2

Joint Optimization of Acoustic Echo Cancellation and Adaptive Beamforming

Wolfgang Herbordt[1], Walter Kellermann[2], and Satoshi Nakamura[1]

[1] ATR – Spoken Language Communication Research Laboratories, Kyoto, Japan
[2] Telecommunications Laboratory, University Erlangen-Nuremberg, Germany

For full-duplex hands-free acoustic human/machine interfaces, often a combination of acoustic echo cancellation and speech enhancement is required to suppress acoustic echoes, local interference, and noise. To optimally exploit positive synergies between acoustic echo cancellation and speech enhancement, various approaches were presented in the literature. However, efficient solutions for situations with high levels of background noise, with time-varying echo paths, and frequent double talk are still a challenging research topic. In this contribution, we address this problem by a joint least-squares (LS) optimization criterion for integrating acoustic echo cancellation and adaptive linearly-constrained minimum variance (LCMV) beamforming. After summarizing the state-of-the-art of this field, we derive the joint acoustic echo cancellation and beamforming system and show its relation to existing approaches. A realization of the joint system integrating a stereophonic acoustic echo canceller (AEC) and a robust generalized sidelobe canceller (RGSC) shows the advantages of the proposed system for high levels of background noise, time-varying echo paths, and frequent double talk. The proposed solution requires only one AEC for an arbitrary number of microphones. A separate adaptation control for the AEC is not necessary. Moreover, for AECs for multiple reproduction channels, the problem of slow convergence due to cross-correlated loudspeaker signals is avoided.

2.1 Introduction

For audio signal acquisition in hands-free human/machine interfaces, adaptive beamforming microphone arrays can be efficiently employed for enhancing a desired signal while suppressing interference and noise [11].

For full-duplex communication systems, not only interference and noise corrupt the desired signal, but also acoustic echoes originating from loudspeakers. For suppressing acoustic echoes, acoustic echo cancellers (AECs) using adaptive filters are the optimum choice since they exploit the reference information provided by the loudspeaker signals [12, 32, 41, 42].

To simultaneously suppress interferers and acoustic echoes, it is thus desirable to combine acoustic echo cancellation with adaptive beamforming in the acoustic human/machine interface. To achieve optimum performance, synergies between the AECs and the beamformer should be maximally exploited while the computational complexity should be kept moderate. When designing such a joint acoustic echo cancellation and beamforming system, it proves necessary to consider especially the time-variance of the acoustic echo path, the background noise level, and the reverberation time of the acoustic environment.

To combine acoustic echo cancellation with beamforming, various strategies were studied in the literature [4, 21, 44, 47, 49, 54, 55, 57, 63, 66, 67], reaching from cascades of AECs and beamformers to integrated solutions. These combinations address aspects such as maximization of the echo and noise suppression for slowly time-varying echo paths and high echo-to-interference ratios (EIRs) [55, 57, 66, 67], strongly time-varying echo paths, and low EIRs [21, 47, 49, 63], or minimization of the computational complexity [4, 44]. Overviews and comparisons of these methods can be found in [48, 58].

In this chapter, we review the state-of-the-art of joint acoustic echo cancellation and beamforming and compare the various approaches. Especially, we analyze the joint acoustic echo cancellation and beamforming system after [47, 49] in more detail. We show that this method, which is based on a joint linearly-constrained minimum variance (LCMV) optimization criterion, is especially efficient for low numbers of microphones ($M = 4 \dots 8$), low and moderate reverberation times in the range of $T_{60} = 50$ ms and 400 ms, low EIRs, and/or strong time-variance of the echo path. A separate adaptation control for the AEC is not required so that the difficult task of designing a robust adaptation control for the AEC is avoided. For multichannel reproduction systems such as, for example, stereophonic or 5.1-channel systems, the commonly known problem of slow convergence due to highly cross-correlated loudspeaker signals [5, 84] is avoided since the system identification problem is reduced to an interference cancellation problem [48].

Our proposed approach is based on the robust generalized sidelobe canceller (RGSC) after [48]. The RGSC provides high suppression of both strongly time-varying interference such as competing speakers and slowly time-varying diffuse noise, (as typical for, e.g., the interior of cars,) while preserving signal integrity of the desired speech, even for relatively small array apertures and limited numbers of microphones, and even in reverberant environments or for a moving desired speaker.

This chapter is organized as follows: In Sec. 2.2, we introduce the concepts of acoustic echo cancellation and of adaptive beamforming and discuss the previously presented combinations of acoustic echo cancellation and beamforming. In Sec. 2.3, the joint LCMV approach to acoustic echo cancellation and beamforming and its realization as a generalized sidelobe canceller (GSC) are presented. Sec. 2.4 outlines a practical realization based on the RGSC. Sec. 2.5 gives experimental results.

2.2 Concepts for Joint Acoustic Echo Cancellation and Adaptive Beamforming

We consider the scenario of an acoustic human/machine front-end with Q loudspeakers and a microphone array with M microphones. The microphones capture the desired speech signal of the user, interference from other sound sources, such as speech of other human talkers, and ambient noise, such as noise from air conditioning or from computer fans. We can thus identify two problems for the acoustic human/machine interface: acoustic echo cancellation for multiple reproduction channels and noise and interference suppression by microphone arrays.

In the following, we provide a concise overview of the problems and solutions for the individual tasks – acoustic echo cancellation (Sec. 2.2.1) and adaptive beamforming (Sec. 2.2.2). The integration into a joint system is discussed in Sec. 2.2.3.

2.2.1 Acoustic Echo Cancellation

With acoustic echo cancellation being considered from several points of view in this book (Chapters 5, 6, 7, and 8), we only review the main aspects of the general multichannel concept. The principle of multichannel acoustic echo cancellation is illustrated in Fig. 2.1. For simplicity, the multichannel AEC is shown only for a single recording channel.

The signals $x_q(n)$, $q = 0, 1, \ldots Q-1$, are played back by the Q loudspeakers and fed back to the microphones, where the signals $x_q(n)$ appear as acoustic echoes $d_q(n)$. With the assumption that the amplifiers and the transducers are linear, a linear model is commonly used for the echo paths between the loudspeaker signals $x_q(n)$ and the microphone signals $y(n)$. See, e.g., [38, 59, 87] and Chapter 7 of this book for the case where nonlinearities of the transducers and of the amplifiers cannot be neglected. To cancel the acoustic echoes in the microphone channel, adaptive filters $\hat{\boldsymbol{h}}_q(n)$, $q = 0, 1, \ldots Q-1$, are placed in parallel to the echo paths between the loudspeakers and the microphones with the loudspeaker signals $x_q(n)$ as references. The adaptive filters form replicas of the echo paths such that the output signals $\hat{d}_q(n)$ of the adaptive filters are replicas of the acoustic echoes. Subtracting the output signals of the adaptive filters from the microphone signal thus suppresses the acoustic echoes. Acoustic echo cancellation is thus a system identification problem, where the echo paths are usually identified by adaptive linear filtering. The design of the adaptation algorithm requires consideration of the nature of the echo paths and of the echo signals:

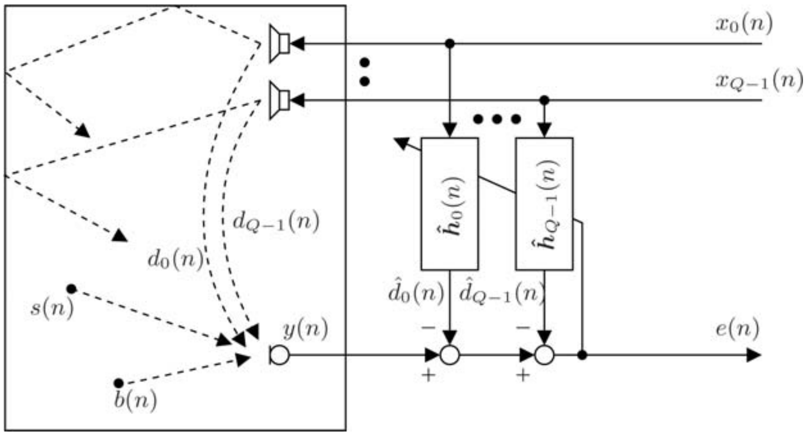

Fig. 2.1. Principle of multichannel acoustic echo cancellation with Q loudspeakers and a single microphone.

Time-variance of acoustic echo paths. The acoustic echo paths may vary strongly over time due to moving sources or changes in the acoustic environment requiring a good tracking performance of the adaptation algorithm [12].

Reverberation time of the acoustic environment. The reverberation time of the acoustic environment typically ranges from, e.g., $T_{60} \approx 50\,\mathrm{ms}$ in passenger cabins of vehicles to $T_{60} > 1\,\mathrm{s}$ in public halls. With

$$N_{\hat{h}} \approx \frac{ERLE}{60} f_\mathrm{s} T_{60} \,, \tag{2.1}$$

where $ERLE$ is the desired echo suppression of the AEC in dB [12], as a rule of thumb it becomes obvious that with many realistic acoustic environments and sampling rates $f_\mathrm{s} = 8 - 48\,\mathrm{kHz}$, FIR filters with several thousands coefficients are needed to achieve $ERLE \approx 20\,\mathrm{dB}$. For environments with long reverberation times, this means that the time for convergence – even for fast converging adaptation algorithms – cannot be neglected and that, after a change of the echo paths, noticeable residual echoes may be present until the adaptation algorithm has re-converged.

Auto- and cross-correlation of loudspeaker signals. The spatial sound impression in multi-loudspeaker systems is often artificially generated by weighting and delaying the spectrally colored source signals according to the position of the sources. This leads to a high auto- and cross-correlation of the loudspeaker signals [5, 14, 15, 30, 84]. With increasing auto- and cross-correlation and with an increasing number of reproduction channels, the condition number of the loudspeaker signals' correlation matrix increases, which reduces the rate of convergence of many adaptation algorithms in turn [43]. To reduce the auto-correlation of the loudspeaker signals, prewhitening filters can be applied [28, 95, 96]. Furthermore, to reduce the cross-correlation of the loudspeaker signals, inaudible nonlinearities [5, 30, 84] or inaudible time-varying filters can be introduced into the loudspeaker channels [2, 53] or inaudible noise with no correlation between the channels can be added to the loudspeaker signals [29, 34].

Double talk. The presence of disturbing sources such as desired speech, interference, or ambient noise may lead to instability and divergence of the adaptive filters. To prevent these instabilities, adaptation control mechanisms are required which adjust the step size of the adaptation algorithm to the present acoustic conditions [12, 42, 64]. With a decrease in the power ratio of acoustic echoes and disturbance a smaller step size becomes mandatory, which increases the time until the adaptive filters have converged to efficient echo path models.

As the discussion about adaptive filtering for acoustic echo cancellation shows, the convergence time of the adaptive filters is a crucial factor in acoustic echo cancellation and limits the performance of AECs in realistic acoustic environments. With the aim of reducing the convergence time while assuring robustness against instabilities and divergence even during double talk, various adaptation algorithms, such as the normalized least mean-squares (NLMS) algorithm, the affine projection algorithm, or the recursive least-squares (RLS) algorithm have been studied for realizations in the time-domain, in the DFT-domain, or in frequency subbands using filterbanks [8, 12, 32, 42, 56, 81, 83]. Acoustic echo cancellation in the DFT-domain or in frequency subbands has the advantage that sparseness of desired speech, interference, and noise can be exploited for selecting the step size of the adaptation algorithm differently for different frequencies as a function of the disturbance level to obtain faster convergence.

Even with fast converging adaptation algorithms, there are typically residual echoes present at the output of the AEC. Furthermore, it is desirable to combine the echo cancellation with noise reduction. Therefore, single-channel echo and noise reduction is often cascaded with the AEC to suppress residual echoes and noise at the AEC output [10, 26, 39, 40, 68, 69]. These methods are typically based on spectral subtraction or Wiener filtering [9, 61] so that estimates of the noise spectrum and of the spectrum of the acoustic echoes at the AEC output are required. These are often difficult to obtain in single-microphone systems for time-varying noise spectra and frequently changing echo paths.

