine interfaces, adaptive beamforming microphone arrays can be efficiently used for enhancing a desired signal while suppressing interference and noise [1]. For full-duplex communication systems, not only local interferers and noise corrupt the desired signal, but also acoustic echoes of loudspeakers. For suppressing acoustic echoes, acoustic echo cancellers (AECs) [2] are the optimum choice since they exploit the available loudspeaker signals as reference information.

For optimally suppressing local interferers and acoustic echos, it is thus desirable to combine acoustic echo cancellation with beamforming in the acoustic human/machine interface. For optimum performance, synergies between the AECs and the beamformer should be maximally exploited while the computational complexity should be kept moderate. This problem may be illustrated with two fundamental concepts for combining adaptive beamforming and AECs as shown in Figure 1 [3].

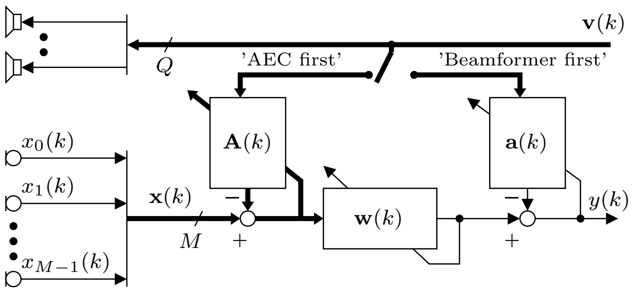

Figure 1: Combinations of AEC and beamforming [3].

The AECs $\mathbf{A}(k)$ can be placed in the sensor channels ('AEC first'), so that the problem of acoustic echo cancellation corresponds to that of the AEC alone.[1] For the adaptive beamformer $\mathbf{w}(k)$, however, positive synergies can be exploited after convergence of the AECs: The acoustic echoes are efficiently suppressed by the AECs, and the adaptive beamformer $\mathbf{w}(k)$ does not depend on the echo signals. Thus, more degrees of freedom are available for the suppression of local interference and noise. Obviously, one AEC is necessary for each sensor channel so that the M-fold complexity, where M is the number of microphones, is required at least for the filtering and for the filter update compared to a single AEC. Even with moderate numbers of microphones ($4 \leq M \leq 8$), this is a limiting factor for the usage of 'AEC first' in cost-sensitive systems [3]. Moreover, for presence of strong local interference and noise, the adaptation of the AECs must be slowed down or even stopped in order to avoid instabilities of the adaptive filters $\mathbf{A}(k)$. This reduces the tracking capability and, consequently, the echo suppression of the AECs for frequently changing echo paths, e.g., for the case, where a frequently moving desired source is located nearby the microphones. Limited echo suppression of the AECs, however, limits the positive synergies for the adaptive beamformer.

Alternatively, the AEC can be placed after the adaptive beamformer ('beamformer first'). Obviously, the complexity is reduced to that of a single AEC. However, positive synergies cannot be exploited for the adaptive beamformer, since the beamformer always 'sees' local interference and acoustic echoes. Furthermore, the AEC $\mathbf{a}(k)$ generally cannot track the time-variance of $\mathbf{w}(k)$ due to the smaller number of filter taps of $\mathbf{w}(k)$, which leads to faster convergence of $\mathbf{w}(k)$ relative to the AEC [3].

Another solution would be the integration of acoustic echo cancellation and adaptive beamforming so that the AEC does not depend on the time-variance of the adaptive beamformer [3]. One possible solution, which is based on the structure of the generalized sidelobe canceller (GSC) [4] is proposed in [5]. For this GSAEC, the AEC is placed in the reference path behind the quiescent beamformer so that the AEC is independent of the time-varying sidelobe cancelling path. However, acoustic echoes may leak through the sidelobe cancelling path when acoustic echoes are efficiently suppressed by the AEC so that the synergies of 'AEC first' cannot be obtained. Moreover, analogously to 'AEC first', the performance of this integrated system for strong local interference and noise and for continuously changing echo paths is limited.

