An Acoustic Human-Machine Front-End for Multimedia Applications

Wolfgang Herbordt

Telecommunications Laboratory, University Erlangen-Nuremberg, Cauerstraße 7, 91058 Erlangen, Germany
Email: herbordt@LNT.de

Herbert Buchner

Telecommunications Laboratory, University Erlangen-Nuremberg, Cauerstraße 7, 91058 Erlangen, Germany
Email: buchner@LNT.de

Walter Kellermann

Telecommunications Laboratory, University Erlangen-Nuremberg, Cauerstraße 7, 91058 Erlangen, Germany
Email: wk@LNT.de

A concept of robust adaptive beamforming integrating stereophonic acoustic echo cancellation is presented which reconciles the need for low-computational complexity and efficient adaptive filtering with versatility and robustness in real-world scenarios. The synergetic combination of a robust generalized sidelobe canceller and a stereo acoustic echo canceller is designed in the frequency domain based on a general framework for multichannel adaptive filtering in the frequency domain. Theoretical analysis and real-time experiments show the superiority of this concept over comparable time-domain approaches in terms of computational complexity and adaptation behaviour. The real-time implementation confirms that the concept is robust and meets well the practical requirements of real-world scenarios, which makes it a promising candidate for commercial products.

Keywords and phrases: hands-free acoustic human-machine front-end, microphone arrays, robust adaptive beamforming, stereophonic acoustic echo cancellation, generalized sidelobe canceller, frequency-domain adaptive filters.

1. INTRODUCTION

With a continuously increasing desire for convenient human-machine interaction, the acoustic interface of any terminal for multimedia or telecommunication services is challenged to allow seamless, hands-free, and untethered audio communication for the benefit of human users.

Audio capture is usually responsible for extracting desired signals for the multimedia device or, in telecommunication applications, for remote listeners. Compared to sound capture by a microphone next to the source, seamless audio interfaces as depicted in Figure 1 cause the desired signals to be impaired by

(a) acoustic echoes from the loudspeaker(s),
(b) local interferers, and
(c) reverberation due to distant talking.

Techniques for acoustic echo cancellation (AEC) evolved over the last two decades [1, 2] and lead to the recent presentation of a five-channel AEC for real-time operation on a personal computer (PC) [3, 4]. If no distortion of the desired signal should be allowed, suppression of local interference is best handled by microphone arrays [5, 6]. Here, robust adaptive beamforming algorithms are necessary to cope with time-varying acoustic environments including moving desired sources. Removing reverberation from the desired signal, ideally, requires blind identification and inversion of the channel(s) from the source to the sensor(s). For realistic time-varying environments, this problem still awaits theoretical solutions with robust implementations out of reach. Consequently, practical dereverberation is limited to the spatial filtering effected by a beamforming microphone array, which suppresses acoustic reflections from undesired directions.

From the above, for practical multimedia terminals, a combination of a beamforming microphone array with AEC is desirable. While the general properties and synergies of such combinations have been studied in [7], we describe here a system which incorporates advanced adaptive filtering techniques for both beamforming and multichannel AEC leading to a highly efficient and robust real-time implementation. For beamforming, a robust generalized sidelobe
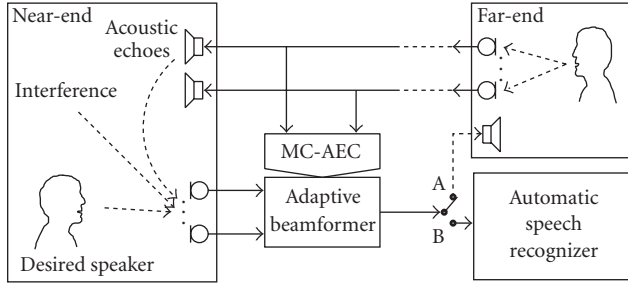
FIGURE 1: Seamless human-machine interface.

canceller (RGSC) [8] serves as a starting point which is discussed in Section 2. For stereo sound reproduction as considered here, the system identification problem of stereo AEC (SAEC) is described in Section 3.

In Section 4, a general framework for multichannel adaptive filtering in the frequency domain (more exactly, discrete Fourier transform (DFT) domain) is presented, which is subsequently used to systematically derive efficient algorithms for both adaptive beamforming and SAEC using the formalism.

In Section 5, as the main contribution of this paper, the embedding of SAEC into an RGSC structure in the frequency domain is described. The algorithms for each of the adaptive building blocks are formulated while including crucial issues of adaptation control.

The functionality and efficiency of the realized system are documented in Section 6. Results for convergence behaviour of the various adaptive components are presented for well-defined real-world simulation scenarios, and main characteristics of the real-time implementation are described.

Compared to [9], (a) the frequency-domain system is derived starting from a description in the time domain and rigorously applying the concept of multiple-input multiple-output (MIMO) frequency-domain adaptive filtering, and (b) the AEC part is integrated such that the system using stereophonic AEC runs in real time on low-cost PC platforms.

2. GENERALIZED SIDELOBE CANCELLER (GSC) FOR NONSTATIONARY BROADBAND SIGNALS

Essentially, beamforming microphone arrays separate desired signals from interference by exploiting spatial information about the source location. Since acoustic environments are strongly time-variant, means of adaptive beamforming are necessary. First approaches only took the time-variance of the interference into account but assumed fixed positions for the desired speaker [10]. Although this yields sufficient interference suppression, it often leads to cancellation of the desired signal for even slightly moving desired sources. Thus, adaptive beamformers are necessary, which track (a) transient interference and (b) moving desired sources. Due to their simplicity, adaptive beamformers realized as GSC structures [11] are especially promising. A RGSC was presented in [8] that explicitly takes the time-variance of the desired

source position into account, which enhances robustness against desired signal cancellation compared to conventional GSCs.

In this section, we describe the RGSC algorithm (see Figure 2). It consists of a fixed-reference path, which is formed by a fixed beamformer (FBF), and an adaptive sidelobe-cancelling path with the adaptive blocking matrix (ABM) and the adaptive interference canceller (AIC). These building blocks are described in the discrete time domain in Sections 2.1, 2.2, and 2.3. In Section 2.4, we show problems of the original RGSC structure and propose solutions by way of a realization in the DFT domain (see Sections 4 and 5).

2.1. Fixed beamformer

Capturing the L_f most recent output samples of M microphone signals $x_m(k)$, $m = 1, 2, \ldots, M$, by vectors

$$\mathbf{x}_m(k) = (x_m(k), x_m(k-1), \ldots, x_m(k - L_f + 1))^T,$$
$$\mathbf{x}(k) = (\mathbf{x}_1^T(k), \mathbf{x}_2^T(k), \ldots, \mathbf{x}_M^T(k))^T, \quad (1)$$

where T denotes a vector or matrix transposition, and describing the FBF impulse responses by vectors

$$\mathbf{w}_m = (w_{0,m}, w_{1,m}, \ldots, w_{L_f-1,m})^T,$$
$$\mathbf{w} = (\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_M^T)^T, \quad (2)$$

where L_f is the number of filter taps, we can write the FBF output signal as

$$y_f(k) = \mathbf{w}^T \mathbf{x}(k), \quad (3)$$

$\mathbf{w}$ defines the beamformer response with respect to (w.r.t.) the signal impinging from the location of the desired source $d(k)$. The weight vector $\mathbf{w}$ is designed such that desired speaker movements within a predefined region are possible without distorting the desired signal (see, e.g., [9]), whereas any interference $n(k)$ arriving from another direction is attenuated.

2.2. Adaptive blocking matrix

The ABM suppresses the desired signal components in the adaptive sidelobe-cancelling path. The M-channel ABM output, ideally, only contains interference components which are used in the AIC to form an estimate of the interference contained in $y_f(k)$.