2.2.2 Adaptive Beamforming

To overcome the limitations of single-channel noise reduction especially for interference and noise with time-varying spectra, beamforming with microphone arrays is promising for many applications as, thereby, the spatial domain supports separation of desired and undesired signals. In practical situations, source positions and signal characteristics change over time so that adaptive, data-dependent beamforming algorithms are preferable over fixed data-independent beamformers [92].

For speech and audio signal processing, adaptive data-dependent beamforming can be classified into LCMV beamforming, minimum mean-squared error (MMSE) beamforming, and maximum-a-posteriori (MAP) beamforming, disregarding special combinations with automatic speech recognition [78, 80].

LCMV beamforming [31,48,51,75]. In LCMV beamforming, with the GSC as one implementation, the variance of the output signal of the beamformer is minimized subject to constraints which prevent distortion of the desired signal. Estimates of the auto-power spectral densities (PSDs) and of the cross-power spectral densities (CPSDs) of interference and noise at the sensors are not required so that the efficient suppression of signals with highly time-varying spectra, such as speech signals, becomes possible. Adaptive differential microphone arrays [25, 89] are a special case of the LCMV beamformer.

However, reverberation of the acoustic environment w.r.t. the desired signal [18, 77, 94], moving desired sources, or array imperfections, such as position errors or gain and phase mismatch of the microphones [16, 35, 52, 97], may lead to distortion of the desired signal by the adaptive LCMV beamformer due to 'leakage' of the desired signal. To resolve this problem, the filter coefficients can be updated only when interference and noise are present [51, 74, 90, 93], quadratic [20, 33, 48, 50–52, 75, 88] or adaptive spatio-temporal constraints [31, 48, 51] can be used, or the speech distortion can be controlled directly [23, 86].

The suppression of ambient noise and 'cocktail-party' noise is limited due to the limited number of spatial degrees of freedom of the microphone array. To overcome this limitation for such noise scenarios, two methods have been proposed: First, LCMV beamformers can be combined with single-channel noise reduction ('post-filtering') [19, 65, 70, 71, 79, 82]. This leads to a structure that is basically equivalent to the MMSE beamformer [24, 82], but which exploits the advantages of the LCMV beamformer. Second, a spatial pre-processor in the structure of the GSC can be combined with single-channel noise reduction [72, 76] or with a so-called 'speech distortion weighted multichannel Wiener filter' [23, 86].

MMSE beamforming [1, 22, 23, 27, 79, 86]. MMSE beamforming is an extension of single-channel noise reduction to the multichannel case. In contrast to adaptive LCMV beamforming, multichannel MMSE beamformers are inherently robust against array imperfections and reverberation of the acoustic environment, so that the problem of cancellation of the desired signal due to signal leakage is avoided. However, the minimization of the mean squared error inherently allows for desire signal distortion which may not be acceptable for applications where high speech quality is required. Moreover, MMSE beamformers require estimates of the CPSDs of interference and noise at the sensors so that – at least from today's point of view– there is limited suppression of noise and interfering signals with highly time-varying PSDs.

MAP beamforming [62]. While the derivation of multichannel MMSE estimators is often difficult, multichannel MAP estimators often provide simpler mathematical descriptions. Thereby, multichannel MAP estimation allows, for example, the use of general statistical models, such as super-Gaussian probability density functions for speech and noise.

Adaptive data-dependent beamformers are generally realized using time-averaging over a finite temporal aperture to estimate the relevant statistics of the sensor data. For directly considering this temporal averaging in the optimization criterion of the MMSE beamformer, the term least-squares error (LSE) beamformer is used in [91]. Following [46, 48], we use in this work the term linearly-constrained least-squares error (LCLSE) beamformer for including this temporal averaging into the optimization criterion of the LCMV beamformer.

2.2.3 Joint Acoustic Echo Cancellation and Adaptive Beamforming

In this section, we briefly discuss solutions to the problem of joint acoustic echo cancellation and adaptive beamforming which were presented previously in the literature, namely 'AEC first', 'beamformer first', AEC integrated into the GSC ('GSAEC'), and a joint system of 'AEC first' and 'beamformer first'.

'AEC first' [13, 21, 44, 48, 54, 57, 58, 66]. The AECs can be captured by a matrix of time-variant impulse responses $\hat{\boldsymbol{H}}(n)$ in the sensor channels. This matrix $\hat{\boldsymbol{H}}(n)$ directly models the echo path between all loudspeakers and all microphones, without interaction with the beamforming (Fig. 2.2). For the adaptive beamformer described by a vector of time-variant impulse responses $\boldsymbol{w}(n)$, positive synergies can be exploited after convergence of the AECs: The acoustic echoes are efficiently suppressed by the AECs, and the adaptive beamformer $\boldsymbol{w}(n)$ does not depend on the echo signals. Thus, all degrees of freedom of the beamformer are available for the suppression of interference and noise. Obviously, one AEC is necessary for each sensor channel so that an M-fold complexity, where M is the number of microphones, is required at least for the filtering and for the filter update in comparison to AEC for a single microphone [57]. Even with a moderate number of microphones ($4 \leq M \leq 8$), this is a limiting factor for the use of 'AEC first' in cost-sensitive systems. Moreover, in the presence of strong interference and noise, the adaptation of the AECs must be slowed down or even stopped in order to avoid instabilities of the adaptive filters $\hat{\boldsymbol{H}}(n)$. This reduces the tracking capability and, consequently, the efficiency of the AECs for frequently changing echo paths. Limited echo suppression of the AECs, however, limits the positive synergies with the adaptive beamformer so that the performance improvement of 'AEC first' relative to an adaptive beamformer alone strongly depends on the acoustic environment.

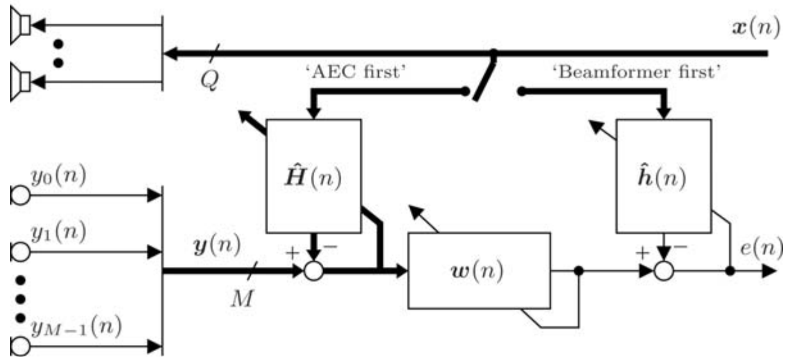

Fig. 2.2. Combinations of AEC and beamforming [58, 66].

'Beamformer first' [4, 44, 48, 54, 57, 58, 66]. Alternatively, the AEC can be placed behind the adaptive beamformer (Fig. 2.2). Obviously, the complexity is reduced to that of AEC for a single microphone. However, positive synergies cannot be exploited for the adaptive beamformer, since the beamformer always 'sees' not only interference but also acoustic echoes. On the other hand, the AEC captured in a vector of time-variant impulse responses $\hat{\boldsymbol{h}}(n)$ generally cannot track the relatively fast time-variance of $\boldsymbol{w}(n)$, which results from the dependency of $\boldsymbol{w}(n)$ on the time-varying spectra of the sensor signals and the generally smaller number of filter taps of $\boldsymbol{w}(n)$ relative to $\hat{\boldsymbol{h}}(n)$ [48].

AEC integrated into the GSC (GSAEC). Another solution would be to integrate acoustic echo cancellation and adaptive beamforming so that the AEC does not depend on the time-variance of the adaptive beamformer [58]. One option, which is based on the structure of the GSC [37] (see Sec. 2.3.2), was proposed in [44]. For this so-called GSAEC, the AEC is placed in the reference path behind the quiescent weight vector $\boldsymbol{w}_{\mathrm{c}}$ of the GSC so that the AEC is independent of the time-varying sidelobe-cancelling path (Fig. 2.3), which consists of the blocking matrix $\boldsymbol{B}(n)$ and the interference canceller $\hat{\boldsymbol{h}}(n)$.

However, acoustic echoes may leak through the sidelobe-cancelling path although they may be efficiently suppressed by the AEC in the reference path, so that the overall performance of 'AEC first' cannot be expected. Moreover, analogously to 'AEC first', the performance of this integrated system is limited for strong interference and noise or for frequently changing echo paths.

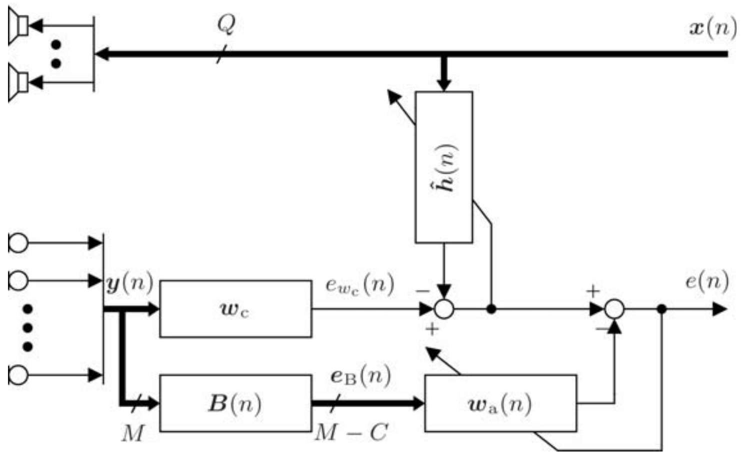

Fig. 2.3. AEC integrated into the GSC (GSAEC) [44].

To overcome the problems of these structures in environments with frequently changing echo paths, frequent double talk, interference, and background noise, we study here the joint optimization of adaptive beamforming and acoustic echo cancellation. We focus on an LCLSE optimization criterion to derive the beamformer weight vector. MMSE/LSE and MAP criteria are not considered since they require estimates of the interference spectra at the microphones, which are difficult to obtain for mixtures of non-stationary signals.

2.3 Joint Optimization of Acoustic Echo Cancellation and Adaptive Beamforming

In contrast to 'beamformer first' in Fig. 2.2, where different signals are used to optimize $\boldsymbol{w}(n)$ and the AEC $\hat{\boldsymbol{h}}(n)$, we propose to use the output signal $e(n)$ to optimize both AEC and the adaptive beamformer as shown in Fig. 2.4. The reference loudspeaker signals $\boldsymbol{x}(n)$ can thus be interpreted as additional input signals for the adaptive beamformer. This idea was first used in [21] for a combination of acoustic echo cancellation and multichannel noise-reduction based on the generalized singular value decomposition (GSVD). In [63], a similar approach is used for a combination of blind source separation with acoustic echo cancellation.

