Chapter 5
Innovative Adaptive System For Optimal Noise and Acoustic Echo Cancellation Using Proposed Adaptive Beamformer and Developed IIR-RLS Filter

5.1 Introduction

In previous chapters, the noise reduction is achieved by using the proposed adaptive beamformer which has shown to yield high noise reduction performance in practical environments. While the acoustic echo can be remedied by using the developed adaptive IIR-RLS filter. However, in many hands-free communication systems both the background noise and a far-end acoustic echo may be present. Therefore, we need to combine the noise reduction unit with the acoustic echo cancellation unit to ensure a high quality of speech transmission. Especially, the technique which we use for noise reduction can easily be modified to incorporate acoustic echo canceller. Keeping the computational complexity of the overall combined adaptive system at very low level. Thus, the proposed adaptive combined system is more promising technique for real-life implementation.

An acoustic echo canceller (AEC) traditionally uses an adaptive filter with typically a large number of filter taps, for instance 1000 taps for a signal sampled at 8kHz. The reason is that it aims to model the acoustic impulse response of the room. Long filters imply a large computational complexity burden and slow convergence rate. While adaptive noise canceller (ANC) typically have shorter filters (e.g. Delay-and-sum beamformer do not model the room impulse response, but are designed to have a certain spatial sensitivity pattern, which can obtained by relatively short filter lengths). We are interested in ANC techniques that use multi-microphone in order to exploit both spatial and spectral characteristics of the desired speech signal and noise signals.

For optimally suppressing noise and acoustic echoes, it is thus desirable to combine acoustic echo cancellation with beamforming microphone array in modern hands-free systems such as teleconferencing systems, mobile phones and computers, car information systems, and home entertainment equipment like voice controlled systems [56, 89, 90, 177-187]. For optimum performance, synergies between the AEC and the beamforming
microphone array should be maximally exploited while the computational complexity should be kept moderate. Therefore, we study in this chapter various combinations of acoustic echo cancellation and beamforming. Each of these is likely have their own merits under different acoustic conditions. Compared to \([56, 89, 90, 91, 177, 188, 189]\), we discuss the different options in more details. Especially, a new structure is presented, which uses the proposed adaptive beamformer and an efficient recursive least squares (RLS) algorithm based on infinite impulse response (IIR) filter for acoustic echo cancellation, which were developed in previous chapters. The performance of this new structure is outperforms the other combinations for noise reduction and acoustic echo cancellation while keeping the computational complexity moderate \([141, 206]\). In Section 5.2, we formulate the problem of present noise and acoustic echoes in reverberant environments. Section 5.3 and Section 5.4 presents two generic combinations of AEC and beamforming, with the echo canceller in front of the beamformer (this scheme is denoted by AEC-first) and with the echo canceller after the beamformer (denoted by BF-first). For the GSC beamformer, two methods for integrating of the AEC into the beamformer are possible (Section 5.5). First, the AEC can be placed in the reference path of the GSC after the quiescent weight vector. Second, the AEC can be integrated into the multi-channel noise canceller. In Section 5.6, we describe an approach for integrating echo canceller based IIR-RLS filter into the proposed adaptive beamformer. The acoustic echo canceller can be placed in the reference path of the proposed adaptive beamformer after the fixed beamformer.

5.2 Model Description

We now assume that both noise and acoustic echoes are present. Since the far-end signals emitted by the loudspeakers are readily available, these signals can be used as a reference for the echo canceller (unlike the noise sources, which no reference is available). In this research work, we will assume that a single far-end echo source is present, although the presented results can be extended to the case of multiple far-end echo sources. Each microphone signal \(y_n(k) \ n = 0, \ldots, N - 1\) consists of a filtered version of the clean speech signal \(s(k)\), additive background noise \(r(k)\) and far-end
signals $f_l(k), l = 0, \ldots, M$, where $M$ is the number of far-end sources (as mentioned, we will take only one far-end source, i.e., $l = 0$). The $n$th microphone signal $y_n(k)$ can then be written as

$$y_n(k) = h_n(k) \ast s(k) + r_n(k) + v_n(k)$$

(5.1)

where $h_n(k)$ is the acoustic impulse response between the speech source $s(k)$ and the $n$th microphone, $\ast$ denotes convolution and $h'_n(k)$ is the acoustic impulse response between the far-end loudspeaker $f(k)$ and the $n$th microphone. $x_n(k)$ and $r_n(k)$ are the speech components and the noise components at $n$th microphone respectively. While $v_n(k)$ represents the far-end echo components at $n$th microphone. The enhanced speech signal can be obtained by filtering the microphone signals $y_n(k)$ with the filters $w_n(k)$ and $w'_n(k)$ as given by

$$\hat{s}(k) = \sum_{n=0}^{N-1} w_n(k) \ast h_n(k) \ast s(k) + \sum_{n=0}^{N-1} w'_n(k) \ast f_n(k)$$