The ABM is realized by M adaptive filters with impulse respons vectors $\mathbf{b}_m(k)$ of length L_b,

$$\mathbf{b}_m(k) = (b_{0,m}(k), b_{1,m}(k), \ldots, b_{L_b-1,m}(k))^T, \quad (4)$$

using the FBF output

$$\mathbf{x}_b(k) = (y_f(k), y_f(k-1), \ldots, y_f(k - L_b + 1))^T \quad (5)$$

as reference signals and the sensor signals

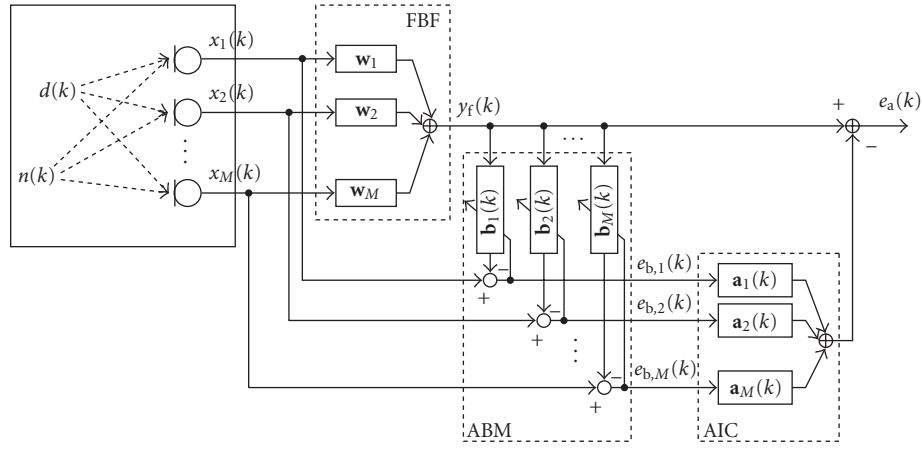$$\mathbf{y}_b(k) = (x_1(k), x_2(k), \ldots, x_M(k))^T \quad (6)$$

FIGURE 2: The structure of the RGSC after [8].

as desired signals (see Figure 2). Defining a matrix $\mathbf{B}(k)$ which captures all adaptive filters $\mathbf{b}_m(k)$,

$$\mathbf{B}(k) = (\mathbf{b}_1(k), \mathbf{b}_2(k), \ldots, \mathbf{b}_M(k)), \qquad (7)$$

we obtain for the ABM output signals in vector notation

$$\mathbf{e}_b(k) = \mathbf{y}_b(k) - \mathbf{B}^T(k)\mathbf{x}_b(k). \qquad (8)$$

In order to cancel the desired signal $d(k)$ by the ABM, $\mathbf{B}(k)$ must be determined such that the ABM output signals $\mathbf{e}_b(k)$ are minimized w.r.t. desired signal components. This can be expressed as an exponentially weighted least-squares problem [12] by

$$\sum_{i=0}^{k} \lambda_b^{k-i} |\mathbf{e}_b(i)|^2 \longrightarrow \min_{\mathbf{B}(k)}, \qquad (9)$$

where λ_b $(0 < \lambda_b < 1)$ is an exponential forgetting factor.

In contrast to fixed blocking matrices, which ensure distortion-free desired signals only for very few predetermined source positions [10], the adaptivity of the ABM allows to track arbitrarily moving desired speakers. Leakage of desired signal components is efficiently prevented by the ABM so that the RGSC is more robust against desired signal cancellation than GSCs using fixed blocking matrices.

2.3. Adaptive interference canceller

The AIC is realized by M adaptive filters with impulse response vectors $\mathbf{a}_m(k)$ of length L_a,

$$\begin{aligned} \mathbf{a}_m(k) &= \left(a_{0,m}(k), a_{1,m}(k), \ldots, a_{L_a-1,m}(k)\right)^T, \\ \mathbf{a}(k) &= \left(\mathbf{a}_1^T(k), \mathbf{a}_2^T(k), \ldots, \mathbf{a}_M^T(k)\right)^T, \end{aligned} \qquad (10)$$

using the ABM outputs $e_{b,m}(k)$,

$$\begin{aligned} \mathbf{e}_{b,m}(k) &= \left(e_{b,m}(k), e_{b,m}(k-1), \ldots, e_{b,m}(k-L_a+1)\right)^T, \\ \mathbf{x}_a(k) &= \left(\mathbf{e}_{b,1}^T, \mathbf{e}_{b,2}^T, \ldots, \mathbf{e}_{b,M}^T\right)^T \end{aligned} \qquad (11)$$

as reference signals and the FBF output $y_f(k)$ as desired signal.

The AIC structure minimizes the interference at the RGSC output $e_a(k)$ by subtracting the produced estimate of the interference from the fixed beamforming path

$$e_a(k) = y_f(k) - \mathbf{a}^T(k)\mathbf{x}_a(k). \qquad (12)$$

Nonstationary interference is efficiently suppressed by determining the optimum AIC filters using again an exponentially weighted least-squares optimization criterion with λ_a $(0 < \lambda_a < 1)$ as an exponential forgetting factor

$$\sum_{i=0}^{k} \lambda_a^{k-i} |e_a(i)|^2 \longrightarrow \min_{\mathbf{a}(k)}. \qquad (13)$$

2.4. Adaptation strategy

The above description reveals the fundamental problem of the RGSC of adapting the adaptive sidelobe-cancelling path. In this section, we first describe this problem, and, second, we show how it can be relieved by transforming the system into the frequency domain.

The adaptive sidelobe-cancelling path consists of two cascaded adaptive modules, ABM and AIC. Although they need to be adapted simultaneously for optimum-tracking of nonstationarities of the desired signal and the interference, they can only be adapted separately, which impairs tracking performance and output signal quality.

The ABM ideally suppresses only desired signal components. Interference must be excluded from the (unconstrained) adaptation of the ABM (see (9)). Therefore, the ABM cannot be adapted during double-talk, tracking performance is impaired, and desired signal may leak to the AIC input.

Ideally, the ABM output only contains interference components, which are used as reference for the AIC for minimizing interference in the GSC output signal (see (13)). Since the ABM output may contain desired signal during double-talk, the AIC cannot be adapted during double-talk for preventing cancellation of the desired signal by the AIC.
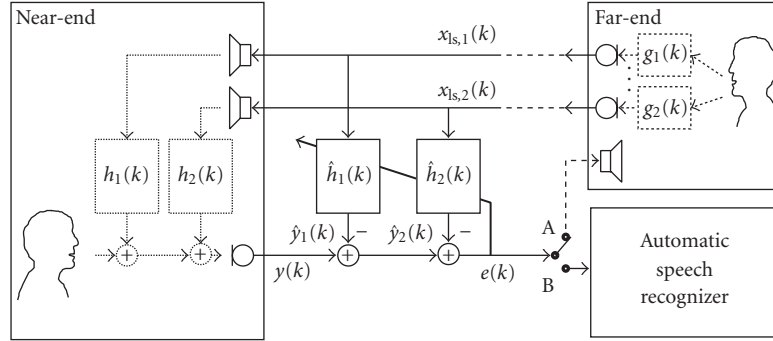
FIGURE 3: Stereophonic AEC.

As a result, the ABM/AIC should only be adapted if the signal-to-noise ratio (SNR) is high/low. Otherwise the adaptation should be stopped for optimum output signal quality. When applying this adaptation strategy for nonstationary signals in the full frequency band, then the adaptive interference cancelling path is nearly inefficient. Interference suppression is reduced to that of the FBF [9].

Application of this adaptation strategy not in the fullband but in narrow subbands allows more flexibility for nonstationary signals. For each of the subbands, the SNR can be estimated independently and, based on this SNR-estimate, adaptation of ABM and AIC can be performed. In Section 6, we illustrate that tracking performance, interference suppression, and output signal quality are improved by our realization in the DFT domain.

3. STEREOPHONIC AEC

The fundamental idea of any two-channel AEC structure (Figure 3) is to use adaptive FIR filters with length-L impulse response vectors

$$\hat{\mathbf{h}}_p(k) = (\hat{h}_{0,p}(k), \dots, \hat{h}_{L-1,p}(k))^T, \quad p = 1, 2, \qquad (14)$$

which identify the truncated (generally time-varying) echo path impulse responses $\mathbf{h}_p(k)$. The filters $\hat{\mathbf{h}}_p(k)$ are stimulated by the loudspeaker signals $x_{\mathrm{ls},p}(k)$ and then the resulting echo estimates $\hat{y}_p(k)$ are subtracted from the microphone signal $y(k)$ to cancel the echoes. The generalization to the multimicrophone case is straightforward, but is disregarded in this section. The residual echo signal reads

$$e(k) = y(k) - \sum_{p=1}^{2} \mathbf{x}_{\mathrm{ls},p}^T(k)\hat{\mathbf{h}}_p(k), \qquad (15)$$

where

$$\mathbf{x}_{\mathrm{ls},p}(k) = (x_{\mathrm{ls},0,p}(k), \dots, x_{\mathrm{ls},L-1,p}(k))^T. \qquad (16)$$

Adaptation of the filters minimizing the power of $e(k)$ is carried out only if there is no activity of the speaker in the receiving room.