We assume that the sensor signals $\boldsymbol{y}(n)$ are given by the superposition of the desired signal $\boldsymbol{s}(n)$, interference and noise $\boldsymbol{b}(n)$, and acoustic echoes $\boldsymbol{d}(n)$,

$$\boldsymbol{y}(n) = \boldsymbol{s}(n) + \boldsymbol{b}(n) + \boldsymbol{d}(n)\,, \tag{2.2}$$

where $\boldsymbol{s}(n)$, $\boldsymbol{b}(n)$, and $\boldsymbol{d}(n)$ are zero-mean and mutually uncorrelated. The output signal $e(n)$ of the combined system can be written as a function of

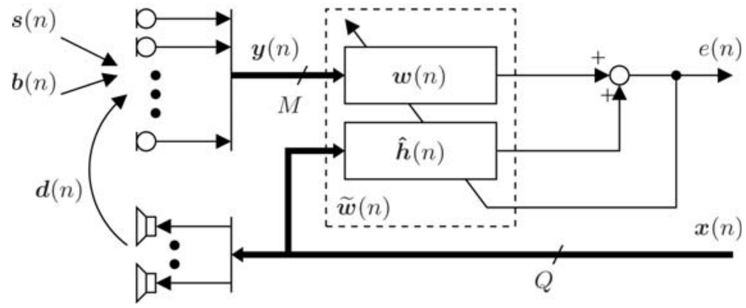

Fig. 2.4. Joint optimization of adaptive beamforming and acoustic echo cancellation.

the sensor signals $\boldsymbol{y}(n)$, the loudspeaker signals $\boldsymbol{x}(n)$, the stacked beamformer weight vector $\boldsymbol{w}(n)$, and the stacked AEC weight vector $\hat{\boldsymbol{h}}(n)$ as

$$e(n) = \boldsymbol{w}^{\mathrm{T}}(n)\boldsymbol{y}(n) + \hat{\boldsymbol{h}}^{\mathrm{T}}(n)\boldsymbol{x}(n)\,, \tag{2.3}$$

where

$$\boldsymbol{y}(n) = \left[\boldsymbol{y}_0^{\mathrm{T}}(n),\, \boldsymbol{y}_1^{\mathrm{T}}(n),\, \ldots,\, \boldsymbol{y}_{M-1}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{2.4}$$

$$\boldsymbol{y}_m(n) = \left[y_m(n),\, y_m(n-1),\, \ldots,\, y_m(n-N_w+1)\right]^{\mathrm{T}}, \tag{2.5}$$

$$\boldsymbol{x}(n) = \left[\boldsymbol{x}_0^{\mathrm{T}}(n),\, \boldsymbol{x}_1^{\mathrm{T}}(n),\, \ldots,\, \boldsymbol{x}_{Q-1}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{2.6}$$

$$\boldsymbol{x}_q(n) = \left[x_q(n),\, x_q(n-1),\, \ldots,\, x_q(n-N_{\hat{h}}+1)\right]^{\mathrm{T}}, \tag{2.7}$$

$$\boldsymbol{w}(n) = \left[\boldsymbol{w}_0^{\mathrm{T}}(n),\, \boldsymbol{w}_1^{\mathrm{T}}(n),\, \ldots,\, \boldsymbol{w}_{M-1}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{2.8}$$

$$\boldsymbol{w}_m(n) = \left[w_{0,m}(n),\, w_{1,m}(n),\, \ldots,\, w_{N_w-1,m}(n)\right]^{\mathrm{T}}, \tag{2.9}$$

$$\hat{\boldsymbol{h}}(n) = \left[\hat{\boldsymbol{h}}_0^{\mathrm{T}}(n),\, \hat{\boldsymbol{h}}_1^{\mathrm{T}}(n),\, \ldots,\, \hat{\boldsymbol{h}}_{Q-1}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{2.10}$$

$$\hat{\boldsymbol{h}}_q(n) = \left[\hat{h}_{0,q}(n),\, \hat{h}_{1,q}(n),\, \ldots,\, \hat{h}_{N_{\hat{h}}-1,q}(n)\right]^{\mathrm{T}}. \tag{2.11}$$

N_w and $N_{\hat{h}}$ are the number of filter coefficients of the beamformer weight vectors $\boldsymbol{w}_m(n)$ and of the AEC filters $\hat{\boldsymbol{h}}_q(n)$, respectively. With stacked vectors

$$\widetilde{\boldsymbol{w}}(n) = \left[\boldsymbol{w}^{\mathrm{T}}(n),\, \hat{\boldsymbol{h}}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{2.12}$$

$$\widetilde{\boldsymbol{x}}(n) = \left[\boldsymbol{y}^{\mathrm{T}}(n),\, \boldsymbol{x}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{2.13}$$

we can write $e(n)$ as

$$e(n) = \widetilde{\boldsymbol{w}}^{\mathrm{T}}(n)\widetilde{\boldsymbol{x}}(n)\,. \tag{2.14}$$

which reflects that the AEC input signals $\boldsymbol{x}(n)$ and the AEC filters $\hat{\boldsymbol{h}}(n)$ can be interpreted as additional channels of a beamformer $\widetilde{\boldsymbol{w}}(n)$.

2.3.1 Linearly-Constrained Least-Squares Error (LCLSE) Minimization

An LCLSE optimization criterion is obtained when we aim at minimizing the windowed sum of squared output signal samples $e^2(n)$ subject to constraints which assure that the desired signal is not distorted by $\widetilde{\boldsymbol{w}}(n)$. That is,

$$\min_{\widetilde{\boldsymbol{w}}(n)} \sum_{i=0}^{n} g_i(n)\, e^2(i) \quad \text{subject to} \quad \widetilde{\boldsymbol{C}}^{\mathrm{T}}(n)\widetilde{\boldsymbol{w}}(n) = \boldsymbol{c}(n)\,. \tag{2.15}$$

The windowing function $g_i(n)$ extracts desired samples from the output signal $y(n)$ which should be included into the optimization.[3] For example, infinite memory with exponential decay is obtained with $g_i(n) = \lambda^{n-i}$ [43]. The constraint matrix $\widetilde{\boldsymbol{C}}(n)$ of size $(MN_w + QN_{\hat{h}}) \times C$ and the constraint column vector $\boldsymbol{c}(n)$ of length C put C spatial constraints onto $\widetilde{\boldsymbol{w}}(n)$ in order to assure unity beamformer response for the direction-of-arrival of the desired signal [91]. Since the Q loudspeaker signals $\boldsymbol{x}(n)$ can safely be assumed to be orthogonal to the desired signal, the constraints are only required for the microphone signals, just as for conventional LCMV beamformers [91]. We can thus write $\widetilde{\boldsymbol{C}}(n)$ as

$$\widetilde{\boldsymbol{C}}(n) = \left[\boldsymbol{C}^{\mathrm{T}}(n),\, \mathbf{0}_{C \times QN_{\hat{h}}}\right]^{\mathrm{T}}, \tag{2.16}$$

where $\boldsymbol{C}(n)$ of size $MN_w \times C$ is a conventional constraint matrix known from LCMV beamforming [91]. We thus obtain with Eq. 2.15 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 us to update the beamformer and the AEC simultaneously without reducing the step size for the AEC – in contrast to the previously discussed combinations, where the adaptation of the AEC at least has to be slowed down if interference, noise, or the desired signal are active. Thereby, the structural problems for tracking in 'AEC first' and the leakage in GSAEC can be avoided. 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. Due to the correlation of $\boldsymbol{y}(n)$ and $\boldsymbol{x}(n)$, however, it must be expected that the conditioning of the optimization problem is worsened relative to the individual optimization problems.

---

[3] The corresponding LCMV optimization criterion is obtained by replacing the windowed sum of squared output signal samples $e^2(n)$ by the expected value of $e^2(n)$. The solution of the LCMV optimization criterion is analogous to that of the LCLSE criterion shown here.

2.3.2 Realization as a Generalized Sidelobe Canceller (GSC)

A direct solution of Eq. 2.15 can be determined using Lagrange multipliers [91]. 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 [17, 37].

To obtain the GSC, the stacked weight vector $\widetilde{\boldsymbol{w}}(n)$ is projected onto two orthogonal subspaces,

$$\widetilde{\boldsymbol{w}}(n) = \Big[\boldsymbol{P}_{\mathrm{c}}(n) + \boldsymbol{P}_{\mathrm{a}}(n)\Big]\, \widetilde{\boldsymbol{w}}(n)\,. \tag{2.17}$$

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

$$\widetilde{\boldsymbol{C}}^{\mathrm{T}}(n)\widetilde{\boldsymbol{w}}_{\mathrm{c}}(n) \overset{!}{=} \boldsymbol{c}(n)\,. \tag{2.18}$$

From (2.16), it follows that $\widetilde{\boldsymbol{w}}_{\mathrm{c}}(n)$ can be chosen as

$$\widetilde{\boldsymbol{w}}_{\mathrm{c}}(n) = \Big[\boldsymbol{w}_{\mathrm{c}}^{\mathrm{T}}(n),\, \boldsymbol{0}_{1\times QN_{\hat{h}}}\Big]^{\mathrm{T}} \tag{2.19}$$

in order to fulfill Eq. 2.18. The weight vector $\boldsymbol{w}_{\mathrm{c}}(n)$ of size $MN_w \times 1$ is known as quiescent weight vector [91]. The quiescent weight vector $\boldsymbol{w}_{\mathrm{c}}(n)$ steers the sensor array to the position of the desired source and enhances the desired signal relative to interference and noise (Fig. 2.5).[4]

The second (orthogonal) subspace is chosen as

$$\boldsymbol{P}_{\mathrm{a}}(n)\, \widetilde{\boldsymbol{w}}(n) := -\widetilde{\boldsymbol{B}}(n)\, \widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)\,, \tag{2.20}$$

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

$$\widetilde{\boldsymbol{C}}^{\mathrm{T}}(n)\, \widetilde{\boldsymbol{B}}(n) \overset{!}{=} \boldsymbol{0}\,. \tag{2.21}$$

The cascade of $\widetilde{\boldsymbol{B}}(n)$ and $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ is termed the sidelobe-cancelling path [37]. From Eq. 2.16, it may be seen that Eq. 2.21 is met for

$$\widetilde{\boldsymbol{B}}(n) = \begin{bmatrix} \boldsymbol{B}(n) & \boldsymbol{0}_{MN_w\times QN_{\hat{h}}} \\ \boldsymbol{0}_{QN_{\hat{h}}\times(M-C)N_{\mathrm{w_a}}} & \boldsymbol{I}_{QN_{\hat{h}}\times QN_{\hat{h}}} \end{bmatrix}\,, \tag{2.22}$$

where $\boldsymbol{I}_{QN_{\hat{h}}\times QN_{\hat{h}}}$ is the identity matrix of size $QN_{\hat{h}} \times QN_{\hat{h}}$ and where $\boldsymbol{B}(n)$ meets $\boldsymbol{C}^{\mathrm{T}}(n)\boldsymbol{B}(n) = \boldsymbol{0}$. Since the constrained subspace generally contains the desired signal, the matrix $\boldsymbol{B}(n)$, which fulfills the requirement that the second subspace is orthogonal to the constrained subspace, suppresses desired signal

[4] Note that we used in Fig. 2.3 for the GSAEC structure a fixed quiescent weight vector. This assumption is relaxed here for generality of the derivation.

components. Therefore, the matrix $\boldsymbol{B}(n)$ is generally referred to as a blocking matrix [91]. The identity matrix assures that acoustic echoes are not cancelled by $\widetilde{\boldsymbol{B}}(n)$. As a consequence, ideally only acoustic echoes, interference, and noise are present at the output of $\widetilde{\boldsymbol{B}}(n)$, so that the weight vector $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ can be determined by unconstrained LS minimization of $e(n)$,

$$\min_{\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)} \sum_{i=0}^{n} g_i(n) \left[\left(\widetilde{\boldsymbol{w}}_{\mathrm{c}}(n) - \widetilde{\boldsymbol{B}}(n) \widetilde{\boldsymbol{w}}_{\mathrm{a}}(n) \right)^{\mathrm{T}} \widetilde{\boldsymbol{x}}(i) \right]^2 . \tag{2.23}$$

Introducing Eqs. 2.19 and 2.22 into Eq. 2.23 and identifying the result with Eq. 2.3, it may be seen that $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ is equivalent to a stacked weight vector consisting of a weight vector $\boldsymbol{w}_{\mathrm{a}}(n)$ and of the AEC $\hat{\boldsymbol{h}}(n)$,

$$\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n) := \left[\boldsymbol{w}_{\mathrm{a}}^{\mathrm{T}}(n),\, \hat{\boldsymbol{h}}^{\mathrm{T}}(n) \right]^{\mathrm{T}} . \tag{2.24}$$

We obtain for the output signal $e(n)$ the expression

$$e(n) = \left[\boldsymbol{w}_{\mathrm{c}}(n) - \boldsymbol{B}(n)\, \boldsymbol{w}_{\mathrm{a}}(n) \right]^{\mathrm{T}} \boldsymbol{y}(n) - \hat{\boldsymbol{h}}^{\mathrm{T}}(n)\, \boldsymbol{x}(n) , \tag{2.25}$$

which can be put into the structure depicted in Fig. 2.5. The combined system thus corresponds to the GSC, where $\boldsymbol{w}_{\mathrm{a}}(n)$ is combined with the AEC $\hat{\boldsymbol{h}}(n)$, and where the loudspeaker signals $\boldsymbol{x}(n)$ are used as additional channels of the sidelobe-cancelling path. $\boldsymbol{w}_{\mathrm{a}}(n)$ is generally called an interference canceller since $\boldsymbol{w}_{\mathrm{a}}(n)$ is optimized to cancel interference and noise at the output of the GSC. Analogously, we refer to $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ as the 'echo and interference canceller' (EIC) and to the combined system of AEC and GSC as the 'generalized echo and interference canceller' (GEIC).