As a new original contribution, we present in this work the combined optimization of adaptive beamforming and acoustic echo cancellation, which especially addresses the presence of strong local interference and noise, and frequently changing echo paths. In Section 2, the optimization criterion is presented. Section 3 derives a computationally efficient realization of the combined system which is based on the GSC structure. Section 4 describes an exemplary

[1]Upper case bold font and lower case bold font denote matrix quantities and vector quantities, respectively. k represents the discrete time index.

realization and illustrates the performance by experimental results for stationary conditions and for transient conditions.

2. OPTIMIZATION CRITERION

In contrast to 'beamformer first' in Figure 1, where different signals are used for the optimization of $\mathbf{w}(k)$ and of the AEC $\mathbf{a}(k)$, we propose to use the output signal $y(k)$ for the optimization of both AEC and adaptive beamformer as shown in Figure 2. The reference loudspeaker signals $\mathbf{v}(k)$ can thus be interpreted as additional input signals for the adaptive beamformer.[2]

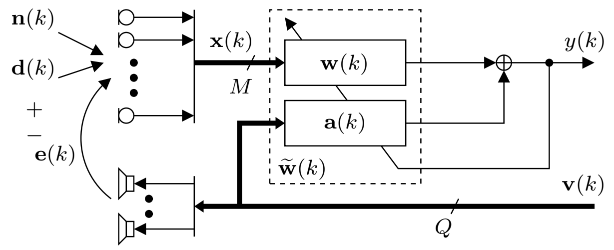

Figure 2: Joint optimization of adaptive beamforming and acoustic echo cancellation.

We now formulate the joint optimization criterion for the adaptive beamformer and for the AEC. We assume that the sensor signals $\mathbf{x}(k)$ are given by the superposition of the desired signal $\mathbf{d}(k)$, local interference $\mathbf{n}(k)$, and acoustic echoes $\mathbf{e}(k)$,

$$\mathbf{x}(k) = \mathbf{d}(k) + \mathbf{n}(k) + \mathbf{e}(k)\,, \tag{1}$$

where $\mathbf{d}(k)$, $\mathbf{n}(k)$, and $\mathbf{e}(k)$ are zero-mean and mutually uncorrelated. The output signal $y(k)$ of the combined system can be written as a function of the sensor signals $\mathbf{x}(k)$, the loudspeaker signals $\mathbf{v}(k)$, the stacked beamformer weight vector $\mathbf{w}(k)$, and the stacked AEC weight vector $\mathbf{a}(k)$ as

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k) + \mathbf{a}^T(k)\mathbf{v}(k)\,, \tag{2}$$

where

$$\begin{aligned}
\mathbf{x}(k) &= (\mathbf{x}_0(k),\, \mathbf{x}_1(k),\, \ldots\,,\, \mathbf{x}_{M-1}(k))^T\,, && (3)\\
\mathbf{x}_m(k) &= (x_m(k),\, x_m(k-1),\, \ldots\,,\, x_m(k-N_\mathbf{w}+1))^T, && (4)\\
\mathbf{v}(k) &= (\mathbf{v}_0(k),\, \mathbf{v}_1(k),\, \ldots\,,\, \mathbf{v}_{Q-1}(k))^T\,, && (5)\\
\mathbf{v}_q(k) &= (v_q(k),\, v_q(k-1),\, \ldots\,,\, v_q(k-N_\mathbf{a}+1))^T\,, && (6)\\
\mathbf{w}(k) &= (\mathbf{w}_0(k),\, \mathbf{w}_1(k),\, \ldots\,,\, \mathbf{w}_{M-1}(k))^T\,, && (7)\\
\mathbf{w}_m(k) &= (w_{0,m}(k),\, w_{1,m}(k),\, \ldots\,,\, w_{N_\mathbf{w}-1,m}(k))^T\,, && (8)\\
\mathbf{a}(k) &= (\mathbf{a}_0(k),\, \mathbf{a}_1(k),\, \ldots\,,\, \mathbf{a}_{Q-1}(k))^T\,, && (9)\\
\mathbf{a}_q(k) &= (a_{0,q}(k),\, a_{1,q}(k),\, \ldots\,,\, a_{N_\mathbf{a}-1,q}(k))^T\,. && (10)
\end{aligned}$$