3.1. Specific problems of SAEC compared to single-channel AEC

The specific problems of SAEC include all those known for single-channel AEC such as colored and nonstationary excitation of very long adaptive filters (e.g., [1]), but in addition to that, SAEC usually has to cope with high crosscorrelation between the loudspeaker signals, which in turn cause correlated echoes that cannot easily be distinguished in the microphone signal [13]. The correlation results from the fact that the signals are usually derived from a common sound source at the far-end, for example, a speaker as shown in Figure 3. Straightforward extension of known mono AEC schemes thus often leads to very slow convergence of the adaptive filter towards the physically true echo paths [13]. If the relation between the signals $x_{\mathrm{ls},p}(k)$ is strictly linear, then there is a fundamental problem of nonuniqueness in the two-channel case as was shown in [13]. In general, convergence to the true echo paths is necessary since otherwise the AEC would have to track not only changes of the echo paths at the near-end but also any changes of the crosscorrelation between the channels of the incoming audio signal, leading to sudden degradation of the echo cancellation performance [13]. To some extent, the problem can be relieved by some nearly inaudible preprocessing of the loudspeaker signals (e.g., [14, 15]) for partial decorrelation of the channels, but in addition, sophisticated adaptation algorithms taking the crosscorrelations into account are still necessary for SAEC. This is discussed next.

3.2. Two-channel and multichannel adaptive filtering for highly cross-correlated excitation signals

Multichannel versions of known adaptation algorithms such as the (normalized) least-mean squares ((N)LMS) or the recursive least-squares (RLS) algorithms can be straightforwardly derived by rewriting (15) using concatenated vectors in the same way as shown in (12). However, due to the high crosscorrelation between the loudspeaker signals, the performance of SAEC is more severely affected by the choice of algorithm than the monophonic counterpart. This is easily recognized since the convergence speed of most adaptive algorithms depends on the condition number of the input signal's covariance matrix. In the stereo case, this condition
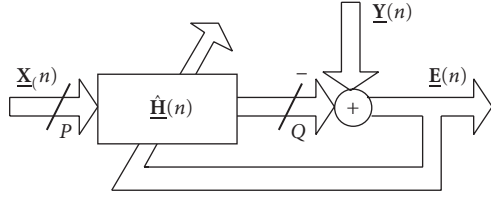
FIGURE 4: Adaptive MIMO filtering in the frequency domain.

number is very high. To cope with such ill-conditioned problems, the RLS algorithm turns out to be the optimum choice since its mean-squared error convergence is completely independent of that condition number [12]. Using concatenated data vectors, the corresponding coefficient update equation reads

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{h}}(k-1) + \mathbf{R}_{xx}^{-1}\mathbf{x}_{\mathrm{ls}}(k)e(k), \qquad (17)$$

where $\mathbf{R}_{xx}$ denotes the $2L \times 2L$ covariance matrix of the loudspeaker signals $x_{\mathrm{ls},p}(k)$. Note that this matrix contains both, estimates of autocorrelations (block matrices on the main diagonal) and crosscorrelations (block matrices on the off-diagonals). Unfortunately, because of the very high computational cost required for the inversion of $\mathbf{R}_{xx}$ and the associated numerical stability problems, this algorithm is not readily suitable for AEC in real-time operation. Therefore, efficient approximations to the multichannel RLS algorithm are needed, which explicitly take the high crosscorrelations into account. Section 4 describes an efficient and systematic adaptive filtering concept to solve this problem in the frequency domain.

4. EFFICIENT MULTICHANNEL FREQUENCY-DOMAIN ADAPTIVE FILTERING

The integrated system presented in Section 5 is solely based on efficient frequency-domain adaptive filtering using the overlap-save method. In the following, we give a compact formulation of a generic adaptive filter structure with P input channels and Q output channels as shown in Figure 4. This formalism will then be applied in Section 5 to our combination of RGSC and SAEC. As it turns out, the following formulation supports a systematic transformation of the entire structure (Sections 2 and 3) into the frequency domain and leads to several desirable properties, such as improved adaptation control for the RGSC and taking into account the crosscorrelation between the loudspeaker signals of the SAEC module. Note that the application of the overlap-save method using DFTs requires block processing of the input and output data streams. In the following, we derive the algorithm for a block length N equal to the filter length L, which yields maximum efficiency. However, to keep the processing delay short and to preserve optimum-tracking behaviour, the data blocks are overlapped in our realization (Section 5). Moreover, we consider here only multichannel frequency-domain adaptive filters in their unconstrained form. A more

general treatment of this class of adaptive algorithms including an in-depth convergence analysis can be found in [4].

4.1. Optimization criterion

To obtain a MIMO algorithm in the frequency domain, we first formulate a block-error signal and a suitable cost function for optimization. According to Figure 4, the error signal of the qth output channel ($q = 1, \ldots, Q$) is

$$e_q(k) = y_q(k) - \sum_{p=1}^{P} \mathbf{x}_p^T(k)\hat{\mathbf{h}}_{p,q} = y_q(k) - \mathbf{x}^T(k)\hat{\mathbf{h}}_q, \qquad (18)$$

where the vectors $\mathbf{x}_p(k)$ and $\hat{\mathbf{h}}_{p,q}$ are defined as in (16) and (14), respectively. The vectors $\mathbf{x}(k)$ and $\hat{\mathbf{h}}_q$ are obtained by concatenating the vectors $\mathbf{x}_p(k)$ and $\hat{\mathbf{h}}_{p,q}$, respectively. For applying L-point DFTs, as the corresponding $L \times 1$ block error signal vector is defined

$$\mathbf{e}_q(n) = (e_q(nL), \ldots, e_q(nL + L - 1))^T, \qquad (19)$$

where n denotes the block index over time. Moreover, the signals of all Q channels are then put together into an $L \times Q$ block-error signal matrix

$$\mathbf{E}(n) = (\mathbf{e}_1(n), \ldots, \mathbf{e}_Q(n)) \qquad (20)$$

which leads to an equivalent matrix formulation of (18) containing the block signal matrix

$$\mathbf{Y}(n) = (\mathbf{y}_1(n), \ldots, \mathbf{y}_Q(n)), \qquad (21)$$

and the $PL \times Q$ matrix of MIMO filter coefficients

$$\hat{\mathbf{H}} = (\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_Q). \qquad (22)$$

The data vector $\mathbf{x}(k)$ in (18) translates into a block-Toeplitz matrix in the block formulation. According to the overlap-save method [4, 16], this matrix can be transformed by appropriate windowing and using DFT matrices $\mathbf{F}$ of size $2L \times 2L$ into a block-diagonal matrix

$$\underline{\mathbf{X}}(n) = (\underline{\mathbf{X}}_1(n), \ldots, \underline{\mathbf{X}}_P(n)), \qquad (23)$$

where

$$\underline{\mathbf{X}}_p(n) = \mathrm{diag}\left\{\mathbf{F}(x_p(nL - L + 1), \ldots, x_p(nL + L))^T\right\}. \qquad (24)$$

It follows for the MIMO block-error matrix

$$\mathbf{E}(n) = \mathbf{Y}(n) - \mathbf{W}_{L \times 2L}^{01}\mathbf{F}^{-1}\underline{\mathbf{X}}(n)\hat{\underline{\mathbf{H}}}, \qquad (25)$$

with the windowing matrices

$$\mathbf{W}_{L \times 2L}^{01} = (\mathbf{0}_{L \times L}, \mathbf{I}_{L \times L}), \qquad \mathbf{W}_{2L \times L}^{10} = (\mathbf{I}_{L \times L}, \mathbf{0}_{L \times L})^T, \qquad (26)$$

and the coefficient matrix in the frequency domain

$$\hat{\underline{\mathbf{H}}} = \mathrm{diag}\left\{\mathbf{F}\mathbf{W}_{2L \times L}^{10}, \ldots, \mathbf{F}\mathbf{W}_{2L \times L}^{10}\right\}\hat{\mathbf{H}}. \qquad (27)$$