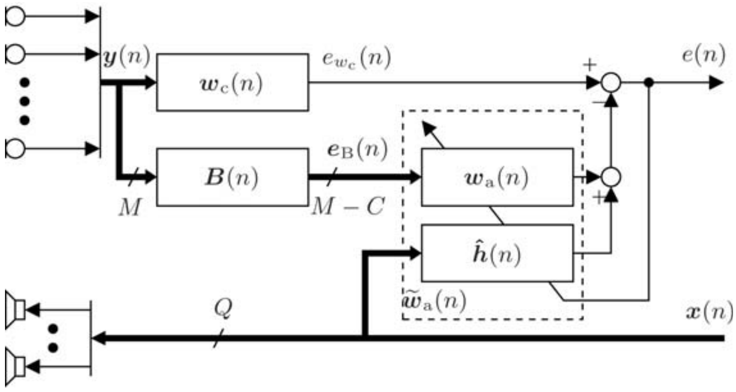

Fig. 2.5. Generalized echo and interference canceller (GEIC).

The optimum weight vector $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ is now obtained by setting the derivative of Eq. 2.23 w.r.t. $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ equal to zero and by solving the obtained system of linear equations for $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$:

$$\widetilde{\boldsymbol{w}}_{\mathrm{a,opt}}(n) = \left[\widetilde{\boldsymbol{B}}^{\mathrm{T}}(n)\,\widetilde{\boldsymbol{\Phi}}(n)\,\widetilde{\boldsymbol{B}}(n)\right]^{+}\widetilde{\boldsymbol{B}}^{\mathrm{T}}(n)\,\widetilde{\boldsymbol{\Phi}}(n)\,\widetilde{\boldsymbol{w}}_{\mathrm{c}}(n)\,, \tag{2.26}$$

$$\widetilde{\boldsymbol{\Phi}}(n) = \sum_{i=0}^{n} g_i(n)\,\widetilde{\boldsymbol{x}}(i)\,\widetilde{\boldsymbol{x}}^{\mathrm{T}}(i) = \begin{bmatrix} \boldsymbol{\Phi}_{yy}(n) & \boldsymbol{\Phi}_{yx}(n) \\ \boldsymbol{\Phi}_{xy}(n) & \boldsymbol{\Phi}_{xx}(n) \end{bmatrix}. \tag{2.27}$$

The $(\cdot)^{+}$ is the pseudoinverse of a matrix, and $\widetilde{\boldsymbol{\Phi}}(n)$ is the sample correlation matrix of the stacked data vector $\widetilde{\boldsymbol{x}}(n)$ [43] for a given windowing function $g_i(n)$. As shown in (2.27), $\widetilde{\boldsymbol{\Phi}}(n)$ can be decomposed into the submatrices

$$\boldsymbol{\Phi}_{yy}(n) = \sum_{i=0}^{n} g_i(n)\,\boldsymbol{y}(i)\,\boldsymbol{y}^{\mathrm{T}}(i)\,, \tag{2.28}$$

$$\boldsymbol{\Phi}_{xx}(n) = \sum_{i=0}^{n} g_i(n)\,\boldsymbol{x}(i)\,\boldsymbol{x}^{\mathrm{T}}(i)\,, \tag{2.29}$$

$$\boldsymbol{\Phi}_{yx}(n) = \sum_{i=0}^{n} g_i(n)\,\boldsymbol{y}(i)\,\boldsymbol{x}^{\mathrm{T}}(i)\,, \tag{2.30}$$

$$\boldsymbol{\Phi}_{xy}(n) = \boldsymbol{\Phi}_{yx}^{\mathrm{T}}(n)\,, \tag{2.31}$$

with the sample correlation matrix of the sensor signals $\boldsymbol{\Phi}_{yy}(n)$, the sample correlation matrix of the loudspeaker signals $\boldsymbol{\Phi}_{xx}(n)$, and the sample cross-correlation matrices between the sensor signals and the loudspeaker signals $\boldsymbol{\Phi}_{xy}(n)$ and $\boldsymbol{\Phi}_{yx}(n)$, respectively. The solution of the optimum weight vector $\widetilde{\boldsymbol{w}}_{\mathrm{a,opt}}(n)$ is formally equivalent to the optimum weight vector of the GSC [17]. Finally introducing Eqs. 2.19, 2.22, and 2.27 into Eqs. 2.26, 2.26 can be written as

$$\begin{bmatrix} \boldsymbol{w}_{\mathrm{a,opt}}(n) \\ \hat{\boldsymbol{h}}_{\mathrm{opt}}(n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^{\mathrm{T}}(n)\,\boldsymbol{\Phi}_{yy}(n)\,\boldsymbol{B}(n) & \boldsymbol{B}^{\mathrm{T}}(n)\,\boldsymbol{\Phi}_{yx}(n) \\ \boldsymbol{\Phi}_{xy}(n)\,\boldsymbol{B}(n) & \boldsymbol{\Phi}_{xx}(n) \end{bmatrix}^{+} \times \begin{bmatrix} \boldsymbol{B}^{\mathrm{T}}(n)\,\boldsymbol{\Phi}_{yy}(n)\,\boldsymbol{w}_{\mathrm{c}}(n) \\ \boldsymbol{\Phi}_{yx}(n)\,\boldsymbol{w}_{\mathrm{c}}(n) \end{bmatrix}. \tag{2.32}$$

Because of the structural equivalence of the GEIC to the GSC, any implementation of the GSC can be used to realize the GEIC. Especially, any linear constraints can be used for designing the quiescent weight vector and the blocking matrix. Furthermore, the echo and interference canceller can be calculated directly employing Eq. 2.32 or iteratively using recursive adaptation algorithms [48,91]. With regard to practical realizations, the matrix inversion in Eq. 2.32 can be avoided by using recursive adaptation algorithms, and, thus, the computational complexity can be reduced.

For the GSC, the number of filter taps $N_{\mathrm{w_a}}$ is generally chosen such that fast convergence of $\boldsymbol{w}_{\mathrm{a}}(n)$ is assured. Typically, $N_{\mathrm{w_a}} = 64 \ldots 512$ for an $f_{\mathrm{s}} = 8\,\mathrm{kHz}$ sampling rate independently of the reverberation time T_{60} of the acoustic environment [48]. The number of filter taps $N_{\hat{h}}$ of the AEC $\hat{\boldsymbol{h}}(n)$, however, is typically chosen as a function of the reverberation time T_{60}, and is typically $N_{\hat{h}} = 256 \ldots 2048$ for $T_{60} = 0.05 \ldots 0.5\,\mathrm{s}$ (see Eq. 2.1). In most cases, the number of filter taps $N_{\hat{h}}$ should thus be greater than $N_{w_{\mathrm{a}}}$ depending on the reverberation time of the acoustic environment (typically $T_{60} \geq 100\,\mathrm{ms}$) in order to assure optimum performance of $\boldsymbol{w}_{\mathrm{a}}(n)$ and $\hat{\boldsymbol{h}}(n)$. However, different numbers of filter taps are problematic for the convergence behavior of $\widetilde{\boldsymbol{w}}_{\mathrm{a}}(n)$ for a time-varying sample correlation matrix $\widetilde{\boldsymbol{\Phi}}(n)$, since the convergence speed of $\boldsymbol{w}_{\mathrm{a}}(n)$ differs from that of $\hat{\boldsymbol{h}}(n)$. Consider as an extreme case $N_{\hat{h}} \rightarrow \infty$: Then, the convergence speed of $\hat{\boldsymbol{h}}(n)$ tends to zero, which yields inefficiency of the AEC. It is thus necessary to limit $N_{\hat{h}}$ to $N_{w_{\mathrm{a}}}$. This may reduce the performance of GEIC relative to 'AEC first' in situations where 'AEC first' does not exhibit tracking or adaptation problems as, for example, for presence of weak interference and noise and/or for slowly time-varying acoustic echo paths. The influence of the acoustic environment on the performance of GEIC will be investigated experimentally in Sec. 2.5.

2.3.3 Simplification to General Sidelope Acoustic Echo Canceller (GSAEC)

The joint optimization of $\hat{\boldsymbol{h}}(n)$ and $\boldsymbol{w}_{\mathrm{a}}(n)$ introduces the off-diagonal matrices into the first correlation matrix on the right side of (2.32). Setting the off-diagonal matrices equal to zero corresponds to separate optimization of $\hat{\boldsymbol{h}}(n)$ and $\boldsymbol{w}_{\mathrm{a}}(n)$, which yields for the optimum weight vector:

$$\boldsymbol{w}_{\mathrm{a,opt}}(n) = \left[\boldsymbol{B}^{\mathrm{T}}(n)\,\boldsymbol{\Phi}_{yy}(n)\,\boldsymbol{B}(n)\right]^{+} \boldsymbol{B}^{\mathrm{T}}(n)\,\boldsymbol{\Phi}_{yy}(n)\,\boldsymbol{w}_{\mathrm{c}}(n)\,, \tag{2.33}$$

$$\hat{\boldsymbol{h}}_{\mathrm{opt}}(n) = \boldsymbol{\Phi}_{xx}^{+}(n)\,\boldsymbol{\Phi}_{xy}(n)\,\boldsymbol{w}_{\mathrm{c}}(n)\,. \tag{2.34}$$

It may be noticed that Eq. 2.33 corresponds to the LS solution of a GSC interference canceller [17] and that Eq. 2.34 is equivalent to the LS solution of an AEC which is located after the quiescent weight vector. Eqs. 2.33 and 2.34 can thus be described by the system depicted in Fig. 2.3, which is recognized as the structure of the GSAEC [44].

Independent optimization of the GSC and of the AEC after the quiescent weight vector allows to choose the number of filter taps of the interference canceller, $N_{w_{\mathrm{a}}}$, and of the AEC, $N_{\hat{h}}$, independently so that the coupling problems of the echo and interference canceller of GEIC can be avoided. However, [48] describes in detail in that efficient cancellation of the acoustic echoes in the reference path of the GSC leads to leakage of acoustic echoes through the sidelobe-cancelling path of the GSC so that the performance of GSAEC is reduced relative to 'AEC first'. Moreover, for the presence of strong interference and noise and/or time-varying echo paths, GSAEC exhibits the same convergence problems as 'AEC first'. Experimental results can be found in Sec. 2.5.

2.4 Implementation

In this section, we describe the practical implementation of the joint acoustic echo cancellation and adaptive beamforming systems examined experimentally in Sec. 2.5, namely GEIC, 'AEC first', GSAEC, and GSC. For all joint acoustic echo cancellation and adaptive beamforming systems, the beamformer is realized as a GSC with an adaptive blocking matrix (RGSC, Sec. 2.4.1). The AEC is implemented as a stereophonic AEC (Sec. 2.4.2). A detailed description including parameter setting can be found in [48].

2.4.1 Robust Generalized Sidelobe Canceller (RGSC)

To realize the adaptive beamformer, it is crucial to obtain (a) tracking of moving sources with time-varying spectra and (b) robustness against cancellation of the desired signal due to reverberation, source movements, and array imperfections. To solve these problems, we choose the RGSC in the discrete Fourier transform (DFT) domain [48] with an adaptive blocking matrix [51] as the adaptive beamformer, as depicted in Fig. 2.6.