Q is the number of loudspeaker channels, and $N_\mathbf{w}$ and $N_\mathbf{a}$ are the number of filter coefficients of the beamformer weight vectors $\mathbf{w}_m(k)$ and of the AEC filters $\mathbf{a}_q(k)$, respectively. With stacked vectors

$$\widetilde{\mathbf{w}}(k) = \left(\mathbf{w}^T(k),\, \mathbf{a}^T(k)\right)^T\,, \tag{11}$$

$$\widetilde{\mathbf{x}}(k) = \left(\mathbf{x}^T(k),\, \mathbf{v}^T(k)\right)^T\,, \tag{12}$$

we can write $y(k)$ as

$$y(k) = \widetilde{\mathbf{w}}^T(k)\widetilde{\mathbf{x}}(k)\,. \tag{13}$$

[2]This idea was first used in [6] for a combination of acoustic echo cancellation and multi-channel noise-reduction based on generalized singular value decomposition (GSVD).

A LS optimization criterion is obtained by minimizing the windowed sum of squared output signal samples $y(k)$ subject to constraints which assure that the desired signal is not distorted by $\widetilde{\mathbf{w}}(k)$. That is,

$$\min_{\widetilde{\mathbf{w}}(k)} \sum_{i=0}^{k} w_i(k)y^2(i) \quad \text{subject to} \quad \widetilde{\mathbf{C}}^T(k)\widetilde{\mathbf{w}}(k) = \mathbf{c}(k)\,. \tag{14}$$

The windowing function $w_i(k)$ extracts desired samples from the output signal $y(k)$ which should be included into the optimization. For example, infinite memory with exponential averaging is obtained with $w_i(k) = \lambda^{k-i}$ [7]. The constraint matrix $\widetilde{\mathbf{C}}(k)$ of size $(MN_\mathbf{w} + QN_\mathbf{a}) \times C$ and the constraint column vector $\mathbf{c}(k)$ of length C put C spatial constraints onto $\widetilde{\mathbf{w}}(k)$ in order to assure unity beamformer response for the direction-of-arrival of the desired signal [8]. Since the Q loudspeaker signals $\mathbf{v}(k)$ are assumed to be uncorrelated with the desired signal, the constraints are only required for the microphone signals, just as for conventional LCMV beamformers [8]. We can thus write $\widetilde{\mathbf{C}}(k)$ as

$$\widetilde{\mathbf{C}}(k) = \left(\mathbf{C}^T(k),\, \mathbf{0}_{C\times QN_\mathbf{a}}\right)^T\,, \tag{15}$$

where $\mathbf{C}(k)$ of size $MN_\mathbf{w} \times C$ is a conventional constraint matrix known from LCMV beamforming [8]. We thus obtain with (14) a formally simple optimization criterion, where only one single error signal needs to be minimized for an arbitrary number of microphones. This combined optimization allows to update the beamformer and the AEC simultaneously – in contrast to the previously discussed combinations, where the AEC can only be updated if local interference and desired signal are not active. As a consequence, the tracking problems of 'AEC first' and the leakage problem of GSAEC are thus resolved. The number of spatial degrees of freedom for interference suppression and for echo cancellation are increased by the number of loudspeakers Q relative to a beamformer alone.

3. TRANSFORMATION TO THE GSC

A direct solution of (14) can be determined using Lagrange multipliers [8]. However, with regard to an efficient realization of this combined system, we transform the constrained optimization problem into an unconstrained one using the structure of the GSC [4, 9].