(5.2)

where $w_n(k)$ and $w'_n(k)$ are the adaptive filter coefficients for noise reduction and for far-end echo cancellation, respectively. The filters $w_n(k)$ generally are finite impulse response (FIR) filters of length $L$, i.e.,

$$w_n(k) = [w_{n0}(k) \quad w_{n1} \quad \ldots w_{NL-1}(k)]^T$$

(5.3)

and similarly the filter $w'_n(k)$ of length $L_{AEC}$ is defined for far-end signal. The goal of combined noise reduction and acoustic echo cancellation is to cancel the far-end echo components $v_n(k)$ using the filter $w'_n(k)$ and also to reduce the noise components $r_n(k)$ (and possibly also far-end echo components to some extent) using the filters $w_n(k)$. Combinations of beamforming and acoustic echo cancellation can be devised in two obvious ways as depicted in Figs. 5.1 and 5.2. The AEC can be placed in the microphone channels in front of the beamformer (this is denoted by AEC-first) or after the beamformer (this denoted by BF-first). Depending on the succession of acoustic echo cancellation and beamforming, interactions have to be considered in order to optimally exploit synergies. As an example for the beamformer, we will use the standard
GSC beamformer and the proposed adaptive beamformer given in Chapter 3. However, the discussion is equally valid for other adaptive beamformer realizations.
5.3 AEC-first Scheme

In the scheme presented in Fig. 5.1, first the echo components \( v_{\alpha}(k) \) in all microphone signals are cancelled using \( N \) filters \( w_{\alpha}(k), \ n = 0...N - 1 \). The goal of each filter \( w_{\alpha}(k) \) is to model the acoustic impulse response \( h_{\alpha}(k) \), which can be achieved by using an adaptive filtering algorithm (see chapter 4). Next, a multi-microphone noise reduction technique (e.g. GSC, multi-channel Wiener filter (MWF), or proposed adaptive beamformer, which are explained in chapter 3) is applied to the echo-free microphone signals, yielding the same noise reduction performance as if no echo source was present. However, the computational complexity of this noise and echo reduction scheme (AEC-first) is quite high, since \( N \) adaptive filters, typically with large filter lengths, are required.

5.3.1 Implications for the AEC and for Fixed Beamformers

A simple delay-and-sum (DS) beamformer as the noise reduction unit has been chosen. The setup of a multi-channel AEC-unit preceding the DS beamformer is discussed in this subsection. With this setup, the AEC-unit is independent of the DS beamformer. The echo suppression length of an AEC’s adaptive filter can be greatly reduced. Resulting that the combined system have much less computational complexity [180]. Adaptive beamformer are discussed next.

5.3.2 Implications for Adaptive Beamformers

For adaptive data-depended beamformers, positive synergies between acoustic echo cancellation and the beamformer should be fulfilled. For \( ERLE_{\text{nag}} = 20\text{dB} \) in an environment with \( T_{60} = 400\text{ms} \), the length of the adaptive filters of the AEC is often as large as \( L_{\text{AEC}} = 1000 \text{ taps} \), at a sampling rate of \( f_s = 8\text{kHz} \) according to Equation (4.5) in Chapter 4. This leads to a slow convergence of the adaptive filters. For data-depended beamformers, however, a typical filter length for noise reduction equal to 15dB in an environment with \( T_{60} = 400\text{ms} \) is \( L_{\text{AEC}} = 256\text{ taps} \) (see chapter 3). Adaptive data-depended beamformers thus converge considerably faster than AECs if comparable adaptation algorithms are used for the AEC and for the beamformer. This effect can be exploited if
the beamformer is placed after the AEC. The beamformer does not feel any adverse effects of the time-variance of the AEC, since the beamformer can track the time-variance of the AEC due to the faster convergence. Whenever acoustic echoes leak through the AEC because of, e.g., change of the echo paths, the beamformer suppresses both acoustic echoes and noise signals (as long as the beamformer adaptation is not impaired due to presence of the desired signal). The beamformer thus uses the spatial degrees of freedom of the microphone array for both echo suppression and noise reduction. The number of spatial degrees of freedom for suppressing noise signals reduces. After convergence of the AEC, the AEC efficiently suppresses acoustic echoes, and the beamformer only sees residual echoes with low energy. The beamformer can concentrate on the suppression of noise signals with all spatial degrees of freedom. The suppression of noise signals increases.