Multiplying (25) by $\mathbf{FW}_{2L \times L}^{01}$, we get the block-error signal matrix in the frequency domain

$$\underline{\mathbf{E}}(n) = \underline{\mathbf{Y}}(n) - \mathbf{G}\underline{\mathbf{X}}(n)\hat{\underline{\mathbf{H}}}, \tag{28}$$

where

$$\underline{\mathbf{E}}(n) = \mathbf{FW}_{2L \times L}^{01}\mathbf{E}(n) = \mathbf{F}\begin{pmatrix} \mathbf{0}_{L \times 1} & \cdots & \mathbf{0}_{L \times 1} \\ \mathbf{e}_1(n) & \cdots & \mathbf{e}_Q(n) \end{pmatrix}, \tag{29}$$

$$\underline{\mathbf{Y}}(n) = \mathbf{FW}_{2L \times L}^{01}\mathbf{Y}(n) = \mathbf{F}\begin{pmatrix} \mathbf{0}_{L \times 1} & \cdots & \mathbf{0}_{L \times 1} \\ \mathbf{y}_1(n) & \cdots & \mathbf{y}_Q(n) \end{pmatrix}, \tag{30}$$

$$\mathbf{G} = \mathbf{FW}_{2L \times 2L}^{01}\mathbf{F}^{-1}, \tag{31}$$

$$\mathbf{W}_{2L \times 2L}^{01} = \mathbf{W}_{2L \times L}^{01}\mathbf{W}_{L \times 2L}^{01} = \begin{pmatrix} \mathbf{0}_{L \times L} & \mathbf{0}_{L \times L} \\ \mathbf{0}_{L \times L} & \mathbf{I}_{L \times L} \end{pmatrix}. \tag{32}$$

Having derived a frequency-domain error matrix, the following frequency-domain criterion [4] is applied for optimizing the coefficient matrix $\hat{\underline{\mathbf{H}}} = \hat{\underline{\mathbf{H}}}(n)$:

$$J_f(n) = (1 - \lambda)\sum_{i=0}^{n} \lambda^{n-i} \text{tr}\{\underline{\mathbf{E}}^H(i)\underline{\mathbf{E}}(i)\}, \tag{33}$$

where H denotes conjugate transpose and λ $(0 < \lambda < 1)$ is an exponential forgetting factor. The criterion (33) is very similar to the one leading to the well-known RLS algorithm. The main advantage of using (33) is to take advantage of the fast Fourier transform (FFT) in order to have low-complexity adaptive filters.

4.2. Adaptive algorithm

An RLS-like algorithm can be straightforwardly derived from the so-called normal equation that is obtained by setting the gradient of (33) w.r.t $\hat{\underline{\mathbf{H}}}$ equal to zero. According to [12] and by noting that $\mathbf{G}^H\mathbf{G} = \mathbf{G}$ and $\mathbf{G}^H\underline{\mathbf{Y}}(i) = \underline{\mathbf{Y}}(i)$, we have for the gradient

$$\begin{aligned}\nabla_{\hat{\underline{\mathbf{H}}}}J_f(n) &= 2\frac{\partial J_f(n)}{\partial \hat{\underline{\mathbf{H}}}^*(n)} \\ &= 2(1 - \lambda)\sum_{i=0}^{n} \lambda^{n-i}\left[-\underline{\mathbf{X}}^H(i)\underline{\mathbf{Y}}(i) + \underline{\mathbf{X}}^H(i)\mathbf{G}\underline{\mathbf{X}}(i)\hat{\underline{\mathbf{H}}}\right].\end{aligned} \tag{34}$$

Setting this gradient equal to zero, we obtain the normal equation

$$\underline{\mathbf{S}}_{xx}(n)\hat{\underline{\mathbf{H}}}(n) = \underline{\mathbf{S}}_{xy}(n), \tag{35}$$

where

$$\begin{aligned}\underline{\mathbf{S}}_{xx}(n) &= (1 - \lambda)\sum_{i=0}^{n} \lambda^{n-i}\underline{\mathbf{X}}^H(i)\mathbf{G}\underline{\mathbf{X}}(i) \\ &= \lambda\underline{\mathbf{S}}_{xx}(n-1) + (1 - \lambda)\underline{\mathbf{X}}^H(n)\mathbf{G}\underline{\mathbf{X}}(n),\end{aligned} \tag{36}$$

$$\begin{aligned}\underline{\mathbf{S}}_{xy}(n) &= (1 - \lambda)\sum_{i=0}^{n} \lambda^{n-i}\underline{\mathbf{X}}^H(i)\underline{\mathbf{Y}}(i) \\ &= \lambda\underline{\mathbf{S}}_{xy}(n-1) + (1 - \lambda)\underline{\mathbf{X}}^H(n)\underline{\mathbf{Y}}(n).\end{aligned} \tag{37}$$

The iterative algorithm, that is, the recursive update of the coefficient matrix $\hat{\underline{\mathbf{H}}}$, is directly derived from (35), (36), and (37). In the recursive equation (37), we replace $\underline{\mathbf{S}}_{xy}(n)$ and $\underline{\mathbf{S}}_{xy}(n-1)$ by formulating (35) in terms of block-time indices n and $n-1$, respectively. We then eliminate $\underline{\mathbf{S}}_{xx}(n-1)$ from the resulting equation using (36). Reintroducing the error signal vector (28), we obtain the adaptive algorithm

$$\underline{\mathbf{E}}(n) = \underline{\mathbf{Y}}(n) - \mathbf{G}\underline{\mathbf{X}}(n)\hat{\underline{\mathbf{H}}}(n-1), \tag{38}$$

$$\hat{\underline{\mathbf{H}}}(n) = \hat{\underline{\mathbf{H}}}(n-1) + (1 - \lambda)\underline{\mathbf{S}}_{xx}^{-1}(n)\underline{\mathbf{X}}^H(n)\underline{\mathbf{E}}(n). \tag{39}$$

Additionally, matrix $\underline{\mathbf{S}}_{xx}(n)$ is estimated by (36).

The above algorithm is equivalent to the RLS algorithm in the sense that its mean-squared error convergence is also independent of the condition number of the input covariance matrix. To reduce the computational complexity of the adaptation drastically, it is shown in [4] that matrix $\mathbf{G}$ in (36) can be well approximated by a diagonal matrix $\mathbf{G} \approx \mathbf{I}/2$. Using this approximation and introducing a diagonal $2L \times 2L$ matrix $\boldsymbol{\mu}$ containing frequency-dependent stepsizes, we may rewrite (36) and (39) as

$$\begin{aligned}\underline{\mathbf{S}}_{xx}'(n) &= \lambda\underline{\mathbf{S}}_{xx}'(n-1) + (1 - \lambda)\underline{\mathbf{X}}^H(n)\underline{\mathbf{X}}(n), \\ \hat{\underline{\mathbf{H}}}(n) &= \hat{\underline{\mathbf{H}}}(n-1) + (1 - \lambda)\boldsymbol{\mu}\underline{\mathbf{S}}_{xx}'^{-1}(n)\underline{\mathbf{X}}^H(n)\underline{\mathbf{E}}(n),\end{aligned} \tag{40}$$

where $\boldsymbol{\mu}$ is a diagonal matrix of stepsizes, with elements $0 \leq \mu_i \leq 2$, $i = 0, 1, \ldots, L - 1$, and optimum stepsize $\boldsymbol{\mu} = 2\mathbf{I}$. Note that prior to inversion of $\underline{\mathbf{S}}_{xx}'(n)$, a proper regularization by adding a suitable diagonal matrix [4] is important to ensure robust convergence behaviour.