For obtaining the GSC, the stacked weight vector $\widetilde{\mathbf{w}}(k)$ is projected onto two orthogonal subspaces,

$$\widetilde{\mathbf{w}}(k) = (\mathbf{P}_\mathrm{c}(k) + \mathbf{P}_\mathrm{a}(k))\,\widetilde{\mathbf{w}}(k)\,. \tag{16}$$

The first subspace $\widetilde{\mathbf{w}}_\mathrm{c}(k) := \mathbf{P}_\mathrm{c}(k)\widetilde{\mathbf{w}}(k)$ (constrained subspace) fulfills the constraint equation. That is,

$$\widetilde{\mathbf{C}}^T(k)\widetilde{\mathbf{w}}_\mathrm{c}(k) \stackrel{!}{=} \mathbf{c}(k)\,. \tag{17}$$

From (15), it follows that $\widetilde{\mathbf{w}}_\mathrm{c}(k)$ can be chosen as

$$\widetilde{\mathbf{w}}_\mathrm{c}(k) = \left(\mathbf{w}_\mathrm{c}^T(k),\, \mathbf{0}_{1\times QN_\mathbf{a}}\right)^T \tag{18}$$

in order to fulfill (17). The weight vector $\mathbf{w}_\mathrm{c}(k)$ of size $MN_\mathbf{w} \times 1$ is known as quiescent weight vector [8].

The second (orthogonal) subspace is chosen as

$$\mathbf{P}_\mathrm{a}(k)\widetilde{\mathbf{w}}(k) := -\widetilde{\mathbf{B}}(k)\widetilde{\mathbf{w}}_\mathrm{a}(k)\,, \tag{19}$$

where the columns of the matrix $\widetilde{\mathbf{B}}(k)$ are orthogonal to the columns of the constraint matrix $\widetilde{\mathbf{C}}(k)$, i.e.,

$$\widetilde{\mathbf{C}}^T(k)\widetilde{\mathbf{B}}(k) \stackrel{!}{=} \mathbf{0}\,. \tag{20}$$

The cascade of $\widetilde{\mathbf{B}}(k)$ and $\widetilde{\mathbf{w}}_\mathrm{a}(k)$ is termed sidelobe cancelling path [4]. From (15), it may be seen that (20) is met for

$$\widetilde{\mathbf{B}}(k) = \begin{pmatrix} \mathbf{B}(k) & \mathbf{0}_{MN_\mathbf{w}\times QN_\mathbf{a}} \\ \mathbf{0}_{QN_\mathbf{a}\times(M-C)N_{\mathbf{w}_\mathrm{a}}} & \mathbf{I}_{QN_\mathbf{a}\times QN_\mathbf{a}} \end{pmatrix}, \qquad (21)$$

where $\mathbf{I}_{QN_\mathbf{a}\times QN_\mathbf{a}}$ is the identity matrix of size $QN_\mathbf{a} \times QN_\mathbf{a}$ and where $\mathbf{B}(k)$ meets $\mathbf{C}^T(k)\mathbf{B}(k) = \mathbf{0}$. Since the constrained subspace generally contains the desired signal, the matrix $\mathbf{B}(k)$, which fulfills the requirement that the second subspace is orthogonal to the constrained subspace, suppresses desired signal components. Therefore, the matrix $\mathbf{B}(k)$ is generally referred to as blocking matrix [8]. The identity matrix assures that acoustic echoes are not cancelled by $\widetilde{\mathbf{B}}(k)$. As a consequence, ideally, only acoustic echoes, local interference, and noise are present at the output of $\widetilde{\mathbf{B}}(k)$, so that the weight vector $\widetilde{\mathbf{w}}_\mathrm{a}(k)$ can be determined by unconstrained LS minimization of $y(k)$,

$$\min_{\widetilde{\mathbf{w}}_\mathrm{a}(k)} \sum_{i=0}^{k} w_i(k) \left[\left(\widetilde{\mathbf{w}}_\mathrm{c}(k) - \widetilde{\mathbf{B}}(k)\widetilde{\mathbf{w}}_\mathrm{a}(k) \right)^T \widetilde{\mathbf{x}}(i) \right]^2 . \qquad (22)$$