5.3.3 Computational Complexity of the AEC-First

At least the filtering and the filter coefficients update of the AEC adaptation will require $N$-fold computational cost compared to a single-channel AEC. Even with continuing growth of the performance-cost ratio of signal processing hardware, this computational load will remain prohibitive in the near future for many cost-sensitive or very large systems employing $N=5,...,512$ microphones [56, 190-194]. One option to alleviate the computational burden is to reduce the length $L_{AEC}$ in Equation (4.5) of the FIR filter models $w_{AEC}(k)$, and to rely on the beamformer for suppressing the residual echoes $e(k)$. Shortening $w_{AEC}(k)$ implies however that the adaptation of the AEC is disturbed by an increased noise component, which is due to the unmodeled tail of the true echo path impulse response $w_{tar}(k)$ [56]. The need for an adaptive filter with long impulse response arises in applications such as acoustic echo cancellation in hands-free systems. We may deal with such a requirement by implementing the adaptive filters either in the frequency domain [60, 128, 195-198], or may build the adaptive process around a linear filter with infinite-duration impulse response (IIR) filter at a computational cost far less than that of the corresponding FIR adaptive filters, as discussed in chapter 4. In this thesis, main attention is paid to deal with adaptive filters using IIR structure with the
double purpose of reducing the computational complexity and of its more suitability to model physical systems due to its pole-zero structure.

5.4 BF-first Scheme

The scheme of BF-first is presented in Fig. 5.2. A single-channel echo canceller $w'(k)$ is used after a multi-microphone noise reduction technique, thereby reducing the computational complexity compared to the AEC-first scheme in Fig. 5.1. A possible implementation has been discussed in [89, 90, 128], where echo cancellation is combined with the GSC. Since the multi-microphone noise reduction technique just considers the far-end echo sources as an additional noise source, it now has to reduce both the noise $r_i(k)$ and far-end echo components $v_{i0}(k)$. Therefore, it is obvious that the noise reduction performance for the noise sources is worse than if no echo source were present. Of course, the noise reduction technique now also reduces the $v_{i0}(k)$ to some extent. The adaptive filter $w'(k)$ has to model $\sum_{n=0}^{N-1} w(n) \otimes h'_{i0}(k)$, with $w_i(k)$, $n = 0...N - 1$, generally also adaptive filters. This cascade of two adaptive filtering algorithms may give rise to problem when the adaptive filter $w'_{i0}(k)$ cannot track the changes of the adaptive filters $w_i(k)$. The BF-first scheme can not be the good choice for combining the noise and acoustic echo cancellation. Hence, an integrated noise and echo reduction approach is called for, which is explained in Section 5.5 and 5.6. It will be shown that by using the proposed adaptive beamformer, the combined adaptive system can obtain the same noise reduction performance for the background noise sources as if no echo source were present, and the far-end echo components can be completely cancelled.

5.4.1 Implications for the Beamformer

The beamformer operates without any repercussions from the AEC. The performance of the beamformer thus corresponds to that discussed in Section 5.3.1. For adaptive signal-depended beamformers, it is possible to stop the adaptation of the beamformer if acoustic echoes are present. The beamformer is thus not optimized for the spatio-
temporal characteristics of the acoustic echoes. On one hand, this limits the echo suppression of the beamformer and reduces the tracking capability of the beamformer since the beamformer is adapted less frequently. On the other hand, however, more spatial degrees of freedom are available for noise suppression. On the contrary, the beamformer can be designed with fixed spatial nulls for the positions of the loudspeakers in order to provide optimum suppression of the direct signal path and of reflections of the acoustic echoes even for strongly time-varying environments. This is specially important for loudspeakers with known positions close to the array where the direct-path-to-reverberation ratio is high [199].

5.4.2 Implications for the AEC

The presence of the beamformer in the signal path of the echo canceller can be viewed from a simple noise reduction perspective and from a system identification perspective.

Influence of the beamformer on the noise reduction. Due to the spatial filtering of the beamformer, the power ratio of acoustic echoes to the desired signal and to noise signals is generally changed at the input of the AEC. A reduction of this power ratio makes the identification of the echo paths for the echo canceller more difficult, which generally reduces the convergence speed of the AEC and the echo suppression in time-varying environments [200]. This aspect has to be especially considered for adaptive beamformers which provide high suppression of noise signals. Consider, for example, acoustic echoes from a loudspeaker in an environment with spatially diffuse noise components, where the echo path is time-varying. An adaptive beamformer tracks the time-variance of the echo path and suppresses the spatially coherent acoustic echoes more efficiently than non-coherent noise. The power ratio of acoustic echoes to the diffuse noise decreases. The convergence speed of the echo canceller decreases, and the echo suppression reduces in turn. In contrast, if the beamformer is not optimized for suppression of acoustic echoes, the background noise is generally more efficient suppressed than the acoustic echoes. The power ratio of acoustic echoes to diffuse noise increases, yielding faster convergence and a higher echo suppression of the AEC.