5. REALIZATION OF RGSC WITH EMBEDDED STEREOPHONIC AEC IN THE DFT DOMAIN

Fundamentally, adaptive beamforming and AEC need to be combined such that advantages are explored and insufficiencies are relieved. Optimum positive synergies between SAEC and GSC are obtained when the SAEC is placed in the sensor channels of the GSC (AEGSC) [7]. Maximum computational efficiency is given if the SAEC is located in the fixed-reference path after the FBF (GSAEC), since the number of SAEC output channels Q is minimized (see Figure 5). In [17], it is shown that most of the synergies are preserved for the latter structure.

In Section 6, we present a frequency-domain GSAEC realization (FGSAEC). Systematic application of multichannel frequency-domain adaptive filters (see Section 4) yields a system that exploits the advantages of multichannel frequency-domain adaptive filtering while preserving positive synergies between GSC and SAEC. Especially, (a) crosscorrelation between the loudspeaker signals is taken into account for fast convergence of the SAEC, (b) adaptation problems of the adaptive sidelobe-cancelling path of the GSC are efficiently resolved (see Section 5.6), and (c) computational complexity is minimized for efficient implementation of the integrated system on low-cost PC platforms for real-time application (see Section 6 and [9]).
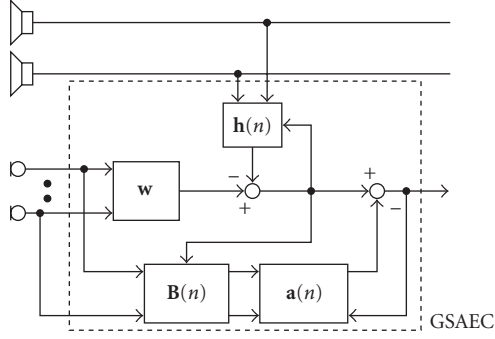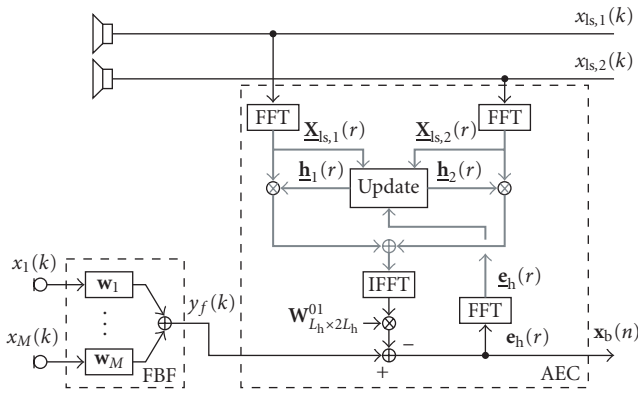
FIGURE 5: System overview.

FIGURE 6: SAEC in the fixed-reference path of the GSC.

5.1. Notations

For optimum performance, we use different DFT lengths $2L_g$ and $2L_h$ for GSC and SAEC, respectively, which yields the DFT matrices $\mathbf{F}_{2L_g \times 2L_g}$ and $\mathbf{F}_{2L_h \times 2L_h}$. The parameters $L_g = L_b = L_a$ and L_h are identical to the number of filter taps of GSC and AEC adaptive filters, respectively. For better tracking behavior of the adaptive filters, block overlaps by factors α_g and α_h are introduced in the GSC and AEC input signal blocks, respectively [18]. This leads to the block-time index $n = k\alpha_g/L_g$. It reflects the discrete time in numbers of blocks of length L_g/α_g. In the sequel, we assume that L_h/α_h is an integer multiple of L_g/α_g which maximizes efficiency. For a better reading, we define $R = L_h\alpha_g/L_g\alpha_h$ and the time index $r = k\alpha_h/L_h$. GSC and AEC adaptive filters are updated at times n and r, respectively.

5.2. FBF

The mainlobe of simple delay&sum beamformer with broadside steered microphone arrays is too narrow at high frequencies. This often leads to cancellation of the desired signal at high frequencies if the desired speaker position and the steering direction do not match.

Dolph-Chebyshev beamformer design [19] allows to arbitrarily choose the first null of the array pattern relative to the steering direction while minimizing the level of the

sidelobes. It allows to design filter&sum beamformers with predefined mainlobe widths that are constant over a wide range of frequencies. This makes this design method especially appropriate for our application since it allows to arbitrarily specify a region where desired signals are not attenuated [9].

5.3. Stereophonic AEC integrated in fixed-reference path of RGSC

Using the notations of Section 4, the basic signal-processing of the SAEC in the fixed-reference path of the GSC can be summarized as follows (see Figure 6). The number of input channels is $P = 2$ and the number of output channels is $Q = 1$. According to (24), we capture the last $2L_h$ samples of the loudspeaker signals $x_{ls,p}(k)$, $p = 1, 2$, in vectors, and we find for the frequency-domain loudspeaker signals,

$$\underline{\mathbf{X}}_{ls,p}(r) = \text{diag}\left\{ \mathbf{F}_{2L_h \times 2L_h} \begin{pmatrix} x_{ls,p}\left(r\frac{L_h}{\alpha_h} - 2L_h + 1\right) \\ x_{ls,p}\left(r\frac{L_h}{\alpha_h} - 2L_h + 2\right) \\ \vdots \\ x_{ls,p}\left(r\frac{L_h}{\alpha_h}\right) \end{pmatrix} \right\}. \quad (41)$$

The loudspeaker signals $x_{ls,p}(k)$ are assumed to be preprocessed by inaudible nonlinearities [14]. Capturing the $2L_h \times 1$ vectors of adaptive filter transfer functions $\underline{\mathbf{h}}_p(r)$ according to (27) in a vector

$$\underline{\mathbf{h}}(r) = \left(\underline{\mathbf{h}}_1^T(r), \underline{\mathbf{h}}_2^T(r) \right)^T, \quad (42)$$

and defining a matrix of loudspeaker signals

$$\underline{\mathbf{X}}_{ls}(r) = \left(\underline{\mathbf{X}}_{ls,1}(r), \underline{\mathbf{X}}_{ls,2}(r) \right), \quad (43)$$

as in (23), we obtain for the $L_h \times 1$ time-domain block error signal vector (see (25))

$$\mathbf{e}_h(r) = \mathbf{y}_f(r) - \mathbf{W}_{L_h \times 2L_h}^{01} \mathbf{F}_{2L_h \times 2L_h}^{-1} \underline{\mathbf{X}}_{ls}(r)\underline{\mathbf{h}}(r - 1), \quad (44)$$

where the $L_h \times 1$ vector of FBF output signal samples is defined as

$$\mathbf{y}_f(r) = \begin{pmatrix} y_f\left(r\frac{L_h}{\alpha_h} - L_h + 1\right) \\ y_f\left(r\frac{L_h}{\alpha_h} - L_h + 2\right) \\ \vdots \\ y_f\left(r\frac{L_h}{\alpha_h}\right) \end{pmatrix}. \quad (45)$$

We define a frequency-domain error signal $\underline{\mathbf{e}}_h(r)$ according to (29) as

$$\underline{\mathbf{e}}_h(r) = \mathbf{F}_{2L_h \times 2L_h} \mathbf{W}_{2L_h \times L_h}^{01} \mathbf{e}_h(r). \quad (46)$$
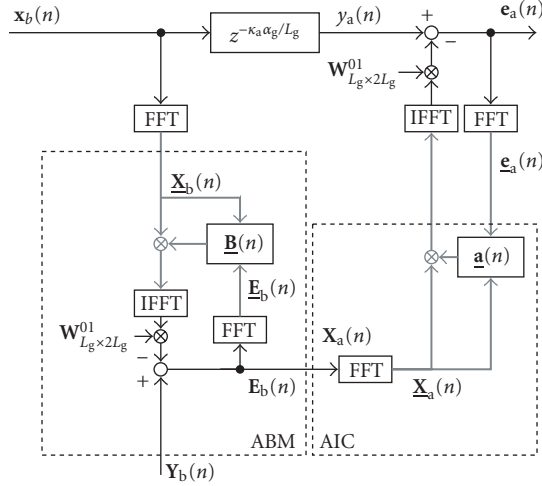
FIGURE 7: FGSAEC: ABM and AIC.