Introducing (18) and (21) into (22) and identifying the result with (2), it may be seen that $\widetilde{\mathbf{w}}_\mathrm{a}(k)$ is equivalent to a stacked weight vector consisting of a weight vector $\mathbf{w}_\mathrm{a}(k)$ and of the AEC $\mathbf{a}(k)$,

$$\widetilde{\mathbf{w}}_\mathrm{a}(k) := (\mathbf{w}_\mathrm{a}^T(k),\, \mathbf{a}^T(k))^T . \qquad (23)$$

We obtain for the output signal $y(k)$ the expression

$$y(k) = (\mathbf{w}_\mathrm{c}(k) - \mathbf{B}(k)\mathbf{w}_\mathrm{a}(k))^T \mathbf{x}(k) - \mathbf{a}^T(k)\mathbf{v}(k) , \qquad (24)$$

which can be put into the structure that is depicted in Figure 3. The combined system thus corresponds to the GSC, where $\mathbf{w}_\mathrm{a}(k)$ is combined with the AEC $\mathbf{a}(k)$, and where the loudspeaker signals $\mathbf{v}(k)$ are used as additional channels of the sidelobe cancelling path. $\mathbf{w}_\mathrm{a}(k)$ is generally called interference canceller since $\mathbf{w}_\mathrm{a}(k)$ is optimized in order to cancel interference and noise at the output of the GSC. Analogously, we refere to $\widetilde{\mathbf{w}}_\mathrm{a}(k)$ as 'echo and interference canceller' (EIC) and to the combined system of AEC and GSC as 'generalized echo and interference canceller' (GEIC).

The optimum weight vector $\widetilde{\mathbf{w}}_\mathrm{a}(k)$ is now obtained by setting the derivative of (22) w.r.t. $\widetilde{\mathbf{w}}_\mathrm{a}(k)$ equal to zero and by solving the obtained system of linear equations for $\widetilde{\mathbf{w}}_\mathrm{a}(k)$:

$$\widetilde{\mathbf{w}}_\mathrm{a,opt}(k) = \left(\widetilde{\mathbf{B}}^T(k)\widetilde{\mathbf{\Phi}}(k)\widetilde{\mathbf{B}}(k) \right)^+ \widetilde{\mathbf{B}}^T(k)\widetilde{\mathbf{\Phi}}(k)\widetilde{\mathbf{w}}_\mathrm{c}(k) , \quad (25)$$

$$\widetilde{\mathbf{\Phi}}(k) = \sum_{i=0}^{k} w_i(k)\widetilde{\mathbf{x}}(i)\widetilde{\mathbf{x}}^T(i) = \begin{pmatrix} \mathbf{\Phi}_\mathbf{xx}(k) & \mathbf{\Phi}_\mathbf{xv}(k) \\ \mathbf{\Phi}_\mathbf{vx}(k) & \mathbf{\Phi}_\mathbf{vv}(k) \end{pmatrix} . \quad (26)$$

$\widetilde{\mathbf{\Phi}}(k)$ is the sample correlation matrix of the stacked data vector $\widetilde{\mathbf{x}}(k)$ [7] for a given windowing function $w_i(k)$ and $(\cdot)^+$ is the pseudoinverse of a given matrix. The solution of the optimum weight vector $\widetilde{\mathbf{w}}_\mathrm{a,opt}(k)$ is formally equivalent to the optimum weight vector of the GSC [9]. Introducing finally (18), (21), and (26) into (25), (25) can be written as

$$\begin{pmatrix} \mathbf{w}_\mathrm{a,opt}(k) \\ \mathbf{a}_\mathrm{opt}(k) \end{pmatrix} = \begin{pmatrix} \mathbf{B}^T(k)\mathbf{\Phi}_\mathbf{xx}(k)\mathbf{B}(k) & \mathbf{B}^T(k)\mathbf{\Phi}_\mathbf{xv}(k) \\ \mathbf{\Phi}_\mathbf{vx}(k)\mathbf{B}(k) & \mathbf{\Phi}_\mathbf{vv}(k) \end{pmatrix}^+ \times$$