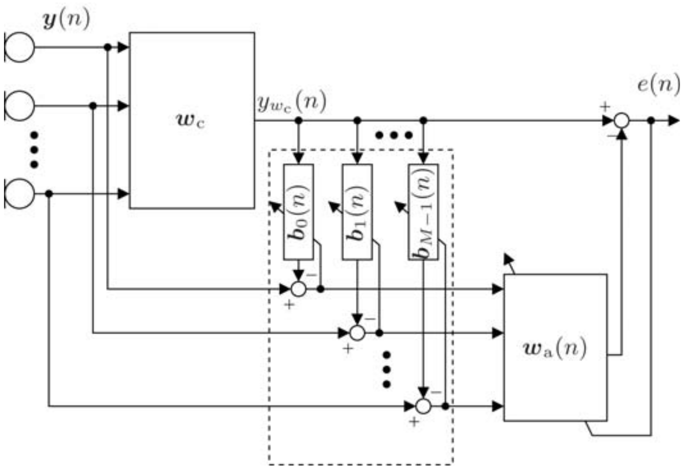

Fig. 2.6. GSC with an adaptive blocking matrix after [51].

For adaptation of the blocking matrix and of the interference canceller, we use computationally efficient multichannel DFT-domain adaptive filters (MC-FDAFs) [6, 7, 15]. Their RLS-like convergence behavior leads to fast convergence and they allow for a frequency-selective adaptation to exploit sparseness of the sensor signals.

2.4.1.1 Quiescent weight vector

The quiescent weight vector is realized as a fixed beamformer $\boldsymbol{w}_{\mathrm{c}}(n) := \boldsymbol{w}_{\mathrm{c}}$. We thus assume that the position of the desired speaker is roughly known, as it can be safely assumed for, for example, laptop PCs or personal digital assistants (PDAs). The width of the mainlobe of the quiescent weight vector needs to be adjusted to the expected variations of the source position.

2.4.1.2 Blocking Matrix

The blocking matrix is realized by adaptive filters $\boldsymbol{b}_m(n)$ between the output of the time-invariant quiescent beamformer $\boldsymbol{w}_{\mathrm{c}}$ and each of the inputs of the interference canceller $\boldsymbol{w}_{\mathrm{a}}(n)$. The adaptive filters $\boldsymbol{b}_m(n)$ use the output of $\boldsymbol{w}_{\mathrm{c}}$ as a reference for the desired signal and subtract the desired signal from the sidelobe-cancelling path. Orthogonality of the reference path and of the sidelobe-cancelling path is thus assured for the desired signal. Since the quiescent beamformer cannot produce an estimate of the desired signal that is free of interference, the filters $\boldsymbol{b}_m(n)$ should only be adapted when the signal-to-interference ratio (SIR) is high in order to prevent suppression of the interference by the blocking matrix [51, 90].

In [48], the adaptive blocking matrix is formally linked to LCLSE beamforming and to the derivation of the GSC in Sec. 2.3.2.

Realization of the blocking matrix by adaptive filters yields greater robustness against distortion of the desired signal than fixed realizations [31, 48, 51, 90]: For the GSC, the distortion results from the interference canceller, which cancels desired signal components leaking through the blocking matrix due to inherent mismatched constraints. The inherent mismatch results from possible array imperfections and especially from the fact that the required exact spatio-temporal information for the desired signal is not perfectly given. Adaptive filters, however, allow tracking of time-varying propagation for the desired source and time-varying array imperfections so that the desired signal is efficiently cancelled by the blocking matrix.

2.4.1.3 Interference Canceller

The interference canceller $\boldsymbol{w}_{\mathrm{a}}(n)$ adaptively subtracts the signal components from the reference path, which are correlated with the output signals of the blocking matrix. However, the blocking matrix – due to limited convergence speed, limited tracking capability, and limited number of filter coefficients – generally does not produce an estimate of the interference which is perfectly free of the desired signal. Therefore, the interference canceller (1) is realized using a (usually quadratic) norm constraint [20, 33, 48, 50–52, 75, 88] and (2) is only adapted when the SIR is low in order to maximally prevent distortion of the desired signal [51, 74, 90, 93].

2.4.1.4 Adaptation Control

The blocking matrix and the interference canceller cannot be adapted simultaneously but should only be adapted when the SIR is high and low, respectively. By exploiting sparseness in the spectra of desired speech and interference, i.e., by considering individual frequency components separately, the blocking matrix and the interference canceller can be adapted more often than adaptation in the fullband enabling a better tracking capability and a better convergence speed to be obtained. Experiments show that the exploitation of sparseness is also necessary for the interference canceller of the RGSC to track the time variance of the adaptive blocking matrix and to efficiently suppress non-stationary interference [45, 48].

Obviously, to exploit the sparseness, an activity detector is required, which detects 'desired signal only' (adaptation of the blocking matrix), 'interference only' (adaptation of the interference canceller), and 'double talk' (no adaptation) in discrete frequency bins [48].

2.4.2 Acoustic Echo Canceller

The design of the AEC – as long as it is realized independently of the beamformer – requires consideration of the tracking performance, of the convergence speed, and of the robustness against double talk. For the joint adaptation of the AEC and beamformer, acoustic echoes can simply be interpreted as additional interference, and these aspects do not need to be explicitly taken into account. However, at the EIC input, the variance of the output signals of the blocking matrix needs to be adjusted to the variance of the loudspeaker signals by an automatic gain control to have similar signal levels.

Especially because of the high convergence speed with moderate computational complexity, we employ MC-FDAFs to realize the AECs. Adaptation of the AECs of 'AEC first' and of GSAEC is controlled by a double talk detector based on a shadow filter [85] with a constant frequency-independent step size during adaptation. The GEIC is realized as an RGSC with additional channels of the interference canceller for the AECs. For the experiments described below, the time-averaged variance of the loudspeaker signals is manually adjusted to the time-averaged variance of the blocking matrix output signals. For all structures, the loudspeaker signals are de-cross-correlated by a simple time-invariant nonlinearity to increase the convergence speed of the adaptive filters [5].

2.4.3 Computational Complexity

The computational complexity of GEIC is compared to that of 'AEC first', GEIC, GSAEC, and RGSC in Fig. 2.7 as a function of the filter length $N_{\hat{h}}$ of the AEC for $M = 4$ microphones (Fig. 2.7a) and $M = 8$ microphones

(Fig. 2.7b) for a stereophonic AEC. The filter length of the interference canceller is $N_{w_a} = 256$ for all systems. For GEIC, the filter length of the AEC is adjusted to the filter length of the interference canceller, i.e., $N_{w_a} = N_{\hat{h}} = 256$. The adaptation control of the AEC and of the RGSC is not taken into account. Furthermore, the filter length N_{w_a} is not changed since experimental results in environments with various reverberation times show that the optimum filter length does not change with the reverberation time in our implementations of the RGSC and the GEIC. The computational complexity is measured as 'real-valued multiplications per output sample' (NRM)[5]. Comparing Fig. 2.7a with Fig. 2.7b, roughly speaking, it may be noticed that doubling the number of sensors doubles NRM. The relative complexity reduction from 'AEC first' to GEIC rises with increasing $N_{\hat{h}}$: For $N_{w_a} = N_{\hat{h}} = 256$, the relative complexity reduction from 'AEC first' to GEIC is 21 % for $M = 4$ and 25 % for $M = 8$, while 59 % ($M = 4$) and 37 % ($M = 8$) for $N_a = 2048$. Obviously, the complexity of the RGSC dominates the complexity of the additional AECs for 'AEC first'.

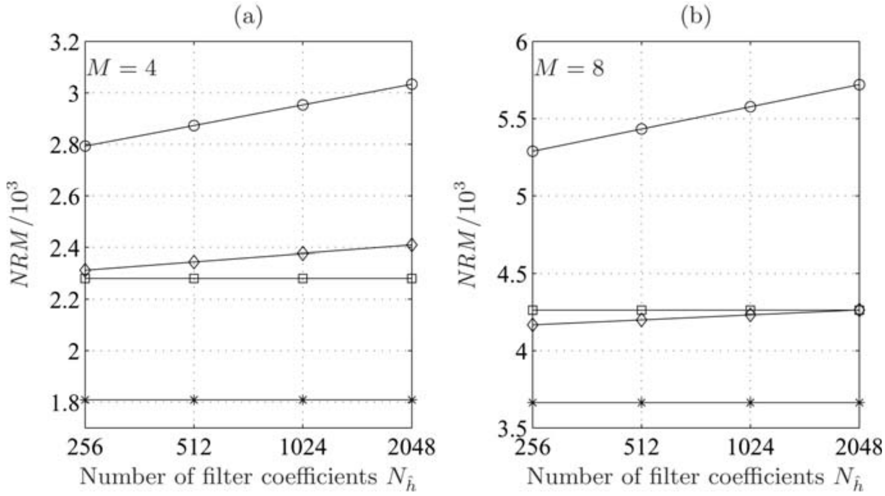

Fig. 2.7. Comparison of the number of real multiplications per sample (NRM) of '○' 'AEC first', '□' GEIC, '◇' GSAEC, and '∗' RGSC for (a) $M = 4$ and for (b) $M = 8$ ($N_{w_a} = 256$ and for GEIC $N_{\hat{h}} = 256$).

[5] The results differ from the results in [48], since, here, NRM includes the inversion of the CPSD matrix of the input signals of the interference canceller and of the EIC. The matrix inversion is assumed to be carried out using the matrix inversion lemma [36].

2.5 Experimental Results

We illustrate the performance of the joint acoustic echo cancellation and adaptive beamforming systems by experiments in the passenger cabin of a car and in an office room. In Sec. 2.5.1, we analyze the performance for time-invariant echo paths, for a fixed position of the desired source, and for variable noise level. In Sec. 2.5.2, we examine the influence of time-varying echo paths and of a time-varying position of the desired source on the performance of joint AEC-beamforming systems. Section 2.5.3 illustrates the influence of the reverberation time on the performance of GEIC.

2.5.1 Time-Invariant Echo Paths and Time-Invariant Source Position

In this section, we study the performance of GEIC for variable SIR and time-invariant echo paths in the passenger cabin of a car relative to the other concepts presented in Sec. 2.2. The interference is slowly time-varying car noise recorded with a microphone array setup inside of the car's passenger cabin (Fig. 2.8).

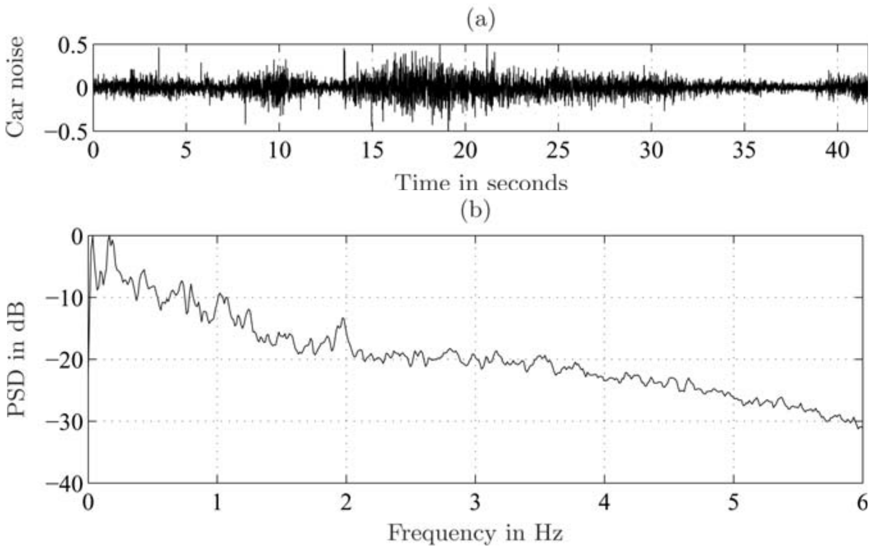

Fig. 2.8. Temporal signal (a) and power spectral density (PSD) of the car noise (b) measured at one of the microphones (before highpass filtering).