Beamformer as a system-identification. The AEC has to model the echo paths and the beamformer weight vector $w_{\text{GSC}}(k)$. This generally increases the length of the AEC filters
Chapter 5 Innovative Adaptive System For Optimal Noise and Acoustic Echo Cancellation

for a given echo suppression compared to AEC-first and shows all problems of longer adaptive filters. For adaptive beamformers, the necessity to model the beamformer weight vector raises severe difficulties, since the AEC cannot track the beamformer satisfactorily. Generally, adaptive beamformers are designed for fast convergence and fast tracking in order to optimally suppress time-varying noise and in order to track moving desired sources with time-varying second-order statistics. Essentially, this means that the length of the adaptive filters is moderate ($L_{\text{AEC}}=64,...,512$) and that fast-convergence adaptation algorithms are used, as e.g., [201, 202]. The length of the AEC filters should be much longer because of the long reverberation time of acoustic environments and because the beamformer has to be modeled in addition to the room impulse responses for BF-first. Typically, the AEC filters are by a factor of 4...8 longer than the beamformer filters for the same environment, which yields the same reduction of the convergence speed of the AEC. Therefore, the AEC converges slower than the beamformer. For the AEC being able to converge, the beamformer has to be time-invariant for a sufficiently long period of time. This, of course, is not acceptable for time-varying acoustic conditions and for non-stationary signals, where the beamformer should be adapted continuously. As a result, the AEC will be almost ineffective for BF-first. Therefore, we will not further consider this scheme.

5.5 Integration of AEC with the GSC Beamformer

The potentially high computational complexity of AEC-first and the tracking problematic for BF-first suggest to integrate the echo canceller into the time-varying beamformer such that optimum synergies between beamforming and acoustic echo cancellation are obtained with minimum computational complexity. For optimum synergies, we demand that the beamformer does not see acoustic echo signals after convergence of the AEC, so that the number of degrees of freedom for suppressing noise is maximum.

The decomposition of the beamformer into a time-invariant (delay and sum beamformer) part parallel to a time-varying (multi-channel noise canceller) part-according to the GSC structure with a time-invariant quiescent weight vector $w_t$-yields two options for integrating the AEC: First, the AEC can be cascaded with the quiescent weight vector,
yielding the GSAEC structure (Section 5.5.1) [56, 90, 128, 188, 189]. Second, the AEC can be combined with the multi-channel noise canceller as additional channels of the sidelobe canceling path, yielding the GEIC structure (Section 5.5.2) [56, 91, 203].

5.5.1 AEC Embedded into GSC Beamformer (GSAEC)

With the AEC placed after the quiescent weight vector $w_c$, we obviously obtain for the AEC the characteristics of BF-first with a time-invariant realization of the beamformer. Especially, the computational complexity is reduced to that of a single AEC for an arbitrary number of microphones. However, from Fig. 5.3, we see that acoustic echoes are still present in the adaptive sidelobe canceling path while they may be efficiently cancelled by the AEC in the reference path. Therefore, the multi-channel noise canceller $w_{m,n}(k)$ has to prevent leakage of acoustic echoes through the sidelobe canceling path in addition to cancel noise signals. The multi-channel noise canceller has to cancel acoustic echoes before the subtraction of the output of the noise canceller from the output of the AEC. Interpreting the weight vector $w_{m,n}(k)$ as a beamformer with the output signals of the blocking matrix $B_r(k)$ as input signals, this means that $w_{m,n}(k)$ has to place spatial nulls for the acoustic echo signals at the output of $B_r(k)$. As a result, the number of spatial degrees of freedom for suppression of noise is generally not increased relative to beamforming alone. However, positive synergies cannot be obtained [204].

![Fig. 5.3 Structures Of The GSAEC Schemes.](image-url)
5.5.2 AEC Integrated with Multi-channel Noise Canceller (GEIC)

The analysis of GSAEC has shown that the independent adaptation of the AEC in the reference path does not increase the number of spatial degrees of freedom for the multi-channel noise canceller, since acoustic echoes are not cancelled in the sidelobe canceling path before they are seen by the multi-channel noise canceller. Joint adaptation of the multi-channel noise canceller and of the AEC resolves this problem [91, 203]. In Fig. 5.4, the GSAEC is redrawn with the filters of the AEC moved parallel to the multi-channel noise canceller. Both units use the same error signal, i.e., the output signal of the robust GSC, for optimization so that the loudspeaker signals can be interpreted as additional channels of the sidelobe canceling path. This combination is denoted by GEIC [91]. The stacked weight vector of AEC and multi-channel noise canceller is defined as [91]

\[ w_{GSAEC}(k) = [w_{AEC}(k), w_{GSC}(k)]^T \]  

(5.4)

However, for using this scheme it is required a necessary preconditions such as the weight vector of the AEC, \( w_{AEC}(k) \), should be adapted simultaneously with \( w_{GSC}(k) \), and the length \( L_{AEC} \) of the filter \( w_{AEC}(k) \) should be equal to the length \( L_{GSC} \) of the filters \( w_{GSC}(k) \). With these assumptions, the rate of convergence of the AEC is equal to the rate of convergence of the noise canceller, and tracking of the time-variance of the beamformer is assured. On one hand, although the GEIC scheme showed the good tracking ability and better echo suppression with compare to AEC-first, for the time-varying echo paths in double-talk situation. On the other hand, due to the limited length of the AEC filters in GEIC scheme, it cannot be expected for converged adaptive filters that the suppression of acoustic echoes of GEIC is as high as for AEC-first, were longer filters for the AEC are possible. The maximum suppression of acoustic echoes, which would be obtained for single-talk situations can hardly be obtained in turn. In addition, a combination of acoustic echo cancellation and multi-microphone noise reduction based on generalized singular value decomposition (GSVD) using same idea, i.e., incorporates the echo reference directly into the GSVD-based optimal filtering technique was investigated in [183, 205]. It shows that the AEC-first scheme always out performance the scheme incorporating the echo reference directly into the GSVD-based optimal filtering technique.
5.6 Integration of AEC Based IIR-RLS Filter into the Proposed Adaptive Beamformer