$$\times \begin{pmatrix} \mathbf{B}^T(k)\mathbf{\Phi}_\mathbf{xx}(k)\mathbf{w}_\mathrm{c}(k) \\ \mathbf{\Phi}_\mathbf{vx}(k)\mathbf{w}_\mathrm{c}(k) \end{pmatrix} . \qquad (27)$$

Note that the combined optimization of $\mathbf{a}(k)$ and $\mathbf{w}_\mathrm{a}(k)$ introduces the off-diagonal matrices into the first correlation matrix on the right side of (27). Separate optimization of $\mathbf{a}(k)$ and $\mathbf{w}_\mathrm{a}(k)$ (with the off-diagonal matrices equal to zero) yields the GSAEC with the AEC behind $\mathbf{w}_\mathrm{c}(k)$ [3, 5].

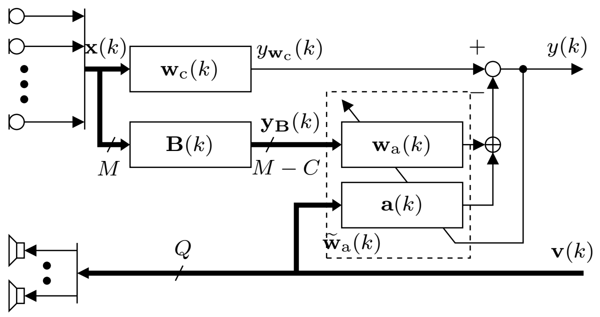

Figure 3: Generalized echo and interference canceller (GEIC).

4. REALIZATION EXAMPLE

For realizing the GEIC for practical applications, we use the implementation of the GSC in the discrete Fourier transform (DFT) domain after [10]. This implementation uses an adaptive blocking matrix [11] for tracking movements of the desired source and for robustness against distortion of the desired signal in reverberant environments. Blocking matrix and EIC are realized using DFT-domain adaptive filtering. The blocking matrix is adapted for presence of only desired signal, the EIC is adapted for presence of local interference and/or acoustic echoes. For controling the adaptation, the time-frequency double-talk detector presented in [12] is used. A separate adaptation control for the AEC as for 'AEC first' is thus not required.

For the realization of the EIC by adaptive filters, three aspects should be considered: First, the AEC $\mathbf{a}(k)$ and the interference canceller $\mathbf{w}_\mathrm{a}(k)$ should have the same filter length $N_\mathbf{a} = N_{\mathbf{w}_\mathrm{a}}$ in order to assure the same convergence speed, although generally $N_\mathbf{a} > N_{\mathbf{w}_\mathrm{a}}$ [3]. Therefore, $\mathbf{a}(k)$ should not be viewed as a conventional AEC but as additional degrees of freedom for the interference canceller. As a result, the echo suppression of GEIC will be smaller than the echo suppression of 'AEC first' for stationary conditions especially for reverberant environments, where generally $N_\mathbf{a} > N_{\mathbf{w}_\mathrm{a}}$. For environments with low reverberation, as, e.g., car environments, where $N_\mathbf{a} \approx N_{\mathbf{w}_\mathrm{a}}$, the performance of GEIC for stationary conditions approaches that of 'AEC first'. Second, the level of the loudspeaker signals $\mathbf{v}(k)$ should be adjusted to the level of the output signals of the blocking matrix in order to avoid large level differences which may lead to instabilities of the adaptive filters. Third, the loudspeaker signals should be temporally synchronized with the output signals of the blocking matrix in order to maximize the effective filter length of $\mathbf{a}(k)$.