The desired source and two loudspeakers are located in broadside direction ($\theta = 90°$) and in the two endfire directions ($\theta = 0°$, $180°$), respectively, at a distance of 60 cm from the array center. The room impulse responses between the two loudspeakers and the microphones and between the desired source position and the microphones are simulated using the image method [3] with a simulated reverberation time $T_{60} = 50\,\mathrm{ms}$. The desired source signal is a subset of 50 utterances of the TIDigits database [60], while the loudspeaker signals are stereophonic pop music. The microphone signals are obtained by convolving the clean source signal with the room impulse responses followed by superposing noise with variable SIR and a fixed signal-to-echo ratio $SER = 7\,\mathrm{dB}$. The microphone array consists of $M = 4$ sensors or $M = 8$ sensors with sensor spacing $d = 4\,\mathrm{cm}$. The frequency range is 200 Hz–4 kHz. The echo suppression $ERLE$ and the interference suppression IR averaged over the whole test data are given in Fig. 2.9 ($M = 4$) and in Fig. 2.10 ($M = 8$). The filter lengths are chosen as follows: 'AEC first', GSAEC: $N_{\hat{h}} = 512$, $N_w = 256$; GEIC, RGSC: $N_{\hat{h}} = N_w = 256$).

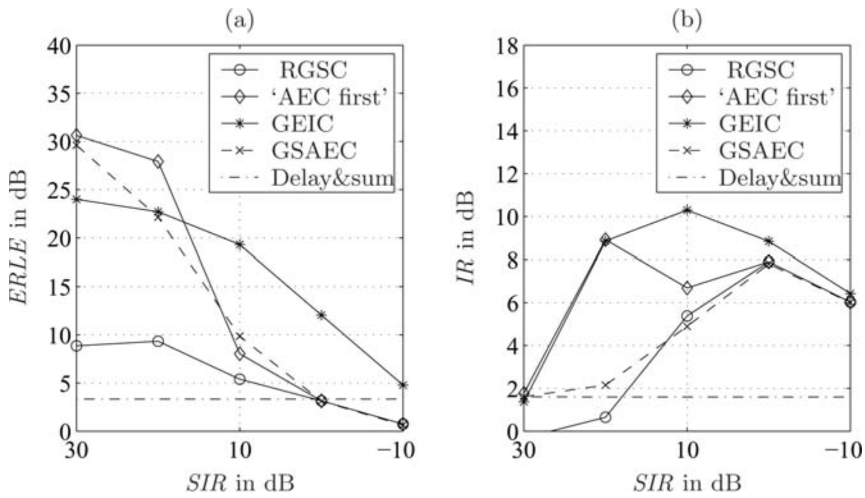

Fig. 2.9. Interference suppression IR and echo suppression $ERLE$ for RGSC alone, 'AEC first', GSAEC, and GEIC for fixed echo paths and fixed source position in the car environment for $M = 4$ (Signal-to-echo ratio $SER = 7\,\mathrm{dB}$).

For high SIR (equivalent to high EIR, since $SER = 7\,\mathrm{dB}$), the AECs of 'AEC first' converge in pauses of the desired speaker and provide high echo suppression, which translates to a greater $ERLE$ and IR of 'AEC first' relative to GSC and GEIC. With decreasing EIR, the echo suppression of the AECs of 'AEC first' decreases until the AECs are inefficient and $ERLE$ and IR of 'AEC first' are equivalent to the RGSC. Here, the GEIC outperforms 'AEC first', since the number of degrees of freedom does not depend on the EIR. Nevertheless, $ERLE$ of GEIC falls with decreasing EIR, since the system concentrates on the suppression of the stronger car noise. For $M = 4$ (Fig. 2.9), it can be noticed that the improvement of $ERLE$ and IR relative to RGSC

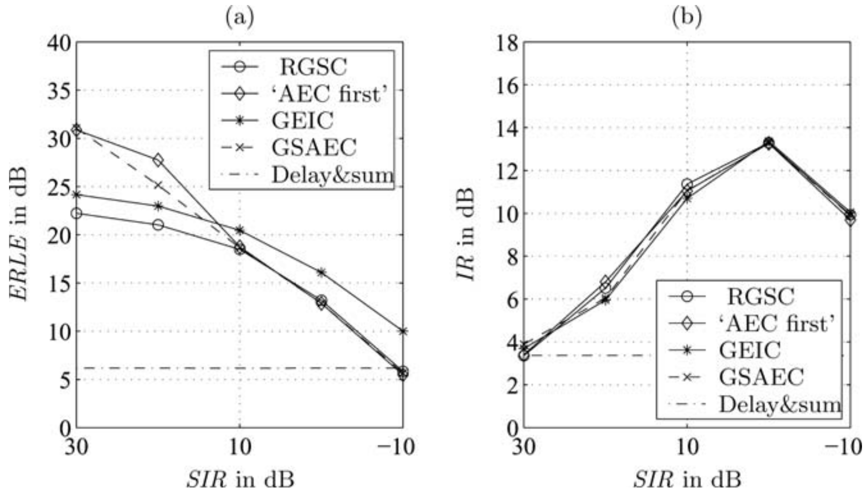

Fig. 2.10. Interference suppression *IR* and echo suppression *ERLE* for RGSC alone, 'AEC first', GSAEC, and GEIC for fixed echo paths and fixed source position in the car environment for $M = 8$ (Signal-to-echo ratio $SER = 7\,\text{dB}$).

is larger than for $M = 8$ (Fig. 2.10). This is due to the RGSC's greater number of degrees of freedom, where the additional degrees of freedom of GEIC due to the AEC yield a relative lower performance improvement. In fact, the improvement of *IR* can even be neglected in this scenario. The performance of GSAEC decreases relative to 'AEC first', since, after convergence of the AEC in the reference path of the GSC (Fig. 2.3), acoustic echoes leak through the sidelobe-cancelling path of the GSC, which leads to reduced echo and interference suppression relative to 'AEC first' [44, 48].

2.5.2 Time-Varying Echo Path and Time-Varying Source Position

In this section, we compare the performance of joint acoustic echo cancellation and adaptive beamforming for a time-varying echo path and a moving desired source. Because of the better tracking during double talk, we expect that the performance gap between echo and noise suppression of GEIC and that of 'AEC first' and GSAEC increases. The position of the desired source is switched randomly for each file of the TIDigits database in the interval $\theta = 80^\circ \ldots 100^\circ$ in steps of 2° with equal probability for all directions. This range corresponds to the 5 dB width of the mainlobe of the uniformly weighted delay&sum beamformer at 4 kHz. The desired signal is thus attenuated by less than 5 dB at 4 kHz. While one of the loudspeakers is located at $\theta = 180^\circ$, the position of the second loudspeaker is switched every 20000 samples between $\theta = 0^\circ$ and $\theta = 60^\circ$. The distance between the sources and the array center is fixed at 60 cm. The interference suppression *IR* and the echo suppression *ERLE* are averaged over the entire data set across all variations of the loudspeaker position.

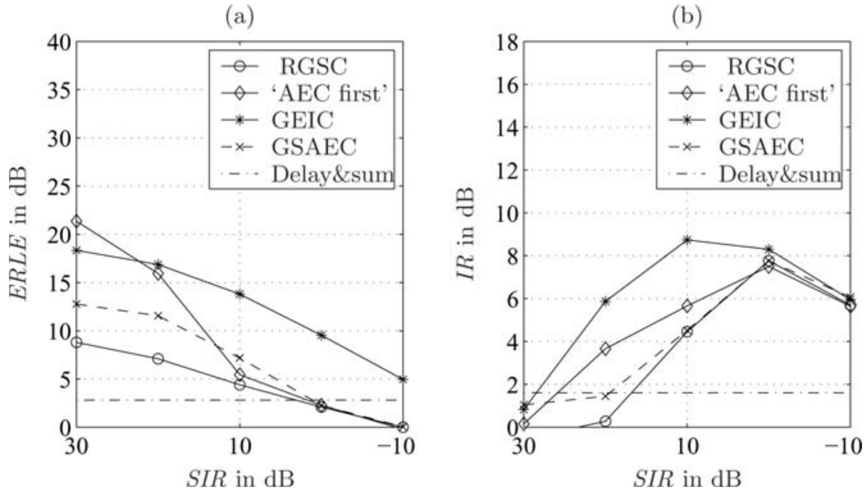

Fig. 2.11. Interference suppression *IR* and echo suppression *ERLE* for RGSC alone, 'AEC first', GSAEC, and GEIC for time-varying echo paths and fixed source position in the car environment for $M = 4$ (Signal-to-echo ratio $SER = 7\,\text{dB}$).

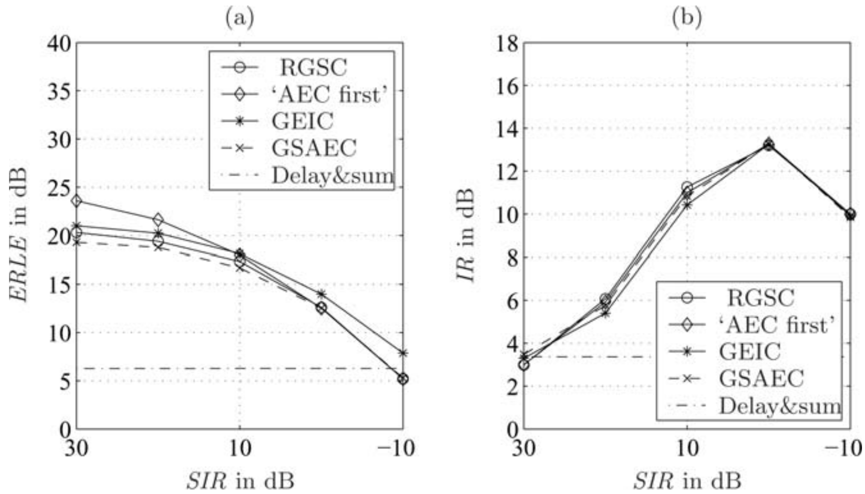

Fig. 2.12. Interference suppression *IR* and echo suppression *ERLE* for RGSC alone, 'AEC first', GSAEC, and GEIC for time-varying echo paths and fixed source position in the car environment for $M = 8$ (Signal-to-echo ratio $SER = 7\,\text{dB}$).

Figs. 2.11 and 2.12 depict the results for $M = 4$ and $M = 8$, respectively. The difference in performance between $M = 4$ and $M = 8$ can be explained similarly as in Figs. 2.3 and 2.9 by the greater number of degrees of freedom of the beamformer for $M = 8$. It may be further noticed that the performance of 'AEC first' and GSAEC is considerably reduced for $SIR \geq 20\,\mathrm{dB}$ relative to fixed echo paths (see Figs. 2.9 and 2.10). This effect can be explained by the reduced efficiency of the AECs of 'AEC first' and of GSAEC due to the missing capability to adapt the AECs while desired speech and acoustic echoes are simultaneously active. The performance loss is mainly related to the time-variance of the echo paths: Experiments showed that the performance for fixed echo paths and the time-varying position of the desired source ($\theta = 80^\circ \ldots 100^\circ$) can almost not be distinguished from the results in Figs. 2.9 and 2.10. As for fixed echo paths and the fixed position of the desired source, the echo suppression and the interference suppression converge with an increasing number of microphones.

Note that the AECs of 'AEC first' and of GSAEC are realized using a frequency-independent double talk detector with a constant step size during adaptation. When using a DFT bin-wise step-size control with variable frequency-dependent step size as, for example, proposed in [26, 73], it is possible to exploit sparseness of desired speech and of interference. The AECs can therefore be adapted more frequently, which improves the performance of 'AEC first' and of GSAEC for time-varying acoustic conditions and high EIRs.