A comparative study of the performance of integrating schemes (GSAEC and GEIC which are mentioned in the preceding section) with the existing conventional robust generalized sidelobe canceling (RGSC) technique [56, 90, 91, 188, 189, 128, 203], has shown that both suffer from disadvantages. In order to overcome these disadvantages, we propose a novel method for combining noise and acoustic echo canceller [206]. The proposed adaptive combined system includes a new adaptive beamformer given in Chapter 3 and an efficient recursive least squares (RLS) algorithm based on infinite impulse response (IIR) filter for acoustic echo cancellation given in Chapter 4. The proposed adaptive beamformer uses a spatial adaptive blocking matrix (BM). This BM is based on linear prediction error filters (LPEFs). It is known that the coefficients of the LPEFs converge such that the prediction error signals (output of BM) becomes white [85]. Since a near-end speech signal and far-end echo signal can be represented as the stationery signals over a short interval of time [145], most of near-end speech signal and far-end echo signal will be predicted (cancelled out) by the linear prediction error filters in the adaptive sidelobe canceling path. On the other hand, when the input signals of the LPEFs is a background noise, the prediction error signals becomes white. Then we can reconstruct the background noise from the prediction error signals by adaptive noise estimation filters (ANEFs) in the multi-channel noise canceller (MC) which is used for inverse modeling.
5.6.1 The Proposed System Details

In this part, we introduce an approach integrating acoustic echo canceller and a noise reduction system which is illustrated in Fig. 5.5. In the proposed adaptive combined system, the AEC is located behind the fixed beamformer (FBF) (this scheme is also denoted by AEC-behind FBF). Obviously, the computational complexity is reduced to that of a single AEC for an arbitrary number of microphones and the AEC does not feel any repercussions from the adaptive beamformer. The AEC is based on IIR-RLS algorithm [165]. The principle of AEC is explained in Chapter 4, Section 4.5. Referring to Fig. 5.5, first, the echo components $v_n(k)$ are efficiently suppressed by the AEC in the reference path. However, from Fig.5.5, it can be seen that the acoustic echo components will also be contained in the adaptive sidelobe canceling path. In order to cancel the acoustic echo $v_n(k)$ as well as the desired speech (near-end) components $x_s(k)$ in the adaptive sidelobe canceling path, we use a spatial adaptive blocking matrix (BM). This blocking matrix is based on linear prediction error filters (LPEFs).

![Diagram of the proposed system](image)

Fig. 5.5 Structure Of The Proposed System (AEC-behind FBF).
Since a desired speech signal and echo signal (which has speech characteristics) is considered to be stationary and periodic signals during a short time interval, most of near-end speech and far-end echo signals can be predicted with linear prediction error filters at the BM outputs. However, in real-time implementation, the residual desired speech and echo might be included in whitened prediction error signals (outputs of BM) which might lead to speech cancellation in the reference path. In order to improve on these negative effect, we introduce ANEFs. The ANEFs in the MC, which are used for reconstructing the background noise from whitened signals (prediction error signals), can not estimate the residual desired speech and echo due to the prediction error signals which does not correlated with the desired speech and echo in the reference path. Moreover, a small step size is used to update tap coefficients of ANEFs, so as not to estimate the residual speech signal. Therefore, the proposed system can reduce the cancellation of desired speech signal effectively. Then the output of the ANEFs represents only the reconstructed background noise signal.

The proposed adaptive beamformer technique has been studied for noise reduction only (see chapter 3). In this chapter, also a far-end echo signal is considered. Therefore, the adaptive blocking matrix of the proposed beamformer is redesigned to cancel out the near-end speech signal and far-end echo signal in the sidelobe canceling path. Referring to Fig. 5.5, the proposed adaptive combined-system consists of a fixed beamformer, an adaptive blocking matrix (BM) with LPEFs and multi-channel noise canceller (MC) with ANEFs used for inverse modeling.