This allows us to write the SAEC filter update equation as

$$\underline{\mathbf{h}}(r) = \underline{\mathbf{h}}(r-1) + \mu_{\mathrm{h}}(1-\lambda_{\mathrm{h}})\underline{\mathbf{S}}_{X_{\mathrm{ls}}X_{\mathrm{ls}}}^{-1}(r)\underline{\mathbf{X}}_{\mathrm{ls}}^{H}(r)\underline{\mathbf{e}}_{\mathrm{h}}(r), \quad (47)$$

where μ_{h} ($0 \le \mu_{\mathrm{h}} \le 2$) is a stepsize parameter, λ_{h} ($0 < \lambda_{\mathrm{h}} < 1$) is an exponential forgetting factor, and $\underline{\mathbf{S}}_{X_{\mathrm{ls}}X_{\mathrm{ls}}}(r)$ is a recursive estimate of the cross-power spectral density matrix of the loudspeaker signals

$$\underline{\mathbf{S}}_{X_{\mathrm{ls}}X_{\mathrm{ls}}}(r) = (1-\lambda_{\mathrm{h}})\underline{\mathbf{S}}_{X_{\mathrm{ls}}X_{\mathrm{ls}}}(r-1) + \lambda_{\mathrm{h}}\underline{\mathbf{X}}_{\mathrm{ls}}^{H}(r)\underline{\mathbf{X}}_{\mathrm{ls}}(r). \quad (48)$$

With the inverse of the cross-power spectral density matrix $\underline{\mathbf{S}}_{X_{\mathrm{ls}}X_{\mathrm{ls}}}(r)$ in the update equation, crosscorrelation of the loudspeaker signals is explicitly taken into account, leading to fast convergence of the adaptive filters.

One block of length $L_{\mathrm{h}}/\alpha_{\mathrm{h}}$ of the AEC output signal is finally given by the last $L_{\mathrm{h}}/\alpha_{\mathrm{h}}$ samples of the error signal $\mathbf{e}_{\mathrm{h}}(r)$, which is by a factor R larger than the signal blocks which are required for the GSC. We split $\mathbf{e}_{\mathrm{h}}(r)$ into R blocks $\mathbf{x}_{\mathrm{b}}(n-i)$, $i = 0, 1, \ldots, R-1$ of length L_{h}/R. Therefore, $R-1$ blocks of $\mathbf{x}_{\mathrm{b}}(n-i)$ are buffered until they are used by the GSC.

5.4. Adaptive blocking matrix

In Figure 7, reference path and adaptive sidelobe-cancelling path are depicted. For the ABM, $P = 1$, $Q = M$. For applying the overlap-save method to the ABM adaptive filter inputs in the frequency domain, we have to transform $2\alpha_{\mathrm{g}}$ subsequent blocks of the AEC output signal $\mathbf{x}_{\mathrm{b}}(n)$ into the frequency domain. That is,

$$\underline{\mathbf{X}}_{\mathrm{b}}(n) = \mathrm{diag}\left\{\mathbf{F}_{2L_{\mathrm{g}}\times 2L_{\mathrm{g}}}\begin{pmatrix}\mathbf{x}_{\mathrm{b}}(n-2\alpha_{\mathrm{g}}+1)\\ \mathbf{x}_{\mathrm{b}}(n-2\alpha_{\mathrm{g}}+2)\\ \vdots\\ \mathbf{x}_{\mathrm{b}}(n)\end{pmatrix}\right\}. \quad (49)$$

With the ABM adaptive filters $\mathbf{B}(n)$, written in the frequency domain according to (27) as a $2L_{\mathrm{g}}\times M$ matrix $\underline{\mathbf{B}}(n)$, the

$L_{\mathrm{g}}\times M$ block error matrix $\mathbf{E}_{\mathrm{b}}(n)$ is obtained from (25) as

$$\mathbf{E}_{\mathrm{b}}(n) = \mathbf{Y}_{\mathrm{b}}(n) - \mathbf{W}_{L_{\mathrm{g}}\times 2L_{\mathrm{g}}}^{01}\mathbf{F}_{2L_{\mathrm{g}}\times 2L_{\mathrm{g}}}^{-1}\underline{\mathbf{X}}_{\mathrm{b}}(n)\underline{\mathbf{B}}(n-1), \quad (50)$$

where the $L_{\mathrm{g}}\times M$ block sensor signal matrix is defined as

$$\mathbf{Y}_{\mathrm{b}}(n) = (\mathbf{y}_{\mathrm{b},1}(n), \mathbf{y}_{\mathrm{b},2}(n), \ldots, \mathbf{y}_{\mathrm{b},M}(n)),$$

$$\mathbf{y}_{\mathrm{b},m}(n) = \begin{pmatrix}x_m\left(n\dfrac{L_{\mathrm{g}}}{\alpha_{\mathrm{g}}} - \kappa_{\mathrm{b}} - L_{\mathrm{g}} + 1\right)\\ x_m\left(n\dfrac{L_{\mathrm{g}}}{\alpha_{\mathrm{g}}} - \kappa_{\mathrm{b}} - L_{\mathrm{g}} + 2\right)\\ \vdots\\ x_m\left(n\dfrac{L_{\mathrm{g}}}{\alpha_{\mathrm{g}}} - \kappa_{\mathrm{b}}\right)\end{pmatrix}. \quad (51)$$

The time delay κ_{b} ensures causality of the ABM adaptive filters. Defining $\underline{\mathbf{E}}_{\mathrm{b}}(n)$ as

$$\underline{\mathbf{E}}_{\mathrm{b}}(n) = \mathbf{F}_{2L_{\mathrm{g}}\times 2L_{\mathrm{g}}}\mathbf{W}_{2L_{\mathrm{g}}\times L_{\mathrm{g}}}^{01}\mathbf{E}_{\mathrm{b}}(n), \quad (52)$$

the update equation for $\underline{\mathbf{B}}(n)$ reads[1]

$$\underline{\mathbf{B}}(n) = \underline{\mathbf{B}}(n-1) + (1-\lambda_{\mathrm{b}})\mathbf{G}\boldsymbol{\mu}_{\mathrm{b}}(n)\underline{\mathbf{S}}_{X_{\mathrm{b}}X_{\mathrm{b}}}^{-1}(n)\underline{\mathbf{X}}_{\mathrm{b}}^{H}(k)\underline{\mathbf{E}}_{\mathrm{b}}(n). \quad (53)$$

The matrix $\mathbf{G}$ is defined according to (31) with L replaced by L_{g}. In contrast to the SAEC (see Section 5.3), circular convolution constraints [4] are required for the ABM since the impulse responses of the ideal ABM filters are generally much longer than the length of the adaptive filters. Thus, circular convolution effects cannot be disregarded.

The $2L_{\mathrm{g}}\times 2L_{\mathrm{g}}$ diagonal matrix $\boldsymbol{\mu}_{\mathrm{b}}(n)$ is a matrix with frequency-dependent stepsizes on the main diagonal, which control the adaptation of the ABM (see Section 5.6).

The diagonal power spectral density matrix is given by

$$\underline{\mathbf{S}}_{X_{\mathrm{b}}X_{\mathrm{b}}}(n) = (1-\lambda_{\mathrm{b}})\underline{\mathbf{S}}_{X_{\mathrm{b}}X_{\mathrm{b}}}(n-1) + \lambda_{\mathrm{b}}\underline{\mathbf{X}}_{\mathrm{b}}^{H}(n)\underline{\mathbf{X}}_{\mathrm{b}}(n). \quad (54)$$

One block of length $L_{\mathrm{g}}/\alpha_{\mathrm{g}}$ of the time-domain AIC input signals $\mathbf{x}_{\mathrm{a},s}(n)$ is obtained by saving the last $L_{\mathrm{g}}/\alpha_{\mathrm{g}}$ samples of the mth column of the block error signal matrix $\mathbf{E}_{\mathrm{b}}(n)$.