2.5.3 Reverberation Time

In this section, we study the dependency of the echo and noise suppression of GEIC on the reverberation time T_{60}. Because of the limited number of filter taps of the EIC, we expect the performance of the GEIC to decrease compared to 'AEC first' and to GSAEC with increasing reverberation time. The experimental setup is the same as in Sec. 2.5.1, except for the fact that the impulse responses between the loudspeakers and the microphones are taken from three different acoustic environments: the environment with $T_{60} = 50\,\mathrm{ms}$ as above, and measured impulse responses from office rooms with $T_{60} = 250\,\mathrm{ms}$ and with $T_{60} = 400\,\mathrm{ms}$. The microphone array with $M = 4$ sensors is used, $SER = 7\,\mathrm{dB}$, and $SIR = 10\,\mathrm{dB}$. The results are depicted in Fig. 2.13.

It can be seen that the average echo suppression $ERLE$ (Fig. 2.13a) decreases with increasing reverberation time from 19.5 dB for $T_{60} = 50\,\mathrm{ms}$ to 15.5 dB for $T_{60} = 400\,\mathrm{ms}$. The interference suppression IR decreases from 10 dB for $T_{60} = 50\,\mathrm{ms}$ to 9.5 dB for $T_{60} = 400\,\mathrm{ms}$. Considering that the number of filter taps of the AEC is only $N_{\hat{h}} = 256$ for a reverberation time $T_{60} = 400\,\mathrm{ms}$, where, according to (2.1), $N_{\hat{h}} = 1240$ is required for $ERLE = 15.5\,\mathrm{dB}$, these results reflect that the AECs within GEIC are better interpreted as interference cancellers than as system identifiers. The usage of GEIC –despite the limitation on the number of filter taps– is thus not

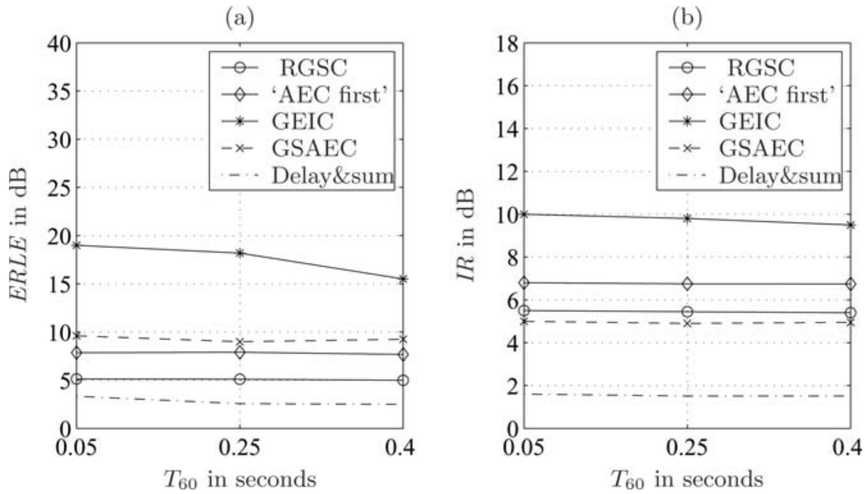

Fig. 2.13. Interference suppression *IR* and echo suppression *ERLE* for RGSC alone, 'AEC first', GSAEC, and GEIC as a function of the reverberation time T_{60} for $M = 4$ (Signal-to-echo ratio $SER = 7$ dB, Signal-to-interference ratio $SIR = 10$ dB).

restricted to environments with low reverberation times but still gives acceptable echo suppression in environments with longer reverberation time such as office or home environments, at least for slowly time-varying conditions. Note, however, that the performance of joint acoustic echo cancellation and adaptive beamforming systems based on the GSC depends on the robustness of the GSC against distortion of the desired signal in reverberant environments. It is thus not assured that all GSC realizations yield an undistorted desired signal.

2.6 Conclusion

We presented a technique for joint optimization of acoustic echo cancellation and adaptive LCMV beamforming. The derivation of the system shows that it can be interpreted as a straightforward extension of the GSC with additional input channels of the interference canceller (GEIC). With a realization example based on the RGSC and a stereophonic AEC, we showed that the GEIC is especially efficient for (a) transient echo paths if frequent double talk between acoustic echoes, local interference, and desired speakers is to be expected and (b) high levels of background noise. For stationary conditions and low levels of background noise, the performance of GEIC is reduced relative to 'AEC first' due to a constraint on the number of filter taps of the weight vector of the AEC. However, the proposed solution requires only one AEC for an arbitrary number of microphones and no separate adaptation control for the

AEC. For acoustic echo cancellation with multiple reproduction channels, the problem of slow convergence due to cross-correlated loudspeaker signals can be avoided, since the system identification problem is reduced to an interference cancellation problem.

References

[1] R. Aichner, W. Herbordt, H. Buchner, W. Kellermann: Least-squares error beamforming using minimum statistics and multichannel frequency-domain adaptive filtering, *Proc. IWAENC '03*, Kyoto, Japan, 223–226, September 2003.

[2] M. Ali: Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation,*Proc. ICASSP '98*, **6**, 3689–3692, Washington, DC, USA, May 1998.

[3] J.B. Allen: Multimicrophone signal-processing technique to remove room reverberation from speech signals, *Journal of the Acoustical Society of America*, **62**(4), 912–915, October 1977.

[4] H.J.W. Belt, C.P. Janse: Method and device for acoustic echo cancellation combined with adaptive beamforming, *United States Patent*, US 2002/0015500 A1, 2002.

[5] J. Benesty, D.R. Morgan, 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*, **6**(2), 156–165, 1998.

[6] J. Benesty, D.R. Morgan: Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation, *Proc. ICASSP '00*, 261-264, Istanbul, Turkey, June 2000.

[7] J. Benesty, D.R. Morgan: Multi-channel frequency-domain adaptive filtering, in S.L. Gay, J. Benesty, (eds.), *Acoustic Signal Processing for Telecommunication*, Chapter 7, 121–133, Boston, MA, USA: Kluwer Academic Publishers, 2000.

[8] J. Benesty, Y. Huang (eds): *Adaptive Signal Processing: Applications to Real-World Problems,* Berlin, Germany: Springer, 2003.

[9] J. Benesty, S. Makino, J. Chen (eds.): *Speech Enhancement,* Berlin, Germany: Springer, 2005.

[10] R. Le Bouquin, P. Scalart, G. Faucon, C. Beaugeant: Combined noise and echo reduction in hands-free systems: A survey, *IEEE Trans. on Speech and Audio Processing*, **9**(8), 808–820, November 2001.

[11] M.S. Brandstein, D.B. Ward (eds.): *Microphone Arrays: Signal Processing Techniques and Applications,* Berlin, Germany: Springer, 2001.

[12] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, J. Tilp: Acoustic echo control - an application

of very-high-order adaptive filters, *IEEE Signal Processing Magazine*, **16**(4), 42–69, July 1999.

[13] H. Buchner, W. Herbordt, W. Kellermann: An efficient combination of multi-channel acoustic echo cancellation with a beamforming microphone array, *Proc. Int. Workshop on Hands-Free Speech Communication*, 55–58, Kyoto, Japan, April 2001.

[14] H. Buchner, W. Kellermann: Acoustic echo cancellation for two or more reproduction channels, *Proc. IWAENC '01,* 99–102, Darmstadt, Germany, September 2001.

[15] H. Buchner, J. Benesty, W. Kellermann: Multichannel frequency-domain adaptive filtering with application to multichannel acoustic echo cancellation, in J. Benesty, Y. Huang (eds.), *Adaptive Signal Processing: Applications to Real-World Problems,* Berlin, Germany: Springer, 2003.

[16] M. Buck: Aspects of first order differential microphone arrays in the presence of sensor imperfections, *European Trans. on Telecommunications*, **13**(2), 115–122, March/April 2002.

[17] K.M. Buckley: Broad-band beamforming and the generalized sidelobe canceller, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **34**(5), 1322–1323, October 1986.

[18] A. Cantoni, L.C. Godara: Resolving the direction of sources in a correlated field incident on an array, *Journal of the Acoustical Society of America*, **67**(4), 1247–1255, April 1980.

[19] I. Cohen, B. Berdugo: Microphone array post-filtering for non-stationary noise suppression, *Proc. ICASSP '02*, **1**, 901–904, Orlando, FL, USA, May 2002.

[20] H. Cox, R.M. Zeskind, T. Kooij: Robust adaptive beamforming, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, **35**(10), 1365–1376, October 1987.

[21] S. Doclo, M. Moonen, E. De Clippel: Combined acoustic echo and noise reduction using GSVD-based optimal filtering, *Proc. ICASSP '00*, **2**, 1051–1054, Istanbul, Turkey, June 2000.

[22] S. Doclo, M. Moonen: GSVD-based optimal filtering for single and multi-microphone speech enhancement, *IEEE Trans. on Signal Processing*, **50**(9), 2230–2244, September 2002.

[23] S. Doclo, A. Spriet, J. Wouters, M. Moonen: Speech distortion weighted multichannel Wiener filtering techniques for noise reduction, in J. Benesty, S. Makino, J. Chen (eds.), *Speech Enhancement,* Chapter 9, 199–228. Berlin, Germany, Springer, 2005.

[24] D.J. Edelblute, J.M. Fisk, G.L. Kinneson: Criteria for optimum-signal-detection theory for arrays, *Journal of the Acoustical Society of America*, **41**(1), 199–205, January 1967.

[25] G.W. Elko: Microphone array systems for hands-free telecommunication, *Speech Communication*, **20**, 229–240, 1996.

[26] G. Enzner, P. Vary: Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter, *Proc. IWAENC '03*, 43–46, Kyoto, Japan, September 2003.

[27] D.A. Florencio, H.S. Malvar: Multichannel filtering for optimum noise reduction in microphone arrays, *Proc. ICASSP '01*, **1**, 197–200, Salt Lake City, UT, USA, May 2001.

[28] R. Frenzel, M. Hennecke: Using prewhitening and step-size control to improve the performance of the LMS algorithm for acoustic echo cancellation, *Proc. ISCAS '92,* **4**, 1930–1932, San Diego, CA, USA, 1992.

[29] T. Gänsler, P. Eneroth: Influence of audio coding on stereophonic acoustic echo cancellation, *Proc. ICASSP '98*, **6**, Washington, DC, USA, 3649–3652, May 1998.

[30] T. Gänsler, J. Benesty: New insight into the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution, *IEEE Trans. on Speech and Audio Processing*, **10**(5), 257–267, July 2002.

[31] S. Gannot, D. Burshtein, E. Weinstein: Signal enhancement using beamforming and nonstationarity with applications to speech, *IEEE Trans. on Signal Processing*, **49**(8), 1614–1626, August 2001.

[32] S.L. Gay, J. Benesty (eds.): *Acoustic Signal Processing for Telecommunications,* Boston, MA, USA: Kluwer Academic Publishers, 2000.

[33] E.N. Gilbert, S.P. Morgan: Optimum design of directive antenna arrays subject to random variables, *Bell Systems Technical Journal*, **34**, 637–663, May 1955.

[34] A. Gilloire, V. Turbin: Using auditory properties to improve the behavior of stereophonic acoustic echo cancellers, *Proc. ICASSP '98*, **6**, 3681–3684, Washington, DC, USA, May 1998.

[35] L.C. Godara: Error analysis of the optimal antenna array processors, *IEEE Trans. on Aerospace and Electronic Systems*, **22**(4), 395–409, July 1986.

[36] G.H. Golub, C.F. van Loan: *Matrix Computations,* 2-nd edition, Baltimore, MD, USA: John Hopkins University Press, 1989.

[37] L.J. Griffiths, C.W. Jim: An alternative approach to linearly constrained adaptive beamforming, *IEEE Trans. on Antennas and Propagation*, **30**(1), 27–34, January 1982.