**Fixed beamformer:** The output signal of the fixed beamformer is defined as
\[
d(k) = W_d^T(k) y(k)
\]
where
\[
y(k) = [y_0(k), y_1(k), ..., y_n(k)]^T
\]
\[
y_n(k) = [y_0(k), y_1(k), ..., y_n(k - M + 1)]^T
\]
and
\[
W_d(k) = [W_{dz}(k), W_{dz}(k), ..., W_{dz}(k)]^T
\]
\[
W_{dz}(k) = [W_{dz}(k), W_{dz}(k - 1), ..., W_{dz}(k - M + 1)]^T
\]
is the fixed beamformer weight vector of the DS beamformer. We demand that the
desired signal processed by a fixed beamformer weight vector $\mathbf{w}_d(k)$ is not distorted at the
output of the proposed adaptive combined system over a data block of $M$ successive
samples. Therefore, the $\mathbf{w}_d(k)$ is chosen such that the following constraint equation is
fulfilled

$$\mathbf{c}(k) = \mathbf{C}^T(k)\mathbf{w}_d(k),$$

where $\mathbf{C}(k) = \text{diag}\{\mathbf{x}(k), \mathbf{x}(k-1), \ldots, \mathbf{x}(k-M+1)\}$

The matrix $\mathbf{C}(k)$ of size $NM \times CM$ is the constraint matrix with linearly independent
columns. The vector $\mathbf{c}(k)$ of size $CM \times 1$ is the constraint vector. The constraint
matrix $\mathbf{C}(k)$ defines the directions where constraint should be put on, while the constraint
vector $\mathbf{c}(k)$ specifies the beamformer response for the constraint directions.

**Blocking matrix (BM):** Consider the following $N \times M$ matrix $\mathbf{B}(K)$

$$\mathbf{B}(k) = [\mathbf{b}_1(k), \mathbf{b}_2(k), \ldots, \mathbf{b}_N(k)]^T$$

with $\mathbf{b}_n(k) = [\mathbf{b}_{n1}(k), \mathbf{b}_{n2}(k), \ldots, \mathbf{b}_{nM}(k)]^T$

creates $N$ so-called noise references

$$e_{n,0}(k) = \mathbf{b}_n^T(k)\mathbf{y}(k) = e_{n,0}^w(k) + e_{n,0}^e(k), \quad n = 0, \ldots, N - 1$$

By using linear prediction error filters (LPEFs) to predict the desired speech and far-
end echo at the outputs of blocking matrix, the whitened noise contribution $e_{n,0}^w(k)$ are
dominant compared to the residual desired speech and far-end echo contributions $e_{n,0}^e(k)$.

In general, the blocking matrix $\mathbf{B}(K)$ is designed to be orthogonal to the reference path
(output of fixed beamformer). Ideally, the noise references do not contain any desired
speech and echo contributions, i.e., $e_{n,0}^w(k) = 0$ for $n = 0, \ldots, N - 1$. The realization of the
blocking matrix is depicted in Fig. 5.5. The input signal of each LPEF is the noisy
microphone signals $\mathbf{y}_n(k)$. The output of the $n$th LPEF is given by

$$e_{n,0}(k) = \mathbf{y}_n(k) - \sum_{i=1}^{M \text{LPEF}} \mathbf{b}_{n,i}(k)\mathbf{y}_n(k-i)$$
where $\mathbf{b}_n(k) = [b_{n1}(k), b_{n2}(k), \ldots, b_{n,M_{LPEF}}(k)]^T$ are the tap coefficients of the $n$th LPEF which are updated by normalized least mean square (NLMS) adaptive algorithm [85]. Since a speech signal can be represented as the stationary signal over a short interval of time [145], most of near-end speech signal and far-end echo signal can be predicted by the LPEFs at the output of blocking matrix. On the other hand, when the input signals of the LPEFs are background noises, the coefficients of the LPEFs will converge such that the signals $e_{n,e}(k)$ becomes white [85].

Then we can reconstruct the background noise from the prediction error signals $e_{n,e}(k)$ by inverse modeling of the transfer functions of LPEFs. This modeling is performed by the ANEFs in the multi-channel noise canceller (MC).

**Multi-channel noise canceller (MC):** By using the GSC-like structure, we have two signals. The first is speech reference, $d(k)$, which contains a desired speech, far-end echo (the fixed beamformer will also perform some echo cancellation) and a residual noise. The second signal is reconstructed background noise signal, $\hat{r}(k)$, which is output signal of MC. The MC consists of an adaptive set of filters $\mathbf{W}_{ANEF}(k)$ that are applied to the whitened noise-only signals $e_{n,e,n}(k)$, $n = 0, \ldots, N-1$. Recall that our goal is to minimize the output power under a constraint on the response at the desired direction. By setting $\mathbf{W}_{e}(k)$ according to (5.10), the constraint is satisfied. Hence, minimization of the output power is achieved by adjusting the filters $\mathbf{W}_{ANEF}(k)$. This is unconstrained minimization, we can implement it by using the multi-channel Wiener filter. The enhanced speech signal, $\hat{s}(k)$, is obtained by subtracting the reconstructed background noise $\hat{r}(k)$ from speech reference $d(k - Q)$, where $Q$ is the number of delay samples for causality.

$$\hat{s}(k) = d(k - Q) - \hat{r}(k)$$

where $\hat{r}(k) = \mathbf{W}_{ANEF}^T(k) \mathbf{B}^T(k) y(k)$

Our goal is to set $\mathbf{W}_{ANEF}(k)$ to minimize

$$\mathbf{e} \left\{ \|d(k) - \mathbf{W}_{ANEF}^T(k) e'(k)\|^2 \right\}$$

(5.19)
where \( \mathbf{e}^{\top}(k) = [e_{r1}^{\top}(k), e_{r2}^{\top}(k), ..., e_{rr}^{\top}(k)] \).