5.5. Adaptive interference canceller

With $P = M$, $Q = 1$, the frequency-domain adaptive filter input matrix $\underline{\mathbf{X}}_{\mathrm{a}}(n)$ of size $2L_{\mathrm{g}}\times 2L_{\mathrm{g}}M$ is given by

$$\underline{\mathbf{X}}_{\mathrm{a}}(n) = (\underline{\mathbf{X}}_{\mathrm{a},1}(n), \underline{\mathbf{X}}_{\mathrm{a},2}(n), \ldots, \underline{\mathbf{X}}_{\mathrm{a},M}(n)), \quad (55)$$

where $\underline{\mathbf{X}}_{\mathrm{a},m}(n)$ is obtained in the same way as in (49) with $\underline{\mathbf{X}}_{\mathrm{b}}(n)$ and $\mathbf{x}_{\mathrm{b}}(n)$ replaced by $\underline{\mathbf{X}}_{\mathrm{a},m}(n)$ and $\mathbf{x}_{\mathrm{a},m}(n)$, respectively.

Writing the AIC adaptive filters $\mathbf{a}(n)$ after (27) in the frequency domain as a $2L_{\mathrm{g}}M \times 1$ vector $\underline{\mathbf{a}}(n)$, then the time-domain block error vector reads

$$\mathbf{e}_{\mathrm{a}}(n) = \mathbf{y}_{\mathrm{a}}(n) - \mathbf{W}_{L_{\mathrm{g}}\times 2L_{\mathrm{g}}}^{01}\mathbf{F}_{2L_{\mathrm{g}}\times 2L_{\mathrm{g}}}^{-1}\underline{\mathbf{X}}_{\mathrm{a}}(n)\underline{\mathbf{a}}(n-1), \quad (56)$$

---

[1]Coefficient constraints for improved robustness against cancellation of desired signal components may be introduced according to [8, 9].

where $\mathbf{y}_a(n)$ is defined as

$$\mathbf{y}_a(n) = \begin{pmatrix} \mathbf{x}_b\left(n - \dfrac{\kappa_a \alpha_g}{L_g} - \alpha_g + 1\right) \\ \mathbf{x}_b\left(n - \dfrac{\kappa_a \alpha_g}{L_g} - \alpha_g + 2\right) \\ \vdots \\ \mathbf{x}_b\left(n - \dfrac{\kappa_a \alpha_g}{L_g}\right) \end{pmatrix}. \quad (57)$$

The time delay κ_a ensures causality. Defining the frequency-domain error signal

$$\underline{\mathbf{e}}_a(n) = \mathbf{F}_{2L_g \times 2L_g} \mathbf{W}^{01}_{2L_g \times L_g} \mathbf{e}_a(n), \quad (58)$$

we obtain the multichannel filter update equation as[2]

$$\underline{\mathbf{a}}(n) = \underline{\mathbf{a}}(n-1) + (1 - \lambda_a)\mathbf{G}\boldsymbol{\mu}_a(n)\underline{\mathbf{S}}^{-1}_{X_a X_a}(n)\underline{\mathbf{X}}^H_a(n)\underline{\mathbf{e}}_a(n), \quad (59)$$

where we introduced the $2L_g \times 2L_g$ diagonal matrix $\boldsymbol{\mu}_a(n)$ with frequency-dependent stepsizes on the main diagonal for controlling the adaptation of the AIC (see Section 5.6). Note that circular convolution is prevented by the matrix $\mathbf{G}$ (see (31)). As for the ABM, ideal impulse responses AIC are much longer than the length of the adaptive filters. The diagonal power spectral density matrix is computed from

$$\underline{\mathbf{S}}_{X_a X_a}(n) = (1 - \lambda_a)\underline{\mathbf{S}}_{X_a X_a}(n-1) + \lambda_a \,\mathrm{diag}\left\{\underline{\mathbf{X}}^H_a(n)\underline{\mathbf{X}}_a(n)\right\}, \quad (60)$$

where $\mathrm{diag}\{\cdot\}$ extracts the main diagonal of the given argument.

Finally, one block of length L_g/α_g of the GSC output signal is obtained by saving the last L_g/α_g samples of $\mathbf{e}_a(n)$.

5.6. Adaptation control

In this system, we modified the GSC adaptation control presented in [8] to a DFT-binwise operation, which increases convergence speed and robustness significantly. It is based on a spatial SNR estimate: the FBF output yields an estimate of the desired signal PSD. A fixed beamformer, which is complementary to the FBF, yields an estimate of the interference PSD. A frequency-dependent SNR estimate is then obtained by the ratio of desired signal PSD and interference PSD. This is used for a bin-wise decision whether the ABM or the AIC is adapted.

We do not consider the stepsize control of the AEC here. Various stepsize-control methods can be found in the literature (see, e.g., [2, 20]). The adaptation control of GSC and SAEC does not need to rely on synergies between both adaptation mechanisms, so that GSC and SAEC can be adapted independently of each other.

---

[2] A norm constraint for improved robustness against desired signal cancellation may be introduced according to [9, 10].

6. REAL-TIME IMPLEMENTATION AND EXPERIMENTAL EVALUATION

For demonstrating the performance of our acoustic human-machine interface in real time, we implemented the FGSAEC algorithm on a PC platform. The multichannel audio capture unit is realized as separate hardware integrating the microphones, the preamplifiers, the A/D conversion, and the microphone calibration. The digitized sensor data is fed into the PC via a standard USB port with specific drivers for the microphone array.

Our experiments were conducted on an Intel Pentium IV 1.8 GHz processor at a sampling rate of 12 kHz. Optimum performance of the frequency-domain RGSC (FGSC) and FGSAEC was obtained with 30% and 52% CPU load, respectively.

For all experiments, we use a linear microphone array with 8 equally spaced, broadside-steered sensors with 4 cm spacing in an office environment with 300 ms reverberation time. The male desired speaker and the male interferer with an average signal power ratio of 0 dB are located in the array look-direction and 30 degrees off the array axis, respectively. The stereophonic loudspeakers emitting music are placed to the left and to the right of the microphone array. All distances to the array center are 60 cm. The frequency band is 300 Hz–5.9 kHz. The FBF is realized by a Dolph-Chebyshev design described in [9]. Typical numerical values for filter lengths and block overlapping factors are $L_g = 128$, $L_h = 2048$ and $\alpha_g = 2$, $\alpha_h = 8$, respectively. In Section 6.1, we study steady-state performance of FGSAEC. In Section 6.2, tracking capability of RGSC with moving desired speaker is illustrated.

6.1. Performance after convergence of the adaptive filters

For evaluating the proposed system after convergence of the adaptive filters, we compare the average interference rejection (IR) and the average echo-return-loss enhancement (ERLE) of FGSAEC, FGSC, frequency-domain AEGSC (FAEGSC), and TGSAEC, (the time-domain equivalent of FGSAEC) for only interference and for double-talk of interference and desired speaker.

Since it is difficult to study IR and ERLE separately for real-time scenarios, we illustrate the results that we obtained with recorded signals in simulations. Audio examples which illustrate the performance of the real-time system can be found in [21].

The results are depicted in Table 1. For only interference, IR and ERLE are higher than for the double-talk case since the ABM is fixed and since the AIC can be adapted permanently over the entire frequency range, yielding optimum-tracking capability of nonstationary interference. The performance of TGSAEC and FGSAEC is identical. During double-talk, IR and ERLE are considerably improved for FGSAEC relative to TGSAEC as controlling the adaptation in individual frequency bins still allows tracking the transient ABM and of nonstationary interference at frequencies with low SNR [9]. FGSAEC clearly improves the suppression of acoustic echoes relative to FGSC; however, optimum performance

TABLE 1: Performance evaluation.