[38] A. Guerin, G. Faucon, R. Le Bouquin-Jeannes: Nonlinear acoustic echo cancellation based on Volterra filters, *IEEE Trans. on Speech and Audio Processing*, **11**(6), 672–683, November 2003.

[39] S. Gustafsson, R. Martin, P. Vary: Combined acoustic echo control and noise reduction for hands-free telephony, *Signal Processing*, **64**(1), 21–32, January 1998.

[40] S. Gustafsson, R. Martin, P. Jax, P. Vary: A psychoacoustic approach to combined acoustic echo cancellation and noise reduction, *IEEE Trans. on Speech and Audio Processing*, **10**(5), 245–256, July 2002.

[41] E. Hänsler: The hands-free telephone problem – an annotated bibliography, *Signal Processing*, **27**(3), 259–271, March 1992.

[42] E. Hänsler, G. Schmidt: *Acoustic Echo and Noise Control,* Hoboken, NJ, USA: John Wiley & Sons, 2004.

[43] S. Haykin: *Adaptive Filter Theory,* 3rd edition, Englewood Cliffs, NJ, USA: Prentice Hall, 1996.

[44] W. Herbordt, W. Kellermann: Acoustic echo cancellation embedded into the generalized sidelobe canceller, *Proc. EUSIPCO '00*, **3**, 1843–1846, Tampere, Finland, September 2000.

[45] W. Herbordt, H. Buchner, W. Kellermann: An acoustic human-machine front-end for multimedia applications, *EURASIP Journal on Applied Signal Processing*, **1**, 1–11, January 2003.

[46] W. Herbordt, W. Kellermann: Adaptive beamforming for audio signal acquisition, in J. Benesty, Y. Huang, (eds.), *Adaptive Signal Processing: Applications to Real-World Problems*, 155–194, Berlin, Germany: Springer, 2003.

[47] W. Herbordt, W. Kellermann, S. Nakamura: Combined optimization of LCMV beamforming and accoustic echo cancellation, *Proc. EUSIPCO '04*, **3**, , 2003–2006, Vienna, Austria, 2004.

[48] W. Herbordt: *Sound Capture for Human/Machine Interfaces: Practical Aspects of Microphone Array Signal Processing*, Berlin, Germany: Springer, 2005.

[49] W. Herbordt, S. Nakamura, W. Kellermann: Joint optimization of LCMV beamforming and acoustic echo cancellation for automatic speech recognition, *Proc. ICASSP '05*, **3**, 77–80, Philadelphia, PA, USA, March 2005.

[50] M.W. Hoffman, K.M. Buckley: Robust time-domain processing of braodband microphone array data, *IEEE Trans. on Speech and Audio Processing*, **3**(3), 193–203, May 1995.

[51] O. Hoshuyama, A. Sugiyama, A. Hirano: A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters, *IEEE Trans. on Signal Processing*, **47**(10), 2677–2684, October 1999.

[52] N.K. Jablon: Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections, *IEEE Trans. on Antennas and Propagation*, **34**(8), 996–1012, August 1986.

[53] Y. Joncour, A. Sugyiama: A stereo echo canceler with pre-processing for correct echo path identification, *Proc. IEEE ICASSP '98*, **6**, 3677–3680, Washington, DC, USA, May 1998.

[54] M. Kallinger, J. Bitzer, K.D. Kammeyer: Study on combining multichannel echo cancellers with beamformers, *Proc. IEEE ICASSP '00*, **2**, 797–800, Istanbul, Turkey, June 2000.

[55] K.D. Kammeyer, M. Kallinger, A. Mertins: New aspects of combining echo cancellers with beamformers, *Proc. ICASSP '05*, **3**, 137–140, Philadelphia, PA, USA, March 2005.

[56] W. Kellermann: Analysis and design of multirate systems for cancellation of acoustical echoes, *Proc. ICASSP '88*, **5**, New York, NY, USA, 2570–2573, April 1988.

[57] W. Kellermann: Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays, *Proc. ICASSP '97*, **1**, 219–222, Munich, Germany, April 1997.

[58] W. Kellermann: Acoustic echo cancellation for beamforming microphone arrays, in M.S. Brandstein, D.B. Ward (eds.), *Microphone Arrays: Signal Processing Techniques and Applications*, Chapter 13, 281–306, Berlin, Germany: Springer, 2001.

[59] F. Küch, W. Kellermann: Partitioned block frequency-domain adaptive second-order Volterra filter, *IEEE Trans. on Signal Processing*, **53**(2), 564–575, February 2005.

[60] R.G. Leonard: A database for speaker independent digit recognition, *Proc. ICASSP '84*, **3**, 42.11.1–42.11.4, San Diego, CA, USA, March 1984.

[61] J.S. Lim: *Speech Enhancement,* Englewood Cliffs, NJ, USA: Prentice Hall 1983.

[62] T. Lotter: Single- and multi-microphone spectral amplitude estimation using super-Gaussian speech models, in J. Benesty, S. Makino, J. Chen (eds.), *Speech Enhancement,* Chapter 4, 67–95, Berlin, Germany: Springer, 2005.

[63] S.Y. Low, S. Nordholm: A blind approach to joint noise and acoustic echo cancellation, *Proc. ICASSP '05*, **3**, 69–72, Philadelphia, PA, USA, March 2005.

[64] A. Mader, H. Puder, G.U. Schmidt: Step-size controls for acoustic echo cancellation filters - an overview, *Signal Processing*, **80**(9), 1697–1719, September 2000.

[65] C. Marro, Y. Mahieux, K.U. Simmer: Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering, *IEEE Trans. on Speech and Audio Processing*, **6**(3), 240–259, May 1998.

[66] R. Martin: *Freisprecheinrichtungen mit mehrkanaliger Echokompensation und Störgeräuschreduktion,* PhD thesis, Aachener Institut für Nachrichtengeräte und Datenkommunikation, 1995 (in German).

[67] R. Martin, P. Vary: Combined acoustic echo cancellation, dereverberation, and noise-reduction: a two microphone approach, *Proc. Int. Workshop on Acoustic Echo Control*, 125–132, Lannion, France, September 1993.

[68] R. Martin, P. Vary: Combined acoustic echo and noise reduction for hands-free telephony – state of the art and perspectives, *Proc. EUSIPCO '96*, 1107–1100, Trieste, Italy, September 1996.

[69] R. Martin, P. Vary: The echo shaping approach to acoustic echo control, *Speech Communication*, **20**(3/4), 181–190, December 1996.

[70] R. Martin: Small microphone arrays with postfilters for noise and acoustic echo cancellation, in M.S. Brandstein, D.B. Ward, (eds.), *Microphone Arrays: Signal Processing Techniques and Applications,* 255–279, Berlin, Germany: Springer, 2001.

[71] I. McCowan, H. Bourlard: Microphone array post-filter for diffuse noise field, *Proc. ICASSP '02*, **1**, 905–908, Orlando, FL, USA, May 2002.

[72] M. Mizumachi, M. Akagi: Noise reduction by paired-microphones using spectral subtraction, *Proc. ICASSP '98*, **2**, 1001–1004, Washington, DC, USA, May 1998.

[73] B. H. Nitsch: A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain, *Signal Processing*, **80**(9), 1733–1745, September 2000.

[74] S. Nordholm, I. Claesson, B. Bengtsson: Adaptive array noise suppression of handsfree speaker input in cars, *IEEE Trans. on Vehicular Technology*, **42**(4), 514–518, November 1993.

[75] S. Nordholm, H.Q. Dam, N. Grbic, S.Y. Low: Adaptive microphone arrays employing spatial quadratic soft constraints and spatial filtering, in J. Benesty, S. Makino, J. Chen (eds.), *Speech Enhancement,* Chapter 9, 229–246, Berlin, Germany: Springer, 2005.

[76] Y. Ohashi, T. Nishikawa, H. Saruwatari, A. Lee, K. Shikano: Noise robust speech recognition based on spatial subtraction array, *Int. Workshop on Nonlinear Signal and Image Processing*, 324–327, Sapporo, Japan, May 2005.

[77] N.L. Owsley: An overview of optimum-adaptive control in sonar array processing, in K.S. Narendra, R.V. Monopoli (eds.), *Applications of Adaptive Control*, 131–164, New York, NY, USA: Academic Press, 1980.

[78] D. Raub, J. McDonough, M. Wölfel: A cepstral domain maximum likelihood beamformer for speech recognition, *Proc. Int. Conf. on Spoken Language Processing*, **2**, 817-820, Jeju Island, Korea, October 2004.

[79] J. Rosca, R. Balan, C. Beaugeant: Multi-channel psychoacoustically motivated speech enhancement, *Proc. ICASSP '03*, **1**, 84–87, April 2003.

[80] M.L. Seltzer: *Microphone array processing for robust speech recognition*, PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, July 2003.

[81] J.J. Shynk: Frequency-domain and multirate adaptive filtering, *IEEE Signal Processing Magazine*, **9**(1), 14–37, January 1992.

[82] K.U. Simmer, J. Bitzer, C. Marro: Post-filtering techniques, in M.S. Brandstein, D.B. Ward (eds.), *Microphone Arrays: Signal Processing Techniques and Applications*, Chapter 3, 39–60, Berlin, Germany: Springer, 2001.

[83] M.M. Sondhi, W. Kellermann: Echo cancellation for speech signals, in S. Furui, M.M. Sondhi (eds.), *Advances in Speech Signal Processing*, 327–356, New York, NY, USA: Marcel Dekker, 1991.

[84] M.M. Sondhi, D.R. Morgan, J.L. Hall: Stereophonic acoustic echo cancellation – an overview of the fundamental problem, *IEEE Signal Processing Letters*, **2**(8), 148–151, August 1995.

[85] W.-J. Song, M.-S. Park: A complementary pair LMS algorithm for adaptive filtering, *Proc. ICASSP '97*, **3**, 2261–2264, Munich, Germany, 1997.

[86] A. Spriet, M. Moonen, J. Wouters: Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids, *Signal Processing,* **84**(12), 2367–2387, December 2004.

[87] A. Stenger: *Kompensation akustischer Echos unter Einfluß von nichtlinearen Audiokomponenten*, PhD thesis, University Erlangen-Nuremberg, Erlangen, Germany, 2000 (in German).

[88] Z. Tian, K.L. Bell, H.L. Van Trees: A recursive least squares implementation for LCMP beamforming under quadratic constraint, *IEEE Trans. on Signal Processing*, **49**(6), 1138–1145, June 2001.

[89] H. Teutsch, G.W. Elko: An adaptive close-talking microphone array, *Proc. WASPAA '01*, 163–166, New Paltz, NY, USA, October 2001.

[90] D. Van Compernolle: Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings, *Proc. ICASSP '90*, **2**, Albuquerque, NM, USA, 833–836, 1990.

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

[92] B.D. Van Veen, K.M. Buckley: Beamforming: A versatile approach to spatial filtering, *IEEE ASSP Magazine*, **5**(2), 4–24, April 1988.

[93] J. Vanden Berghe, J. Wouters: An adaptive noise canceller for hearing aids using two nearby microphones, *Journal of the Acoustical Society of America*, **103**(6), 3621–3626, June 1998.

[94] B. Widrow, K.M. Duvall, R.P. Gooch, W.C. Newman: Signal cancellation phenomena in adaptive antennas: Causes and cures, *IEEE Trans. on Antennas and Propagation*, **30**(3), 469–478, May 1982.

[95] S. Yamamoto, S. Kitayama: An adaptive echo canceller with linear predictor, *IEICE Trans. on Information and Systems*, **62**(12), 851–857, December 1979.

[96] H. Yasukawa, S. Shimada: An acoustic echo canceller using subband sampling and decorrelation methods, *IEEE Trans. on Signal Processing*, **41**(2), 926–930, February 1993.

[97] W.S. Youn, C.K. Un: Robust adaptive beamforming based on the eigenstructure method, *IEEE Trans. on Signal Processing*, **42**(6), 1543–1547, June 1994.