(5.20)

\[
\mathbf{e}_{e,n}^{\top}(k) = [e_e^{\top}(k), e_e^{\top}(k-1), ..., e_e^{\top}(k-M_{\text{LPEF}} + 1)]^\top
\]

(5.21)

Then, the multi-channel Wiener filter is given by

\[
\mathbf{w}_{\text{MCf}}(k) = \mathbf{e}^{\top}(k) \mathbf{e}^\top(k) + \mathbf{e}^{\top}(k) \mathbf{e}^\top(k) \mathbf{d}(k-Q) \mathbf{d}(k-Q)^\top.
\]

(5.22)

The coefficients of the \( \mathbf{w}_{\text{MCf}}(k) \) are updated by NLMS algorithm. Note that the NLMS algorithm in the MC will perform very well due to whitened input signals. On the other hand, in order to minimize the output power at the output of the proposed adaptive combined system, it is required to reconstruct the background noise at the output of MC from the whitened noise. This can be achieved by identifying the MC with an optimum multi-input single-output (MISO) system as shown in Fig. 3.12, Chapter 3. Concerning the background noises only (we assume presence of \( J \) mutually uncorrelated noises). Recall the equations (3.96), (3.97), (3.98), (3.99). The microphone signals are thus calculated by the matrix-vector product \( \mathbf{H}(\zeta)^T \mathbf{r}(\zeta) \). For minimizing the contribution of the noises to the output signal \( \mathbf{i}(\zeta) \) of the proposed adaptive combined system, the filters \( \mathbf{w}_{\text{MCf}}(k) \) will model the combined system of the inverse of the cascade of \( \mathbf{H}(\zeta) \) and \( \mathbf{B}(\zeta) \) in the lower signal path of Fig. 3.12 and the cascade of \( \mathbf{H}(\zeta) \) and of \( \mathbf{W}(\zeta) \) in the upper signal path according to the multiple input/output inverse theorem (MINT)[3].

Note that the background noises are whitened at the output of blocking matrix, then we can reconstruct them by modeling inverses of \( \mathbf{H}(\zeta) \) and of \( \mathbf{B}(\zeta) \) in the lower signal path of Fig. 3.12. This modeling is performed by the filters \( \mathbf{w}_{\text{MCf}}(k) \) in the MC.

The relationship between input and output in the lower signal path of Fig. 3.12 is given as follows

\[
\hat{\mathbf{R}}(\zeta) = \mathbf{H}_{\text{MCf}}(\zeta)[1 - \mathbf{H}_{\text{LPEF}}(\zeta)]\mathbf{H}(\zeta)\mathbf{R}(\zeta)
\]

(5.23)

where \( \mathbf{R}(\zeta) \) and \( \hat{\mathbf{R}}(\zeta) \) represents the Z-transform of the background noise \( \mathbf{r}(\zeta) \) and the reconstructed background noise \( \hat{\mathbf{r}}(\zeta) \), respectively. \( \mathbf{H}_{\text{MCf}}(\zeta), \mathbf{H}_{\text{LPEF}}(\zeta), \) and \( \mathbf{H}(\zeta) \) are defined as the transfer function of \( \mathbf{w}_{\text{MCf}}(k) \), LPEFs in the BM, and impulse response \( \mathbf{H}(\zeta) \), respectively. When \( \hat{\mathbf{R}}(\zeta) \) is equal to \( \mathbf{R}(\zeta) \),

158
Chapter 5 Innovative Adaptive System For Optimal Noise and Acoustic Echo Cancellation

\[ H_{\text{ANEF}}(z) = \frac{1}{\{1 - H_{\text{LPEF}}(z)\}H(z)} \]  

(5.24)

is obvious. The background noise is reconstructed by estimating the inverse filters of circuit including LPEFs and impulse responses between the position of the \( j \)th noise and the \( n \)th microphone. Fig. 5.6 shows an example of the predicting and whitening process by the LPEFs in the blocking matrix and of reconstruction background noise by the \( w_{\text{ANEF}}(k) \) in the MC (the experimental set up is explained in Chapter 6, Section 6.4). Fig. 5.6 (a) and (b) shows the power spectral densities (PSD) of the noisy signal at first microphone and the output of the first LPEF, respectively. From the whiten signal, it can be observe that the most of near-end and echo components are cancelled (and may be also whitened) by the LPEF, thereby the near-end speech signal and far-end echo signal leakage at the sidelobe canceling path are minimized.