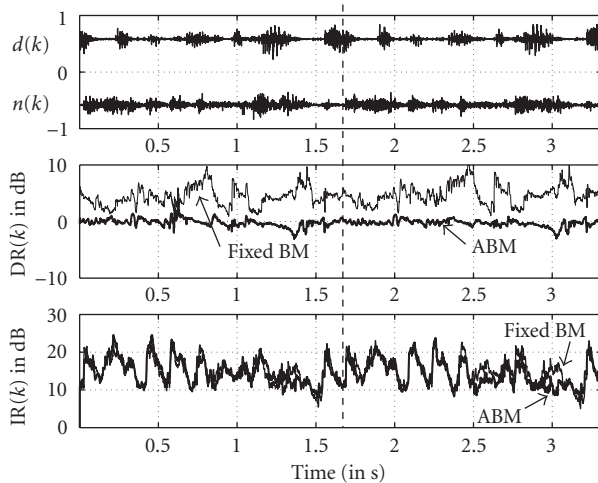| (in dB)	Interference only		Double-talk	
	IR	ERLE	IR	ERLE
TGSAEC	22.4	26.0	5.6	12.3
FGSAEC	21.1	25.6	14.7	21.0
FGSC	20.7	21.9	14.5	14.7
FAEGSC	22.0	30.5	14.9	28.1

FIGURE 8: Tracking performance of the ABM in comparison with the BM after [11], position change of the desired speaker at 1.66 s.

of FAEGSC cannot be obtained due to leakage effects across the GSC sidelobe-cancelling path [17].

6.2. Tracking of the ABM

For illustrating the tracking capability of the ABM, desired signal rejection (DR) and interference rejection of the FGSC over time are measured for changing desired speaker position. Both rejections are estimated by the ratio of recursively averaged squared sensor signals and beamformer output signals w.r.t. the desired signal components and interference components.

Figure 8 depicts the results for the ABM in comparison with a fixed blocking matrix (BM) after [11]. Parameters are chosen to have the same IR(k) for ABM and fixed BM. For controlling the adaptation of both GSC realizations, knowledge about the true sensor SIR is assumed. For controlling the adaptation of ABM and AIC, knowledge of the true sensor SNR is assumed.

At 1.66 s, the desired speaker switches from broadside (0 degrees) to 10 degrees; neither interference suppression nor desired signal quality are impaired due to fast-tracking capability of FGSC. The fixed blocking matrix is designed to suppress signals from a single-propagation path. Due to reverberation, it leads to considerable desired signal distortion before and after changing the desired speaker position.

7. CONCLUSIONS

The presented signal-processing algorithms describe an example of efficient integration of adaptive beamforming and multichannel AEC which meets well the practical requirements regarding the suppression of interference and acoustic echoes for seamless acoustic human-machine interfaces. Without structural changes, it can be extended to more reproduction channels and even multichannel recording. Moving the implementation from the PC platform to a more specialized hardware will be smooth as long as efficient and numerically sound implementations of basic signal-processing algorithms such as fast Fourier transforms are assured and as long as block processing and control loops pose no obstacles.

REFERENCES

[1] C. Breining, P. Dreiseitel, E. Hänsler, et al., "Acoustic echo control—an application of very-high-order adaptive filters," IEEE Signal Processing Magazine, vol. 16, no. 4, pp. 42–69, 1999.

[2] S. L. Gay and J. Benesty, Eds., Acoustic Signal Processing for Telecommunications, Kluwer Academic Publishers, Boston, Mass, USA, 2000.

[3] H. Buchner and W. Kellermann, "Improved Kalman gain computation for multichannel frequency-domain adaptive filtering and application to acoustic echo cancellation," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, pp. 1909–1912, Orlando, Fla, USA, May 2002.

[4] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel frequency-domain adaptive filtering with application to acoustic echo cancellation," in Adaptive Signal Processing: Application to Real-World Problems, J. Benesty and Y. Huang, Eds., Springer-Verlag, Berlin, Germany, January 2003, to appear.

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

[6] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in Adaptive Signal Processing: Application to Real-World Problems, J. Benesty and Y. Huang, Eds., Springer-Verlag, Berlin, Germany, January 2003, to appear.

[7] 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., pp. 281–306, Springer-Verlag, Berlin, Germany, 2001.

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

[9] W. Herbordt and W. Kellermann, "Frequency-domain integration of acoustic echo cancellation and a generalized sidelobe canceller with improved robustness," *European Trans. on Telecommunications*, vol. 13, no. 2, pp. 123–132, 2002.

[10] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

[11] 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, 1982.

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

[13] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic echo cancellation—An overview of the fundamental problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148–151, 1995.

[14] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech, and Audio Processing*, vol. 6, no. 2, pp. 156–165, 1998.

[15] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo canceler with correct echo-path identification based on an input-sliding technique," *IEEE Trans. Signal Processing*, vol. 49, no. 11, pp. 2577–2587, 2001.

[16] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, Upper Saddle River, NJ, USA, 1996.

[17] W. Herbordt and W. Kellermann, "Limits for generalized sidelobe cancellers with embedded acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Salt Lake City, Utah, USA, May 2001.

[18] E. Moulines, O. A. Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: structure and convergence analysis," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 14–28, 1995.

[19] C. L. Dolph, "A current distribution for broadside arrays which optimize the relationship between beam width and side-lobe level," *Proceedings of the I.R.E. and Waves and Electrons*, vol. 34, no. 6, pp. 335–348, 1946.

[20] A. Mader, H. Puder, and G. U. Schmidt, "Step-size controls for acoustic echo cancellation filters—an overview," *Signal Processing*, vol. 80, no. 9, pp. 1697–1719, 2000.

[21] W. Herbordt, *http://www.LNT.de/~herbordt*.

Wolfgang Herbordt received the Dipl.-Ing. degree in electrical engineering from the University of Erlangen-Nuremberg in 1999 after studying at INSA Rennes, France, and at the University Erlangen-Nuremberg, Germany. Since 1999, he is a Researcher at the Chair of Multimedia Communications and Signal Processing at the University of Erlangen-Nuremberg. He has been involved in signal processing for hands-free acoustic human-machine interfaces, that is, array signal-processing, blind source separation, multichannel AEC, noise-robust speech recognition, and acoustic source localization. He received the Best Student Paper Award at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul 2000. In 2000, he was a visiting Scientist at Intel Corp., Hillsboro, OR, and in 2001 and 2002 at Intel Corp., Beijing, China.

Herbert Buchner is a member of the research staff at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Germany. He received the Dipl.-Ing. (FH) and the Dipl.-Ing. Univ. degrees in electrical engineering from the University of Applied Sciences, Regensburg in 1997, and the University of Erlangen-Nuremberg in 2000, respectively. In 1995, he was a visiting Researcher at the Colorado Optoelectronic Computing Systems Center (OCS), Boulder/Ft. Collins, Colorado, where he worked in the field of microwave technology. From 1996 to 1997, he did research at the Cyber Space Labs (former Human Interface Labs) of the R&D division of Nippon Telegraph and Telephone Corp. (NTT), Tokyo, Japan, working on adaptive filtering for teleconferencing. In 1997/1998, he was with the Driver Information Systems Department of Siemens Automotive in Regensburg, Germany. His current areas of interest include efficient multichannel algorithms for adaptive digital filtering, and their applications for acoustic human-machine interfaces such as multichannel AEC, beamforming, blind source separation, and dereverberation. He received the VDI Award in 1998 for his Dipl.-Ing. (FH) thesis from the Verein Deutscher Ingenieure and an IEEE Best Student Paper Award in 2001.

Walter Kellermann is a Professor of communications at the Chair of Multimedia Communications and Signal Processing of the University of Erlangen-Nuremberg, Germany. He received the Dipl.-Ing. (univ.) degree in electrical engineering from the University of Erlangen-Nuremberg in 1983, and the Dr.-Ing. degree from the Technical University Darmstadt, Germany, in 1988. From 1989 to 1990, he was a Postdoctoral Member of technical staff at AT&T Bell Laboratories, Murray Hill, NJ. From 1990 to 1993, he was with Philips Kommunikations Industrie, Nuremberg, Germany. From 1993 to 1999, he was a Professor at the Fachhochschule Regensburg, and in 1997, he became a Director of the Institute of Applied Research of the Fachhochschule Regensburg. In 1999, he cofounded DSP Solutions, a consulting firm in digital signal-processing, and he joined the University Erlangen-Nuremberg as a professor and head of the audio research laboratory. Dr. Kellermann authored and coauthored five book chapters and more than 35 papers in journals and conference proceedings. He served as a Guest Editor for various journals and presently serves as an Associate Editor of IEEE Transactions on Speech and Audio Processing. His current research interests include speech signal-processing, array signal-processing, and adaptive filtering and its applications to acoustic human/machine interfaces.