4.1 Stationary conditions

We now compare the GEIC with 'AEC first' and with the GSC for stationary conditions. We use a sensor array with $M = 4$ and with an aperture of 28 cm in an office room with $T_{60} = 250$ ms reverberation time. The desired speaker is located in broadside direction at a distance of 60 cm from the array center. Two loudspeakers are placed in the two endfire directions at the same distance. A local interferer is located at 60 degrees off the array axis at 1.2 m. All source signals are highpass-filtered mutually uncorrelated white noise signals (cut-off frequency 200 Hz). The sampling rate is 12 kHz, $N_{\mathbf{w}_\mathrm{a}} = 256$, $N_\mathbf{a} = N_{\mathbf{w}_\mathrm{a}}$ for GEIC, $N_\mathbf{a} = 1024$ for 'AEC first'.

In Figure 4, the interference suppression *IR* and the echo suppression *ERLE* are depicted as a function of the echo-to-interference ratio *EIR* at the sensors. The echo suppression of the AEC for 'AEC first' is $ERLE_\mathrm{AEC} = 20$ dB. As we expected, *IR* and *ERLE* for the GEIC is greater than for the GSC but less than for 'AEC first'. Especially for low *EIR*, *ERLE* is considerably smaller for GEIC than for 'AEC first'.

However, for low *EIR*, acoustic echoes are masked by local interference, and this performance improvement is obtained by using longer AEC filters in each of the sensor channels. With increasing *EIR*, where the echo suppression is more and more important, *ERLE* of GEIC increases rapidly, so that *ERLE* is less than 7 dB smaller for GEIC than for 'AEC first' ($EIR > 0$ dB) in this experiment.

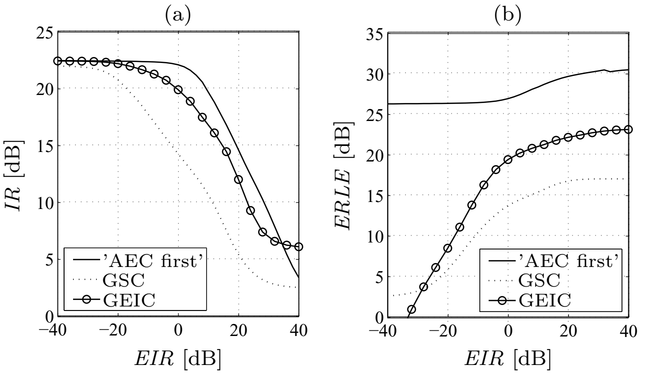

Figure 4: Interference rejection *IR* and echo suppression *ERLE* for GSC alone, 'AEC first', and GEIC after convergence of the adaptive filters.

4.2 Transient conditions

Next, we study the behavior of GEIC and of 'AEC first' for transient echo paths. The results are depicted in Figure 5. The geometrical setup corresponds to Section 4.1. Instead of mutually uncorrelated white noise signals, we now use male speech as desired signal and stereophonic (pop) music from mp3-files as acoustic echo signals (Figures 5 (a)-(c)). At 12.5 s, the loudspeakers are moved symmetrically to 20 degrees off the array axis in order to simulate a change of the echo paths. Figure 5 (d) shows which modules are adapted at a given time-instant. Figure 5 (e), (f) illustrate $IR(k)$ and $ERLE(k)$ for GEIC and for 'AEC first', respectively. It may be seen that 'AEC first' outperforms GEIC during Phases II and III, where the AECs for 'AEC first' are converged. However, during Phases IV and V, where acoustic echoes, local interference, and the desired speaker are active simultaneously, the GEIC outperforms 'AEC first', since the AECs of 'AEC first' cannot be adapted while the EIC can be adapted. $IR(k)$ and $ERLE(k)$ of 'AEC first' are reduced to that of the GSC. During Phase VI, where again only acoustic echoes are present, the AECs of 'AEC first' reconverge.