Fig. 5.6 PSD of Microphone Signals, Output of LPEF, Actual Background Noise, and Reconstructed Noise by ANEFs.
Fig 5.6 (c) and (d) shows the PSD of the output signal of filter $W_{an}(k)$, $\hat{r}(k)$, and the PSD of original actual background noise signals at first microphone, $r_i(k)$, which are contributions from computer fan and two voltage transformer devices, respectively. It can be seen the high accuracy of the estimated (reconstructed) noise signal by filters $W_{an}(k)$.

5.7 Summary

In this chapter, we have discussed five principle methods for integrating acoustic echo cancellation into beamformers including the proposed system. Though the combination of AECs with fixed beamformers is studied, too, we put more focus on combinations of AECs and adaptive realizations of data-dependent optimum beamformers. The results are summarized in Table 5.1. While the structure with AEC following the adaptive beamformer (BF-first) can hardly be applied, the usage of the other structures depends on the acoustic conditions and on the available computational resources.

AEC-first scheme, provides maximum echo suppression and noise reduction if the acoustic echo paths do not continuously change and if frequent double-talk between near-end speech signal, background noise, and acoustic echoes is not expected, since the AEC can only be adapted if acoustic echoes predominate. Continuously changing echo paths cannot be tracked during double-talk so that the performance reduces in strongly time-varying acoustic environments. The computational complexity may be prohibitively high, since at least the filtering and the filter coefficient update are required for each microphone channel.

The GSAEC scheme with the AEC in the reference path of a GSC structure requires the AEC only for a single recording channel. Thus, computational complexity is considerably reduced relative to AEC-first scheme. GSAEC scheme has almost the same echo suppression capability as AEC-first for predominant acoustic echoes. However, theoretical studies and experimental results showed that the performance of this system is limited since acoustic echoes are still present in the sidelobe canceling path while they may be efficiently cancelled by the AEC in the reference path. Echo suppression and
noise reduction are greater than for the beamformer alone but lower than for AEC-first. GSAEC scheme provides the same tracking capability as AEC-first scheme.

While the GEIC scheme integrate AEC into the multi-channel noise canceller of the beamformer with a GSC structure. The GEIC provides the same number of spatial degrees of freedom as AEC-first, while only the equivalent of an AEC for a single recording channel is required. The AEC can be adapted even during continuous double-talk of noise and acoustic echoes. Therefore, good tracking capability of the AEC was obtained with GEIC scheme. However, the filters for the AEC should not be longer than the filters of the multi-channel noise canceller. Therefore, the echo suppression after convergence of the AEC is limited.

**TABLE 5.1**
Overview of the Properties of the Combinations of Acoustic Echo Cancellation and Beamforming.

<table>
<thead>
<tr>
<th></th>
<th>AEC-first</th>
<th>BF-first</th>
<th>GSAEC</th>
<th>GEIC</th>
<th>Proposed System</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computational Complexity</strong></td>
<td>High</td>
<td>moderate</td>
<td>moderate</td>
<td>moderate</td>
<td>moderate</td>
</tr>
<tr>
<td><strong>Necessity of a separate adaptation control for the AEC</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Not necessary</td>
</tr>
<tr>
<td><strong>Possibility of exploitation of all degree of freedom</strong></td>
<td>Yes</td>
<td>Only at the cost of highly reduced tracking or for WSS conditions</td>
<td>Only for WSS noise sources</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Echo cancellation and tracking capability for single-talk situation</strong></td>
<td>Very high if echo paths are slowly time-varying</td>
<td>Little</td>
<td>High, if echo paths are slowly time-varying</td>
<td>Very little</td>
<td>Very high</td>
</tr>
<tr>
<td><strong>Echo cancellation and tracking capability for double-talk situation</strong></td>
<td>Little</td>
<td>Very little</td>
<td>Moderate</td>
<td>Very high at the cost of less suppression of acoustic echoes for WSS conditions</td>
<td>Moderate</td>
</tr>
<tr>
<td><strong>Improvement of suppression of acoustic echoes and noise relative to beamformer alone</strong></td>
<td>High for slowly time-varying echo paths</td>
<td>Only for special cases</td>
<td>High for predominant echoes and slowly time-varying echo paths</td>
<td>High for strongly time-varying echo paths</td>
<td>High</td>
</tr>
</tbody>
</table>
Finally, we present a novel method for combining noise reduction and acoustic echo canceller in hands-free communication systems. The proposed adaptive combined system includes a new adaptive beamformer with spatial adaptive blocking matrix. The blocking matrix efficiently cancels out the near-end speech signal and acoustic echo in the sidelobe canceling path. While acoustic echo is efficiently suppressed by the AEC in the reference path. By placing the acoustic echo canceller behind fixed beamformer of the proposed adaptive beamformer, we minimize computational complexity by minimizing the number of echo cancellers and we assure that the identification of the echo path is not complicated by the time-variance of the adaptive beamformer.

In the next Chapter it is propose to carry out performance evaluation of the proposed adaptive combined system for optimal background noise and acoustic echo cancellation making use of the experimental setup given in Chapter 3.