5. CONCLUSIONS

We presented a joint optimization for acoustic echo cancellation and adaptive LCMV beamforming, which leads to an efficient combination of AECs and adaptive beamformers. With a realization example based on a robust GSC, we showed that this structure is especially efficient for transient echo paths if frequent double-talk between acoustic echoes, local interference, and desired speakers is to expected. Our system requires only one AEC for an arbitrary number of microphones and no separate adaptation control for the AEC.

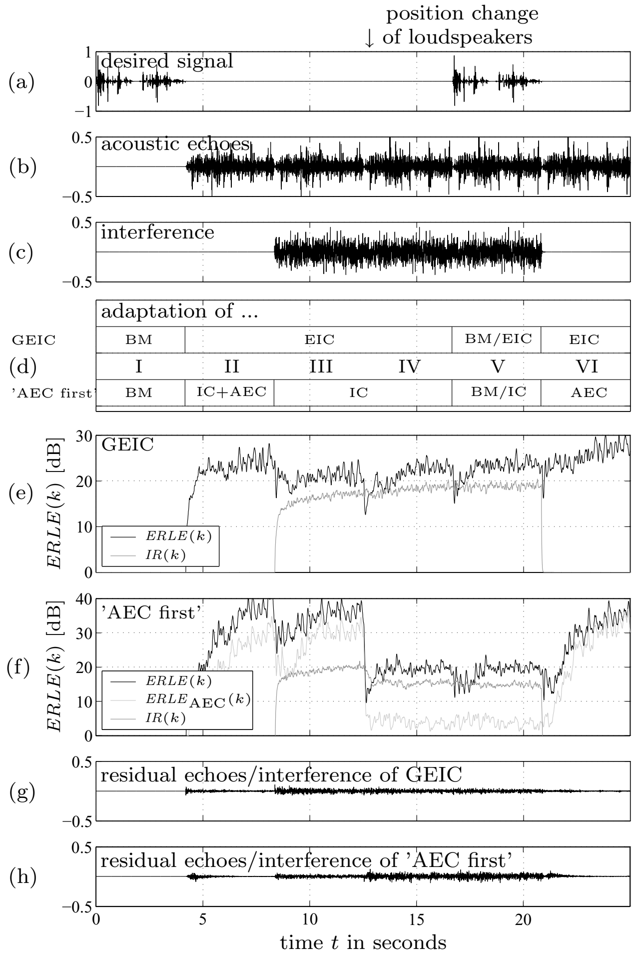

Figure 5: Transient behavior of GEIC and of 'AEC first'.

REFERENCES

[1] M.S. Brandstein and D.B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin, 2001.

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

[3] W. Kellermann, "Acoustic echo cancellation for beamforming microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M.S. Brandstein and D.B. Ward, Eds., chapter 13, pp. 281–306. Springer, Berlin, 2001.

[4] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, January 1982.

[5] W. Herbordt and W. Kellermann, "Acoustic echo cancellation embedded into the generalized sidelobe canceller," *Proc. EURASIP European Signal Processing Conference*, vol. 3, pp. 1843–1846, September 2000.

[6] S. Doclo, M. Moonen, and E. De Clippel, "Combined acoustic echo and noise reduction using GSVD-based optimal filtering," *Proc. IEEE ICASSP*, vol. 2, pp. 1051–1054, June 2000.

[7] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New Jersey, 3rd edition, 1996.

[8] H.L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*, John Wiley & Sons, Inc., New York, 2002.

[9] K.M. Buckley, "Broad-band beamforming and the generalized sidelobe canceller," *IEEE Trans. on Acoustics, Speech. and Signal Processing*, vol. 34, no. 5, pp. 1322–1323, October 1986.

[10] W. Herbordt, H. Buchner, and W. Kellermann, "An acoustic human-machine front-end for multimedia applications," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 1–11, January 2003.

[11] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2677–2684, October 1999.

[12] W. Herbordt, T. Trini, and W. Kellermann, "Robust spatial estimation of the signal-to-interference ratio for non-stationary mixtures," *Proc. Int. Workshop on Acoustic Echo and Noise Control*, pp. 247–250, September 2003.